

Chromosomal fragile sites, genome instability and human diseases

Edited by

Qing Hu, Advaita Madireddy, Chun-Long Chen
and Stefano Gnan

Published in

Frontiers in Genetics



FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-83251-418-4
DOI 10.3389/978-2-83251-418-4

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Chromosomal fragile sites, genome instability and human diseases

Topic editors

Qing Hu — University of Texas Southwestern Medical Center, United States

Advaitha Madireddy — Rutgers, The State University of New Jersey, United States

Chun-Long Chen — Institut Curie, France

Stefano Gnan — Institut Curie, France

Citation

Hu, Q., Madireddy, A., Chen, C.-L., Gnan, S., eds. (2023). *Chromosomal fragile sites, genome instability and human diseases*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-83251-418-4

Table of contents

- 04 **Editorial: Chromosomal fragile sites, genome instability and human diseases**
Qing Hu, Advaita Madireddy, Stefano Gnan and Chun-Long Chen
- 06 **DNA-PKcs: A Multi-Faceted Player in DNA Damage Response**
Xiaoqiao Yue, Chenjun Bai, Dafei Xie, Teng Ma and Ping-Kun Zhou
- 18 **DNA Repair Gene Polymorphisms and Chromosomal Aberrations in Exposed Populations**
Yasmeen Niazi, Hauke Thomsen, Bozena Smolkova, Ludmila Vodickova, Sona Vodenkova, Michal Kroupa, Veronika Vymetalkova, Alena Kazimirova, Magdalena Barancokova, Katarina Volkovova, Marta Staruchova, Per Hoffmann, Markus M. Nöthen, Maria Dusinska, Ludovit Musak, Pavel Vodicka, Kari Hemminki and Asta Försti
- 27 **Folate Deficiency Triggers the Abnormal Segregation of a Region With Large Cluster of CG-Rich Trinucleotide Repeats on Human Chromosome 2**
Lorenza Garribba, Ivan Vogel, Mads Lerdrup, Marisa M. Gonçalves Dinis, Liqun Ren and Ying Liu
- 43 **Large Intronic Deletion of the Fragile Site Gene *PRKN* Dramatically Lowers Its Fragility Without Impacting Gene Expression**
Sebastian H. N. Munk, Vasileios Voutsinos and Vibe H. Oestergaard
- 52 **Common Threads: Aphidicolin-Inducible and Folate-Sensitive Fragile Sites in the Human Genome**
Rachel Adihe Lokanga, Daman Kumari and Karen Usdin
- 60 **RNF4 Regulates the BLM Helicase in Recovery From Replication Fork Collapse**
Nathan Ellis, Jianmei Zhu, Mary K Yagle, Wei-Chih Yang, Jing Huang, Alexander Kwako, Michael M. Seidman and Michael J. Matunis
- 75 **Transcription-Replication Collisions and Chromosome Fragility**
Wei Wu, Jing Na He, Mengjiao Lan, Pumin Zhang and Wai Kit Chu
- 83 **Characterization of Chromosomal Instability in Glioblastoma**
Elisa Balzano, Elena Di Tommaso, Antonio Antocchia, Franca Pelliccia and Simona Giunta
- 95 **Epigenomic signatures associated with spontaneous and replication stress-induced DNA double strand breaks**
Sravan Kodali, Silvia Meyer-Nava, Stephen Landry, Arijita Chakraborty, Juan Carlos Rivera-Mulia and Wenyi Feng



OPEN ACCESS

EDITED AND REVIEWED BY

Jared C. Roach,
Institute for Systems Biology (ISB),
United States

*CORRESPONDENCE

Qing Hu,
✉ Qing.Hu@UTSouthwestern.edu
Chun-Long Chen,
✉ Chunlong.Chen@curie.fr

SPECIALTY SECTION

This article was submitted to Human
and Medical Genomics,
a section of the journal
Frontiers in Genetics

RECEIVED 08 December 2022

ACCEPTED 13 December 2022

PUBLISHED 10 January 2023

CITATION

Hu Q, Madireddy A, Gnan S and
Chen CL (2023), Editorial:
Chromosomal fragile sites, genome
instability and human diseases.
Front. Genet. 13:1119532.
doi: 10.3389/fgene.2022.1119532

COPYRIGHT

© 2023 Hu, Madireddy, Gnan and Chen.
This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License](#)
(CC BY). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Editorial: Chromosomal fragile sites, genome instability and human diseases

Qing Hu ^{1*}, Advaita Madireddy ², Stefano Gnan ³ and
Chun-Long Chen ^{3*}

¹Department of Pathology, University of Texas Southwestern Medical Center, Dallas, TX, United States,
²Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, United States, ³Institut Curie, Dynamics
of Genetic Information, CNRS UMR3244, PSL Research University, Sorbonne Université, Paris, France

KEYWORDS

fragile sites, replication stress, genome instability, human disease, DNA damage repair

Editorial on the Research Topic

Chromosomal Fragile Sites, Genome Instability and Human Diseases

Chromosomal fragile sites are specific regions of the human genome that are normally stable, but exhibit breaks or gaps on metaphase chromosomes under conditions of stress. As an important source of genome instability, fragile sites are associated with human diseases such as cancer and mental retardation. In this Research Topic, we collected nine contributions from different perspectives, which cover the features and causes of fragile sites, the characterization of fragile sites, as well as the cellular mechanisms at play to maintain their stability.

Based on the frequency of fragility, fragile sites are generally categorized as common fragile sites (CFSs) and rare fragile sites (RFSs). Lokanga et al. summarized recent work about CFSs and RFSs and discussed the similarities and differences between them. CFSs are an intrinsic part of normal chromosome structures and are present in all individuals. Moreover, CFSs are appeared to be conserved throughout mammalian evolution, for example, human FHIT/FRA3B and mouse Fhit/Fra14A2 are orthologs. RFSs are found in a minority of the human population and are inherited in a Mendelian manner. They are often associated with the expansion of repeat elements. RFSs can be further categorized into two groups: folate-sensitive RFSs that are highly sensitive to folate deficiency, and non-folate-sensitive RFSs that are induced by distamycin A or bromodeoxyuridine (BrdU). As the feature of folate-sensitive RFSs, CGG trinucleotide repeats (TNRs) could form secondary structures and perturb DNA replication, thus contributing to their fragility. Garribba et al. tested whether folate deficiency could cause instability at other genomic regions containing CG-rich repeat sequences and showed that a region at Chr2p11.2 displayed an unusual conformation under folate deprivation, leading to the mis-segregation of this locus and Chr2 Aneuploidy.

Although human CFSs have been primarily mapped in lymphocytes and fibroblasts, it has been shown that different CFSs are expressed in different cell types at various

frequencies. Characterization of CFSs in other cell types and investigation of CFS expression in different genetic backgrounds are, therefore, important for us to understand the basis of their fragility. In this Research Topic, Balzano et al. assessed CFS expression in human glioblastoma cell lines and found two CFSs that are specific to glioblastoma. They also showed that the fragility of these sites is related to impaired DNA replication. As the major cause underlying CFS fragility, replication stress can arise from various sources, such as oncogene activation, nucleotide depletion, and Transcription-Replication Conflicts (TRCs). A large majority of the highly expressed CFSs host large genes spanning over megabases and transcription-associated replication stress has been proposed as the prominent mechanisms leading to fragility at these large genes. Transcription can either suppress replication initiation to generate large regions that are poor in replication initiation or can generate direct collision with DNA replication machinery. Wu et al. reviewed recent findings on how conflicts during transcription and replication affect chromosome fragility. This mechanism is also supported by the recent study of Munk et al., who deleted 80 kb intron sequences of an extremely large gene *PRKN* and showed that this deletion resulted in a twofold reduction of the *PRKN* fragility without affecting its expression. The mechanism underlying CFS fragility is not restricted to TRCs. Other features of CFSs such as late replication timing, DNA secondary structure formation, and chromatin modification, can also lead to replication stress and contribute to their fragility. Kodali et al. compared the epigenomic signatures associated with spontaneous and replication stress-induced DNA double-strand breaks (DSBs) and demonstrated the correlation of aphidicolin-induced DSBs with histone 3 lysine 36 trimethylation, which is a marker for active transcription. More interestingly, their result suggested that DSBs were not enriched in the CFS core sequences and rather demarcated the CFS core region. Their analysis suggested that altered replication dynamics are responsible for CFS formation under a relatively higher level of replication stress.

Cellular pathways and signaling components that are involved in mitigating replication stress are important to maintain genome stability. In response to replication stress, the stalled replication forks require stabilization and remodeling to facilitate fork restart. Many proteins that participate in these processes have been shown to affect CFS expression, such as ATR, DNA-PKcs, and some DNA helicases. The function of DNA-PKcs in DNA replication stress is reviewed by Yue et al. The Bloom syndrome DNA helicase BLM belongs to the RECQ family and has been shown to participate in fork restart. Here, Ellis et al. showed that RNF4 facilitates replication fork recovery by regulating BLM. Stalled forks that failed to

restart will collapse and use the DNA damage repair (DDR) pathway to recover. Niazi et al. identified 14 DNA repair genes that are associated with chromosomal aberrations, further emphasizing the importance of DDR in promoting genome stability. Deficiencies in DNA damage mechanisms are, therefore, an important source of genome instability and render cells more sensitive to replication stress.

Our understanding of the mechanisms governing chromosomal fragile sites is constantly evolving. In addition to clearly defining the current state of the field, the articles included in this Research Topic present valuable novel insights into the determinants of chromosomal fragility, which contribute to many human diseases.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

AM was supported by COCR22PRG002 to AM and W81XWH-21-1-0935 to AM. CLC was supported by the YPI program of I. Curie; the ATIP-Avenir program from Centre national de la recherche scientifique (CNRS) and Plan Cancer (grant number ATIP/AVENIR: N°18CT014-00); the Agence Nationale pour la Recherche (ANR) (grant number ReDeFINE—19-CE12-0016-02, TELOCHROM—19-CE12-0020-02, SMART—21-CE12-0033-02); and Institut National du Cancer (INCa) (grant number PLBIO19-076).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.



DNA-PKcs: A Multi-Faceted Player in DNA Damage Response

Xiaoqiao Yue^{1,2}, Chenjun Bai², Dafei Xie², Teng Ma^{3*} and Ping-Kun Zhou^{2*}

¹ School of Public Health, University of South China, Hengyang, China, ² Department of Radiation Biology, Beijing Key Laboratory for Radiobiology, Beijing Institute of Radiation Medicine, Beijing, China, ³ Department of Cellular and Molecular Biology, Beijing Chest Hospital, Capital Medical University/Beijing Tuberculosis and Thoracic Tumor Research Institute, Beijing, China

OPEN ACCESS

Edited by:

Chunlong Chen,
Institut Curie, France

Reviewed by:

Xingzhi Xu,
Shenzhen University, China
Shan Zha,
Columbia University, United States

*Correspondence:

Teng Ma
mateng82913@163.com
Ping-Kun Zhou
zhoupk@bmi.ac.cn;
birm4th@163.com

Specialty section:

This article was submitted to
Human and Medical Genomics,
a section of the journal
Frontiers in Genetics

Received: 17 September 2020

Accepted: 01 December 2020

Published: 23 December 2020

Citation:

Yue X, Bai C, Xie D, Ma T and
Zhou PK (2020) DNA-PKcs:
A Multi-Faceted Player in DNA
Damage Response.
Front. Genet. 11:607428.
doi: 10.3389/fgene.2020.607428

DNA-dependent protein kinase catalytic subunit (DNA-PKcs) is a member of the phosphatidylinositol 3-kinase related kinase family, which can phosphorylate more than 700 substrates. As the core enzyme, DNA-PKcs forms the active DNA-PK holoenzyme with the Ku80/Ku70 heterodimer to play crucial roles in cellular DNA damage response (DDR). Once DNA double strand breaks (DSBs) occur in the cells, DNA-PKcs is promptly recruited into damage sites and activated. DNA-PKcs is auto-phosphorylated and phosphorylated by Ataxia-Telangiectasia Mutated at multiple sites, and phosphorylates other targets, participating in a series of DDR and repair processes, which determine the cells' fates: DSBs NHEJ repair and pathway choice, replication stress response, cell cycle checkpoints, telomeres length maintenance, senescence, autophagy, etc. Due to the special and multi-faceted roles of DNA-PKcs in the cellular responses to DNA damage, it is important to precisely regulate the formation and dynamic of its functional complex and activities for guarding genomic stability. On the other hand, targeting DNA-PKcs has been considered as a promising strategy of exploring novel radiosensitizers and killing agents of cancer cells. Combining DNA-PKcs inhibitors with radiotherapy can effectively enhance the efficacy of radiotherapy, offering more possibilities for cancer therapy.

Keywords: DNA-PKcs, DNA damage response, DNA repair, genomic instability, radiosensitization

INTRODUCTION

Eukaryotic cells are constantly encountering various endogenous (e.g., DNA replication errors) and exogenous (ionizing radiation, chemical carcinogens, and UV) stresses of genomic DNA damage (Jeggo et al., 2016). DNA damage includes base damage, pyrimidine dimer formation, single or double strand breaks, etc. Among them, DNA double strand break (DSB) is the most serious and dangerous threat that affects the stability of the genome and cell fate (Jackson and Bartek, 2009; Huang and Zhou, 2020). Failure to make timely and precise repairs can lead to accumulation of residual DNA damage, mutations, rearrangements, and/or loss of chromosomes, which can lead to a series of cellular consequences such as cell death, senescence, transformation, mutagenesis, or carcinogenesis (Her and Bunting, 2018). Fortunately, cells have evolved several precise DNA damage response (DDR) and repair machineries to deal with various types of DNA damages.

DDRs include DNA damage sensing, initiating DNA damage signaling cascades, remodeling and relaxing chromatin around DSBs, recruiting DNA repair proteins to the damaged site, activating cell cycle checkpoints, and repairing DSB (Jackson and Bartek, 2009; Huang and Zhou, 2020). DNA repair defects are closely related to a series of human diseases and aging (Vijg and Suh, 2013; Longerich et al., 2014; Madabhushi et al., 2014; Shimizu et al., 2014; Chow and Herrup, 2015; Konstantinopoulos et al., 2015; Fang et al., 2016; Uryga et al., 2016; Chatzidoukaki et al., 2020). The DSBs repair pathways of eukaryotic cells mainly include the non-homologous end joining (NHEJ), homologous recombination (HR) and alternative end-joining (Lieber, 2008; Chiruvella et al., 2013; Chang et al., 2017; Shibata and Jeggo, 2020).

DNA-dependent protein kinase (DNA-PK) is an important player in the NHEJ pathway and was also found to function in multiple nodes of DDRs. DNA-activated/or DNA-dependent protein kinase was first reported in 1985 by Walker et al. (1985). They discovered this protein by chance, finding that addition of double-stranded DNA (dsDNA) into the cell extracts increased the phosphorylation of certain proteins. In 1990, Carter et al. (1990) and Lees-Miller et al. (1990) identified the DNA-PK catalytic subunit (DNA-PKcs) with hsp90 or casein as phosphorylation substrate bait. DNA-PKcs is an abundant protein with 50,000–100,000 molecules per cell in humans (Anderson and Lees-Miller, 1992; Blunt et al., 1995; Miller et al., 1995; van der Burg et al., 2009; Woodbine et al., 2013; Chang and Lieber, 2016). *PRKDC* gene mutations in patients or expression of kinase-dead DNA-PKcs protein in mice causes severe combined immunodeficiency (SCID). DNA-PKcs, Ataxia-Telangiectasia Mutated (ATM), and ATM and RAD3 related (ATR) belong to the family of the phosphatidylinositol 3-kinase related kinase (PIKK), which plays significant roles in DNA damage repair (Blackford and Jackson, 2017; Kantidze et al., 2018). They all harbor similar domain compositions, including a N-terminal HEAT-repeat rich segment followed by the conserved FRAP-ATM-TRRAP (FAT) domain, the kinase domain, the PIKK regulatory domain (PRD) and the FAT C-terminal motif (FATC; Blackford and Jackson, 2017). DNA-PKcs, ATM, and ATR preferentially phosphorylate the S/T-Q motif (serine or threonine residue followed by a glutamine) (Kim et al., 1999). DNA-PKcs and ATM mainly mediate the repair of DNA double-strand breaks through NHEJ and HR, respectively, while ATR responds to the stalled DNA replication forks and DNA single-strand breaks (Falck et al., 2005; Menolfi and Zha, 2020). Whereas, a series of reports indicated that DNA-PKcs is also required for optimal replication stress response (Lin et al., 2014, 2018; Wang et al., 2015; Ying et al., 2016; Kantidze et al., 2018; Joshi et al., 2019; Villafanez et al., 2019).

This review mainly summarized and discussed the updated accomplishments about DNA-PKcs research, including post-translational modifications and activity regulation of DNA-PKcs and its involvements in DDRs. Furthermore, considering the special role of DNA-PKcs in the DDR, we have also reviewed the progress on exploration of DNA-PKcs inhibitors as radiosensitizers for cancer radiotherapy.

Post-translational Modifications and Activation of DNA-PKcs Triggered by DNA Damage Signaling

DNA-dependent protein kinase catalytic subunit belongs to the PIKKs family, and is a type of DNA-activated serine/threonine protein kinase with a molecular weight of approximately 469 kD encoded by the *PRKDC* or *XRCC7* gene (Hartley et al., 1995; Siple et al., 1995; Lees-Miller, 1996). It forms a holoenzyme DNA-dependent kinase (DNA-PK) with the heterodimer regulatory subunits of Ku70 and Ku80 (referred to as Ku together) (Jin and Weaver, 1997; Singleton et al., 1999). Ku70 is encoded by *XRCC6* gene and Ku80 is encoded by *XRCC5*. Ku70 and Ku80 have a strong affinity for DNA ends, and they also provide docking sites for other proteins during DDR (Featherstone and Jackson, 1999).

The DNA-PKcs structure consists of an N-terminal region, a circular cradle unit, and a head unit with the kinase domain between FAT (FRAP, ATM, and TRRAP) and FATC domains (Sibanda et al., 2010, 2017; Sharif et al., 2017; Yin et al., 2017). Both the N-terminal region and the cradle unit contain HEAT (Huntingtin, Elongation factor 3, regulatory subunit A of PP2A, and TOR1) repeats. Both the ABCDE and PQR phosphorylation clusters fall into the cradle unit (Figure 1; Spagnolo et al., 2006; Sibanda et al., 2010; Morris et al., 2011).

The activity and function of DNA-PKcs in end-ligation or NHEJ are tightly regulated by phosphorylation modification. The N-terminal domain has many phosphorylation sites, DNA-PKcs can be auto-phosphorylated or phosphorylated by ATM and ATR (Chen et al., 2007; Dobbs et al., 2010). The FAT and FATC domains surround the catalytic domain, stabilize the conformational changes of the catalytic center and regulate kinase activity (Rivera-Calzada et al., 2005; Jiang et al., 2006; Spagnolo et al., 2006). As shown in Figure 1, the phosphorylation sites of ABCDE clusters are T2609, S2612, T2620, S2624, T2638, and T2647, and the phosphorylation sites of PQR clusters are S2023, S2029, S2041, S2051, S2053, and S2056 (Douglas et al., 2002). Among them, S2056 and T2609 are two prominent autophosphorylation sites of DNA-PKcs, both are crucial for its activity in DNA repair (Ding et al., 2003; Block et al., 2004; Nagasawa et al., 2011, 2017). The current model suggests that DNA-PKcs S2056 phosphorylation causes conformational changes, thereby promoting DNA-PK disassembly from the DSB site, allowing DNA end ligation (Jiang et al., 2015), while the ABCDE cluster phosphorylation is required for DNA end resection (Shibata et al., 2011). Another autophosphorylation site, T3950, is located in the kinase domain and its phosphorylation was suggested to shut down the activity of DNA-PKcs kinase (Douglas et al., 2007).

In addition to phosphorylation modification, there are also other forms of post-translational modifications on DNA-PKcs (Table 1). Proteomics studies have revealed that DNA-PKcs are widely acetylated (Mori et al., 2016) and ubiquitinated (Ho et al., 2014). Lysine acetylation is an important form of post-translational modifications and is also the most widely studied post-translational modification on histones. Proteomics showed that at least 16 lysine residues were acetylated in the DNA-PKcs.

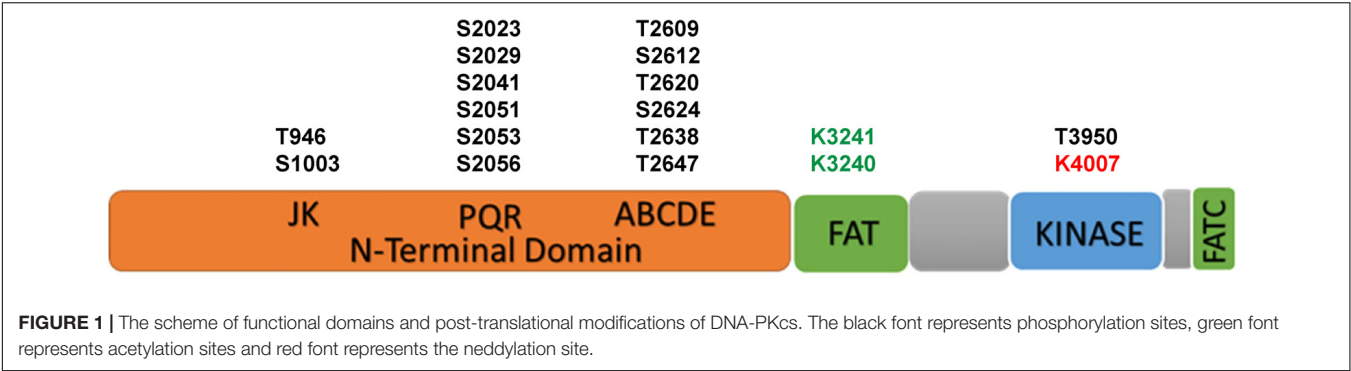


TABLE 1 | The modification sites of DNA-PKcs.

Modification sites	Modification type	Function	References
T2609	phosphorylation	Activates Artemis-mediated endonuclease activity; Promoting cNHEJ repair through end processing; DSB repair and cellular sensitivity to gamma radiation; Facilitates telomere leading strand maturation	Liu et al., 2012; Lin et al., 2014; Longerich et al., 2014; Lin et al., 2018
S2056	phosphorylation	DSB repair and cellular sensitivity to gamma radiation; Promotes end ligation in cNHEJ; Increases expression of RBX1 in G1 stage	Ma et al., 2002; Goodarzi et al., 2006; Lu et al., 2007; Reichert et al., 2011; Lin et al., 2014; Madabhushi et al., 2014; Jiang et al., 2019
T2647	phosphorylation	Activates Artemis-mediated endonuclease activity	Goodarzi et al., 2006; Lin et al., 2018
T3950	phosphorylation	Shuts down the activity of DNA-PKcs kinase	Konstantinopoulos et al., 2015
K3241 K3260	acetylation	Maintains genome stability and radiation resistance	Kotsantis et al., 2018
K4007	neddylation	Promotes autophosphorylation of DNA-PKcs at Ser2056	Guo et al., 2020

Mori et al. (2016) selected eight of them for further research and have confirmed that K3241 and K3260 were acetylated on DNA-PKcs. In addition, our research group also found that both PARYlation (Han et al., 2019) and neddylation (Guo et al., 2020) can regulate DNA-PK activity, a neddylation site was identified at K4007 within the kinase domain. PARP1-dependent DNA-PKcs PARYlation can be induced by DNA damage signals and affects DNA-PKcs Ser2056 phosphorylation in cells (Han et al., 2019). DNA-PKcs neddylation occurs in the kinase domain and is catalyzed by HUWE1 ligase (Guo et al., 2020). In 2014, the ring finger protein 144A (RNF144A) was discovered as the first E3 ubiquitin ligase of cytoplasmic DNA-PKcs. Ho et al. (2014) RNF144A is induced in a p53-dependent manner during DNA damage and targets cytoplasmic DNA-PKcs for ubiquitination and degradation.

Considering the close relationship between DNA-PKcs and DSB, it is particularly important to determine the mechanism controlling DNA-PKcs activation. In the cellular response to DSB, Ku heterodimer recognizes and localizes DNA damage sites, and promptly recruits DNA-PKcs. Once DNA-PKcs is recruited, it is activated through phosphorylation in a DNA-dependent manner. The C-terminal 178 amino acid residues of Ku80 is required for Ku80/DNA-PKcs interaction and indispensable for DNA-PKcs activation (Singleton et al., 1999). DNA-PKcs pushes Ku protein inwards onto the DNA, and then phosphorylates the other components nearby, including its own phosphorylation. DNA-PKcs autophosphorylation can be regulated by N-terminal conformational changes of protein (Meek et al., 2012). In addition to the automatic regulation of DNA-PKcs, many other

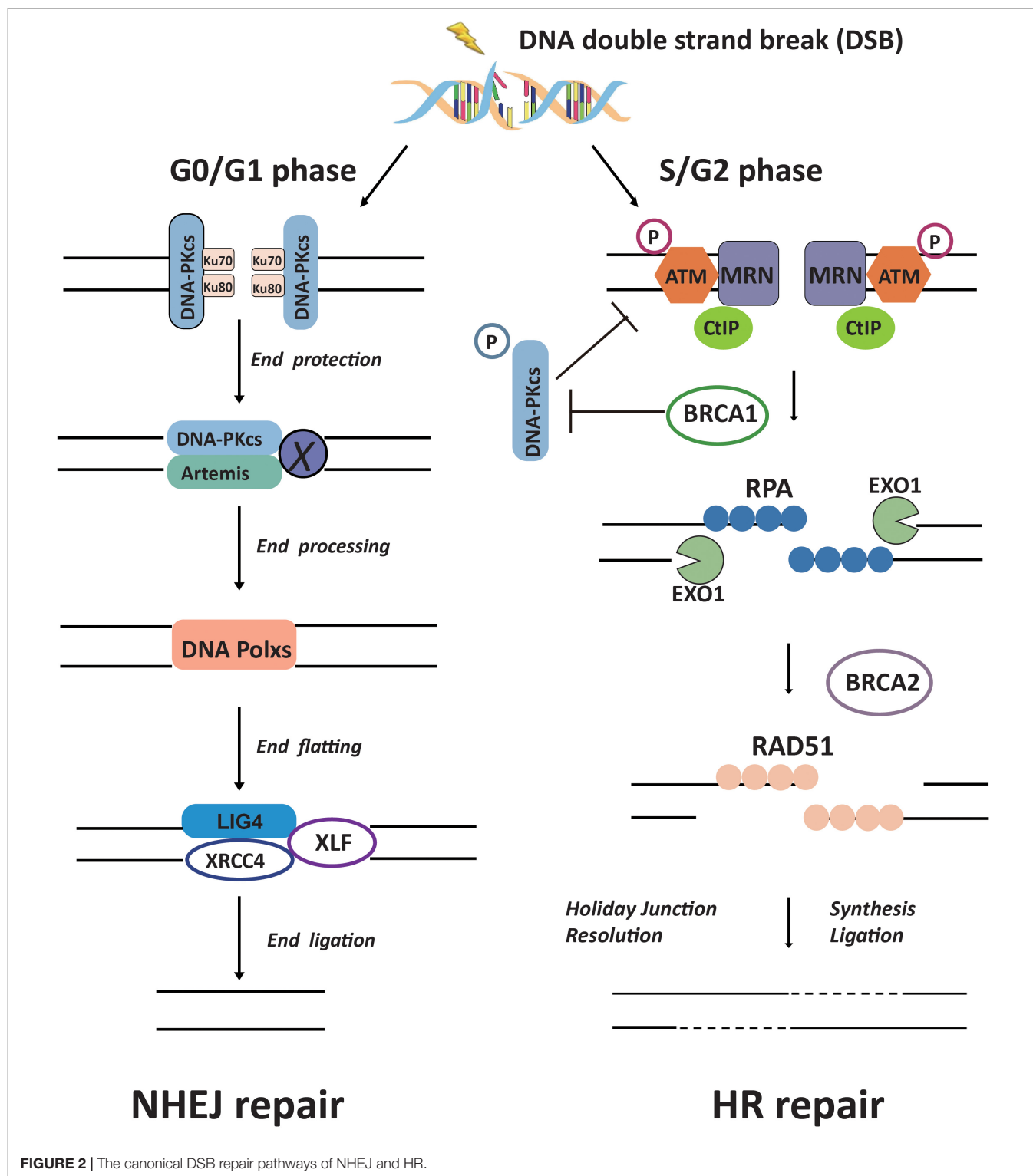
factors are also suggested to be key regulators of DNA-PKcs activity. For example, epidermal growth factor (EGFR) can bind DNA-PKcs and enhance the activity of DNA-PKcs to deal with the damage (Liccardi et al., 2011; Javvadi et al., 2012). The non-kinase regulator protein phosphatase 6 (PP6) and protein phosphatase 1(PP1) is recruited to the DSB site and promotes DNA-PKcs activity through direct interaction and dephosphorylation action (Douglas et al., 2010; Zhu et al., 2017). Thus, the activation of DNA-PKcs is not only the result of binding to the broken DNA ends, but also a complex process affected by many factors. The multiple mechanistic ways of activation suggest that there are many unknown functions waiting to be explored in DNA-PKcs in addition to participating in double-strand break repair.

Regulation of DNA Damage Response by DNA-PKcs

Current knowledge displays that DNA-PKcs can function in multiple pathways of cellular DDRs to maintain the genome stability and cell survival.

DNA-PKcs in Non-homologous End Joining

There are two major repair pathways for DSB: NHEJ and HR (Figure 2). NHEJ repair is an error-prone repair pathway and theoretically executes its function throughout the cell cycle, but it is most important during G1 when no homologous template for recombination is available. HR performs under the guidance



of an intact homologous template DNA, and it is known as an error-free repair pathway. HR is most active in the S/G2 phase and nearly absent in G1 phase (Mao et al., 2008; Menon and Povirk, 2016). As a main DSB repair pathway in mammals, NHEJ is initiated by the circular Ku70/Ku80 heterodimer binding to

the broken DNA ends. Ku has a high abundance and strong affinity for free DNA. Within a second or less, Ku proteins can associate with any DSB that occurs in the nuclei genomic DNA. The interaction of Ku and broken DNA recruits DNA-PKcs to form DNA-PK complex (Gomes et al., 2017), then a

set of NHEJ downstream factors including Artemis, XRCC4, DNA polymerase X family, and DNA ligase IV to join the broken DNA ends. DNA-PKcs is the only active protein kinase described in the NHEJ pathway, it is autophosphorylated in the presence of DNA termini, Mg^{2+} and ATP. DNA-PKcs forms a tight complex with Artemis to stimulate the nuclease activity of the latter (Kamoi and Mochizuki, 2010; Chang and Lieber, 2016). In this process, DNA-PKcs phosphorylates the C-terminal inhibitory region of Artemis, facilitating the dissociation of the inhibitory region from the N-terminal catalytic domain (Ma et al., 2002; Pannunzio et al., 2018). Ionizing radiation-induced DNA DSB broken ends are usually described as “dirty ends” and incompatibility for ends direct ligation caused by chemical modifications and 5′ or 3′ mismatching overhangs. The activated Artemis processes these “dirty ends” for subsequent ligation. DNA-PKcs kinase activity becomes necessary for direct end-ligation of NHEJ in the presence of DNA-PK protein. The DNA-PKcs^{KD/KD} (kinase-dead mutant) cells show severe end-ligation defects (Jiang et al., 2015).

Usually, NHEJ takes the first action to respond to the DSB (Shibata et al., 2011). If NHEJ cannot be completed, then DSB “cuts,” where one strand of the DNA duplex is degraded to produce a single-stranded DNA overhang suitable for alternative repair pathways, HR occurs (Sallmyr and Tomkinson, 2018). Experiments have shown that NHEJ is much faster than HR and occurs within 30 min (while HR requires 7 h or more), accounting for approximately 75% of repair events (Zhang and Matlashewski, 2019). NHEJ repair does not use sequence homology, regardless of the position or sequence, the DNA break ends are brought together, so this type of repair is prone to errors.

Regulation of DNA-PKcs on the Pathway Choice of DSBs Repair

Homologous recombination repair is initiated by DSB end resection. MRE11 and CtIP together with EXO1 fulfill the DNA end resection process (Syed and Tainer, 2018). It is usually acknowledged that the interaction between 53BP1 and BRCA1 is an important factor to control the end resection (Daley and Sung, 2014; Zimmermann and de Lange, 2014; Mirman and de Lange, 2020). BRCA1 antagonizes certain functions of 53BP1 and promotes HR by reducing 53BP1-mediated NHEJ. After end resection, HR uses the undamaged homologous DNA as a template, and recombinase RAD51 invades the homologous DNA strand, resulting in precise repair (Sun et al., 2020).

How cells select the pathway to execute DSBs repair between HR and NHEJ is a critical issue for efficient and precise repair of DSBs, whereas its mechanism is not fully understood yet (Ceccaldi et al., 2016; Scully et al., 2019). The initiation of DNA end resection is a definite factor that enables the cells to perform HR repair to prevent NHEJ repair (Scully et al., 2019). There are many factors that affect the initiation of end resection, but DNA-PKcs phosphorylation status is a clear factor that impacts cells to choose NHEJ or HR (Neal et al., 2011). A series of reports indicated that DNA-PKcs can regulate the choice of DSBs repair pathways at multiple biochemical nodes

in association with the cell cycle. In the G₂ phase of cell cycle, autophosphorylation of DNA-PKcs promotes DNA-PKcs dissociation from the DSBs sites, and facilitates the recruitment of end resection enzymes such as EXO1 (Shibata et al., 2011; Sallmyr and Tomkinson, 2018). DNA-PKcs also promotes end-resection from *in vitro* analyses (Deshpande et al., 2020). The DNA-PKcs mutant that makes autophosphorylation defective at the ABCDE cluster (DNA-PKcs ABCDE 6A) binds DSBs but precludes the completion of NHEJ, significantly reducing DSB end resection at all DSBs (Shibata et al., 2011). Whereas in developing lymphocytes, robust end-resection is detected in both DNA-PKcs kinase-dead or the phosphorylation (ABCDE) mutant (Crowe et al., 2018, 2020). ATM kinase activity can compensate for DNA-PKcs autophosphorylation when DNA-PKcs activity is inhibited and promote resection. The Mre11-Rad50-Nbs1 (MRN) complex further stimulates resection in the presence of Ku and DNA-PKcs by recruiting EXO1 and enhancing DNA-PKcs autophosphorylation, and it also inhibits DNA ligase IV/XRCC4-mediated end rejoining (Zhou and Paull, 2013). In the S phase of cell cycle, BRCA1 interacts with DNA-PKcs and directly blocks DNA-PKcs autophosphorylation, thus priming DNA end resection for HR and HR factors loading on DSBs (Davis et al., 2014; **Figure 2**). The interaction of TIP60 histone acetyltransferase and DNA-PKcs prompts the autophosphorylation and activation of DNA-PKcs (Jiang et al., 2006). We recently revealed that an increased SUMO2 modification of TIP60 K430 mediated by PISA4 E3 ligase hinders its interaction with DNA-PKcs in S phase cells, leading to impediment of DNA-PKcs S2056 autophosphorylation and preferentially employing the HR pathway for DSBs repair in S phase (**Figure 3**). TIP60 K430R mutation can recover the interaction of DNA-PKcs and TIP60, resulting in abnormally increased phosphorylation of DNA-PKcs S2056 in S phase and dramatical inhibition of HR efficiency (Gao et al., 2020).

Meanwhile, HR activity in G₁ is restricted. Zhou et al. (2017) found that DNA-PKcs phosphorylates ATM directly, inhibiting ATM activity and ATM signaling upon DNA damage, providing a possible mechanism of HR restriction in G₁ phase. Moreover, our recent report demonstrated that EXO1 protein is strictly suppressed in G₁ phase. Increased expression of RBX1 protein prompts the neddylation and activity of cullin1, a key component of the Skp1-Cullin1-F-box (SCF) ubiquitin 3 ligase, and consequently mediates the ubiquitination degradation of EXO1 in G₁ phase. Increased DNA-PKcs activity is responsible for increased RBX1 protein expression, and limiting the formation of DSB-end ssDNA by EXO1 and suppressing the HR repair pathway in G₁ cells (**Figure 3**; Xie et al., 2020).

DNA-PKcs in DNA Replication Stress

In addition to the role in DSB repair, DNA-PKcs is also involved in DNA replication stress response. Cells are particularly susceptible to DNA damage during the process of DNA replication. Almost all forms of DNA damage can disturb the DNA replication and cause replication stress (Bass et al., 2012).

Replication stress is defined as any DNA replication barrier that hinders, prevents or terminates DNA synthesis, and can activate the replication stress response to resolve the damage

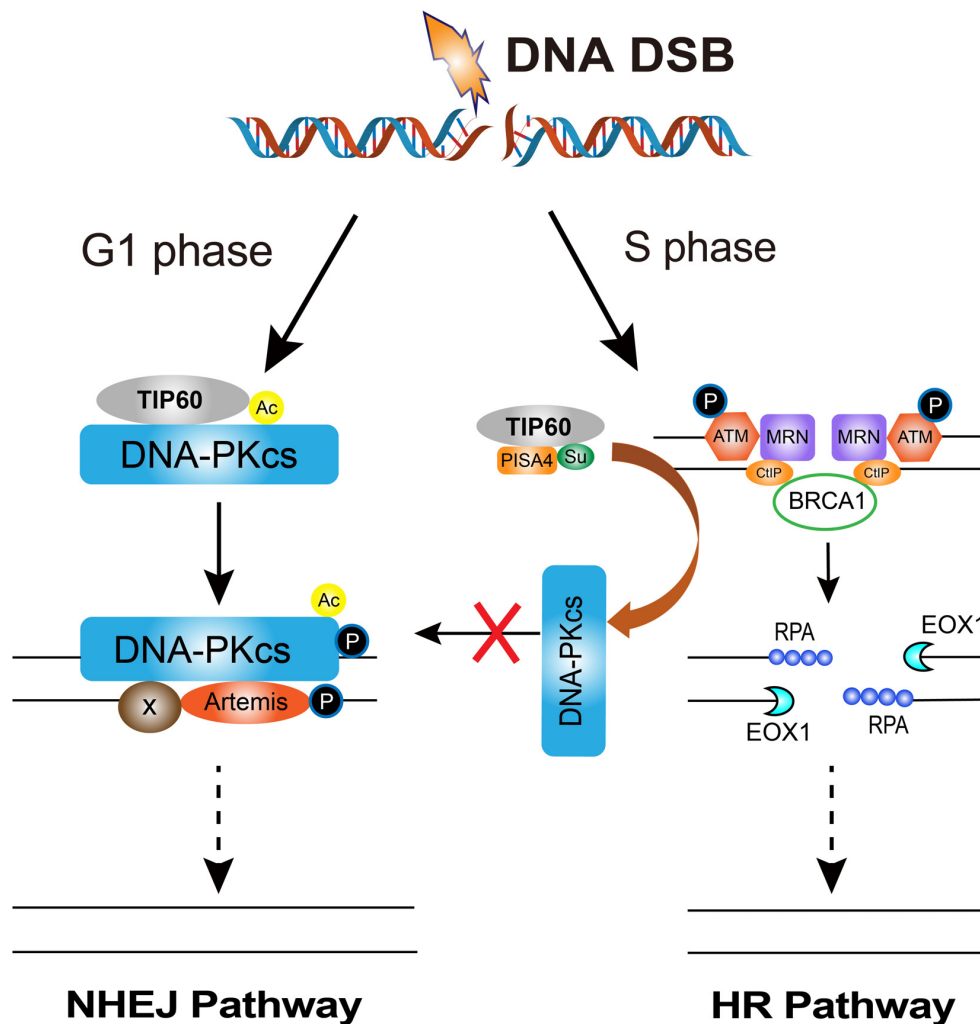


FIGURE 3 | The regulation of DNA-PKcs-interacting TIP60 on the cell cycle-dependent choice of DNA DSBs repair pathway. Upon DNA DSBs signaling, TIP60 interacts with and acetylates DNA-PKcs to activate DNA-PKcs through autophosphorylation at S2056 for executing NHEJ in G1 phase. Whereas, in S phase cells, SUMO2 modification (SU) of TIP60 at K430 is mediated by PISA4 E3 ligase, which hinders the interaction between TIP60 and DNA-PKcs, consequently inactivating DNA-PKcs to give way to HR proteins and facilitates HR repair.

(Ubhi and Brown, 2019). Replicating stress checkpoint (S-phase checkpoint) involves the gradual activation of damage sensors, mediators and effectors (Allen et al., 2011). Replication protein A (RPA) binds to single-stranded DNA (ssDNA), recruiting numerous sensor proteins including ataxia telangiectasia and Rad3-related (ATR)-interacting protein (ATRIP), the 9-1-1 DNA clamp complex (RAD9-RAD1-HUS1), topoisomerase II binding protein 1 (TOPBP1) and Ewing tumor-associated antigen 1 (ETAA1), etc., triggers ataxia telangiectasia and Rad3-related (ATR) and phosphorylation of RPA32 (a subunit of RPA), leading to Chk1 activation (Zeman and Cimprich, 2014; Kotsantis et al., 2018).

DNA-dependent protein kinase catalytic subunit is necessary for ATR-Chk1 signal transduction. Upon replication stress, the DNA-PKcs is phosphorylated by ATR at the stalled replication forks, leading to transcriptional activation of Claspin expression

and Chk1-Claspin complex stability, which is required for the optimal activation of intra S-phase checkpoint (Lin et al., 2014). The apoptosis mediator p53-induced protein with a death domain (PIDD) mediates DNA-PKcs recruitment at the stalled replication forks, and promotes the ATR signaling pathway in the cellular response to replication pressure and the cellular resistance to replication pressure (Lin et al., 2018). Phosphorylation of RPA32 Ser4/Ser8 is a key early step to fully activate ATR in response to replication stress and subsequent replication checkpoint stagnation. DNA-PK is the main kinase that targets Ser4/Ser8 in replication stress. The PIKK phosphorylation of RPA32 plays a key role in replication checkpoint activation, and DNA-PK was considered as an important contributor to this response (Liu et al., 2012). In addition, it was reported that ATR was critical in the early S phase, especially in cells under high replication stress; however, ATR

can be circumvented by DNA-PKcs and Chk1 under moderate replication stress. They concluded distinct but concerted roles of ATR, DNA-PK, and Chk1 in countering replication stress (Buisson et al., 2015).

DNA-PKcs in Autophagy

Autophagy is a physiological degradation mechanism of cells through which autophagic vesicles can deliver unfolded proteins and damaged organelles to lysosomes to eliminate them (Leidal et al., 2018). Autophagy plays an important role in different DDR pathways (Hewitt and Korolchuk, 2017). Inhibition of DNA-PKcs by bromovanin, a vanillin derivative, exhibited a potent antiproliferation effect through induction of both apoptosis and autophagy in HepG2 cells (Yan et al., 2007). Daido et al. (2005) reported that relative low-dose IR induced massive autophagic cell death in M059J cells that lack DNA-PKcs. The treatment of M059K cells with DNA-PKcs antisense oligonucleotides caused radiation-induced autophagy and radiosensitized the cells. Similar observations were also obtained by Puustinen et al. (2020) study. Inactivation or depression of DNA-PKcs increases ionizing radiation-induced cell autophagy as well as occurrence of cell apoptosis. It was found that DNA-PKcs interacts with the AMP-dependent protein kinase (AMPK) complex and phosphorylates nucleotide-sensing $\gamma 1$ subunit (protein kinase AMP-activated non-catalytic subunit gamma) PRKAG1/AMPK $\gamma 1$ at Ser192 and Thr284, thereby facilitating the AMPK complex activation by STK11 at lysosomal and autophagy.

The Association of DNA-PKcs With Telomeres

Telomeres are DNA-protein complexes located at the ends of chromosomes in eukaryotic cells. The major function of telomeres is to maintain the stability and integrity of chromosomes (Roake and Artandi, 2020). The telomere sequence has unique characteristics, making telomeres susceptible to various DNA damages induced by the external environment and genotoxins (Lazzerini-Denchi and Sfeir, 2016). Any abnormality in telomere function may lead to cell aging/senescence and canceration. Telomeres can sense both intrinsic and extrinsic stresses and its dysfunction drives cell senescence (Vitorelli and Passos, 2017). DNA-PKcs plays a role in telomere maintenance and is necessary for telomere capping (Goytisolo et al., 2001; Zhang et al., 2016; Sui et al., 2020). Sui et al. (2015) reported that hTR-mediated DNA-PKcs stimulation and subsequent hnRNP A1 phosphorylation affect the cell cycle-dependent distribution of telomeric repeat-containing RNA (TERRA) on telomeres by promoting the removal of TERRA in telomeres, which is important for the S-phase process, thereby promoting efficient telomere replication and capping (Ting et al., 2009). In cells lacking DNA-PKcs, uncapped telomeres are inappropriately detected and processed as DSB, and therefore not only participate in spontaneous telomere-telomere fusion, but also participate in ionizing radiation induction Telomere-DSB fusion event. The association between accelerated telomere shortening and decreased expression of DDR genes, including *DNA-PKcs*, *Mre11*, *Xrcc4*, etc, was found in the accelerated thymic aging of a rat model of developmental programming (Tarry-Adkins et al.,

2019), suggesting DNA-PKcs may affect aging by maintaining telomeres integrity.

DNA-PKcs Functions in Cell Cycle Checkpoints

When DDR occurs, the PIKK kinases ATM, ATR and DNA-PKcs coordinately play the role of maintaining genome integrity (Blackford and Jackson, 2017). These kinases control the repair of the broken DNA ends and transmit the damage signal through the tumor suppressor p53, CHK1/CHK2 and so on to induce cell cycle arrest, apoptosis, or aging. In this process, DNA-PKcs works by affecting G₂/M DNA damage checkpoint in ATM deficient cells (Lee et al., 1997; Arlander et al., 2008; Tomimatsu et al., 2009; Shang et al., 2010). In ATM knockdown human mammary epithelial cells, either DNA-PK inhibitor treatment or RNAi knockdown of DNA-PKcs significantly attenuated G2 checkpoint (Arlander et al., 2008). In ATM-deficient AT5BIVA cells, DNA-PKcs inhibition led to a prolonged G₂/M arrest (Shang et al., 2010). This phenotype could also be an indirect effect of no repair due to DNA-PKcs inhibitory role in ATM phosphorylation in DDR (Zhou et al., 2017). Inactivation of DNA-PKcs strikingly attenuated the ionizing radiation-induced phosphorylation of Chk1 or Chk2/T68 in ATM-deficient cells.

Ataxia-Telangiectasia Mutated, ATR, and DNA-PKcs can directly phosphorylate p53. ATM and ATR phosphorylate p53 through checkpoint kinases 2 (CHK2) and checkpoint kinases 1 (CHK1), respectively (Roos and Kaina, 2006). p53 stability is mainly regulated by oncoprotein MDM2, a key negative regulator of p53 (Haupt et al., 1997; Kubbutat et al., 1997). Under non-stress conditions, MDM2 and p53 N-terminal domain form a stable complex. MDM2-mediated proteasome degradation keeps p53 at a low level.

Under DNA damage stresses, p53 and MDM2 dissociate and p53 accumulates in the nucleus. p53 functions as a transcription factor to promote the expression of genes involved in DNA repair, cell cycle arrest, apoptosis, and aging to maintain genome integrity. Achanta et al. (2001) found that the DNA-PKcs/Ku/p53 complex may act as a sensor for DNA strand breaks caused by the incorporation of drug molecules, and then transduce signals to trigger apoptosis (Wang et al., 2000). Boehme et al. (2008) found that the activation of DNA-PKcs led to the phosphorylation and activation of Akt/PKB, which subsequently led to the inactivation of GSK-3 β . As a result, MDM2 is phosphorylated and p53 accumulation increases. In this regard, DNA-PKcs modulates the p53-dependent apoptosis after DNA damage. At the same time, there are two negative feedback loops between p53 and MDM2 and wild-type p53-induced phosphatase 1 (Wip1) to regulate the level of p53 protein (Lu et al., 2007). Wip1 dephosphorylates MDM2 and downregulates p53 protein levels by stabilizing MDM2, facilitating its access to p53.

Targeting DNA-PKcs to Reprogram the Cellular Radiosensitivity

In the past century, radiation therapy has become the main treatment for cancer (Baskar et al., 2012). Statistically, about 50–70% of malignant tumor patients in the world are receiving radiotherapy (Huang and Zhou, 2020). However, the existence of

TABLE 2 | The commonly used DNA-PK inhibitors.

Inhibitor	Target	Cell line or animal model	IC50(DNA-PK)	Clinical Trial	References
NU7026	DNA-PK, PI3K	K562, ML1 FeBALB/C mice, HeLa	0.23 μ M		Nutley et al., 2005; Zimmermann and de Lange, 2014; Zhang et al., 2016
NU7441 (KU-57788)	DNA-PK, PI3K, mTOR	SW620, LoVo, V3-YAC, V3 cells, Ferude mice bearing SW620 xenografts	14 nM		Zhang and Matlashewski, 2019
LY3023414	PI3K, mTOR, DNA-PK	Athymic nude mice, CD-1 nude mice and NMRI athymic nude mice	4.24 nM	Phase 2	Zhou et al., 2017 Zhu et al., 2017
AZD-7648	DNA-PK	A549, H1299, BT-474, DLD1, FaDu, HCC70, HCC1806, HCC1937, HT-29, JEKO-1, MDA-MB-231, MDA-MB-436, MDA-MB-468, SKOV3, SUM149PT, TOV21G, UWB1.289	0.6 nM	Phase 1/2a	Zhao et al., 2006
M3814	DNA-PK	HCT116, FaDu, NCI-H460, A549, Capan-1, BxPC3	< 3 nM	Phase 1/2	Zhou and Paull, 2013
CC-115	mTOR, DNA-PK	Chronic lymphocytic leukemia (CLL) cells	13 nM	Phase 2	Zhu et al., 2009

radioresistance of cancer cells greatly hampers the efficacy and application of tumor radiotherapy (Kim et al., 2015). Radiation resistance is conducive to normal cells to escape the damage caused by ionizing radiation, but it is not conducive to the radiotherapy of malignant tumors.

The ability of cancer cells to repair DNA damage is a crucial factor that determines their sensitivity to radiation therapy or chemotherapy (Zhu et al., 2009). Inhibiting the repair ability of cancer cells offers a strategy to reduce the radiation resistance of cells and improve the efficacy of radiotherapy.

DNA-dependent protein kinase has become an attractive therapeutic target in various cancer treatments, especially when used in combination with genotoxic chemotherapy or ionizing radiation (Huang and Zhou, 2020; Medová et al., 2020). Studies have shown that targeting DNA-PKs with various inhibitors can effectively enhance radiotherapy, and many small molecule inhibitors of DNA-PKs have been developed and shown to be effective radiosensitizers *in vitro* (Table 2). One of the earliest identified inhibitors was wortmannin, which was isolated in 1957. It was defined as an effective non-competitive PI3K irreversible inhibitor in 1993, but it also targets other members of the PIKK family, including DNA-PKs, ATM, ATR (Ui et al., 1995). The DNA-PK inhibitor NU7026 was reported to potentiate topo II poisons in treatment of leukemia through inhibition of NHEJ and G₂/M checkpoint arrest (Willmore et al., 2004). In preclinical evaluation of DNA-PK inhibitor NU7441, it showed sufficient chemosensitization and radiosensitization with either etoposide, doxorubicin or ionizing radiation (Zhao et al., 2006). A highly selective DNA-PK inhibitor, AZD7648 showed efficient sensitization on radiation or doxorubicin in xenograft and patient-derived xenograft (PDX) models (Fok et al., 2019). M3814 inhibits DNA-PK catalytic activity and sensitizes various cancer cell lines to ionizing radiation (IR) and DSB inducers (Zenke et al., 2020). In addition, LY3023414 and CC-115 are both mTOR inhibitors, but they also have inhibitory effects on DNA-PKs and other PI3KK (Smith et al., 2016; Thijssen et al., 2016; Bendell et al., 2018). It can be envisioned that the inhibition of DNA-PK shows considerable prospects in anti-tumor resistance. However, DNA-PK inhibitors are usually

limited by poor pharmacokinetics: these compounds have poor solubility and unstable metabolism in the body, resulting in short serum half-life. Development of new compounds with better performance is a key step in obtaining effective anti-cancer drugs.

FUTURE PERSPECTIVES

Over the decades, DNA-PKs has been revealed to play multi-faceted roles and has been proved to be an essential regulator in the processes of DDR, however, there are still many unresolved questions. How does the interaction of DNA-PKs and Ku-DNA complex leads to activation of DNA-PK protein kinase activity? There are a large number of sites and forms of post-translation modifications (PTMs) on DNA-PKs, how these PTMs are inter-related and regulated, and how they influence DNA-PKs functions. Are there any other DNA-PK substrates yet to be discovered? It is worth to draw the panorama of function-related DNA-PKs interaction networks. The use of DNA-PK inhibitors for cancer therapy will not only inhibit the activity of DNA-PK in cancer cells, but also inhibit the activity of DNA-PK in normal cells. How to solve this problem? Ongoing exploration of DNA-PKs' multi-facet roles in DDR and besides will greatly contribute to our overall understanding of cancer and the discovery of new therapies.

AUTHOR CONTRIBUTIONS

P-KZ conceived and designed this study. XY drafted the initial manuscript. CB and DX reviewed and commented the manuscript. TM and P-KZ critically revised and finalized the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by grants from the National Natural Science Foundation of China (11705283 and 31870847).

REFERENCES

- Achanta, G., Pelicano, H., Feng, L., Plunkett, W., and Huang, P. (2001). Interaction of p53 and DNA-PK in response to nucleoside analogues: potential role as a sensor complex for DNA damage. *Cancer Res.* 61, 8723–8729.
- Allen, C., Ashley, A. K., Hromas, R., and Nickoloff, J. A. (2011). More forks on the road to replication stress recovery. *J. Mol. Cell Biol.* 3, 4–12. doi: 10.1093/jmcb/mjq049
- Anderson, C. W., and Lees-Miller, S. P. (1992). The nuclear serine/threonine protein kinase DNA-PK. *Crit. Rev. Eukaryot. Gene Expr.* 2, 283–314.
- Arlander, S. J., Greene, B. T., Innes, C. L., and Paules, R. S. (2008). DNA protein kinase-dependent G2 checkpoint revealed following knockdown of ataxia-telangiectasia mutated in human mammary epithelial cells. *Cancer Res.* 68, 89–97. doi: 10.1158/0008-5472.Can-07-0675
- Baskar, R., Lee, K. A., Yeo, R., and Yeoh, K. W. (2012). Cancer and radiation therapy: current advances and future directions. *Int. J. Med. Sci.* 9, 193–199. doi: 10.7150/ijms.3635
- Bass, K. L., Murray, J. M., and O'Connell, M. J. (2012). Brc1-dependent recovery from replication stress. *J. Cell Sci.* 125(Pt 11), 2753–2764. doi: 10.1242/jcs.103119
- Bendell, J. C., Varghese, A. M., Hyman, D. M., Bauer, T. M., Pant, S., Callies, S., et al. (2018). A first-in-human phase I study of LY3023414, an oral PI3K/mTOR dual inhibitor, in patients with advanced cancer. *Clin. Cancer Res.* 24, 3253–3262. doi: 10.1158/1078-0432.Ccr-17-3421
- Blackford, A. N., and Jackson, S. P. (2017). ATM, ATR, and DNA-PK: the trinity at the heart of the DNA damage response. *Mol. Cell* 66, 801–817. doi: 10.1016/j.molcel.2017.05.015
- Block, W. D., Yu, Y., Merkle, D., Gifford, J. L., Ding, Q., Meek, K., et al. (2004). Autophosphorylation-dependent remodeling of the DNA-dependent protein kinase catalytic subunit regulates ligation of DNA ends. *Nucleic Acids Res.* 32, 4351–4357. doi: 10.1093/nar/gkh761
- Blunt, T., Finnie, N. J., Taccioli, G. E., Smith, G. C., Demengeot, J., Gottlieb, T. M., et al. (1995). Defective DNA-dependent protein kinase activity is linked to V(D)J recombination and DNA repair defects associated with the murine scid mutation. *Cell* 80, 813–823. doi: 10.1016/0092-8674(95)90360-7
- Boehme, K. A., Kulikov, R., and Blattner, C. (2008). p53 stabilization in response to DNA damage requires Akt/PKB and DNA-PK. *Proc. Natl. Acad. Sci. U.S.A.* 105, 7785–7790. doi: 10.1073/pnas.0703423105
- Buisson, R., Boisvert, J. L., Benes, C. H., and Zou, L. (2015). Distinct but concerted roles of ATR, DNA-PK, and Chk1 in countering replication STRESS during S phase. *Mol. Cell* 59, 1011–1024. doi: 10.1016/j.molcel.2015.07.029
- Carter, T., Vancurová, I., Sun, I., Lou, W., and DeLeon, S. (1990). A DNA-activated protein kinase from HeLa cell nuclei. *Mol. Cell. Biol.* 10, 6460–6471. doi: 10.1128/mcb.10.12.6460
- Ceccaldi, R., Rondinelli, B., and D'Andrea, A. D. (2016). Repair pathway choices and consequences at the double-strand break. *Trends Cell Biol.* 26, 52–64. doi: 10.1016/j.tcb.2015.07.009
- Chang, H. H. Y., Pannunzio, N. R., Adachi, N., and Lieber, M. R. (2017). Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell Biol.* 18, 495–506. doi: 10.1038/nrm.2017.48
- Chang, H. H., and Lieber, M. R. (2016). Structure-specific nuclease activities of Artemis and the Artemis: DNA-PKcs complex. *Nucleic Acids Res.* 44, 4991–4997. doi: 10.1093/nar/gkw456
- Chatzidoukaki, O., Goulielmaki, E., Schumacher, B., and Garinis, G. A. (2020). DNA damage response and metabolic reprogramming in health and disease. *Trends Genet.* 36, 777–791. doi: 10.1016/j.tig.2020.06.018
- Chen, B. P., Uematsu, N., Kobayashi, J., Lerenthal, Y., Krempler, A., Yajima, H., et al. (2007). Ataxia telangiectasia mutated (ATM) is essential for DNA-PKcs phosphorylations at the Thr-2609 cluster upon DNA double strand break. *J. Biol. Chem.* 282, 6582–6587. doi: 10.1074/jbc.M611605200
- Chiruvella, K. K., Liang, Z., and Wilson, T. E. (2013). Repair of double-strand breaks by end joining. *Cold Spring Harb. Perspect. Biol.* 5:a012757. doi: 10.1101/cshperspect.a012757
- Chow, H. M., and Herrup, K. (2015). Genomic integrity and the ageing brain. *Nat. Rev. Neurosci.* 16, 672–684. doi: 10.1038/nrn4020
- Crowe, J. L., Shao, Z., Wang, X. S., Wei, P. C., Jiang, W., Lee, B. J., et al. (2018). Kinase-dependent structural role of DNA-PKcs during immunoglobulin class switch recombination. *Proc. Natl. Acad. Sci. U.S.A.* 115, 8615–8620. doi: 10.1073/pnas.1808490115
- Crowe, J. L., Wang, X. S., Shao, Z., Lee, B. J., Estes, V. M., and Zha, S. (2020). DNA-PKcs phosphorylation at the T2609 cluster alters the repair pathway choice during immunoglobulin class switch recombination. *Proc. Natl. Acad. Sci. U.S.A.* 117, 22953–22961. doi: 10.1073/pnas.2007455117
- Daido, S., Yamamoto, A., Fujiwara, K., Sawaya, R., Kondo, S., and Kondo, Y. (2005). Inhibition of the DNA-dependent protein kinase catalytic subunit radiosensitizes malignant glioma cells by inducing autophagy. *Cancer Res.* 65, 4368–4375. doi: 10.1158/0008-5472.Can-04-4202
- Daley, J. M., and Sung, P. (2014). 53BP1, BRCA1, and the choice between recombination and end joining at DNA double-strand breaks. *Mol. Cell. Biol.* 34, 1380–1388. doi: 10.1128/mcb.01639-13
- Davis, A. J., Chi, L., So, S., Lee, K. J., Mori, E., Fattah, K., et al. (2014). BRCA1 modulates the autophosphorylation status of DNA-PKcs in S phase of the cell cycle. *Nucleic Acids Res.* 42, 11487–11501. doi: 10.1093/nar/gku824
- Deshpande, R. A., Myler, L. R., Soniat, M. M., Makharashvili, N., Lee, L., Lees-Miller, S. P., et al. (2020). DNA-dependent protein kinase promotes DNA end processing by MRN and CtIP. *Sci. Adv.* 6:eay0922. doi: 10.1126/sciadv.aay0922
- Ding, Q., Reddy, Y. V., Wang, W., Woods, T., Douglas, P., Ramsden, D. A., et al. (2003). Autophosphorylation of the catalytic subunit of the DNA-dependent protein kinase is required for efficient end processing during DNA double-strand break repair. *Mol. Cell. Biol.* 23, 5836–5848. doi: 10.1128/mcb.23.16.5836-5848.2003
- Dobbs, T. A., Tainer, J. A., and Lees-Miller, S. P. (2010). A structural model for regulation of NHEJ by DNA-PKcs autophosphorylation. *DNA Repair (Amst.)* 9, 1307–1314. doi: 10.1016/j.dnarep.2010.09.019
- Douglas, P., Cui, X., Block, W. D., Yu, Y., Gupta, S., Ding, Q., et al. (2007). The DNA-dependent protein kinase catalytic subunit is phosphorylated in vivo on threonine 3950, a highly conserved amino acid in the protein kinase domain. *Mol. Cell. Biol.* 27, 1581–1591. doi: 10.1128/mcb.01962-06
- Douglas, P., Sapkota, G. P., Morrice, N., Yu, Y., Goodarzi, A. A., Merkle, D., et al. (2002). Identification of in vitro and in vivo phosphorylation sites in the catalytic subunit of the DNA-dependent protein kinase. *Biochem. J.* 368(Pt 1), 243–251. doi: 10.1042/bj20020973
- Douglas, P., Zhong, J., Ye, R., Moorhead, G. B., Xu, X., and Lees-Miller, S. P. (2010). Protein phosphatase 6 interacts with the DNA-dependent protein kinase catalytic subunit and dephosphorylates gamma-H2AX. *Mol. Cell. Biol.* 30, 1368–1381. doi: 10.1128/mcb.00741-09
- Falck, J., Coates, J., and Jackson, S. P. (2005). Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage. *Nature* 434, 605–611. doi: 10.1038/nature03442
- Fang, E. F., Scheibye-Knudsen, M., Chua, K. F., Mattson, M. P., Croteau, D. L., and Bohr, V. A. (2016). Nuclear DNA damage signalling to mitochondria in ageing. *Nat. Rev. Mol. Cell Biol.* 17, 308–321. doi: 10.1038/nrm.2016.14
- Featherstone, C., and Jackson, S. P. (1999). Ku, a DNA repair protein with multiple cellular functions? *Mutat. Res.* 434, 3–15. doi: 10.1016/s0921-8777(99)00006-3
- Fok, J. H. L., Ramos-Montoya, A., Vazquez-Chantada, M., Wijnhoven, P. W. G., Follia, V., James, N., et al. (2019). AZD7648 is a potent and selective DNA-PK inhibitor that enhances radiation, chemotherapy and olaparib activity. *Nat. Commun.* 10:5065. doi: 10.1038/s41467-019-12836-9
- Gao, S. S., Guan, H., Yan, S., Hu, S., Song, M., Guo, Z. P., et al. (2020). TIP60 K430 SUMOylation attenuates its interaction with DNA-PKcs in S-phase cells: facilitating homologous recombination and emerging target for cancer therapy. *Sci. Adv.* 6:eaba7822. doi: 10.1126/sciadv.aba7822
- Gomes, L. R., Menck, C. F. M., and Leandro, G. S. (2017). Autophagy roles in the modulation of DNA repair pathways. *Int. J. Mol. Sci.* 18:2351. doi: 10.3390/ijms18112351
- Goodarzi, A. A., Yu, Y., Riballo, E., Douglas, P., Walker, S. A., Ye, R., et al. (2006). DNA-PK autophosphorylation facilitates Artemis endonuclease activity. *Embo J.* 25, 3880–3889. doi: 10.1038/sj.emboj.7601255
- Goytisolo, F. A., Samper, E., Edmonson, S., Taccioli, G. E., and Blasco, M. A. (2001). The absence of the DNA-dependent protein kinase catalytic subunit in mice results in anaphase bridges and in increased telomeric fusions with normal telomere length and G-strand overhang. *Mol. Cell. Biol.* 21, 3642–3651. doi: 10.1128/mcb.21.11.3642-3651.2001

- Guo, Z., Wang, S., Xie, Y., Han, Y., Hu, S., Guan, H., et al. (2020). HUWE1-dependent DNA-PKcs neddylation modulates its autophosphorylation in DNA damage response. *Cell Death Dis.* 11:400. doi: 10.1038/s41419-020-2611-0
- Han, Y., Jin, F., Xie, Y., Liu, Y., Hu, S., Liu, X. D., et al. (2019). DNA-PKcs PARylation regulates DNA-PK kinase activity in the DNA damage response. *Mol. Med. Rep.* 20, 3609–3616. doi: 10.3892/mmr.2019.10640
- Hartley, K. O., Gell, D., Smith, G. C., Zhang, H., Divecha, N., Connelly, M. A., et al. (1995). DNA-dependent protein kinase catalytic subunit: a relative of phosphatidylinositol 3-kinase and the ataxia telangiectasia gene product. *Cell* 82, 849–856. doi: 10.1016/0092-8674(95)90482-4
- Haupt, Y., Maya, R., Kazaz, A., and Oren, M. (1997). Mdm2 promotes the rapid degradation of p53. *Nature* 387, 296–299. doi: 10.1038/387296a0
- Her, J., and Bunting, S. F. (2018). How cells ensure correct repair of DNA double-strand breaks. *J. Biol. Chem.* 293, 10502–10511. doi: 10.1074/jbc.TM118.000371
- Hewitt, G., and Korolchuk, V. I. (2017). Repair, reuse, recycle: the expanding role of autophagy in genome maintenance. *Trends Cell Biol.* 27, 340–351. doi: 10.1016/j.tcb.2016.11.011
- Ho, S. R., Mahanic, C. S., Lee, Y. J., and Lin, W. C. (2014). RNF144A, an E3 ubiquitin ligase for DNA-PKcs, promotes apoptosis during DNA damage. *Proc. Natl. Acad. Sci. U.S.A.* 111, E2646–E2655. doi: 10.1073/pnas.1323107111
- Huang, R. X., and Zhou, P. K. (2020). DNA damage response signaling pathways and targets for radiotherapy sensitization in cancer. *Signal Transduct. Target. Ther.* 5:60. doi: 10.1038/s41392-020-0150-x
- Jackson, S. P., and Bartek, J. (2009). The DNA-damage response in human biology and disease. *Nature* 461, 1071–1078. doi: 10.1038/nature08467
- Javvadi, P., Makino, H., Das, A. K., Lin, Y. F., Chen, D. J., Chen, B. P., et al. (2012). Threonine 2609 phosphorylation of the DNA-dependent protein kinase is a critical prerequisite for epidermal growth factor receptor-mediated radiation resistance. *Mol. Cancer Res.* 10, 1359–1368. doi: 10.1158/1541-7786.Mcr-12-0482-t
- Jeggo, P. A., Pearl, L. H., and Carr, A. M. (2016). DNA repair, genome stability and cancer: a historical perspective. *Nat. Rev. Cancer* 16, 35–42. doi: 10.1038/nrc.2015.4
- Jiang, W., Crowe, J. L., Liu, X., Nakajima, S., Wang, Y., Li, C., et al. (2015). Differential phosphorylation of DNA-PKcs regulates the interplay between end-processing and end-ligation during nonhomologous end-joining. *Mol. Cell* 58, 172–185. doi: 10.1016/j.molcel.2015.02.024
- Jiang, W., Estes, V. M., Wang, X. S., Shao, Z., Lee, B. J., Lin, X., et al. (2019). Phosphorylation at S2053 in murine (S2056 in human) DNA-PKcs is dispensable for lymphocyte development and class switch recombination. *J. Immunol.* 203, 178–187. doi: 10.4049/jimmunol.1801657
- Jiang, X., Sun, Y., Chen, S., Roy, K., and Price, B. D. (2006). The FATC domains of PIKK proteins are functionally equivalent and participate in the Tip60-dependent activation of DNA-PKcs and ATM. *J. Biol. Chem.* 281, 15741–15746. doi: 10.1074/jbc.M513172200
- Jin, S., and Weaver, D. T. (1997). Double-strand break repair by Ku70 requires heterodimerization with Ku80 and DNA binding functions. *Embo J.* 16, 6874–6885. doi: 10.1093/emboj/16.22.6874
- Joshi, R. R., Ali, S. I., and Ashley, A. K. (2019). DNA ligase IV prevents replication fork stalling and promotes cellular proliferation in triple negative breast cancer. *J. Nucleic Acids* 2019:9170341. doi: 10.1155/2019/9170341
- Kamoi, K., and Mochizuki, M. (2010). Phaco forward-chop technique for managing posterior nuclear plate of hard cataract. *J. Cataract Refract. Surg.* 36, 9–12. doi: 10.1016/j.jcrs.2009.07.047
- Kantidze, O. L., Velichko, A. K., Luzhin, A. V., Petrova, N. V., and Razin, S. V. (2018). Synthetically lethal interactions of ATM, ATR, and DNA-PKcs. *Trends Cancer* 4, 755–768. doi: 10.1016/j.trecan.2018.09.007
- Kim, B. M., Hong, Y., Lee, S., Liu, P., Lim, J. H., Lee, Y. H., et al. (2015). Therapeutic implications for overcoming radiation resistance in cancer therapy. *Int. J. Mol. Sci.* 16, 26880–26913. doi: 10.3390/ijms161125991
- Kim, S. T., Lim, D. S., Canman, C. E., and Kastan, M. B. (1999). Substrate specificities and identification of putative substrates of ATM kinase family members. *J. Biol. Chem.* 274, 37538–37543. doi: 10.1074/jbc.274.53.37538
- Konstantinopoulos, P. A., Ceccaldi, R., Shapiro, G. I., and D'Andrea, A. D. (2015). Homologous recombination deficiency: exploiting the fundamental vulnerability of ovarian cancer. *Cancer Discov.* 5, 1137–1154. doi: 10.1158/2159-8290.Cd-15-0714
- Kotsantis, P., Petermann, E., and Boulton, S. J. (2018). Mechanisms of oncogene-induced replication stress: jigsaw falling into place. *Cancer Discov.* 8, 537–555. doi: 10.1158/2159-8290.Cd-17-1461
- Kubbutat, M. H., Jones, S. N., and Vousden, K. H. (1997). Regulation of p53 stability by Mdm2. *Nature* 387, 299–303. doi: 10.1038/387299a0
- Lazzerini-Denchi, E., and Sfeir, A. (2016). Stop pulling my strings – what telomeres taught us about the DNA damage response. *Nat. Rev. Mol. Cell Biol.* 17, 364–378. doi: 10.1038/nrm.2016.43
- Lee, S. E., Mitchell, R. A., Cheng, A., and Hendrickson, E. A. (1997). Evidence for DNA-PK-dependent and -independent DNA double-strand break repair pathways in mammalian cells as a function of the cell cycle. *Mol. Cell Biol.* 17, 1425–1433. doi: 10.1128/mcb.17.3.1425
- Lees-Miller, S. P. (1996). The DNA-dependent protein kinase, DNA-PK: 10 years and no ends in sight. *Biochem. Cell Biol.* 74, 503–512. doi: 10.1139/o96-054
- Lees-Miller, S. P., Chen, Y. R., and Anderson, C. W. (1990). Human cells contain a DNA-activated protein kinase that phosphorylates simian virus 40 T antigen, mouse p53, and the human Ku autoantigen. *Mol. Cell Biol.* 10, 6472–6481. doi: 10.1128/mcb.10.12.6472
- Leidal, A. M., Levine, B., and Debnath, J. (2018). Autophagy and the cell biology of age-related disease. *Nat. Cell Biol.* 20, 1338–1348. doi: 10.1038/s41556-018-0235-8
- Liccardi, G., Hartley, J. A., and Hochhauser, D. (2011). EGFR nuclear translocation modulates DNA repair following cisplatin and ionizing radiation treatment. *Cancer Res.* 71, 1103–1114. doi: 10.1158/0008-5472.Can-10-2384
- Lieber, M. R. (2008). The mechanism of human nonhomologous DNA end joining. *J. Biol. Chem.* 283, 1–5. doi: 10.1074/jbc.R700039200
- Lin, Y. F., Shih, H. Y., Shang, Z. F., Kuo, C. T., Guo, J., Du, C., et al. (2018). PIDD mediates the association of DNA-PKcs and ATR at stalled replication forks to facilitate the ATR signaling pathway. *Nucleic Acids Res.* 46, 1847–1859. doi: 10.1093/nar/gkx1298
- Lin, Y. F., Shih, H. Y., Shang, Z., Matsunaga, S., and Chen, B. P. (2014). DNA-PKcs is required to maintain stability of Chk1 and Claspin for optimal replication stress response. *Nucleic Acids Res.* 42, 4463–4473. doi: 10.1093/nar/gku116
- Liu, S., Opiyo, S. O., Manthey, K., Glanzer, J. G., Ashley, A. K., Amerin, C., et al. (2012). Distinct roles for DNA-PK, ATM and ATR in RPA phosphorylation and checkpoint activation in response to replication stress. *Nucleic Acids Res.* 40, 10780–10794. doi: 10.1093/nar/gks849
- Longerich, S., Li, J., Xiong, Y., Sung, P., and Kupfer, G. M. (2014). Stress and DNA repair biology of the Fanconi anemia pathway. *Blood* 124, 2812–2819. doi: 10.1182/blood-2014-04-526293
- Lu, X., Ma, O., Nguyen, T. A., Jones, S. N., Oren, M., and Donehower, L. A. (2007). The Wip1 Phosphatase acts as a gatekeeper in the p53-Mdm2 autoregulatory loop. *Cancer Cell* 12, 342–354. doi: 10.1016/j.ccr.2007.08.033
- Ma, Y., Pannicke, U., Schwarz, K., and Lieber, M. R. (2002). Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* 108, 781–794. doi: 10.1016/s0092-8674(02)00671-2
- Madabhushi, R., Pan, L., and Tsai, L. H. (2014). DNA damage and its links to neurodegeneration. *Neuron* 83, 266–282. doi: 10.1016/j.neuron.2014.06.034
- Mao, Z., Bozzella, M., Seluanov, A., and Gorbunova, V. (2008). DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. *Cell Cycle* 7, 2902–2906. doi: 10.4161/cc.7.18.6679
- Medová, M., Medo, M., Hovhannisyan, L., Maldonado, C. M., Aebbersold, D. M., and Zimmer, Y. (2020). DNA-PK in human malignant disorders: mechanisms and implications for pharmacological interventions. *Pharmacol. Ther.* 215:107617. doi: 10.1016/j.pharmthera.2020.107617
- Meek, K., Lees-Miller, S. P., and Modesti, M. (2012). N-terminal constraint activates the catalytic subunit of the DNA-dependent protein kinase in the absence of DNA or Ku. *Nucleic Acids Res.* 40, 2964–2973. doi: 10.1093/nar/gkr1211
- Menolfi, D., and Zha, S. (2020). ATM, ATR and DNA-PKcs kinases-the lessons from the mouse models: inhibition \neq deletion. *Cell Biosci.* 10:8. doi: 10.1186/s13578-020-0376-x
- Menon, V., and Povirk, L. F. (2016). End-processing nucleases and phosphodiesterases: an elite supporting cast for the non-homologous end joining pathway of DNA double-strand break repair. *DNA Repair (Amst.)* 43, 57–68. doi: 10.1016/j.dnarep.2016.05.011

- Miller, R. D., Hogg, J., Ozaki, J. H., Gell, D., Jackson, S. P., and Riblet, R. (1995). Gene for the catalytic subunit of mouse DNA-dependent protein kinase maps to the scid locus. *Proc. Natl. Acad. Sci. U.S.A.* 92, 10792–10795. doi: 10.1073/pnas.92.23.10792
- Mirman, Z., and de Lange, T. (2020). 53BP1: a DSB escort. *Genes Dev.* 34, 7–23. doi: 10.1101/gad.333237.119
- Mori, E., Davis, A. J., Hasegawa, M., and Chen, D. J. (2016). Lysines 3241 and 3260 of DNA-PKcs are important for genomic stability and radioresistance. *Biochem. Biophys. Res. Commun.* 477, 235–240. doi: 10.1016/j.bbrc.2016.06.048
- Morris, E. P., Rivera-Calzada, A., da Fonseca, P. C., Llorca, O., Pearl, L. H., and Spagnolo, L. (2011). Evidence for a remodelling of DNA-PK upon autophosphorylation from electron microscopy studies. *Nucleic Acids Res.* 39, 5757–5767. doi: 10.1093/nar/gkr146
- Nagasawa, H., Lin, Y. F., Kato, T. A., Brogan, J. R., Shih, H. Y., Kurimasa, A., et al. (2017). Inhibition of the Ser2056 and Thr2609 clusters of DNA-PKcs in regulating gamma rays and extremely low fluencies of alpha-particle irradiation to G(0)/G(1) phase cells. *Radiat. Res.* 187, 259–267. doi: 10.1667/rr14679.1
- Nagasawa, H., Little, J. B., Lin, Y. F., So, S., Kurimasa, A., Peng, Y., et al. (2011). Differential role of DNA-PKcs phosphorylations and kinase activity in radiosensitivity and chromosomal instability. *Radiat. Res.* 175, 83–89. doi: 10.1667/rr2092.1
- Neal, J. A., Dang, V., Douglas, P., Wold, M. S., Lees-Miller, S. P., and Meek, K. (2011). Inhibition of homologous recombination by DNA-dependent protein kinase requires kinase activity, is titratable, and is modulated by autophosphorylation. *Mol. Cell. Biol.* 31, 1719–1733. doi: 10.1128/mcb.01298-10
- Nutley, B. P., Smith, N. F., Hayes, A., Kelland, L. R., Brunton, L., Golding, B. T., et al. (2005). Preclinical pharmacokinetics and metabolism of a novel prototype DNA-PK inhibitor NU7026. *Br. J. Cancer* 93, 1011–1018. doi: 10.1038/sj.bjc.6602823
- Pannunzio, N. R., Watanabe, G., and Lieber, M. R. (2018). Nonhomologous DNA end-joining for repair of DNA double-strand breaks. *J. Biol. Chem.* 293, 10512–10523. doi: 10.1074/jbc.TM117.000374
- Puustinen, P., Keldsbo, A., Corcelle-Termeau, E., Ngoei, K., Sønder, S. L., Farkas, T., et al. (2020). DNA-dependent protein kinase regulates lysosomal AMP-dependent protein kinase activation and autophagy. *Autophagy* 16, 1871–1888. doi: 10.1080/15548627.2019.1710430
- Reichert, S., Rödel, C., Mirsch, J., Harter, P. N., Tomicic, M. T., Mittelbronn, M., et al. (2011). Survivin inhibition and DNA double-strand break repair: a molecular mechanism to overcome radioresistance in glioblastoma. *Radiother. Oncol.* 101, 51–58. doi: 10.1016/j.radonc.2011.06.037
- Rivera-Calzada, A., Maman, J. D., Spagnolo, L., Pearl, L. H., and Llorca, O. (2005). Three-dimensional structure and regulation of the DNA-dependent protein kinase catalytic subunit (DNA-PKcs). *Structure* 13, 243–255. doi: 10.1016/j.str.2004.12.006
- Roake, C. M., and Artandi, S. E. (2020). Regulation of human telomerase in homeostasis and disease. *Nat. Rev. Mol. Cell Biol.* 21, 384–397. doi: 10.1038/s41580-020-0234-z
- Roos, W. P., and Kaina, B. (2006). DNA damage-induced cell death by apoptosis. *Trends Mol. Med.* 12, 440–450. doi: 10.1016/j.molmed.2006.07.007
- Sallmyr, A., and Tomkinson, A. E. (2018). Repair of DNA double-strand breaks by mammalian alternative end-joining pathways. *J. Biol. Chem.* 293, 10536–10546. doi: 10.1074/jbc.TM117.000375
- Scully, R., Panday, A., Elango, R., and Willis, N. A. (2019). DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat. Rev. Mol. Cell Biol.* 20, 698–714. doi: 10.1038/s41580-019-0152-0
- Shang, Z. F., Huang, B., Xu, Q. Z., Zhang, S. M., Fan, R., Liu, X. D., et al. (2010). Inactivation of DNA-dependent protein kinase leads to spindle disruption and mitotic catastrophe with attenuated checkpoint protein 2 Phosphorylation in response to DNA damage. *Cancer Res.* 70, 3657–3666. doi: 10.1158/0008-5472.Can-09-3362
- Sharif, H., Li, Y., Dong, Y., Dong, L., Wang, W. L., Mao, Y., et al. (2017). Cryo-EM structure of the DNA-PK holoenzyme. *Proc. Natl. Acad. Sci. U.S.A.* 114, 7367–7372. doi: 10.1073/pnas.1707386114
- Shibata, A., and Jeggo, P. A. (2020). Canonical DNA non-homologous end-joining: capacity versus fidelity. *Br. J. Radiol.* 93:20190966. doi: 10.1259/bjr.20190966
- Shibata, A., Conrad, S., Birraux, J., Geuting, V., Barton, O., Ismail, A., et al. (2011). Factors determining DNA double-strand break repair pathway choice in G2 phase. *Embo J.* 30, 1079–1092. doi: 10.1038/emboj.2011.27
- Shimizu, I., Yoshida, Y., Suda, M., and Minamino, T. (2014). DNA damage response and metabolic disease. *Cell Metab.* 20, 967–977. doi: 10.1016/j.cmet.2014.10.008
- Sibanda, B. L., Chirgadze, D. Y., and Blundell, T. L. (2010). Crystal structure of DNA-PKcs reveals a large open-ring cradle comprised of HEAT repeats. *Nature* 463, 118–121. doi: 10.1038/nature08648
- Sibanda, B. L., Chirgadze, D. Y., Ascher, D. B., and Blundell, T. L. (2017). DNA-PKcs structure suggests an allosteric mechanism modulating DNA double-strand break repair. *Science* 355, 520–524. doi: 10.1126/science.aak9654
- Singleton, B. K., Torres-Arzuayus, M. I., Rottinghaus, S. T., Taccioli, G. E., and Jeggo, P. A. (1999). The C terminus of Ku80 activates the DNA-dependent protein kinase catalytic subunit. *Mol. Cell. Biol.* 19, 3267–3277. doi: 10.1128/mcb.19.5.3267
- Sipley, J. D., Menninger, J. C., Hartley, K. O., Ward, D. C., Jackson, S. P., and Anderson, C. W. (1995). Gene for the catalytic subunit of the human DNA-activated protein kinase maps to the site of the XRCC7 gene on chromosome 8. *Proc. Natl. Acad. Sci. U.S.A.* 92, 7515–7519. doi: 10.1073/pnas.92.16.7515
- Smith, M. C., Mader, M. M., Cook, J. A., Iversen, P., Ajamie, R., Perkins, E., et al. (2016). Characterization of LY3023414, a novel PI3K/mTOR dual inhibitor eliciting transient target modulation to impede tumor growth. *Mol. Cancer Ther.* 15, 2344–2356. doi: 10.1158/1535-7163.Mct-15-0996
- Spagnolo, L., Rivera-Calzada, A., Pearl, L. H., and Llorca, O. (2006). Three-dimensional structure of the human DNA-PKcs/Ku70/Ku80 complex assembled on DNA and its implications for DNA DSB repair. *Mol. Cell* 22, 511–519. doi: 10.1016/j.molcel.2006.04.013
- Sui, J., Lin, Y. F., Xu, K., Lee, K. J., Wang, D., and Chen, B. P. (2015). DNA-PKcs phosphorylates hnRNP-A1 to facilitate the RPA-to-POT1 switch and telomere capping after replication. *Nucleic Acids Res.* 43, 5971–5983. doi: 10.1093/nar/gkv539
- Sui, J., Zhang, S., and Chen, B. P. C. (2020). DNA-dependent protein kinase in telomere maintenance and protection. *Cell. Mol. Biol. Lett.* 25:2. doi: 10.1186/s11658-020-0199-0
- Sun, Y., McCorvie, T. J., Yates, L. A., and Zhang, X. (2020). Structural basis of homologous recombination. *Cell. Mol. Life Sci.* 77, 3–18. doi: 10.1007/s00018-019-03365-1
- Syed, A., and Tainer, J. A. (2018). The MRE11-RAD50-NBS1 complex conducts the orchestration of damage signaling and outcomes to stress in DNA replication and repair. *Annu. Rev. Biochem.* 87, 263–294. doi: 10.1146/annurev-biochem-062917-012415
- Tarry-Adkins, J. L., Aiken, C. E., Ashmore, T. J., Fernandez-Twinn, D. S., Chen, J. H., and Ozanne, S. E. (2019). A suboptimal maternal diet combined with accelerated postnatal growth results in an altered aging profile in the thymus of male rats. *Faseb J.* 33, 239–253. doi: 10.1096/fj.201701350RR
- Thijssen, R., Ter Burg, J., Garrick, B., van Bochove, G. G., Brown, J. R., Fernandes, S. M., et al. (2016). Dual TORC/DNA-PK inhibition blocks critical signaling pathways in chronic lymphocytic leukemia. *Blood* 128, 574–583. doi: 10.1182/blood-2016-02-700328
- Ting, N. S., Pohorelec, B., Yu, Y., Lees-Miller, S. P., and Beattie, T. L. (2009). The human telomerase RNA component, hTR, activates the DNA-dependent protein kinase to phosphorylate heterogeneous nuclear ribonucleoprotein A1. *Nucleic Acids Res.* 37, 6105–6115. doi: 10.1093/nar/gkp636
- Tomimatsu, N., Mukherjee, B., and Burma, S. (2009). Distinct roles of ATR and DNA-PKcs in triggering DNA damage responses in ATM-deficient cells. *EMBO Rep.* 10, 629–635. doi: 10.1038/embor.2009.60
- Ubhi, T., and Brown, G. W. (2019). Exploiting DNA replication stress for cancer treatment. *Cancer Res.* 79, 1730–1739. doi: 10.1158/0008-5472.Can-18-3631
- Ui, M., Okada, T., Hazeki, K., and Hazeki, O. (1995). Wortmannin as a unique probe for an intracellular signalling protein, phosphoinositide 3-kinase. *Trends Biochem. Sci.* 20, 303–307. doi: 10.1016/s0968-0004(00)89056-8
- Uryga, A., Gray, K., and Bennett, M. (2016). DNA damage and repair in vascular disease. *Annu. Rev. Physiol.* 78, 45–66. doi: 10.1146/annurev-physiol-021115-105127
- van der Burg, M., Ijspeert, H., Verkaik, N. S., Turul, T., Wiegant, W. W., Morotomi-Yano, K., et al. (2009). A DNA-PKcs mutation in a radiosensitive T-B- SCID

- patient inhibits Artemis activation and nonhomologous end-joining. *J. Clin. Invest.* 119, 91–98. doi: 10.1172/jci37141
- Victorelli, S., and Passos, J. F. (2017). Telomeres and cell senescence – size matters not. *EBioMedicine* 21, 14–20. doi: 10.1016/j.ebiom.2017.03.027
- Vijg, J., and Suh, Y. (2013). Genome instability and aging. *Annu. Rev. Physiol.* 75, 645–668. doi: 10.1146/annurev-physiol-030212-183715
- Villafanez, F., Garcia, I. A., Carbajosa, S., Pansa, M. F., Mansilla, S., Llorens, M. C., et al. (2019). AKT inhibition impairs PCNA ubiquitylation and triggers synthetic lethality in homologous recombination-deficient cells submitted to replication stress. *Oncogene* 38, 4310–4324. doi: 10.1038/s41388-019-0724-7
- Walker, A. I., Hunt, T., Jackson, R. J., and Anderson, C. W. (1985). Double-stranded DNA induces the phosphorylation of several proteins including the 90 000 mol. wt. heat-shock protein in animal cell extracts. *Embo J.* 4, 139–145. doi: 10.1002/j.1460-2075.1985.tb02328.x
- Wang, C. Y., Huang, E. Y., Huang, S. C., and Chung, B. C. (2015). DNA-PK/Chk2 induces centrosome amplification during prolonged replication stress. *Oncogene* 34, 1263–1269. doi: 10.1038/onc.2014.74
- Wang, S., Guo, M., Ouyang, H., Li, X., Cordon-Cardo, C., Kurimasa, A., et al. (2000). The catalytic subunit of DNA-dependent protein kinase selectively regulates p53-dependent apoptosis but not cell-cycle arrest. *Proc. Natl. Acad. Sci. U.S.A.* 97, 1584–1588. doi: 10.1073/pnas.97.4.1584
- Willmore, E., de Caux, S., Sunter, N. J., Tilby, M. J., Jackson, G. H., Austin, C. A., et al. (2004). A novel DNA-dependent protein kinase inhibitor, NU7026, potentiates the cytotoxicity of topoisomerase II poisons used in the treatment of leukemia. *Blood* 103, 4659–4665. doi: 10.1182/blood-2003-07-2527
- Woodbine, L., Neal, J. A., Sasi, N. K., Shimada, M., Deem, K., Coleman, H., et al. (2013). PRKDC mutations in a SCID patient with profound neurological abnormalities. *J. Clin. Invest.* 123, 2969–2980. doi: 10.1172/jci67349
- Xie, Y., Liu, Y. K., Guo, Z. P., Guan, H., Liu, X. D., Xie, D. F., et al. (2020). RBX1 prompts degradation of EXO1 to limit the homologous recombination pathway of DNA double-strand break repair in G1 phase. *Cell Death Differ.* 27, 1383–1397. doi: 10.1038/s41418-019-0424-4
- Yan, Y. Q., Zhang, B., Wang, L., Xie, Y. H., Peng, T., Bai, B., et al. (2007). Induction of apoptosis and autophagic cell death by the vanillin derivative 6-bromine-5-hydroxy-4-methoxybenzaldehyde is accompanied by the cleavage of DNA-PKcs and rapid destruction of c-Myc oncoprotein in HepG2 cells. *Cancer Lett.* 252, 280–289. doi: 10.1016/j.canlet.2007.01.007
- Yin, X., Liu, M., Tian, Y., Wang, J., and Xu, Y. (2017). Cryo-EM structure of human DNA-PK holoenzyme. *Cell Res.* 27, 1341–1350. doi: 10.1038/cr.2017.110
- Ying, S., Chen, Z., Medhurst, A. L., Neal, J. A., Bao, Z., Mortusewicz, O., et al. (2016). DNA-PKcs and PARP1 bind to unresected stalled DNA replication forks where they recruit XRCC1 to mediate repair. *Cancer Res.* 76, 1078–1088. doi: 10.1158/0008-5472.Can-15-0608
- Zeman, M. K., and Cimprich, K. A. (2014). Causes and consequences of replication stress. *Nat. Cell Biol.* 16, 2–9. doi: 10.1038/ncb2897
- Zenke, F. T., Zimmermann, A., Sirrenberg, C., Dahmen, H., Kirkin, V., Pehl, U., et al. (2020). Pharmacologic inhibitor of DNA-PK, M3814, potentiates radiotherapy and regresses human tumors in mouse models. *Mol. Cancer Ther.* 19, 1091–1101. doi: 10.1158/1535-7163.Mct-19-0734
- Zhang, S., Matsunaga, S., Lin, Y. F., Sishc, B., Shang, Z., Sui, J., et al. (2016). Spontaneous tumor development in bone marrow-rescued DNA-PKcs(3A/3A) mice due to dysfunction of telomere leading strand deprotection. *Oncogene* 35, 3909–3918. doi: 10.1038/onc.2015.459
- Zhang, W. W., and Matlaszewski, G. (2019). Single-strand annealing plays a major role in double-strand DNA break repair FOLLOWING CRISPR-Cas9 cleavage in *Leishmania*. *mSphere* 4:e00408–19. doi: 10.1128/mSphere.00408-19
- Zhao, Y., Thomas, H. D., Batey, M. A., Cowell, I. G., Richardson, C. J., Griffin, R. J., et al. (2006). Preclinical evaluation of a potent novel DNA-dependent protein kinase inhibitor NU7441. *Cancer Res.* 66, 5354–5362. doi: 10.1158/0008-5472.Can-05-4275
- Zhou, Y., and Paull, T. T. (2013). DNA-dependent protein kinase regulates DNA end resection in concert with Mre11-Rad50-Nbs1 (MRN) and ataxia telangiectasia-mutated (ATM). *J. Biol. Chem.* 288, 37112–37125. doi: 10.1074/jbc.M113.514398
- Zhou, Y., Lee, J. H., Jiang, W., Crowe, J. L., Zha, S., and Paull, T. T. (2017). Regulation of the DNA damage response by DNA-PKcs inhibitory phosphorylation of ATM. *Mol. Cell* 65, 91–104. doi: 10.1016/j.molcel.2016.11.004
- Zhu, S., Fisher, L. A., Bessho, T., and Peng, A. (2017). Protein phosphatase 1 and phosphatase 1 nuclear targeting subunit-dependent regulation of DNA-dependent protein kinase and non-homologous end joining. *Nucleic Acids Res.* 45, 10583–10594. doi: 10.1093/nar/gkx686
- Zhu, Y., Hu, J., Hu, Y., and Liu, W. (2009). Targeting DNA repair pathways: a novel approach to reduce cancer therapeutic resistance. *Cancer Treat. Rev.* 35, 590–596. doi: 10.1016/j.ctrv.2009.06.005
- Zimmermann, M., and de Lange, T. (2014). 53BP1: pro choice in DNA repair. *Trends Cell Biol.* 24, 108–117. doi: 10.1016/j.tcb.2013.09.003

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Yue, Bai, Xie, Ma and Zhou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



DNA Repair Gene Polymorphisms and Chromosomal Aberrations in Exposed Populations

Yasmeen Niazi^{1,2,3*}, Hauke Thomsen^{1,4}, Bozena Smolkova⁵, Ludmila Vodickova^{6,7,8}, Sona Vodenkova⁶, Michal Kroupa^{6,8}, Veronika Vymetalkova^{6,7,8}, Alena Kazimirova⁹, Magdalena Barancokova⁹, Katarina Volkovova⁹, Marta Staruchova⁹, Per Hoffmann^{10,11}, Markus M. Nöthen¹⁰, Maria Dusinska¹², Ludovit Musak¹³, Pavel Vodicka^{6,7,8}, Kari Hemminki^{1,8,14*†} and Asta Försti^{1,2,3†}

OPEN ACCESS

Edited by:

Stefano Gnan,
Institut Curie, France

Reviewed by:

Vladimir Druzhinin,
Kemerovo State University, Russia
Alexander V. Rubanovich,
Vavilov Institute of General Genetics,
Russian Academy of Sciences, Russia

*Correspondence:

Yasmeen Niazi
y.niazi@kitz-heidelberg.de
Kari Hemminki
K.Hemminki@dkfz-heidelberg.de

†These authors share senior
authorship

Specialty section:

This article was submitted to
Human and Medical Genomics,
a section of the journal
Frontiers in Genetics

Received: 07 April 2021

Accepted: 30 April 2021

Published: 16 June 2021

Citation:

Niazi Y, Thomsen H, Smolkova B,
Vodickova L, Vodenkova S,
Kroupa M, Vymetalkova V,
Kazimirova A, Barancokova M,
Volkovova K, Staruchova M,
Hoffmann P, Nöthen MM, Dusinska M,
Musak L, Vodicka P, Hemminki K and
Försti A (2021) DNA Repair Gene
Polymorphisms and Chromosomal
Aberrations in Exposed Populations.
Front. Genet. 12:691947.
doi: 10.3389/fgene.2021.691947

¹ Department of Molecular Genetic Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany, ² Hopp Children's Cancer Center (KITZ), Heidelberg, Germany, ³ Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), German Cancer Consortium (DKTK), Heidelberg, Germany, ⁴ GeneWerk GmbH, Heidelberg, Germany, ⁵ Department of Molecular Oncology, Cancer Research Institute, Biomedical Research Center of the Slovak Academy of Sciences, Bratislava, Slovakia, ⁶ Department of Molecular Biology of Cancer, Institute of Experimental Medicine, Czech Academy of Sciences, Prague, Czechia, ⁷ First Faculty of Medicine, Institute of Biology and Medical Genetics, Charles University, Prague, Czechia, ⁸ Faculty of Medicine and Biomedical Center in Pilsen, Charles University in Prague, Prague, Czechia, ⁹ Department of Biology, Faculty of Medicine, Slovak Medical University, Bratislava, Slovakia, ¹⁰ Institute of Human Genetics, School of Medicine and University Hospital Bonn, University of Bonn, Bonn, Germany, ¹¹ Division of Medical Genetics, Department of Biomedicine, University of Basel, Basel, Switzerland, ¹² Health Effects Laboratory, Department of Environmental Chemistry, NILU-Norwegian Institute for Air Research, Kjeller, Norway, ¹³ Jessenius Faculty of Medicine, Biomedical Center Martin, Comenius University in Bratislava, Bratislava, Slovakia, ¹⁴ Division of Cancer Epidemiology, German Cancer Research Centre (DKFZ), Heidelberg, Germany

DNA damage and unrepaired or insufficiently repaired DNA double-strand breaks as well as telomere shortening contribute to the formation of structural chromosomal aberrations (CAs). Non-specific CAs have been used in the monitoring of individuals exposed to potential carcinogenic chemicals and radiation. The frequency of CAs in peripheral blood lymphocytes (PBLs) has been associated with cancer risk and the association has also been found in incident cancer patients. CAs include chromosome-type aberrations (CSAs) and chromatid-type aberrations (CTAs) and their sum CATot. In the present study, we used data from our published genome-wide association studies (GWASs) and extracted the results for 153 DNA repair genes for 607 persons who had occupational exposure to diverse harmful substances/radiation and/or personal exposure to tobacco smoking. The analyses were conducted using linear and logistic regression models to study the association of DNA repair gene polymorphisms with CAs. Considering an arbitrary cutoff level of 5×10^{-3} , 14 loci passed the threshold, and included 7 repair pathways for CTA, 4 for CSA, and 3 for CATot; 10 SNPs were eQTLs influencing the expression of the target repair gene. For the base excision repair pathway, the implicated genes *PARP1* and *PARP2* encode poly(ADP-ribosyl) transferases with multiple regulatory functions. *PARP1* and *PARP2* have an important role in maintaining genome stability through diverse mechanisms. Other candidate genes with known roles for CSAs included *GTF2H* (general transcription factor IIH

subunits 4 and 5), Fanconi anemia pathway genes, and *PMS2*, a mismatch repair gene. The present results suggest pathways with mechanistic rationale for the formation of CAs and emphasize the need to further develop techniques for measuring individual sensitivity to genotoxic exposure.

Keywords: chromosomal aberrations, association study, DNA repair, exposure, polymorphism

INTRODUCTION

Human cancers are often associated with chromosomal instability with complex numerical and structural chromosomal aberrations (CAs), which may be causative events in the process of malignant transformation (Futreal et al., 2004; Rajagopalan and Lengauer, 2004; Mitelman et al., 2007; Burrell et al., 2013). Structural CAs may be specific, such as translocations and inversions, or non-specific, such as chromatid breaks, fragmented or missing parts of chromosomes, and fusions resulting in dicentric and ring chromosomes (Bignold, 2009). The former are often recurrent and they are currently analyzed by molecular cytogenetic methods while the latter are scored by classical cytogenetic techniques, which are able to recognize chromosome-type aberrations (CSAs) and chromatid-type aberrations (CTAs) according to morphological changes (Hagmar et al., 2004). CTAs are formed due to insufficiently repaired double-strand breaks (DSBs) during the late S or G2 phase of the cell cycle (Natarajan and Palitti, 2008; Bignold, 2009; Durante et al., 2013), whereas CSAs are the result of direct DNA damage due to radiation, chemical mutagens, or shortening of telomeres during the G0/G1 phase (Albertini et al., 2000; Jones et al., 2012). Non-specific CAs have been used in the monitoring of populations occupationally exposed to potential carcinogenic chemicals and radiation and an increased frequency of CAs in peripheral blood lymphocytes (PBLs) has been associated with cancer risk and the association has also been found in incident cancer patients (Rossner et al., 2005; Vodicka et al., 2010; Vodenkova et al., 2015).

Unrepaired or insufficiently repaired DSBs, as well as telomerase dysfunction, represent the mechanistic bases for the formation of structural CAs (Natarajan and Palitti, 2008; Bignold, 2009; Durante et al., 2013; Vodicka et al., 2018; Srinivas et al., 2020). However, even other types of DNA repair pathways may contribute to CA formation as these are found in inherited syndromes manifesting DNA repair gene mutations (Rahman, 2014). Eukaryotic cells have four conserved but distinct pathways of DSB repair: non-homologous DNA end joining (NHEJ), alternate end joining (a-EJ), homologous recombination (HR), and single-strand annealing (SSA) (Sung, 2018). In non-malignant cells, the majority of DSBs are removed *via* either NHEJ or HR, with minor contribution of a-EJ and SSA. Repair *via* HR may be error-free while the three other DSB repairs are error-prone, particularly the rare a-EJ and SSA. Repair errors emerge as mutations and CAs with smaller or larger DNA sequence losses. The role of telomerase dysfunction has been emerging more recently, with growing evidence that shorter telomeres are associated with increased frequency of CAs, particularly of the CSA type (Li et al., 2013;

Hemminki et al., 2015). Telomeres become shorter at each round of replication and critically shortened telomeres may be poorly end-capped and may be recognized as DSBs by repair machinery that may result as CAs (Maser and DePinho, 2002; Meeker et al., 2004; Gostissa et al., 2011; Jones and Jallepalli, 2012; Maciejowski et al., 2015). It has been shown that telomere shortening is associated with a decreased capacity to repair DSBs in multiple types of cancer (Kroupa et al., 2017).

In the present study, we used data from our published genome-wide association studies (GWASs) (Niazi et al., 2018, 2019) and extracted the results for 153 DNA repair genes to find out the association between CA frequency and DNA repair pathways. The population was occupationally exposed to diverse harmful substances/radiation and/or personally exposed to tobacco smoking. The analyses were conducted for the types of CAs (CA_{tot}, CSAs, and CTAs) using linear and logistic regression models.

MATERIALS AND METHODS

Our cohort comprised 607 individuals recruited from the Czech and Slovak Republics. The subjects were investigated for chromosomal abnormalities in previous occupational exposure-related epidemiological studies or as regular medical monitoring in factories with exposure to genotoxic compounds. These studies involved individuals with defined exposure to small organic compounds, heavy metals, radiations, and asbestos and other mineral fibers as well as unexposed controls (Vodicka et al., 2004a,b; Dusinska et al., 2004a,b, 2012; Musak et al., 2008; Kazimirova et al., 2009). Prior to blood sampling, study subjects were informed according to the rules of Helsinki declaration and written approval was obtained. Ethics Committees of the Slovak Medical University, the Jessenius Faculty of Medicine, the Comenius University Bratislava, the Institute for Clinical and Experimental Medicine in Slovakia, and the Thomayer Hospital and the General University Hospital in the Czech Republic approved the study design.

The study population (**Table 1**) contained about 60% males and 40% females. All individuals included in the study were either exposed to genotoxic compounds due to their occupation and/or they were smokers. About half of the individuals (52.1%) had a history of occupational exposure to genotoxic organic compounds while 12.7% were exposed to heavy metals, mineral fibers, and low levels of radiations. All subjects filled a questionnaire listing beside the type of job and periods of exposure other exogenous factors such as smoking, radiation exposure, and dietary dispositions. About 66% of the individuals included in the study were smokers. Age of the participants

TABLE 1 | Descriptive attributes of the study cohort and exposure-based distribution.

Study cohort		Covariate effect (<i>P</i> -value) ^e	
Age (years)	Median	43	0.56
	Range	19–80	
Gender (%)	Females	40.5	0.05
	Males	59.5	
Smoking status (%)	Smokers	66.1	5.55E-05
	Non-smokers	33.9	
Occupational exposure (n)	Small organic compounds	316	2.42E-05
	Heavy metals	6	
	Radiation (pilots)	6	
	Asbestos	19	
	Stone wool	28	
	Glass fibers	18	
	Others ^a	214	
No. of individuals with	High CATot freq ^b	342	
	Low CATot freq	265	
	High CTA freq ^c	345	
	Low CTA freq	262	
	High CSA freq ^d	321	
	Low CSA freq	286	

^aOffice workers and blood donors who were reported as smokers.

^bHigh CATot freq = ≥ 2 CAs/100 cells.

^cHigh CTA freq = ≥ 1 CA/100 cells.

^dHigh CSA freq = ≥ 1 CA/100 cells.

^e*P*-value indicates the association of the covariates (age, gender, smoking status, and occupational exposure) with CAs.

ranged from 19 to 80 years with a median age of 43 years. Cytogenetic analysis was done in PBLs that were stimulated to grow and cultured for 48 h (Vodicka et al., 2010). About 100 mitoses per person were evaluated to score the frequency of CSAs and CTAs and they were summed up to CATot (i.e., CSA + CTA = CATot).

For GWAS genotyping, Illumina HumanOmniExpressExome8v1.3 chip arrays were used and the quality control (QC) criteria were implemented according to the predetermined benchmarks (Niazi et al., 2018, 2019). Samples were included on the basis of successful genotyping $\geq 95\%$. Duplicates and related individuals were excluded by identity-by-state (IBS) score. Population outliers determined by the principal component analysis were removed. After prephasing with SHAPEIT v2.12 (Delaneau et al., 2011), imputation was performed using UK10K (Walter et al., 2015) and 1,000 genomes (phase 3, October 2014) (1000 Genomes Project Consortium, Auton et al., 2015) as reference panels with IMPUTE2 v2.3.2 software (Howe et al., 2011). Prior to analysis, SNPs were filtered according to call rate ($<95\%$), Hardy–Weinberg equilibrium (HWE) ($P < 1.0 \times 10^{-5}$), minor allele frequency (MAF) (<0.05), and imputation quality (Info <0.70).

Association analysis between CA frequency and SNPs in DNA repair genes was conducted using PLINK version 1.90b3.30 (Purcell et al., 2007) using logistic (binary) and linear regression

analyses on three phenotypes CATot, CSAs, and CTAs. For binary logistic regression analysis, individuals were divided into high and low CA frequency groups. For CATot analysis, individuals with $\geq 2\%$ CAs were included in the high-frequency group, while for CSAs and CTAs, the threshold for inclusion into the high-frequency group was $\geq 1\%$ (Dusinska et al., 2004a; Vodicka et al., 2010). The analyses were adjusted for gender, age, smoking status, and occupational exposure. GWAS summary statistics were then used for our gene-based study that included a list of 170 DNA repair genes (Wood et al., 2001, 2005; Friedberg et al., 2006; Lange et al., 2011; **Table 2**). For these genes, coordinates were extracted from UCSC genome browser's hg19 assembly, which gave a list of genes with chromosome number and transcription start and end position. Genes on the X chromosome were excluded from the analysis as well as those with no match found in NCBI RefSeq list, leaving 153 genes for the analysis (Rosenbloom et al., 2015). On the basis of the gene coordinates, a region including the gene of interest with 100 kb upstream and 100 kb downstream regions was selected, and all the SNPs in this window were analyzed. In total, about 40,000 SNPs from the repair genes' regions were analyzed, with about 2000 independent loci among them as determined by using PLINK's linkage disequilibrium-based pruning. These regions were plotted in LocusZoom (Pruim et al., 2010) and SNPs with P -value 5×10^{-3} or below were further studied for their capacity to influence the functional aspects of the corresponding DNA repair genes. This threshold was set to only select the SNPs above the background level of association in the analysis. *In silico* tools utilized in this analysis were Haploreg, GTEx, and RegulomDB 2.0 (Ward and Kellis, 2012; Boyle et al., 2012; GTEx Consortium., 2013). These were used to ascertain linkage disequilibrium (LD) between the SNPs from the same locus identified by different phenotypic analysis as well as location [intergenic, 3' and 5' untranslated regions (UTRs)], intronic or expression quantitative trait locus (eQTL, minimal P -value of 10^{-5}), and effect (synonymous, missense, and non-sense) of the genetic variation. Regulome DB version2.0 provided chromatin state, information about changed motifs, transcription factors, and DNase accessibility.

RESULTS

We identified 14 independent loci associated with CA frequency from six different analyses (two regression models, namely, linear and logistic for each of the three phenotypes, CATot, CSA, and CTA) below the applied cutoff, P -value 5×10^{-3} ; note that the *REV3L* (REV3 like, DNA-directed polymerase zeta catalytic subunit) SNP was detected by both the linear and logistic models in CTA analysis (**Table 3**). All the SNPs that remained after cutoff P -value 5×10^{-3} in all phenotypes' logistic and linear models are given in **Supplementary Material**. If one would consider the analysis of 153 genes, and assume one association per gene, the Bonferroni type of corrected significance level would have a P -value of 3.2×10^{-4} . SNPs that remained significantly associated with CAs after applying this criterion are indicated in bold in **Table 3**. Among CATot

TABLE 2 | Total 153 studied genes grouped based on DNA repair type (genes where the SNPs were associated with CAs are in bold letters).

Base excision repair (BER)	Other BER and strand break joining factors	Poly(ADP-ribose) polymerase (PARP) enzymes that bind to DNA	Direct reversal of damage	Repair of DNA-protein crosslinks	Mismatch excision repair (MMR)	Nucleotide excision repair (NER)	Nucleotide excision repair (NER)	Homologous recombination (HR)	Non-homologous end-joining (NHEJ)	Fanconi anemia	DNA polymerases (catalytic subunits)	Editing and processing nucleases	Ubiquitination and modification	Chromatin structure	Genes defective in diseases associated with sensitivity to DNA damaging agents	Other identified genes with known or suspected DNA repair function	Other conserved DNA damage response genes
UNG	APEX1 (APE1)	PARP1 (ADPRT)	MGMT	TDP1	MSH2	CDK7	XPC	RAD51	XRCC6 (Ku70)	FANCA	POLB	FEN1 (DNase IV)	UBE2B (RAD6B)	H2AFX (H2AX)	BLM	DCLRE1A (SNM1)	ATR
SMUG1	LIG3	PARP2 (ADPRTL2)	ALKBH2 (ABH2)		MSH3	CCNH	RAD23B	DMC1	XRCC5 (Ku80)	FANCC	POLG	FAN1 (MTMR15)	RAD18	CHAF1A (CAF1)	WRN	DCLRE1B (SNM1B)	MDC1
MBD4	XRCC1	PARP3 (ADPRTL3)	ALKBH3 (DEPC1)		MSH6	MINAT1	RAD23A	XRCC2	PRKDC	BRCA2 (FANCD1)	POLD1	TREX1 (DNase III)	SHPRH	SETMAR (METNASE)	RECQL4	RECQL (RECQ1)	RAD1
TDG	PNKP				MLH1	ERCC5 (XPG)	XPA	XRCC3	LIG4	FANCD2	POLE	EXO1 (HEX1)	HLTF (SMARCA3)		ATM	RECQL5	RAD9A
OGG1	APLF (C2ORF13)				PMS2	ERCC1	DDB1	RAD52	XRCC4	FANCE	PCNA	APTX (apratatin)	RNF168		TTDN1 (C7orf11)	HELQ (HEL308)	HUS1
MUTYH (MYH)					MSH4	ERCC4 (XPF)	DDB2	RAD54L	DCLRE1C (Artemis)	FANCF	REV3L (POLZ)	SPO11	RNF8			RDM1 (RAD52B)	RAD17 (RAD24)
NTHL1 (NTH1)					MSH5	LIG1	RPA1	RAD54B	NHEJ1 (XLF, Cernunnos)	FANCG (XRCC9)	MAD2L2 (REV7)	FLJ35220 (ENDOV)	RNF4				CHEK1
MPG					MLH3	ERCC8 (CSA)	RPA2	BRCA1	NUDT1 (MTH1)	FANCI (KIAA1794)	POLH		UBE2V2 (MMS2)				CHEK2
NEIL1					PMS1	ERCC6 (CSB)	RPA3	RAD50	DUT	BRIP1 (FANCIJ)	POLI (RAD30B)		UBE2N (UBC13)				TP53
NEIL2						XAB2 (HCNP)	ERCC3 (XPB)	NBN (NBS1)	RRM2B (p53R2)	FANCL	POLQ						TP53BP1 (53BP1)
NEIL3						MMS19L (MMS19)	ERCC2 (XPD)	RBBP8 (CtIP)		FANCM	POLK (DINB1)						ATRIP
						GTF2H1	GTF2H3	MUS81		PALB2 (FANCD1)	POLL						TOPBP1
						GTF2H2	GTF2H4			RAD51C (FANCO)	POLM						CLK2
							GTF2H5 (TTDA)			FAAP24 (C19orf40)	POLN (POL4P)						PER1

TABLE 3 | SNP associations with P -value $\leq 5 \times 10^{-3}$ from logistic and linear regression analyses of three CA types (CA_{tot}, CTA, and CSA).

CA _{tot} -logistic	DNA repair Gene	Type of DNA repair	SNP	CHR	BP	A1	OR	95% CI	P	In silico
	<i>GTF2H4</i>	Nucleotide excision repair (NER)	rs3130780	6	30874308	T	1.89	1.36–2.64	1.77E-04	1.7 kb 5' of GTF2H4
	<i>PARP1</i>	Base excision repair (BER) PARP enzymes	rs1341334	1	226605024	G	1.56	1.21–2.00	5.16E-04	9.2 kb 5' of PARP1/eQTL
CA _{tot} -linear	DNA repair Gene	Type of DNA repair	SNP	CHR	BP	A1	Beta	95% CI	P	In silico
	<i>MGMT</i>	Direct reversal of DNA damage	rs12247555	10	131370520	C	0.09	0.03–0.15	2.78E-03	Intronic/eQTL
CTA-logistic	DNA repair Gene	Type of DNA repair	SNP	CHR	BP	A1	OR	95% CI	P	In silico
	<i>NEIL3</i>	Base excision repair (BER)	rs10009807	4	178229925	A	0.69	0.54–0.89	4.62E-03	1.1 kb 5' of NEIL3/histone marks
	<i>REV3L</i>	DNA polymerases (catalytic subunits)	rs7742724	6	111839019	A	1.63	1.23–2.16	6.42E-04	eQTL
	<i>BRIP1</i>	Fanconi anemia	rs17542001	17	59915590	C	0.64	0.48–0.85	1.86E-03	Intronic/eQTL
CTA-linear	DNA repair Gene	Type of DNA repair	SNP	CHR	BP	A1	Beta	95% CI	P	In silico
	<i>FANCC</i>	Fanconi anemia	rs13292454	9	97995075	A	0.16	0.05–0.28	3.93E-03	Intronic
	<i>MDC1</i>	Conserved DNA damage response genes	rs3094090	6	30669956	C	0.19	0.07–0.32	2.33E-03	Intronic
	<i>REV3L</i>	DNA polymerases (catalytic subunits)	rs7742724	6	111839019	A	0.16	0.09–0.24	2.47E-05	eQTL
	<i>XRCC4</i>	Non-homologous end-joining (NHEJ)	rs301286	5	82602955	C	–0.14 (–0.22)–(–0.05)		1.42E-03	Intronic/histone marks
	<i>TP53BP1</i>	Conserved DNA damage response genes	rs28702649	15	43648629	T	0.12	0.05–0.19	1.31E-03	eQTL
CSA-logistic	DNA repair Gene	Type of DNA repair	SNP	CHR	BP	A1	OR	95% CI	P	In silico
	<i>GTF2H5</i>	Nucleotide excision repair (NER)	rs1744178	6	158496856	T	1.74	1.31–2.31	1.50E-04	eQTL
	<i>PARP2</i>	Base excision repair (BER) PARP enzymes	rs2318861	14	20758949	G	0.54	0.38–0.78	9.85E-04	eQTL
CSA-linear	DNA repair Gene	Type of DNA repair	SNP	CHR	BP	A1	Beta	95% CI	P	In silico
	<i>FANCD2</i>	Fanconi anemia	rs61429272	3	10037320	C	0.15	0.05–0.24	4.13E-03	eQTL
	<i>PMS2</i>	Mismatch excision repair (MMR)	rs12702464	7	6041506	C	–0.13 (–0.21)–(–0.04)		4.64E-03	Intronic/eQTL

ORs (in logistic regression analysis), Beta values (in linear regression analysis), and their corresponding P -values and in silico predictions are shown. SNP associations that survived Bonferroni correction for multiple testing are marked in bold. SNP single nucleotide polymorphism; CHR chromosome; OR odds ratios; A1 The allele for which beta and OR are calculated.

associated loci, SNP rs3130780 from logistic regression analysis had a P -value that was significant according to such a correction (OR 1.89, 95% CI 1.36–2.64, P -value 1.77×10^{-4}). The SNP is located 1.7 kb 5' to *GTF2H4* (general transcription factor IIH subunit 4), which belongs to the nucleotide excision repair (NER) pathway. In the same analysis for gene *PARP1* [poly(ADP-ribose polymerase 1), rs1341334 at 1q42.12 with OR 1.56, 95% CI 1.21–2.00, and P -value 5.16×10^{-4} also

came close to the significance threshold. From the linear regression analysis for CA_{tot}, no significant association was identified and the only SNP above the background level was an intronic variant in *MGMT* (O6-methylguanine-DNA methyltransferase) gene. This gene is involved in the direct reversal of DNA damage.

A SNP marking the gene *REV3L* was found to be associated with the CTA phenotype in both linear and binary logistic

regression models. This variant, rs7742724, exhibited a notable association in the linear model with β 0.16, 95% CI 0.09–0.24, and P -value 2.47×10^{-5} and a similar tendency in the binary model but with an elevated P -value of 6.42×10^{-4} . The other variants from the CTA analysis had P -values ranging from 1.31×10^{-3} to 4.62×10^{-3} . These included intronic SNPs in the genes *BRIP1* (BRCA1 interacting protein C-terminal helicase 1), *FANCC* (Fanconi anemia complementation group C), *MDC1* (mediator of DNA damage checkpoint 1), and *XRCC4* (X-ray repair cross-complementing protein 4) as well as in *TP53BP1* (TP53-binding protein 1). *BRIP1* and *FANCC* belong to the Fanconi anemia pathway while *MDC1* and *TP53BP1* are conserved DNA damage response genes. *XRCC4* is an NHEJ gene. A variant rs10009807 located at 1.1 kb 5' to *NEIL3* (nei like DNA glycosylase 3), which is a base excision repair (BER) pathway gene, was also among the associations identified from the CTA analysis.

The CSA group presented a total of four associations from both models; SNP at 6q25.3 and eQTL to *GTF2H5* (general transcription factor IIH subunit 5) was the best candidate with OR 1.74, 95% CI 1.31–2.31, and P -value 1.50×10^{-4} . *GTF2H5* is a member of the NER pathway. rs2318861, an eQTL SNP, for *PARP2* [poly(ADP-ribose) polymerase 2], which ADP-ribosylates DNA by acting on terminal phosphates at DNA strand breaks, had a P -value of 9.85×10^{-4} . In the linear regression analysis for CSAs, two associations included a Fanconi anemia gene *FANCD2* (Fanconi anemia complementation group D2) and an intronic variant for mismatch repair (MMR) pathway gene *PMS2* (PMS1 homolog 2, mismatch repair system component).

In **Table 4**, the candidate SNPs were annotated using RegulomeDB and GTex; some eQTL data were also retrieved using Haploreg. SNPs linked to genes *PARP1*, *NEIL3*, *FANCC*, *XRCC4*, and *FANCD2* show DNase accessibility in blood and all the selected variants were located in either the region of strong transcription or in transcription start sites (TSSs) and enhancers in blood and many other tissues. The eQTLs summarized in **Table 4** each target the linked DNA repair gene. The SNP linked to *PARP1* was a strong eQTL in cultured fibroblasts at 9×10^{-23} , and the one linked to the *MGMT* gene was a strong eQTL in the whole blood at 1.4×10^{-33} . The SNPs associated with *MDC1*, *GTF2H5*, *PARP2*, and *FANCD2* were eQTLs in the whole blood/cultured fibroblasts. The SNP for *PMS2* was an eQTL in the aorta, whereas those linked to genes *REV3L* and *BRIP1* were eQTLs in the brain and the tibial nerve, respectively.

DISCUSSION

Genetic variation can be the cause of inter-individual differences in susceptibility to CAs and susceptibility to cancer (Vodicka et al., 2004a). In a previous analysis of 11 DNA repair genes in mixed population of occupationally exposed and unexposed individuals, associations with CAs were found with *XPD* and *RAD54L* polymorphisms (Vodicka et al., 2015). In the present study, SNPs from a total of 153 DNA repair genes were tested on an exposed population of 607 individuals. It can be expected that DNA repair is more critical in persons exposed to high apparent exposure vs. background environmental exposure and

the distribution of CAs has been shown to be skewed to higher damage levels in the exposed population (Niazi et al., 2019). This population difference together with a more stringent significance threshold in the present study might be the reason for the different outcomes of these two studies. While CSAs and CTAs are assumed to be independent markers of damage, arising at different phases of the cell cycle, CATot is a composite measure as the sum of CSAs and CTAs. For the presentation of the results, we selected an arbitrary cutoff level of 5×10^{-3} , which appeared to be stringent as only 14 SNPs passed the threshold and, with one exception, these results from the linear and logistic models were different. We considered the Bonferroni type of adjusted P -value as 3.2×10^{-4} based on the 153 genes tested; even though many more independent LD regions were considered, the sample size for rarer SNPs afforded a limited power (with a MAF of 10%, only six homozygous variants were to be expected). Credibility to the findings is supported by the chromatin state and eQTL data. All SNPs were located at a site of strong transcription, enhancer, or TSS, and five SNPs were located at DNase-accessible sites in blood. Ten of the 14 candidate SNPs influenced the expression of the target DNA repair gene, and for 5, the data were obtained from whole blood.

Most positive associations at P -value below 5×10^{-3} were found for CTAs ($N = 7$), followed by four for CSAs and three for CATot. SNPs for the NER pathway emerged two times, for BER and Fanconi anemia repair pathways three times, while SNPs in other pathways were unique. For the BER pathway, the implicated SNPs were eQTLs to the target genes *PARP1* and *PARP2*, two homologs encoding chromatin-associated enzymes, poly(ADP-ribosyl)transferases, which modify various nuclear proteins by poly(ADP-ribosylation) with multiple downstream regulatory functions (Azarm and Smith, 2020). *PARP1* may be associated with xeroderma pigmentosum, complementation group A through interactions with XPA, and the related susceptibility to skin cancer. *PARP1* is involved in the synthesis of telomere C-strand (Azarm and Smith, 2020). *PARP2* has partially overlapping biochemical functions with *PARP1*. *PARP1* and *PARP2* function in both single- and double-strand DNA repair, and they have an important role in maintaining genome stability through diverse mechanisms. PARP inhibitors are being used as anticancer agents in *BRCA1/2* mutated cancers (Boussios et al., 2020).

The NER pathway genes *GTF2H4* and *GTF2H5* encode different subunits of general transcription factor IIH and both associations were highly significant (Rimel and Taatjes, 2018; Hill and Theos, 2019; Kolesnikova et al., 2019). The proteins share structural and functional homology and they are associated with NER enzymes XPB and XPD (Kolesnikova et al., 2019). Syndromes associated with *GTF2H4/5* include xeroderma pigmentosum, complementation groups B and D, Cockayne syndrome, and trichothiodystrophy, all of which are characterized by extreme sensitivity to ultraviolet radiation and development of other sun-related problems such as excessive freckling and skin cancer. These syndromes as well as Fanconi anemia germline mutations display genomic instability and CAs (Chan and Ngeow, 2017; Hill and Theos, 2019). The related SNP targeted *GTF2H5* as an eQTL.

TABLE 4 | Regulome DB 2.0/GTex *in silico* analysis of associated variants.

Chr	SNP	Gene	Accessibility (DNase and FAIRE) in tissues/cell types	Chromatin state	Motifs	eQTL			ChIP
						Tissue	Normalized effect size*	P-value	
6	rs3130780	GTF2H4	—	Blood + 102	BCL6, MEF2A	—	—	—	ZNF792, MIXL1
1	rs1341334	PARP1	Blood (K562) + 9 other tissues	Enhancer in blood	BCL6, NANOG	Cultured fibroblasts	−0.31	9.00E-23	ZFX,ZNF770
10	rs12247555	MGMT	—	Blood + 16	—	Whole blood	−0.29	1.40E-33	—
4	rs10009807	NEIL3	Blood, B cells, T cells	Enhancer in blood + 30	—	—	—	—	POLR2A
17	rs17542001	BRIP1	—	Blood + 62	—	Nerve-tibial	−0.41	7.30E-15	—
9	rs13292454	FANCC	Blood + 3 others	Blood + 48	FOXJ3	—	—	—	—
6	rs3094090	MDC1	Tibial nerve	Blood + 123	IRF3	Whole blood	0.13	6.40E-06	ZNF664
6	rs7742724	REV3L	Mammary glands	Blood + 23	—	Brain-Cerebellum	0.28	1.40E-06	SMARCA4 #
5	rs301286	XRCC4	Blood + 6 others	Active TSS in blood + 6	—	—	—	—	STAT5A#, STAT3#, TBP# + 10
15	rs28702649	TP53BP1	Placenta, H9, OCI-LY7	Blood + 109	SOX1	Cultured fibroblasts	−0.13	4.10E-08	CTCF#, RAD21, ZBTB33#
6	rs1744178	GTF2H5	H7-hESC, Lower leg	Blood + 118	—	*Whole blood	*	2.8783E-17	—
14	rs2318861	PARP2	—	Blood + 124	—	*Whole blood	*	6.61974E-07	ZBTB40
3	rs61429272	FANCD2	Naïve B cell + 9 other tissues	Blood + 99	—	Whole blood	−0.17	2.70E-09	—
7	rs12702464	PMS2	—	Blood + 124	—	Artery-aorta	−0.44	4.50E-12	—

*eQTL values from Westra et al. (2013).

#In blood.

In summary, the present study on DNA repair gene polymorphisms in a healthy population with occupational and personal genotoxic exposures revealed SNP associations with CA frequency at the *P*-value level of 5×10^{-3} within 14 different genes, many of which with key roles in maintaining genomic integrity and thus plausibly associated with mechanisms leading to CAs. More than half of the implicated SNPs were eQTLs to the target DNA repair genes. Although the recent interest in measuring random CAs has decreased because of cumbersome techniques, the present results suggest that the results may have understandable mechanistic links. If the current techniques cannot be improved, there will be a need to provide alternative approaches for measuring individual sensitivity to genotoxic exposure that may lead to increased risk of cancer.

DATA AVAILABILITY STATEMENT

All the relevant data presented in the study is given in **Supplementary Material**, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committees of the Slovak Medical University, the Jessenius Faculty of Medicine, the Comenius University Bratislava, the Institute for Clinical and Experimental Medicine in Slovakia and the Thomayer Hospital and the General University Hospital in the Czech Republic. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AF, KH, YN designed the study. HT and YN analyzed the data. YN and BS performed the genotyping. PV, LV, LM, MD, and BS provided the samples. SV, MK, VV, KV, AK, and MS collected subject information and performed cytogenetic analysis. PH and MN were responsible for the GWAS. AF, KH, PV, LV, MD, and BS critically revised the manuscript. YN wrote the first draft. KH and AF supervised the study. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

In the Czech Republic, the KH was supported by the European Union's Horizon 2020 Research and Innovation Programme, Grant No. 856620 (Chaperon), the National Science Foundation (18-09709S, 19-10543S), Charles University in Prague (PROGRE Q 28), Medical Faculty in Pilsen, Charles University in Prague, National Sustainability Programme I (Nr.LO 1503), Charles University Research Centre program (UNCE/MED/006), and European Commission contracts (QLK4-CT-1999-01629, ERBICI 15-CT96-1012 and CIPA-CT94-0129). In Slovakia, support came from Slovak Grant Agency (APVT-21 013202, APVT-21-017704), Ministry of Health, Slovak Republic (2005/43-SZU-21, 2006/07-SZU-02 MZ SR, 2005/42-SZU-20), SZU and Competence Center for Research

and Development in the Field of Diagnostics and Therapy of Oncological Diseases Slovakia (ITMS code: 26220220153), project Biomedical Center Martin, Slovakia co-financed from EU sources (ITMS code: 26220220187), the project "Carcinogenic and toxic metals in working environment" co-financed by EU sources and the European Regional Development Fund, Slovakia (ITMS: 26220220111), and the Research and Development Support Agency Slovakia (APVV-15-0217).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.691947/full#supplementary-material>

REFERENCES

- 1000 Genomes Project Consortium, Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., et al. (2015). A global reference for human genetic variation. *Nature* 526:68. doi: 10.1038/nature15393
- Albertini, R. J., Anderson, D., Douglas, G. R., Hagmar, L., Hemminki, K., Merlo, F., et al. (2000). IPCS guidelines for the monitoring of genotoxic effects of carcinogens in humans. International Programme on Chemical Safety. *Mutat. Res.* 463, 111–172. doi: 10.1016/s1383-5742(00)00049-1
- Azarm, K., and Smith, S. (2020). Nuclear PARPs and genome integrity. *Genes Dev.* 34, 285–301. doi: 10.1101/gad.334730.119
- Bignold, L. P. (2009). Mechanisms of clastogen-induced chromosomal aberrations: a critical review and description of a model based on failures of tethering of DNA strand ends to strand-breaking enzymes. *Mutat. Res.* 681, 271–298. doi: 10.1016/j.mrrev.2008.11.004
- Boussios, S., Abson, C., Moschetta, M., Rassy, E., Karathanasi, A., Bhat, T., et al. (2020). Poly (ADP-Ribose) polymerase inhibitors: Talazoparib in ovarian cancer and beyond. *Drugs R D* 20, 55–73.
- Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A., Kasowski, M., et al. (2012). Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 22, 1790–1797. doi: 10.1101/gr.137323.112
- Burrell, R. A., McClelland, S. E., Endesfelder, D., Groth, P., Weller, M. C., Shaikh, N., et al. (2013). Replication stress links structural and numerical cancer chromosomal instability. *Nature* 494, 492–496. doi: 10.1038/nature11935
- Chan, S. H., and Ngeow, J. (2017). Germline mutation contribution to chromosomal instability. *Endocr. Relat. Cancer* 24, T33–T46.
- Delaneau, O., Marchini, J., and Zagury, J. F. (2011). A linear complexity phasing method for thousands of genomes. *Nat. Methods* 9, 179–181. doi: 10.1038/nmeth.1785
- Durante, M., Bedford, J. S., Chen, D. J., Conrad, S., Cornforth, M. N., Natarajan, A. T., et al. (2013). From DNA damage to chromosome aberrations: joining the break. *Mutat. Res.* 756, 5–13. doi: 10.1016/j.mrgentox.2013.05.014
- Dusinska, M., Barancokova, M., Kazimirova, A., Harrington, V., Volkovova, K., Staruchova, M., et al. (2004a). Does occupational exposure to mineral fibres cause DNA or chromosome damage? *Mutat. Res.* 553, 103–110. doi: 10.1016/j.mrfmmm.2004.06.029
- Dusinska, M., Collins, A., Kazimirova, A., Barancokova, M., Harrington, V., Volkovova, K., et al. (2004b). Genotoxic effects of asbestos in humans. *Mutat. Res.* 553, 91–102. doi: 10.1016/j.mrfmmm.2004.06.027
- Dusinska, M., Staruchova, M., Horska, A., Smolkova, B., Collins, A., Bonassi, S., et al. (2012). Are glutathione S transferases involved in DNA damage signalling? interactions with DNA damage and repair revealed from molecular epidemiology studies. *Mutat. Res.* 736, 130–137. doi: 10.1016/j.mrfmmm.2012.03.003
- Friedberg, E. C., Walker, G. C., Siede, W., Wood, R. D., Schultz, R. A., and Ellenberger, T. (2006). *DNA Repair and Mutagenesis*, Second Edn. Washington, DC: American Society for Microbiology Press. 273–.
- Futreal, P. A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., et al. (2004). A census of human cancer genes. *Nat. Rev. Cancer* 4, 177–183.
- Gostissa, M., Alt, F. W., and Chiarle, R. (2011). Mechanisms that promote and suppress chromosomal translocations in lymphocytes. *Annu. Rev. Immunol.* 29, 319–350.
- GTEX Consortium. (2013). The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* 45, 580–585.
- Hagmar, L., Stromberg, U., Tinnerberg, H., and Mikoczy, Z. (2004). Epidemiological evaluation of cytogenetic biomarkers as potential surrogate end-points for cancer. *IARC Sci. Publ.* 157, 207–215.
- Hemminki, K., Rachakonda, S., Musak, L., Vymetalkova, V., Halasova, E., Forsti, A., et al. (2015). Telomere length in circulating lymphocytes: association with chromosomal aberrations. *Genes Chromosomes Cancer* 54, 194–196.
- Hill, C. R., and Theos, A. (2019). What's new in genetic skin diseases. *Dermatol. Clin.* 37, 229–239.
- Howie, B., Marchini, J., and Stephens, M. (2011). Genotype imputation with thousands of genomes. *G3 (Bethesda)* 1, 457–470.
- Jones, C. H., Pepper, C., and Baird, D. M. (2012). Telomere dysfunction and its role in haematological cancer. *Br. J. Haematol.* 156, 573–587.
- Jones, M. J., and Jallepalli, P. V. (2012). Chromothripsis: chromosomes in crisis. *Dev. Cell* 23, 908–917.
- Kazimirova, A., Barancokova, M., Dzapinkova, Z., Wsolova, L., and Dusinska, M. (2009). Micronuclei and chromosomal aberrations, important markers of ageing: possible association with XPC and XPD polymorphisms. *Mutat. Res.* 661, 35–40.
- Kolesnikova, O., Radu, L., and Poterszman, A. (2019). TFIIF: A multi-subunit complex at the cross-roads of transcription and DNA repair. *Adv. Protein Chem. Struct. Biol.* 115, 21–67.
- Kroupa, M., Polivkova, Z., Rachakonda, S., Schneiderova, M., Vodenkova, S., Buchler, T., et al. (2017). Bleomycin-induced chromosomal damage and shortening of telomeres in peripheral blood lymphocytes of incident cancer patients. *Genes Chromosomes Cancer* 57, 61–69.
- Lange, S. S., Takata, K., and Wood, R. D. (2011). DNA polymerases and cancer. *Nat. Rev. Cancer* 11, 96–110.
- Li, H., Hilmarsen, H. T., Hossain, M. B., Bjork, J., Hansteen, I. L., Albin, M., et al. (2013). Telomere length and LINE1 methylation is associated with chromosomal aberrations in peripheral blood. *Genes Chromosomes Cancer* 52, 1–10.
- Maciejowski, J., Li, Y., Bosco, N., Campbell, P. J., and de Lange, T. (2015). Chromothripsis and Kataegis induced by telomere crisis. *Cell* 163, 1641–1654.
- Maser, R. S., and DePinho, R. A. (2002). Connecting chromosomes, crisis, and cancer. *Science* 297, 565–569.
- Meeker, A. K., Hicks, J. L., Iacobuzio-Donahue, C. A., Montgomery, E. A., Westra, W. H., Chan, T. Y., et al. (2004). Telomere length abnormalities occur early in the initiation of epithelial carcinogenesis. *Clin. Cancer Res.* 10, 3317–3326.

- Mitelman, F., Johansson, B., and Mertens, F. (2007). The impact of translocations and gene fusions on cancer causation. *Nat. Rev. Cancer* 7, 233–245.
- Musak, L., Soucek, P., Vodickova, L., Naccarati, A., Halasova, E., Polakova, V., et al. (2008). Chromosomal aberrations in tire plant workers and interaction with polymorphisms of biotransformation and DNA repair genes. *Mutat. Res.* 641, 36–42.
- Natarajan, A. T., and Palitti, F. (2008). DNA repair and chromosomal alterations. *Mutat. Res.* 657, 3–7.
- Niazi, Y., Thomsen, H., Smolkova, B., Vodickova, L., Vodenkova, S., Kroupa, M., et al. (2018). Genetic variation associated with chromosomal aberration frequency: a genome-wide association study. *Environ. Mol. Mutagen.* 60, 17–28.
- Niazi, Y., Thomsen, H., Smolkova, B., Vodickova, L., Vodenkova, S., Kroupa, M., et al. (2019). Distinct pathways associated with chromosomal aberration frequency in a cohort exposed to genotoxic compounds compared to general population. *Mutagenesis* 34, 323–330.
- Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., et al. (2010). LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 26, 2336–2337.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575.
- Rahman, N. (2014). Realizing the promise of cancer predisposition genes. *Nature* 505, 302–308.
- Rajagopalan, H., and Lengauer, C. (2004). Aneuploidy and cancer. *Nature* 432, 338–341.
- Rimel, J. K., and Taatjes, D. J. (2018). The essential and multifunctional TFIIH complex. *Protein Sci.* 27, 1018–1037.
- Rosenbloom, K. R., Armstrong, J., Barber, G. P., Casper, J., Clawson, H., Diekhans, M., et al. (2015). The UCSC genome browser database: 2015 update. *Nucleic Acids Res.* 43, D670–D681.
- Rossner, P., Boffetta, P., Ceppi, M., Bonassi, S., Smerhovsky, Z., Landa, K., et al. (2005). Chromosomal aberrations in lymphocytes of healthy subjects and risk of cancer. *Environ. Health Perspect.* 113, 517–520.
- Srinivas, N., Rachakonda, S., and Kumar, R. (2020). Telomeres and telomere length: a general overview. *Cancers (Basel)* 12:558.
- Sung, P. (2018). Introduction to the thematic minireview series: DNA double-strand break repair and pathway choice. *J. Biol. Chem.* 293, 10500–10501.
- Vodenkova, S., Polivkova, Z., Musak, L., Smerhovsky, Z., Zoubkova, H., Sytarova, S., et al. (2015). Structural chromosomal aberrations as potential risk markers in incident cancer patients. *Mutagenesis* 30, 557–563.
- Vodicka, P., Kumar, R., Stetina, R., Musak, L., Soucek, P., Haufroid, V., et al. (2004a). Markers of individual susceptibility and DNA repair rate in workers exposed to xenobiotics in a tire plant. *Environ. Mol. Mutagen.* 44, 283–292.
- Vodicka, P., Musak, L., Frank, C., Kazimirova, A., Vymetalkova, V., Barancokova, M., et al. (2015). Interactions of DNA repair gene variants modulate chromosomal aberrations in healthy subjects. *Carcinogenesis* 36, 1299–1306.
- Vodicka, P., Musak, L., Vodickova, L., Vodenkova, S., Catalano, C., Kroupa, M., et al. (2018). Genetic variation of acquired structural chromosomal aberrations. *Mutat. Res. Genet. Toxicol. Environ. Mutagen.* 836(Pt A), 13–21.
- Vodicka, P., Polivkova, Z., Sytarova, S., Demova, H., Kucerova, M., Vodickova, L., et al. (2010). Chromosomal damage in peripheral blood lymphocytes of newly diagnosed cancer patients and healthy controls. *Carcinogenesis* 31, 1238–1241.
- Vodicka, P., Tuimala, J., Stetina, R., Kumar, R., Manini, P., Naccarati, A., et al. (2004b). Cytogenetic markers, DNA single-strand breaks, urinary metabolites, and DNA repair rates in styrene-exposed lamination workers. *Environ. Health Perspect.* 112, 867–871.
- Walter, K., Min, J. L., Huang, J., Crooks, L., Memari, Y., McCarthy, S., et al. (2015). The UK10K project identifies rare variants in health and disease. *Nature* 526, 82–90.
- Ward, L. D., and Kellis, M. (2012). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 40, D930–D934.
- Westra, H.-J., Peters, M. J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., et al. (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* 45, 1238–1243. doi: 10.1038/ng.2756
- Wood, R. D., Mitchell, M., and Lindahl, T. (2005). Human DNA repair genes, 2005. *Mutat. Res.* 577, 275–283.
- Wood, R. D., Mitchell, M., Sgouros, J., and Lindahl, T. (2001). Human DNA repair genes. *Science* 291, 1284–1289.

Conflict of Interest: HT was an employee at GeneWerk GmbH, which had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Niazi, Thomsen, Smolkova, Vodickova, Vodenkova, Kroupa, Vymetalkova, Kazimirova, Barancokova, Volkovova, Staruchova, Hoffmann, Nöthen, Dusinska, Musak, Vodicka, Hemminki and Försti. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Folate Deficiency Triggers the Abnormal Segregation of a Region With Large Cluster of CG-Rich Trinucleotide Repeats on Human Chromosome 2

OPEN ACCESS

Edited by:

Chunlong Chen,
Institut Curie, France

Reviewed by:

Lin Deng,
Shenzhen Bay Laboratory, China
Richard H. Finnell,
Baylor College of Medicine,
United States

*Correspondence:

Ying Liu
ying@sund.ku.dk
orcid.org/0000-0003-3703-2498

† Present address:

Lorenza Garribba,
Department of Experimental
Oncology, European Institute
of Oncology, Milan, Italy
Liqun Ren,
Basic Medical Research Institute,
Chengde Medical University,
Chengde, China

Specialty section:

This article was submitted to
Human and Medical Genomics,
a section of the journal
Frontiers in Genetics

Received: 14 April 2021

Accepted: 08 June 2021

Published: 01 July 2021

Citation:

Garribba L, Vogel I, Lerdrup M,
Gonçalves Dinis MM, Ren L and Liu Y
(2021) Folate Deficiency Triggers
the Abnormal Segregation of a Region
With Large Cluster of CG-Rich
Trinucleotide Repeats on Human
Chromosome 2.
Front. Genet. 12:695124.
doi: 10.3389/fgene.2021.695124

Lorenza Garribba[†], Ivan Vogel, Mads Lerdrup, Marisa M. Gonçalves Dinis, Liqun Ren[†]
and Ying Liu^{*}

Center for Chromosome Stability and Center for Healthy Aging, Department of Cellular and Molecular Medicine, University of Copenhagen, Copenhagen, Denmark

Folate deficiency is associated with a broad range of human disorders, including anemia, fetal neural tube defects, age-associated dementia and several types of cancer. It is well established that a subgroup of rare fragile sites (RFSs) containing expanded CGG trinucleotide repeat (TNR) sequences display instability when cells are deprived of folate. However, given that folate sensitive RFSs exist in a very small percentage of the population, they are unlikely to be the cause of the widespread health problems associated with folate deficiency. We hypothesized that folate deficiency could specifically affect DNA replication at regions containing CG-rich repeat sequences. For this, we identified a region on human chromosome 2 (Chr2) comprising more than 300 CG-rich TNRs (termed “FOLD1”) by examining the human genome database. Via the analysis of chromosome shape and segregation in mitosis, we demonstrate that, when human cells are cultured under folate stress conditions, Chr2 is prone to display a “kink” or “bending” at FOLD1 in metaphase and nondisjunction in anaphase. Furthermore, long-term folate deprivation causes Chr2 aneuploidy. Our results provide new evidence on the abnormalities folate deficiency could cause in human cells. This could facilitate future studies on the deleterious health conditions associated with folate deficiency.

Keywords: aneuploidy, fluorodeoxyuridine, fragile site, FRAXA, nondisjunction, MiDAS

INTRODUCTION

Folate is a B vitamin that functions as a carrier for one-carbon units that are essential for DNA synthesis (Ducker and Rabinowitz, 2017). Humans cannot synthesize folate and therefore rely on a dietary supply of this nutrient. In human populations where folic acid supplementation is absent, folate deficiency is frequently observed (Bailey et al., 1982; Senti and Pilch, 1985; Roman Vinas et al., 2011; Mensink et al., 2013). This deficiency is known to be associated with anemia (Green and Miller, 1999), fetal neural tube defects (No authors listed, 1991), infertility in men and women (Altmae et al., 2010; Gaskins et al., 2012; Sansone et al., 2018), age-associated dementia, psychological disorders (Araujo et al., 2015), and several types of cancer (including breast, pancreatic, colon, esophageal, and gastric)

(Giovannucci, 2002; Larsson et al., 2006; Ericson et al., 2007). Notably, it is established that a subgroup of so-called rare fragile sites (RFSs), which contain CGG trinucleotide repeat (TNR) sequences (Durkin and Glover, 2007), are particularly sensitive to the intracellular level of folate. Among them, FRAXA is the most widely studied RFS and is associated with fragile X syndrome (FXS) when the TNR sequence expands beyond a critical length (Pieretti et al., 1991; Santoro et al., 2012). RFSs appear on metaphase chromosomes as gaps or breaks in otherwise fully condensed chromatin (called RFS “expression”). However, because they are present in a very small subset of individuals (less than 5%), RFSs containing abnormally expanded CGG-TNR sequences are unlikely to be the cause of the more widespread health problems associated with folate deficiency.

It is known that folate deficiency has a negative impact on genome stability. Previous studies showed that cells deprived of folate exhibit DNA replication-associated DNA breakage (Lamm et al., 2015) and chromosome instability (e.g., chromosome 8 aneuploidy) (Ni et al., 2010). In agreement with this, our previous work demonstrated that, in response to folate stress, the FRAXA locus is dramatically missegregated in mitosis (Bjerregaard et al., 2018). Importantly, long-term folate deprivation not only exacerbates FRAXA instability but also leads to chromosome X aneuploidy (Bjerregaard et al., 2018). However, it remains unknown whether other “non-RFS” regions are also susceptible to folate deprivation. Considering that all of the 24 folate sensitive RFSs contain abnormally long CGG-TNRs (Durkin and Glover, 2007), we hypothesized that folate deficiency might specifically affect genomic regions containing apparently non-pathological CG-rich repeat sequences. It is conceivable that folate deficiency could specifically interfere with the replication of these CG-rich sequences, which potentially can form DNA secondary structures due to the slowing of replication forks in response to a lack of thymidine and/or purines (Fry and Loeb, 1994; Usdin and Woodford, 1995). The resulting replication perturbation would then interfere with the timely locus condensation in early mitosis and the subsequent sister chromatid disjunction in anaphase, as is observed at FRAXA (Bjerregaard et al., 2018). This, in turn, would lead to heritable alterations at CG-rich TNR regions, and potentially explain the diverse health problems associated with folate deficiency.

To address this hypothesis, we conducted a bioinformatic analysis of human genome and identified a region at 2p11.2 that contains a cluster of more than 300 CG-rich TNRs (termed FOLD1 for “folate-deficiency-induced bending 1”). Our results demonstrated that folate deprivation leads to distinct mitotic abnormalities at FOLD1, possibly due to a difficulty in replicating the CG-rich TNRs in this region when cells are subjected to folate stress.

MATERIALS AND METHODS

Cell Lines

Epstein-Barr virus immortalized human B-lymphocyte GM09237 (male, with 931–940 CGG repeats (Bjerregaard et al., 2018) at the FRAXA locus) was from the Coriell Biorepository.

Untransformed foreskin fibroblast Hs68 cells and osteosarcoma cell line U2OS were from the ATCC.

Cell Culture

B-lymphocytes were maintained in RPMI 1640 Medium (Gibco), while Hs68 cells were maintained in Dulbecco's modified Eagle's medium (DMEM). In both cases, the medium was supplemented with 15% fetal bovine serum (FBS, Thermo Fisher Scientific). U2OS cells were maintained in DMEM supplemented with 10% FBS. The cells were maintained at 37°C in a humidified atmosphere with 5% CO₂ and were routinely subjected to mycoplasma testing (using MycoAlert; Lonza). Only mycoplasma-free cells were analyzed.

Cell Synchronization and Treatment

To obtain metaphase chromosomes, asynchronous cells were cultured under several different conditions including untreated (Unt), with 0.5 μM 5-fluorodeoxyuridine (FdU; Abcam), 0.4 μM aphidicolin (APH; Sigma Aldrich) or 0.4 mM hydroxyurea (HU; Sigma Aldrich) for 22 h and were synchronized in metaphase by being cultured with colcemid (100 ng/ml; Thermo Fisher Scientific) for the last 5 h of the 22-h treatment. For “No folate” condition, cells were cultured in RPMI 1640 medium without folic acid (Gibco), or DMEM medium without folic acid (Cell Culture Technologies) for 3 or 5 days and were then synchronized at metaphase by being cultured with colcemid (100 ng/ml; Thermo Fisher Scientific) for the last 5 h before harvesting.

To analyze mitotic DNA synthesis (MiDAS) on metaphase chromosomes, asynchronous cells under untreated or FdU treated conditions were arrested in late G2 using 12 μM RO3306 (12 μM; Merck) for 9 h. Cells were then released into medium containing 20 μM 5-ethynyl-2'-deoxyuridine (EdU; A10044 Thermo Fisher Scientific) and colcemid (100 ng/ml; Thermo Fisher Scientific) and were then incubated for 1 h before harvesting.

To harvest anaphase/telophase cells, asynchronous cultures were treated with RO3306 for 9 h to induce an arrest at the G2/M boundary. Cells were then released from RO3306, incubated for 30 or 45 min in pre-warmed cell culture medium (37°C), and then seeded onto poly L-Lysine coated slides (Sigma Aldrich) prior to fixation. For cells cultured in RPMI 1640 without folic acid (“No folate”), RO3306 was added to the medium after 3- or 5-day incubation, and the cells were then released after 9 h as indicated above.

To assess chromosome nondisjunction, a previously published protocol was used (Bjerregaard et al., 2018). Briefly, cells were released from FdU into medium containing 10 μM EdU (Thermo Fisher Scientific) for 3 h to label the cells that had gone through S phase in the presence of FdU. Cells were then released into cytochalasin B (3 μg/ml; Sigma-Aldrich) for 16 h and seeded onto poly L-Lysine slides (Sigma-Aldrich) for further analysis.

Fluorescence *in situ* Hybridization

To analyze metaphase chromosomes, cells in metaphase were harvested using standard procedures. Briefly, cells were swollen in 75 mM KCl (at 37°C), fixed in methanol:acetic acid (3:1) and then dropped onto glass slides. To analyze anaphase

and cytokinesis-blocked cells, cells were seeded on poly-L-Lysine slides (Sigma-Aldrich) and fixed in methanol:acetic acid (3:1) for fluorescence *in situ* hybridization (FISH) analysis. BAC clones were used to prepare FISH probes targeting the regions of interests: RP11-439L14 (GenBank: AC012454.8) for FOLD1, RP11-464P18 (GenBank: AC073105.1) for Chr2Q, RP11-799C6 (GenBank: AC068519.1) for Chr1CCEN, RP11-383P16 (GenBank: AC233288.1) for FRAXA (**Supplementary Figure 1A**). Probes were labeled using the BioNick labeling system (Thermo Fisher Scientific) or DIG-nick translation mix (Sigma Aldrich). FISH was carried out using standard procedures, as previously published (Bjerregaard et al., 2018). Slides were mounted using Vectashield mounting medium with DAPI (Vector Labs). Images were captured using an Olympus BX63 microscope. Images were analyzed using CellSens (Olympus) or Fiji/ImageJ software. In the analysis of “bent” shape of human chromosome 2 (Chr2), all of the images were analyzed independently by two researchers.

Karyotype Analysis

Karyotypes of metaphase cells were obtained using conventional G-banding and were analyzed using a Leica light microscope equipped with CytoVision software.

FISH/EdU Detection

To detect EdU incorporation at FOLD1, revealed by FISH, we followed a previously described protocol (Garribba et al., 2018) with minor modifications. Briefly, at the end of the FISH procedure, slides were fixed in 4% paraformaldehyde (PFA) in PBS for 4 min at room temperature (RT) and then blocked in 3% bovine serum albumin (BSA) in PBS + 0.5% Triton-X100 for 10 min at RT. Then, EdU detection was carried out using Click-iT® Plus EdU Alexa Fluor® 594 Imaging Kit (C10339 Thermo Fisher Scientific). Slides were incubated with the reaction mix for 1 h at RT in the dark and were then washed in 3% BSA in PBS + 0.5% Triton-X100 (3 × 20 min). Slides were mounted using Vectashield mounting medium with DAPI (Vector Labs). Images were captured using an Olympus BX63 microscope and analyzed using CellSens (Olympus) or Fiji/ImageJ software.

Statistical Analysis

At least three independent experiments were carried out to generate each data set. Statistical significance in each case was calculated using Chi-squared or Fisher's exact test for all of the data sets.

Genome Sequencing and Bioinformatic Analysis

Extraction of Simple Repeats

AT or CG rich simple repeats were obtained from the recent version of the human genome assembly (GRCh38/hg38) curated at UCSC Genome Browser Database (Rosenbloom et al., 2015). The number of nucleotides of repeat unit ranges from 1 to 14. Considering that the average GC-content in human genomes ranges from 35 to 60%, with a mean of 41% (Piovesan et al., 2019), we reasoned that, if a region has more than 60% C/G

repeats, then it will have more C/G than the other parts of the genome. Therefore, we used 60% as a threshold to filter out the regions with simple repeats. Three databases, namely “RepeatMasker Database,” “Tandem Repeat Finder,” and “Microsatellites,” were used to extract repeats. To identify TNRs, the RepeatMasker database was used, while the repeats were concatenated and calculated with the other two databases. To extract and organize potential repeat sequences, genomic features including chromosome position, length of the repeat unit, number of repetitions of the repeat unit, total span of the repeat chain, the overall similarity of the consensus sequence to the repeat chain, consensus sequence, and database source were included (**Supplementary Tables 1, 2**). The extraction process was performed using an automated computational pipeline that includes custom-made Bash and Python scripts. The scripts can be accessed at github: <https://github.com/puko818/trinucl-hunter>.

Overlap Analysis With Genomic Features

To analyze the overlap between repeat sequences and genomic features, we obtained BED files from the following databases (Quinlan and Hall, 2010): (1) CpG islands (database table cpGIslandExt) (hg38); (2) Cytogenetic bands (database table cytoBand) (hg38); (3) Promoters (database table knownGene, with positions 2000 bases upstream and 500 bases downstream of the transcription start site) (hg38); (4) Exons (database table knownGene) (only positions of the exons from the downloaded coordinates) (hg38); (5) introns (database table knownGene) (only positions of the introns from the downloaded coordinates) (hg38); (6) transposable elements (TEs; database table rmsk) Records labeled as simple repeats were skipped as those were already included in the AT/CG-rich database (hg38); (7) transcription factor binding sites (TFBS; database table wgEncodeRegTfbsClusteredV3) (hg19) (with coordinates being converted to hg38 coordinates using the liftOver tool) (Rosenbloom et al., 2015); (8) COSMIC database of somatic mutations – database table COSMIC (hg19) (with coordinates being converted to hg38 using the liftOver tool); (9) polymorphisms (database of single nucleotide polymorphisms-database table snp147Common from hg38 and database of genomic variants - database table dgvMerged from hg38). BEDOPS package (Neph et al., 2012) was used to analyze the overlaps. The criterion for designation of an overlap was 1 bp. The expected or observed representation of the repeats in certain genomic features was calculated using the following formula:

$$\text{Expected representation} : \frac{\sum_{i=1}^n |repeat_i| \cdot \sum_{j=1}^m |feature_j|}{|genome|^2}$$

$$\text{Observed representation} : \frac{\sum_{j=1}^m \sum_{i=1}^n |repeat_i \cup feature_j|}{|genome|}$$

$|repeat_i|$ is the length of the *i*-th repeat from the group of AT-rich or CG-rich repeat database of size *n*, $|feature_j|$ is the length of the *j*-th feature from the particular database of genomic features of size *m*. In detail, feature span corresponds

A

	AT repeats (>60% content)	CG repeats (>60% content)	fold difference (AT/CG)
Total number in the genome	331293	11582	29
Length (bp)	6 – 6490	6 – 997	up to 6.5
Copy number	0.85 – 3245	0.75 - 396	up to 8.19

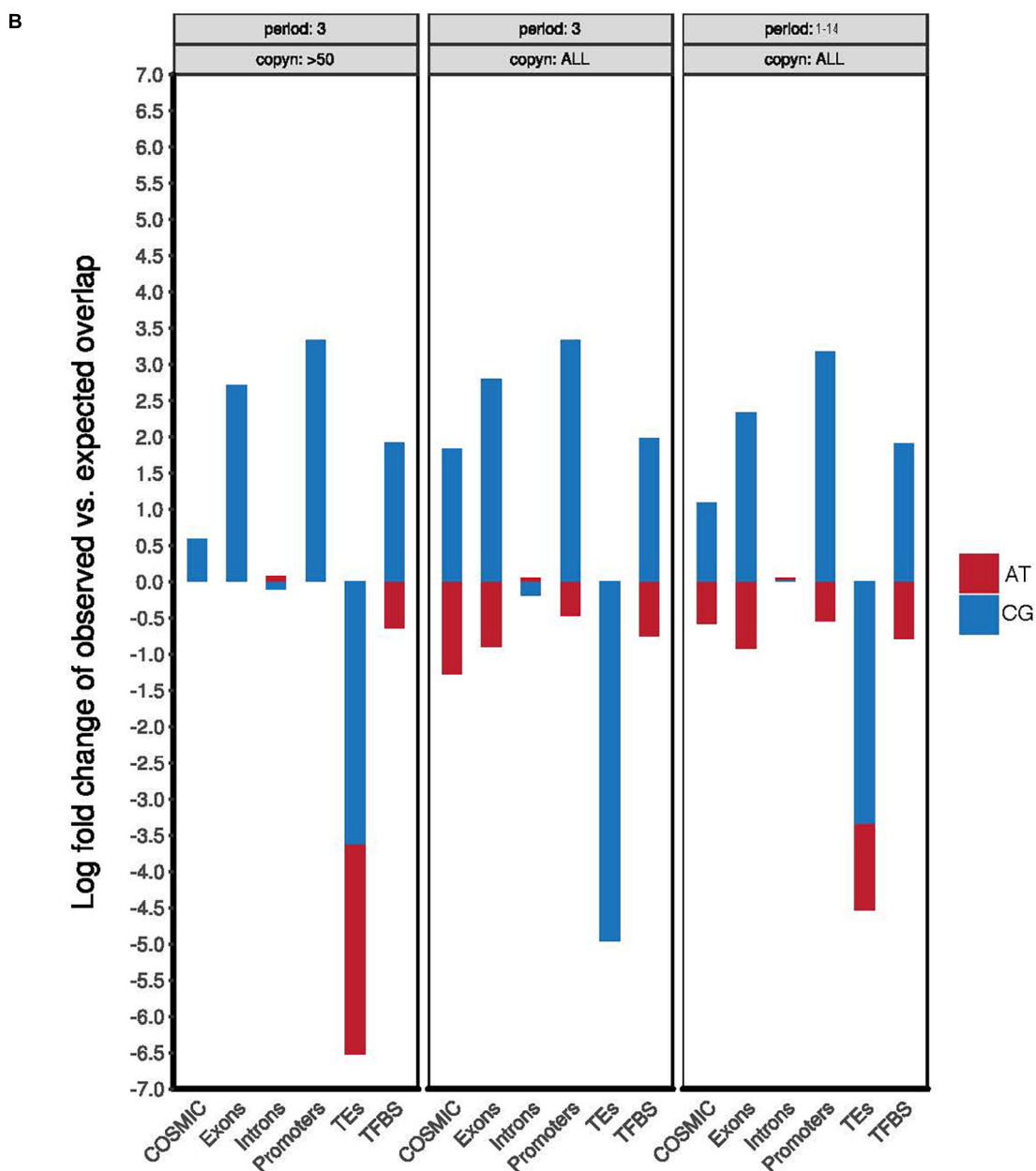


FIGURE 1 | Continued

FIGURE 1 | CG-rich TNRs are frequently located at “functional genomic regions.” **(A)** Summary of the total number, variation of length, and copy number of AT and CG repeats in human genome derived from **Supplementary Table 1**. **(B)** Quantification of the observed AT-rich (red) or CG-rich (blue) repeats in selected genomic features using theoretically predicted distribution frequency as a baseline. The quantification was performed respectively for repeats of period of 3 bp and with more than 50 copies (left), repeats of period of 3 bp (middle), or repeats of period length ranging from 1 to 14 bp (right). COSMIC is the Catalogue of Somatic Mutations in Cancer database. Copyn: copy numbers. TES: transposable elements. TFBS: transcription factor binding sites.

to $\sum_{j=1}^m |feature_j|$ and repeat span corresponds to $\sum_{j=1}^n |repeat_j|$. These are both normalized to the length of the genome ($|genome|^2$ in the denominator). Repeatspan refers to the total lengths for all such repeats within the genome. Featurespan is the total length for all such features across the genome. Nonrandom association of specific features with repeat categories was assessed based on observed association compared to an expectation of random distribution. Subsequently the expected and observed representations of overlapped genomic features with repeats were analyzed using ggplot2 and displayed as Log fold changes.

Genome Sequencing and Data Analysis

Secondary generation of whole-genome sequencing

Two DNA samples extracted from GM09237 cell line cultured with either normal medium (untreated; labeled as “YL2.1.18”) or medium with no folic acid for 5 days (no-folate-5-day; labeled as “YL2.2.18.mixed”) were sequenced by BGI PE100 platform (BGI whole genome 100 bp paired-end sequencing 60×; BGISEQ PE100). These two whole-genome sequencing (WGS) data can be accessed at: <https://ega-archive.org/datasets/EGAD00001007732>. WGS data quality was assessed using FastQC (available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2015), “FastQC,” <https://qubeshub.org/resources/fastqc>), MultiQC (Ewels et al., 2016) and Fastq Screen (Wingett and Andrews, 2018). New WGS data were trimmed using “fastp” version 0.20.0 (Chen et al., 2018) with the settings of “detect_adapter_for_pe -correction -trim_poly_x -cut_tail -trim_front1 2 -trim_front2 2,” and mapped using “bowtie 2” (version 2.2.9) (Langmead and Salzberg, 2012) with the setting of “local,” and converted to bam files using Samtools (version 1.10) (Li et al., 2009) with the settings of “view - Sb.” The WGS data for the U2OS cell line was obtained from Akan et al. (2012) as bam files and assembled accordingly.

Evaluation of CG content in the genome

While optimizing the mapping at the hard to map regions including FOLD1 and FRAXA, we tested a range of different mapping settings including “-k 2” or “-k10000” to allow more multimapping regions in the output as well as “bwa mem” (Li, 2013). In addition, we generated fastq files consisting of 100 bp pseudoreads based on the reference genome sequence of the regions and their surroundings. While the pseudoreads could be mapped with high success-rates using the standard pipeline described above, none of the more permissive settings resulted in mapped reads within the regions. We therefore concluded that the lack of reads mapped to FOLD1 is not due to repetitive sequences or low “mappability” in general, but likely due to the exceptionally high GC content of the locus (92.6%), which may inhibit its PCR amplification, cluster formation in the flow cell and/or the sequencing reaction. Indeed,

under standard library preparation procedures, the abundance of regions with a GC content > 65% are reduced about hundred times compared to mid-GC reference regions after just ten PCR cycles (Aird et al., 2011).

To validate this in detail, FASTA files of hg38 genome sequence were downloaded from UCSC (Karolchik et al., 2004) and used to derive a bedgraph file with the C/G percentage in 100 bp in windows, as well as bed files with coordinates for each CpG dinucleotide or GCC/GGC trinucleotide. These files and bam files with mapped sequencing reads were imported as “datasets” in EaSeq (Lerdrup et al., 2016) using default settings, except for WGS data, where the option to filter for unique reads was disabled. A list of chromosome lengths in hg38 was imported into EaSeq as a Regionset and used to generate 1,000 bp windows covering the entire reference genome using the “Modify”-tool in the Regionset-panel with the setting “Homogenize and fragment regions to 1,000 bp.” The number of CpG and GCC/GGC coordinates as well as average C/G-percentage in each window was obtained for this set of regions using the “Quantify”-tool with all normalization options disabled. Genome-wide data from this set of regions were visualized using the “Scatter” plot type in EaSeq, with scales adjusted as shown in the plots and 100 bins on each axis. Tracks were visualized using the “FillTrack” plot type with Y-axis settings adjusted as indicated in the figures and with normalization and smoothing disabled. Plots were exported as pdfs from EaSeq and lay-outed in Adobe Illustrator CS6.

Third generation of WGS

Two DNA samples extracted from GM09237 cell line cultured with either normal medium (untreated; labeled as “YL2.1.18”) or medium with no folic acid for 5 days (no-folate-5-day; labeled as “YL2.2.18.mixed”) were sequenced by PacBio Sequel system (SMARTbell; performed by BGI). Raw data of this sequencing can be accessed at: <https://ega-archive.org/datasets/EGAD00001007732>. Long reads were mapped using two different algorithms: Minimap2 version 2.17 with the settings -ax map-pb (Li, 2018) and NGMLR version 0.2.7 (Sedlazeck et al., 2018). Reads which mapped using either algorithm were further excised from SAM output using awk and realigned for visualization using BLAST and Sequencher (Gene Codes Corporation, United States).

RESULTS

CG-Rich TNRs Are Frequently Associated With “Functional Genomic Regions”

To examine whether folate deprivation might adversely affect CG-rich regions of the human genome that have not been implicated as disease-causing, we performed a bioinformatic

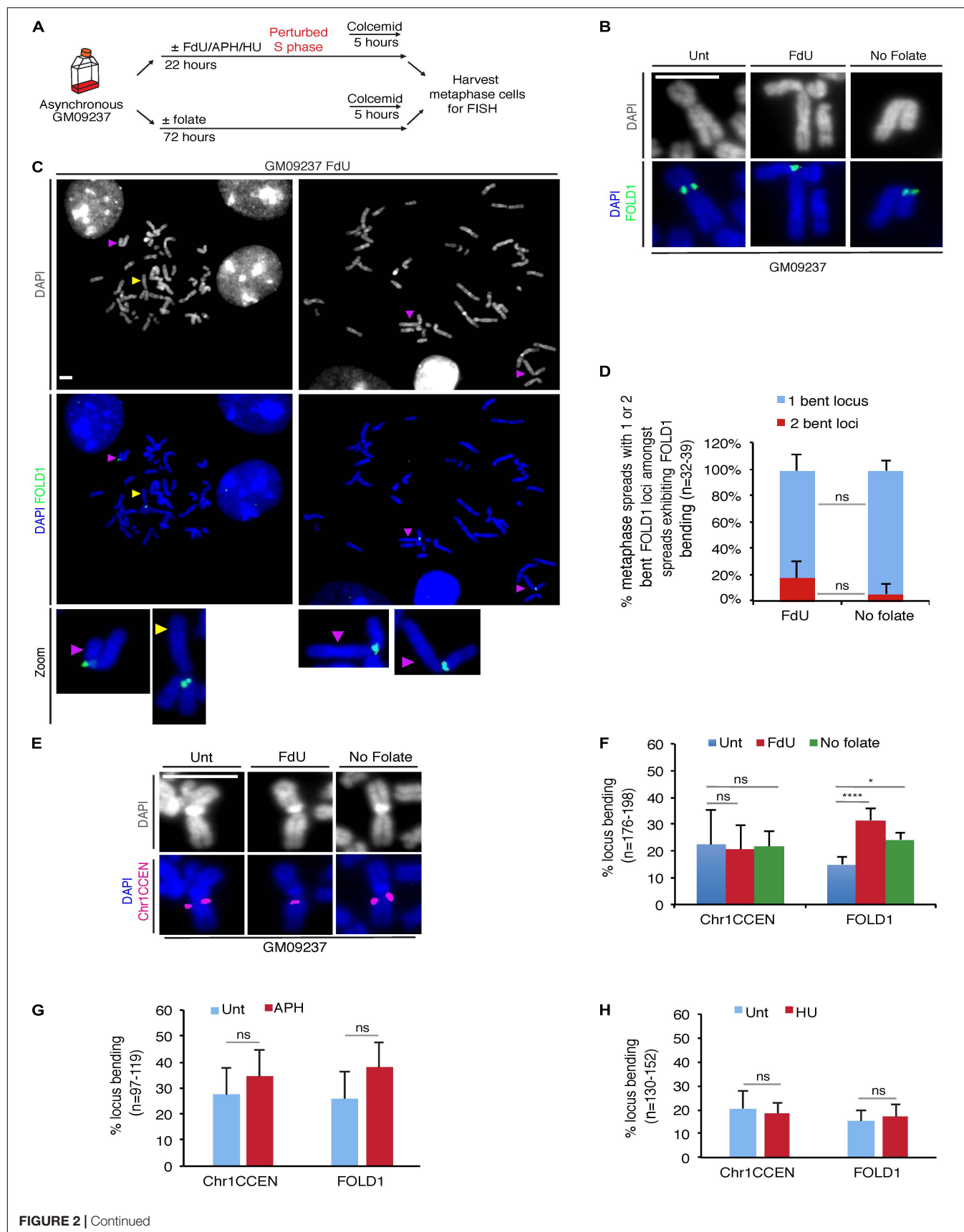


FIGURE 2 | Chr2 exhibits a “kink” at FOLD1 following FdU treatment or folate deprivation. **(A)** Experimental workflow for the analysis of metaphase cells following FdU treatment for 22 h (FdU) or folate deprivation for 3 days (No folate) in GM09237 cells. **(B)** Representative images of kinked Chr2 at FOLD1 (green) in GM09237 cells. **(C)** Representative images and **(D)** quantifications of metaphase spreads with either 1 or 2 bent Chr2 at FOLD1 (green) in cells exhibiting Chr2 bending. An “un-bent” Chr2 is indicated by a yellow arrowhead, and “bent” Chr2s are indicated by purple arrowheads. Selected chromosomes are shown as zoomed images below. *N*: number of metaphase spreads analyzed. **(E)** Representative images of Chr1 at Chr1CCEN. The location of Chr1CCEN (magenta) is revealed by FISH analysis. **(F)** Quantifications of kinked Chr2 at FOLD1 or Chr1 at Chr1CCEN in GM09237 cells. *N*: number of Chr2s or Chr1s analyzed. **(G,H)** Quantification of kinked Chr2 at FOLD1 or Chr1CCEN under aphidicolin (APH) or hydroxyurea (HU) in GM09237 cells. *N*: number of Chr2s or Chr1s analyzed. Scale bar, 5 μ m. Data are means of at least three independent experiments. Error bars represent SD (ns, not significant; * $p < 0.05$; **** $p < 0.0001$).

analysis of the GRCh38/hg38 human genome assembly¹ to identify regions characterized by the presence of extended AT-rich or CG-rich simple repeats. In agreement with earlier work on the composition of the human genome (Burge et al., 1992; Nakashima et al., 1997; Lander et al., 2001; Venter et al., 2001), our analyses illustrated two marked differences between AT-rich and CG-rich repeats: (1) the number of AT-rich repeats in the human genome far exceeds that of CG-rich repeats (nearly 30-fold higher); (2) AT-rich repeats display a much wider range of repeat sequence classes and copy numbers than that of the CG-rich repeats (**Supplementary Table 1** and **Figure 1A**). This relatively lowered abundance and diversity of CG-rich repeats likely reflects the deamination of 5-methylcytosine during evolution (Tomso and Bell, 2003; Fryxell and Moon, 2005; Walsh and Xu, 2006). When focusing on the TNR sequences we, however, found that the copy numbers of AT-rich or CG-rich repeats was very similar, ranging from 1 to 161 or 1 to 332, respectively (**Supplementary Table 1**). This suggested that CG-rich TNRs have been retained during evolution. Possibly, this reflects codons or the local aggregation of CpG-dinucleotides in thousands of CpG-islands, which are characterized by the overall low cytosine methylation levels, and the consequent protection from 5-methylcytosine deamination (Gardiner-Garden and Frommer, 1987; Antequera et al., 1989; Deaton and Bird, 2011; Long et al., 2016). In view of this, we compared the frequency of AT- or CG-rich repeats located in functional genomic regions, e.g., exons, promoters, TEs, TFBS, and regions mutated in cancer (COSMIC) (**Figure 1B**). This analysis showed that all of the CG-rich simple repeats, including TNRs, are more significantly enriched in functional regions including exons, promoters, TFBS, and regions mutated in cancer (**Figure 1B** and **Supplementary Tables 1, 2**). This is consistent with the notion that GC-rich sequences are preferentially retained in gene-rich and highly transcribed regions in the genome of warm-blood animals due to its thermostability, which helps in maintaining the advanced level of genomic organization in these animals (the “transcription/grade” concept) (Vinogradov, 2003). Taken together, this analysis indicates that CG-rich TNRs have a similar copy number range to that of AT-rich TNRs, and are enriched in functional genomic regions in the human genome.

Folate Deficiency Causes “Bending” at Chr2p11.2, FOLD1

Amongst the CG-rich TNR sequences, we identified a region at Chr2p11.2, that we termed FOLD1, which contains the

largest cluster of CG-rich TNRs (~333) in the human genome (**Supplementary Table 1**). To test our hypothesis that folate deficiency could affect the stability of regions containing large number of repeats, we started our analysis with a FXS mutant cell line GM09237, which is known to contain more than 900 CGG repeats at the FRAXA locus (Vincent et al., 1991; Seneca et al., 2012) and display missegregation in mitosis when it is cultured under folate stress conditions (Bjerregaard et al., 2018). First of all, we confirmed that either of the two types of widely-used folate stress-inducing treatments could induce fragility at FRAXA. For this, we treated the cells with FdU (a thymidylate synthase inhibitor that disrupts thymidine production (Wilson et al., 2014)) for 22 h, or deprived the cells of folate for 3 days (Bjerregaard et al., 2018), and then harvested metaphase cells for FISH using DNA probes targeting the FRAXA locus (**Supplementary Figure 1**). We then performed the same analysis with a DNA probe targeting the FOLD1 region (**Supplementary Figure 1A** and **Supplementary Text 1**). Our results showed that, although FOLD1 region did not exhibit a gap or break as observed at fragile sites, there was a recognizable structural abnormality at this locus in response to folate stress, which took the form of a “bending” or “kink” centered at FOLD1 with an angle less than 90°. We have subsequently quantified this bending shape of Chr2 in metaphase spreads obtained from different samples. Our results indicated that this unusual bending of Chr2 is consistently associated with folate stress (**Figures 2A,B,F**). In addition, we could observe that, amongst the metaphase spreads exhibiting Chr2 bending, a proportion of them have both copies of Chr2 showing distinct bending (19% versus 6% in “FdU” and “No-folate” conditions, respectively; **Figures 2C,D**). This suggests that the bending of Chr2 is not exclusive to either the maternal or the paternal copy of Chr2. Given that bending occurs on approximately 30% of the copies of Chr2, it would be expected statistically that, in average, 9% of the cells would show bending on both copies of Chr2, as was observed.

To exclude the possibility that this phenotype was due to the large size of Chr2, we studied a region on the largest human chromosome, Chr1. This region is located at Chr1p12-21 (henceforth referred as Chr1CCEN) and lies proximal to the centromere on the shorter arm of Chr1 at a very similar position to that of FOLD1 on Chr2 (**Supplementary Figure 1A**). However, Chr1CCEN differs from FOLD1 in lacking any long CG-rich TNRs. We observed that, unlike FOLD1, the Chr1CCEN region was not associated with any consistent structural abnormality in response to folate deprivation or FdU treatment, suggesting that specific forms of chromosome “bending” are not a universal feature of large

¹ https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.26/

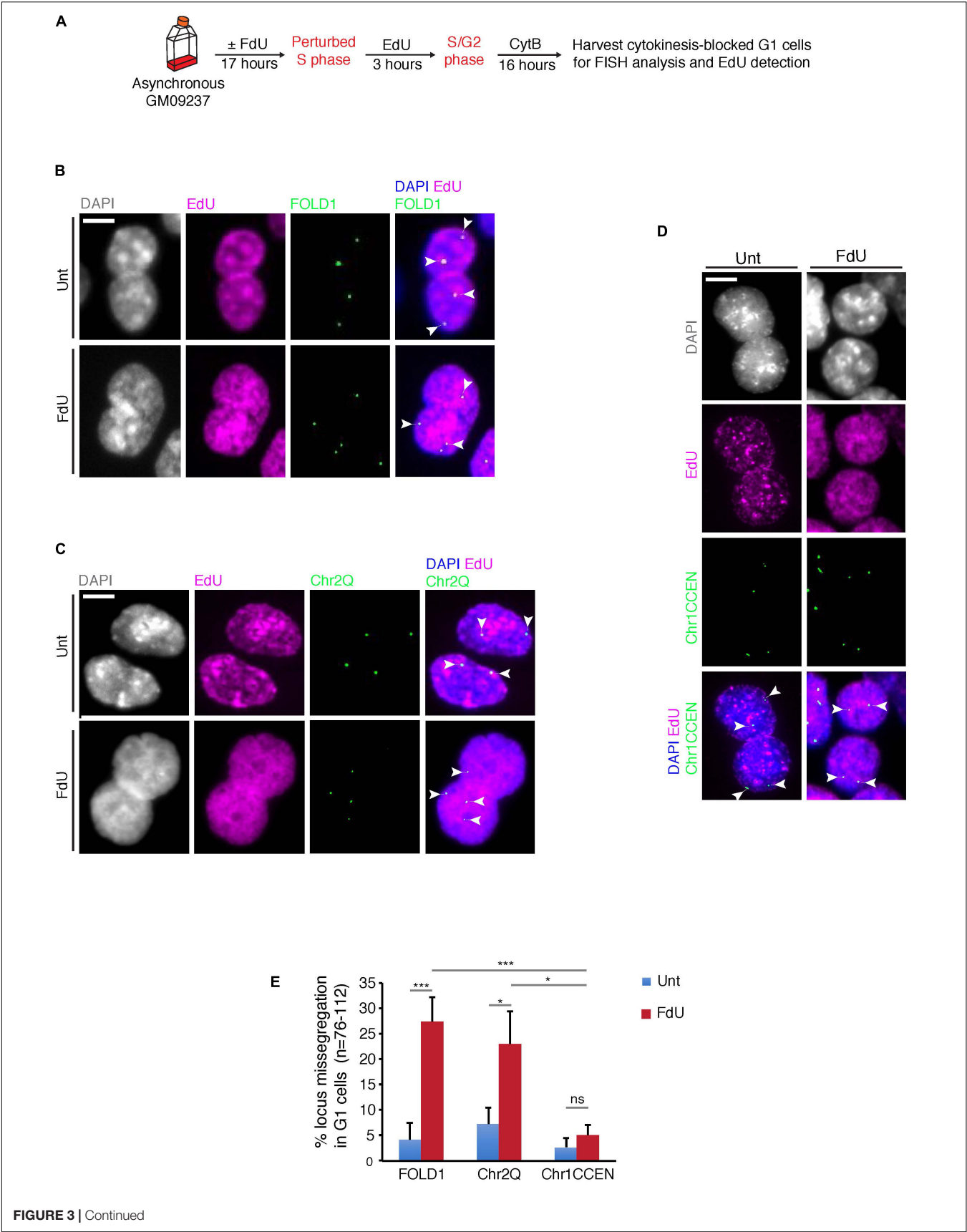


FIGURE 3 | FOLD1 displays frequent nondisjunction in G1 cells following FdU treatment. **(A)** Experimental workflow for quantifying FOLD1 nondisjunction in newly-born G1 cells following FdU treatment for 17 h in GM09237. **(B–D)** Representative images and **(E)** quantification of cytokinesis-blocked binucleated (G1) cells following FdU treatment of GM09237 cells. White arrowheads in panels **(B–D)** denote the location of FISH probes. *N*: number of pairs of newly-born G1 cells analyzed. Scale bar, 5 μ m. Data are means of at least three independent experiments. Error bars represent SDs (* $p < 0.05$; *** $p < 0.005$).

chromosomes *per se* (Figures 2E,F). Furthermore, we could confirm that Chr2 bending at FOLD1 was not induced by APH or HU (Figures 2G,H), which are known to perturb replication and induce common fragile site (CFS) (Glover et al., 1984) or early replicating fragile site (Barlow et al., 2013) expression, respectively.

We subsequently performed the same analysis on two more cell lines: a non-cancerous fibroblast cell line (Hs68) and an osteosarcoma cell line (U2OS). Our results indicated that, in these two cell lines, under folate stress conditions, Chr2 is also prone to bend at FOLD1, while Chr1 does not show increased bending at Chr1CCEN, which is similar to what we observed in the GM09237 cell line (Supplementary Figures 2A,B). We have also carried out conventional karyotyping to study the shape of Chr2 and Chr1 in the GM09237 and Hs68 cell lines, both of which are non-cancerous and diploid. The data from this analysis indicated that Chr2 indeed has a more pronounced bending shape when the cells are cultured under folate stress conditions, which is not the case for Chr1s or Chr3s, the two other largest human chromosomes (Supplementary Figures 2C,D).

Next, we asked whether MiDAS, which serves as a rescue pathway to complete locus duplication in mitosis, and is known to occur at CFSs (Minocherhomji et al., 2015), telomeres (Minocherhomji et al., 2015; Dilley et al., 2016; Min et al., 2017; Ozer et al., 2018) and FRAXA (Garribba et al., 2020), could occur at FOLD1 following folate stress treatment. To address this, we investigated whether mitotic EdU incorporation could be observed at the sites of bending FOLD1 following folate stress using an established protocol (Garribba et al., 2018) in GM09237 cells (Supplementary Figure 3). Although we could indeed see MiDAS at FRAXA as we observed previously, we could not detect MiDAS at FOLD1 after examining more than 150 metaphase spreads obtained from three replicate experiments. This suggests that, unlike FRAXA, the FOLD1 region is most likely fully replicated before mitosis.

FOLD1 Is Missegregated Under Folate Stress Conditions

Next, we investigated whether, similar to FRAXA (Bjerregaard et al., 2018), FOLD1 is missegregated in mitosis under folate stress conditions. To this end, we treated cells with FdU for 17 h, or deprived them of folate for 3 days, synchronized them in late G2 with the CDK1 inhibitor RO3306, and then released them into mitosis to harvest anaphase/telophase cells (Supplementary Figure 4A). We observed that FOLD1 was unevenly segregated in over 20% of the anaphase cells in response to folate stress but not under control conditions (Supplementary Figures 4B,C). To define whether this missegregation was in some way selective for FOLD1 or whether instead could affect the entire Chr2, we

used a previously described FISH-based protocol to examine the fate of two regions on Chr2: FOLD1 and a region on the long arm of Chr2 (henceforth referred as Chr2Q), in cytokinesis-blocked binucleated cells (pseudo-G1 cells) (Figure 3A and Supplementary Figure 1A). We observed a fivefold increase in the frequency of nondisjunction of both FOLD1 and Chr2Q, indicative of whole Chr2 missegregation within one cell cycle of FdU treatment (Figures 3B,C,E). To evaluate the frequency of nondisjunction in other genomic regions that do not contain CG-rich TNRs, we conducted the same analysis with a FISH probe targeting the Chr1CCEN region. Our data indicated that the segregation of Chr1CCEN is not significantly affected by FdU treatment (Figures 3D,E). Intriguingly, and in contrast to what we observed previously through the analysis of FRAXA (Bjerregaard et al., 2018), FOLD1 did not localize to DNA bridges or laggards in anaphase cells (Supplementary Figure 4), or to micronuclei in the pseudo-G1 cells following the folate stress treatment (Figure 3).

Long-Term Folate Deprivation Causes Chr2 Aneuploidy

Our previous study reported severe FRAXA and ChrX instability in response to prolonged folate deprivation (Bjerregaard et al., 2018). To investigate whether this was also the case for FOLD1 and Chr2, we cultured the cells in the absence of folate for 5 days and analyzed the segregation of FOLD1 using FISH probes targeting FOLD1 or Chr1CCEN in anaphase cells (Supplementary Figure 5A). We observed that FOLD1 was either lost or missegregated in over 25% of the cells, while Chr1CCEN was segregated normally (Supplementary Figures 5B,C). Intriguingly, we could not observe FOLD1 located in between the two separating daughter DNA masses, indicating that the missegregation of FOLD1 was caused by the nondisjunction of two sister chromatids. Nonetheless, these findings are consistent with the observed missegregation of FOLD1 in cells treated with FdU or cultured in the absence of folate for 3 days (Supplementary Figures 4B,C).

To define more specifically whether the abnormal segregation of FOLD1 was associated with whole or partial Chr2 missegregation, we analyzed metaphase chromosome preparations from cells that had been cultured for 5 days in the absence of folate (Supplementary Figure 5D and Figure 4A). Using FISH probes targeting FOLD1, Chr2Q or Chr1CCEN in the metaphase spread analysis, we observed that FOLD1 and Chr2Q were frequently lost or gained (Figures 4B,C), while the Chr1CCEN region did not show an increased frequency of copy number changes upon prolonged folate deprivation (Supplementary Figures 5E,F). Taken together, these results demonstrate that long-term folate deficiency has a specific effect on the frequency of Chr2 aneuploidy (Figure 4).

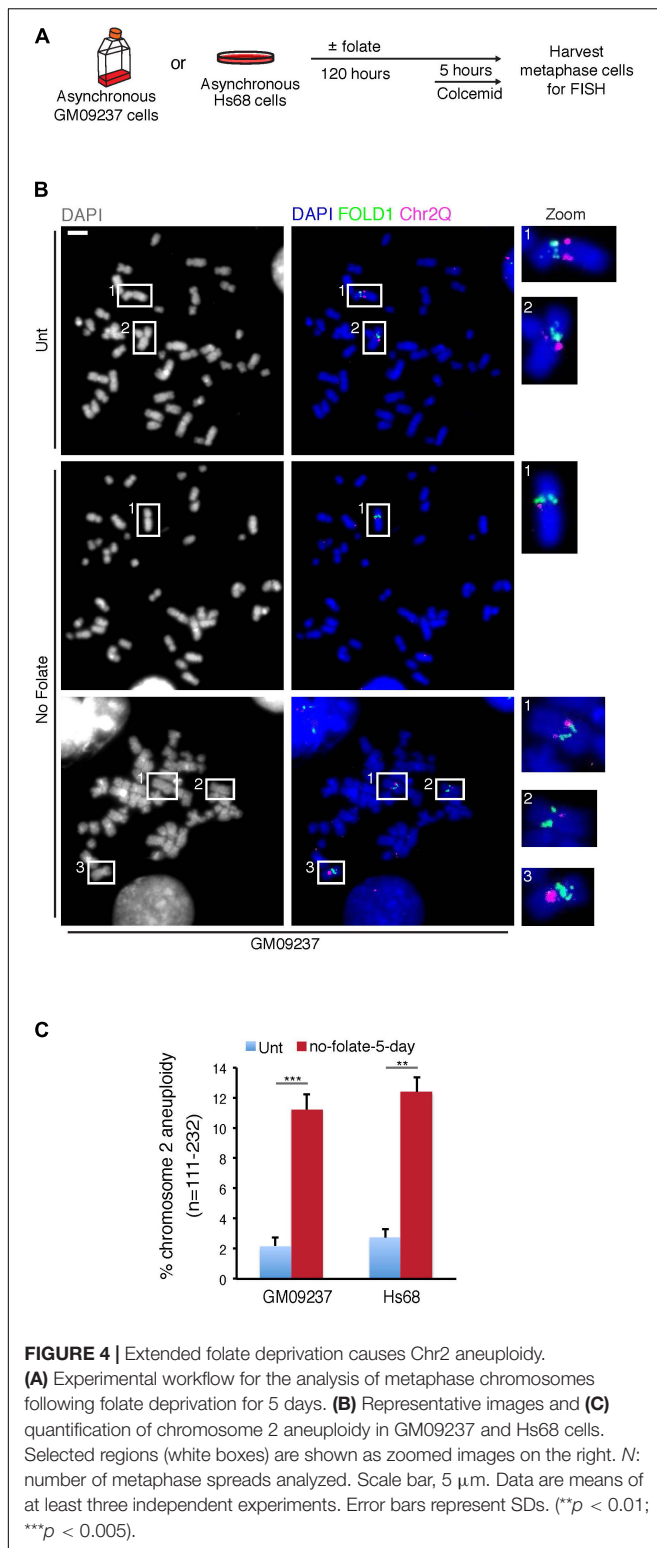


FIGURE 4 | Extended folate deprivation causes Chr2 aneuploidy.

(A) Experimental workflow for the analysis of metaphase chromosomes following folate deprivation for 5 days. **(B)** Representative images and **(C)** quantification of chromosome 2 aneuploidy in GM09237 and Hs68 cells. Selected regions (white boxes) are shown as zoomed images on the right. *N*: number of metaphase spreads analyzed. Scale bar, 5 μ m. Data are means of at least three independent experiments. Error bars represent SDs. (** $p < 0.01$; *** $p < 0.005$).

Sequence Analysis of the FOLD1 Region

To verify the sequence of FOLD1 region, we attempted to use PCR to amplify and then sequence this region using the Sanger method. However, after numerous attempts, we could not

obtain reasonable PCR products to pursue Sanger sequencing. We therefore performed shot-gun WGS and long-range PacBio sequencing of DNA samples extracted from GM09237 cells cultured with either normal medium (untreated) or medium with no folate for 5 days (no-folate-5-day).

The data generated from shot-gun WGS demonstrated that, first, it was not possible to locate repeat rich sequences at the central part of the FOLD1 region in either of the samples, despite the fact we could successfully map the “pseudoreads” generated from the reference genome at this region using our algorithm settings (**Figures 5A–C**). We reasoned that this is because of the high CG content of the sequences of this region, which could inhibit the amplification and sequencing reactions during the sequencing process. We have also analyzed the FRAXA locus in the same fashion. We found that it is impossible to locate the CGG repeat sequences originating from the central part of the FRAXA region in the two samples either (**Figures 5A–C**). Second, the sequences of the reads that did map to the region and possible unmapped overhangs were very similar in the untreated and no-folate samples (**Figure 5D**). Therefore, there was no indication of obvious deletions or insertions surrounding the FOLD1 region in the “no-folate-5-day” samples, which is consistent with our data of MiDAS not occurring at FOLD1 (**Supplementary Figure 3**). Furthermore, we investigated whether the sequencing data mapped to these regions in our study were equal to those obtained in other WGS data. For this, we analyzed published WGS data of U2OS, A431, and U251MG cells (Akan et al., 2012) in the same way as above. Our analysis confirmed that, indeed, standard shot-gun WGS could not provide the detailed sequences in the FOLD1 or FRAXA regions in these cell lines either (**Figure 5E**).

The data generated by PacBio sequencing corresponds to 7 \times or 4 \times coverage for untreated or no-folate-5-day samples respectively (using hg38 as the reference genome). Because of the low coverage read, we could only locate 10 reads partially or fully covering the FOLD1 region, amongst which 4 reads cover the whole FOLD1 region with 60–82% match (**Supplementary Table 3** and **Supplementary Figure 6**). In general, the reads from either untreated or no-folate-5-day samples contained various mismatched sequences, which were single nucleotide changes, small deletions or insertions. It is well-known that PacBio Sequel sequencing has frequent read errors. This is usually mitigated by making consensus sequences from a substantial number of reads, which cannot be achieved in our case due to the low magnitude of sequencing. Therefore, we could not validate whether the mismatches in the Pacbio reads mapped to FOLD1 were due to sequencing errors, or the actual mutations caused by folate deficiency. In the meantime, as a control, we could identify four reads mapped to the FRAXA region, with one read covering the whole the FRAXA repeat locus (**Supplementary Table 3** and **Supplementary Figure 6**). Our data indicated that the FRAXA locus in GM09237 cells has 931 CGG repeats, which is consistent with previous reports based on Southern blot analysis or allele specific PCRs (Vincent et al., 1991; Seneca et al., 2012).

We conclude that it is very difficult to PCR amplify or obtain the sequences of regions harboring long CG rich repeats (e.g., FOLD1 or FRAXA) using shot-gun WGS. Instead,

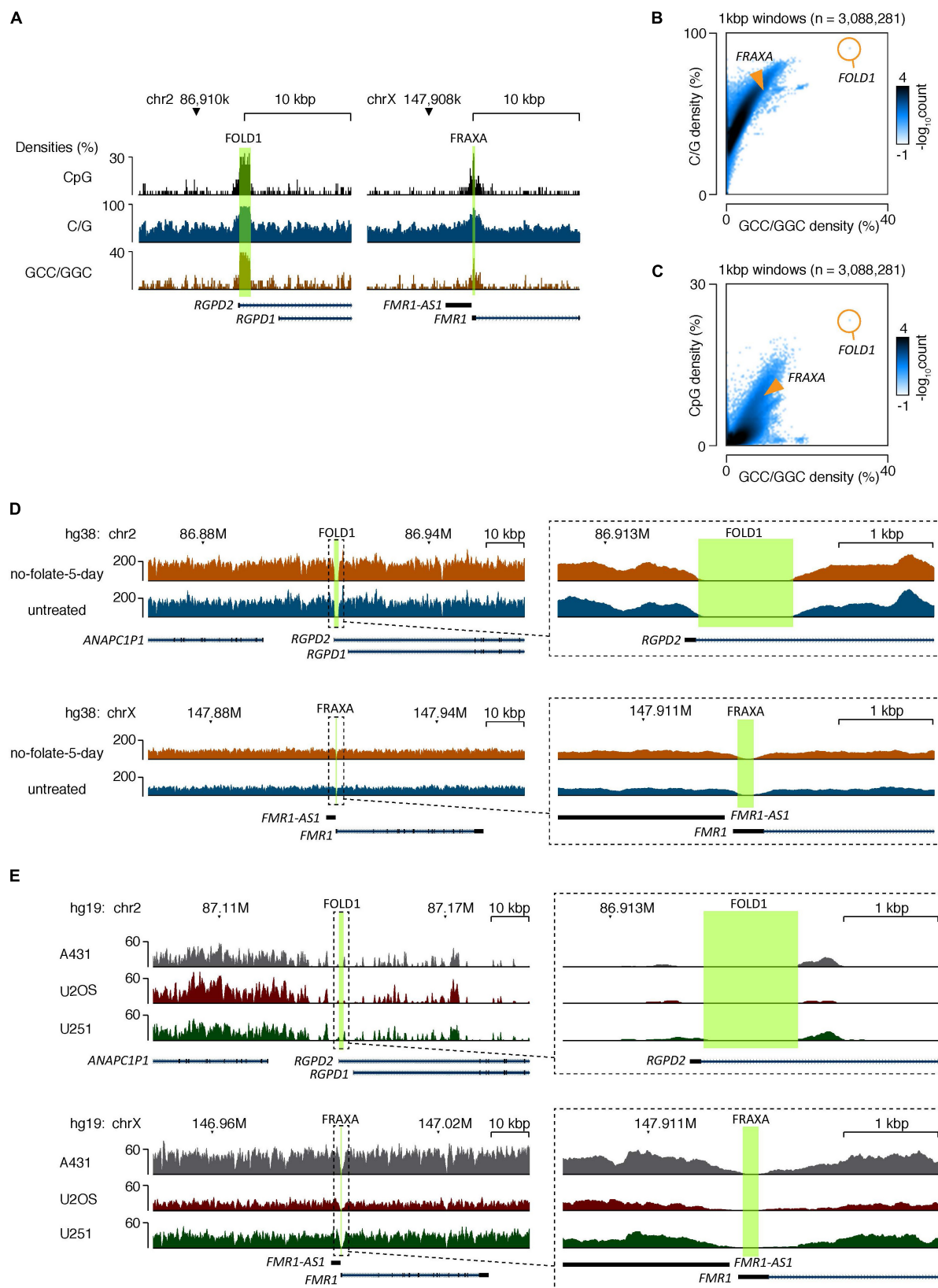


FIGURE 5 | Continued

FIGURE 5 | The CG-rich repeats at FOLD1 or FRAXA regions are difficult to be sequenced by second generation of sequencing due to high C/G content.

(A) Genome browser tracks showing the densities of CpG dinucleotides (top), G/C mononucleotides (middle), and GCC/GGC trinucleotides (bottom) at the FOLD1 (left) and FRAXA (right) regions. The hg38 reference database was used. **(B,C)** 2D-histograms showing the genome-wide relationship between densities of G/C mononucleotides **(B)** or CpG dinucleotides **(C)** on Y-axes and GCC/GGC trinucleotides on X-axes. Genomic sequences were analyzed in 1 kbp windows, and the number of windows with a particular combination are illustrated by the coloring. The composition of the FRAXA and FOLD1 regions is illustrated with orange overlays. The hg38 reference database was used. **(D)** Genome browser tracks showing the densities of mapped reads from whole-genome sequencing (WGS) of GM09237 cells treated with no folate for 5 days (orange) and untreated cells (blue) at the FOLD1 (top) and FRAXA (bottom) regions. The areas covered by the regions are highlighted in green, and the right-side inserts show zoomed tracks of the regions and their immediate surroundings. The hg38 genome database was used as reference. **(E)** Genome browser tracks showing the densities of mapped reads at the FOLD1 (top) and FRAXA (bottom) regions based on published WGS of A431, U2OS, and U251 cell lines. Data processing and mapping setting was preserved from the published data (Akan et al., 2012). The hg19 genome database was used as reference. The areas covered by the regions analyzed are highlighted in green. The right-side inserts show zoomed tracks of the regions and their immediate surroundings.

deep, long-range sequencing (such as Pacbio) is required to verify the sequences at these regions. In addition, based on our Pacbio sequencing data and the Chr2 analysis via FISH (Figure 2), where a portion of cells have two “bent” Chr2s upon folate stress, we propose that GM09237 cell line has homozygous FOLD1 sequence.

DISCUSSION

Based on the previous knowledge of fragility and missegregation at the abnormally expanded FRAXA locus under folate stress (Santoro et al., 2012; Bjerregaard et al., 2018; Garribba et al., 2020), we investigated whether folate deficiency could perturb other genomic regions containing CG-rich repeat sequences. To address this, we analyzed the stability of the FOLD1 region on Chr2, which contains the largest cluster of CG-rich TNRs present in the human genome under folate stress conditions. By analyzing metaphase chromosomes, we observed a significant increase in the frequency of “bending” or “kinking” (but not “gap” or “break”) of Chr2 at the FOLD1 region in response to folate deprivation. To our knowledge, this phenotype has not been described previously in the human genome. Importantly, this effect was not due to the large size of Chr2, as this phenomenon was not observed at a region close to the centromere of Chr1, the largest human chromosome. We could also confirm this “bending Chr2” phenotype in two other human cell lines (Hs68 and U2OS).

Moreover, the “bending” phenomenon was most frequently observed at FOLD1 under conditions of folate stress, but very rarely in untreated cells or cells exposed to other replication-stress-inducing agents such as APH or HU. Thus, the bending at FOLD1 appears to be induced specifically by folate stress. Taking into the account the knowledge gained from previous studies on CFSs and FRAXA, we propose that this increased frequency of “bending” of Chr2 at FOLD1 is likely to be driven by the under-condensation of the DNA due to delayed replication at this region under folate stress condition. It is possible that, because CG-rich TNRs tend to form DNA secondary structures (Fry and Loeb, 1994; Usdin and Woodford, 1995) and assemble into nucleosomes with low efficiency after replication (Wang et al., 1996), folate stress exacerbates the replication-related difficulties of the FOLD1 region, thus causing its delayed condensation in early mitosis under this condition. Nevertheless, we could not rule out the possibility that folate deficiency could cause

changes in replication timing and transcription of genes located in the FOLD1 region (e.g., the *RGPD2* gene), which could then contribute to the delayed condensation at this region. Future work is needed to validate these possibilities.

Interestingly, our data also suggested that FOLD1 differs from FRAXA in that it is fully replicated when cells enter mitosis. This is based on the following observations: (1) FOLD1 does not localize to DNA bridges/laggards in anaphase, or to micronuclei in the newly-born G1 cells; (2) MiDAS is apparently absent from bent FOLD1; and (3) both sister chromatids look very similar at bent FOLD1, whereas the expanded FRAXA locus generally shows the conventional appearance of fragility, with a break/gap on a single sister chromatid only (Bjerregaard et al., 2018). To explain this apparent difference, it is important to note that the CG-rich TNRs at FOLD1, despite being over 900 nucleotides in length, are interrupted by A/T base pairs, with the longest continuous CGG repeat cluster comprising only 17 repeat units (Supplementary Text 1). It was shown previously that interruptions in the TNR sequence suppress repeat instability (Yrigollen et al., 2014), and more than 100 continuous CGG repeats are generally required to cause replication fork blockade (Voineagu et al., 2009). It is likely, therefore, that the CG-rich repeats at FOLD1 lead only to a temporary replication fork perturbation, but not an extended fork blockade. In agreement with this, in response to FdU or 3-day folate deprivation, although FOLD1 was frequently missegregated, we could always observe four copies of this region (in total) distributed between the two daughter cells (Supplementary Figure 4B and Figure 3B), which was not the case for FRAXA (Bjerregaard et al., 2018). Taken together, we propose that the bending (or “kink”) at FOLD1 simply reflects the perturbed condensation of a region due to delayed replication in interphase under folate stress conditions. In comparison with FRAXA, which appears fragile in metaphase and localizes on DNA bridges in anaphase upon folate stress (Bjerregaard et al., 2018), FOLD1 displays a milder phenotype in this respect.

Despite the fact that FOLD1 appears less susceptible to folate stress than the expanded FRAXA locus, we observed marked missegregation of the FOLD1 region, as well as the whole Chr2. In this regard, it might be significant that FOLD1 is adjacent to the Chr2 centromere, which might influence its mobility in mitosis. For example, as discussed above, the delayed condensation of FOLD1 might influence accurate microtubule

attachment or the release of cohesin from Chr2 centromeres, which would then lead to nondisjunction of sister chromatids of this chromosome. A more direct effect of FOLD1 is also possible, based on the likelihood that unfinished replication of this locus might precipitate a failure to decatenate the two Chr2 sister-chromatids in a timely fashion, leading to Chr2 non-disjunction. Due to limited type and number of cells analyzed, it remains to be clarified whether the frequencies of missegregation of FOLD1 region or the whole Chr2 is the same in a larger collection of different cell types. Nonetheless, in line with our data on FOLD1's copy number changes, it was observed that there is an increase of copy number of Chr2p11.2 in several tumor types (Mastracci et al., 2006; Oh et al., 2012). In addition, Chr2p11.2 is also the location of frequent deletions or insertions, some of which, particularly large deletions, are associated with human diseases (Firth et al., 2009) (**Supplementary Tables 4, 5**). Judging by the data from this study, we reckon this would most likely be caused by the mistakes occurring in S phase when this region is being replicated in the cells. Future studies are warranted to analyze the cellular functions of the genes located in this region.

In summary, we have demonstrated that nondisjunction of Chr2 is promoted by folate deficiency via a CG-rich TNR region (**Supplementary Figure 7**). Future studies are warranted to identify the proteins or pathways underlying this missegregation, and how this could lead to the altered expression of the genes located at this region. Considering that CG-rich TNRs are significantly enriched at genomic regions associated with the regulation of gene expression, our findings provide a new line of evidence to explain why folate deficiency is associated with a wide range of human disorders including cancers. In addition, because such chromosomal changes induced by folate deficiency could potentially be irreversible, our results highlight the importance of maintaining optimal folate levels irrespective of age or gender in human population.

DATA AVAILABILITY STATEMENT

The datasets presented in the study are deposited in the European Genome-phenome Archive (EGA) repository, accession number (EGAD00001007732).

AUTHOR CONTRIBUTIONS

LG, MMGD, and LR performed the experiments and data analysis. IV and ML performed the bioinformatic analysis. LG and YL designed the experiments and interpreted results. LG, IV, ML, and YL wrote the manuscript. All authors edited it.

FUNDING

The work in YL laboratory is supported by the European Union (H2020/Marie Skłodowska-Curie Actions; 859853) and The Danish National Research Foundation (DNRF115).

ACKNOWLEDGMENTS

We thank Ian D. Hickson, Anna H. Bizard, Eva R. Hoffmann, Asli Silahtaroglu in the Department of Cellular and Molecular Medicine, University of Copenhagen (ICMM), and Zeynep Tümer in Kennedy Center, Denmark, for inspiring discussions. We thank Malgorzata Clausen (ICMM) for technical assistance and Viki Lee (BGI Hong Kong) for processing sequencing samples.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.695124/full#supplementary-material>

Supplementary Figure 1 | (Related to **Figures 2–4**) The background information of this study. **(A)** A diagram depicting the genomic location of the BAC clones used in FISH analyses. **(B)** Experimental workflow for the analysis of FRAXA fragility upon FdU treatment on metaphase spreads of GM09237 cells. **(C)** Representative images of the FRAXA locus (green) on metaphase chromosomes in GM09237 cells. The yellow arrow indicates an example of a fragile FRAXA locus. **(D)** Quantification of fragility at FRAXA following FdU treatment. *N*: number of ChrX analyzed. Scale bar, 5 μ m. Error bars represent SDs from at least three experiments. Error bars represent SD. (*****p* < 0.001).

Supplementary Figure 2 | (Related to **Figure 2**) Chr2 exhibits bending at FOLD1 in Hs68 and U2OS cell lines upon folate stress. **(A)** Experimental workflow for the analysis of metaphase cells by FISH or karyotyping following FdU treatment for 22 h (FdU) or folate deprivation for 3 days (No folate) in GM09237, Hs68, or U2OS. U2OS cells were not included in karyotyping analysis since it is an aneuploid cell line. **(B)** Quantification of Chr2 or Chr1 bending at either FOLD1 or Chr1CCEN, respectively in Hs68 and U2OS cell lines under "FdU" or "No folate" conditions. *N*: number of Chr2s or Chr1s identified and analyzed by FISH probes targeting FOLD1 or Chr1CCEN respectively. **(C)** Representative karyotype and **(D)** quantification of Chr2 bending at the Chr2p11.2 that contains FOLD1 region (arrowed) in GM09237 and Hs68 cells. *N*: number of Chr2s analyzed in karyotype analysis. Chr1 and Chr3 were analyzed in parallel. Error bars represent SDs from at least three experiments (ns, not significant; **p* < 0.05; ***p* < 0.01; *****p* < 0.001).

Supplementary Figure 3 | Mitotic DNA synthesis does not occur at FOLD1 following FdU treatment. **(A)** Experimental workflow for the analysis of EdU incorporation at FOLD1 in mitosis upon FdU treatment in GM09237 cells. **(B)** Representative images of bent FOLD1 (green) negative for EdU incorporation (magenta). The FRAXA locus was used as a positive control. Scale bar, 5 μ m.

Supplementary Figure 4 | (Related to **Figure 3**) FOLD1 missegregation in anaphase following FdU treatment or folate deprivation. **(A)** Experimental workflow for the analysis of missegregation in anaphase cells following FdU treatment for 17 h (FdU) or folate deprivation for 3 days (No folate) in GM09237 cells. **(B)** Representative images and **(C)** quantification of FOLD1 missegregation in anaphase GM09237 cells. White arrowheads indicate the location of FOLD1. *N*: number of anaphase cells analyzed. Scale bar, 5 μ m. Data are means of at least three independent experiments. Error bars represent SD (ns, not significant).

Supplementary Figure 5 | (Related to **Figure 4**) Long-term effects of folate deprivation in anaphase cells and metaphase spreads. **(A)** Experimental workflow for analysis of anaphase cells following folate deprivation for 5 days. **(B)** Representative images and **(C)** quantification of missegregation/loss of FOLD1 in GM09237 cells. The Chr1CCEN region was used as a control. White arrowheads indicate the location of FOLD1. *N*: number of anaphase cells analyzed. **(D)** Experimental workflow for the analysis of metaphase chromosomes following folate deprivation for 5 days. **(E)** Representative images and **(F)** quantification of

copy number changes of FOLD1 in metaphase chromosomes in GM09237 cells. The Chr1CCEN region was used as a control. Selected regions (numbered white boxes) are shown below as zoomed images. *N*: number of metaphase spreads analyzed. Scale bar, 5 μ m. Data are means of at least three independent experiments. Error bars represent SDs (* $p < 0.05$; *** $p < 0.005$; **** $p < 0.001$).

Supplementary Figure 6 | (Related to **Figure 5**) Examples of PacBio reads containing sequences of FOLD1 or FRAXA obtained from GM09237 cells. **(A)** A diagram of the alignment of a PacBio sequencing read (m54282_190714_174122/27394352/7413_34584), a part of Chr2 (hs38) and a FOLD1 sequence (hg38). The sequence of read m54282_190714_174122/27394352/7413_34584 matching to FOLD1 is marked in purple and shown in the lower panel. **(B)** A diagram of the alignment of a PacBio sequencing read (m54281_190714_174033/27198327/934_17319) and the sequences of 5' or 3' side of the FRAXA CGG repeats located in FMR1 exon 1 (hs38). The abnormally expanded 900 CGG repeats in the m54281_190714_174033/27198327/934_17319 read is marked in purple and shown in the lower panel. The alignment diagrams were created by Sequencher program.

Supplementary Figure 7 | A graphic summary of the key findings. When cells are grown under folate stress conditions, replication at FOLD1 is delayed, but is nevertheless completed in interphase. This, however, causes a delay in the condensation at FOLD1 in early mitosis, leading to the "bent" shape of Chr2 with the inflection point at the position of the FOLD1 region. During anaphase, the uncondensed FOLD1 region negatively influences the accuracy of sister chromatid separation, possibly due to its close association with the Chr2 centromere. This leads to FOLD1 and Chr2 nondisjunction in the newly born G1 cells. Because the copy number changes of Chr2 will likely not immediately affect cell growth, several rounds of cell proliferation under folate stress conditions can occur, leading to an increasing number of cells displaying Chr2 aneuploidy.

REFERENCES

- Aird, D., Ross, M. G., Chen, W. S., Danielsson, M., Fennell, T., Russ, C., et al. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* 12:R18. doi: 10.1186/gb-2011-12-2-r18
- Akan, P., Alexeyenko, A., Costea, P. I., Hedberg, L., Solnestam, B. W., Lundin, S., et al. (2012). Comprehensive analysis of the genome transcriptome and proteome landscapes of three tumor cell lines. *Genome Med.* 4:86. doi: 10.1186/gm387
- Altnae, S., Stavreus-Evers, A., Ruiz, J. R., Laanpere, M., Syvanen, T., Yngve, A., et al. (2010). Variations in folate pathway genes are associated with unexplained female infertility. *Fertil. Steril.* 94, 130–137. doi: 10.1016/j.fertnstert.2009.02.025
- Antequera, F., Macleod, D., and Bird, A. P. (1989). Specific protection of methylated CpGs in mammalian nuclei. *Cell* 58, 509–517. doi: 10.1016/0092-8674(89)90431-5
- Araujo, J. R., Martel, F., Borges, N., Araujo, J. M., and Keating, E. (2015). Folate and aging: Role in mild cognitive impairment, dementia and depression. *Ageing Res. Rev.* 22, 9–19. doi: 10.1016/j.arr.2015.04.005
- Bailey, L. B., Wagner, P. A., Christakis, G. J., Davis, C. G., Appledorf, H., Araujo, P. E., et al. (1982). Folic acid and iron status and hematological findings in black and Spanish-American adolescents from urban low-income households. *Am. J. Clin. Nutr.* 35, 1023–1032. doi: 10.1093/ajcn/35.5.1023
- Barlow, J. H., Faryabi, R. B., Callen, E., Wong, N., Malhowski, A., Chen, H. T., et al. (2013). Identification of early replicating fragile sites that contribute to genome instability. *Cell* 152, 620–632. doi: 10.1016/j.cell.2013.01.006
- Bjerregaard, V. A., Garribba, L., McMurray, C. T., Hickson, I. D., and Liu, Y. (2018). Folate deficiency drives mitotic missegregation of the human FRAXA locus. *Proc. Natl. Acad. Sci. U.S.A.* 115, 13003–13008. doi: 10.1073/pnas.1808371115
- Burge, C., Campbell, A. M., and Karlin, S. (1992). Over- and under-representation of short oligonucleotides in DNA sequences. *Proc. Natl. Acad. Sci. U.S.A.* 89, 1358–1362. doi: 10.1073/pnas.89.4.1358
- Supplementary Table 1 |** (Related to **Figure 1**) An excel file of the genomic information of all of the AT-, or CG- rich repeats (period ranging from 1 to 14 bp) in human genome database, and the genomic information of AT-, or CG- rich trinucleotide repeats that have more than 50 copies in human genome database. This file contains four data containing sheets. The headings for these sheets are: "AT-rich_with_features," "GC-rich_with_features," "AT-rich_period3_copynumber > 50," and "GC-rich_period3_copynumber > 50." The categories of the information included are: the location of the repeat (chromosome number, chromStart – chromEnd, cytogenetic Band of the repeat, cytoBand period length, copy Numbers of the repeat, total Length of the repeat, perfect Match percentage, sequence of the repeat region, database source), and the overlap with functional features of the genome as described in **Figure 1**. The link to access this file is at: <https://drive.google.com/open?id=1Mk3oFO4dxUNEosf-4pPPmJfzbugEi2l>.
- Supplementary Table 2 |** (Related to **Figure 1B**) The data table for the plots shown in **Figure 1B**.
- Supplementary Table 3 |** (Related to **Figure 5**) The sequencing reads generated by Pacbio sequencing containing partial or full FOLD1 or FRAXA regions. This file contains 2 data containing sheets. The headings for these sheets are: "Positive_PacBio_reads_for_FOLD1" and "Positive_PacBio_reads_for_FRAXA."
- Supplementary Table 4 |** A summary of the polymorphisms at FOLD1 (chr2:86913957-86914954) present in the general population.
- Supplementary Table 5 |** A summary of the mutations in patient reported in DECIPHER database containing FOLD1 (chr2:86913957-86914954) (DECIPHER: <https://decipher.sanger.ac.uk/>).
- Supplementary Text 1 |** Sequence of the FOLD1 CG-rich trinucleotide region (chr2:86913957-86914954; hg38). A/T interruptions are highlighted in bold.
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 34, i884–i890. doi: 10.1093/bioinformatics/bty560
- Deaton, A. M., and Bird, A. (2011). CpG islands and the regulation of transcription. *Genes Dev.* 25, 1010–1022. doi: 10.1101/gad.2037511
- Dilley, R. L., Verma, P., Cho, N. W., Winters, H. D., Wondisford, A. R., and Greenberg, R. A. (2016). Break-induced telomere synthesis underlies alternative telomere maintenance. *Nature* 539, 54–58. doi: 10.1038/nature20099
- Ducker, G. S., and Rabinowitz, J. D. (2017). One-carbon metabolism in health and disease. *Cell Metab.* 25, 27–42. doi: 10.1016/j.cmet.2016.08.009
- Durkin, S. G., and Glover, T. W. (2007). Chromosome fragile sites. *Annu. Rev. Genet.* 41, 169–192. doi: 10.1146/annurev.genet.41.042007.165900
- Ericson, U., Sonestedt, E., Gullberg, B., Olsson, H., and Wirfalt, E. (2007). High folate intake is associated with lower breast cancer incidence in postmenopausal women in the Malmö Diet and Cancer cohort. *Am. J. Clin. Nutr.* 86, 434–443. doi: 10.1093/ajcn/86.2.434
- Ewels, P., Magnusson, M., Lundin, S., and Kaller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048. doi: 10.1093/bioinformatics/btw354
- Firth, H. V., Richards, S. M., Bevan, A. P., Clayton, S., Corpes, M., Rajan, D., et al. (2009). DECIPHER: database of chromosomal imbalance and phenotype in humans using ensembl resources. *Am. J. Hum. Genet.* 84, 524–533. doi: 10.1016/j.ajhg.2009.03.010
- Fry, M., and Loeb, L. A. (1994). The fragile X syndrome d(CGG)n nucleotide repeats form a stable tetrahelical structure. *Proc. Natl. Acad. Sci. U.S.A.* 91, 4950–4954. doi: 10.1073/pnas.91.11.4950
- Fryxell, K. J., and Moon, W. J. (2005). CpG mutation rates in the human genome are highly dependent on local GC content. *Mol. Biol. Evol.* 22, 650–658. doi: 10.1093/molbev/msi043
- Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. *J. Mol. Biol.* 196, 261–282. doi: 10.1016/0022-2836(87)90689-9
- Garribba, L., Bjerregaard, V. A., Dinis, M. M. G., Ozer, O., Wu, W., Sakellariou, D., et al. (2020). Folate stress induces SLX1- and RAD51-dependent mitotic DNA synthesis at the fragile X locus in human cells. *Proc. Natl. Acad. Sci. U.S.A.* 117, 16527–16536. doi: 10.1073/pnas.1921219117

- Garribba, L., Wu, W., Ozer, O., Bhowmick, R., Hickson, I. D., and Liu, Y. (2018). Inducing and detecting mitotic DNA synthesis at difficult-to-replicate loci. *Methods Enzymol.* 601, 45–58. doi: 10.1016/bs.mie.2017.11.025
- Gaskins, A. J., Mumford, S. L., Chavarro, J. E., Zhang, C., Pollack, A. Z., Wactawski-Wende, J., et al. (2012). The impact of dietary folate intake on reproductive function in premenopausal women: a prospective cohort study. *PLoS One* 7:e46276. doi: 10.1371/journal.pone.0046276
- Giovannucci, E. (2002). Epidemiologic studies of folate and colorectal neoplasia: a review. *J. Nutr.* 132, 2350S–2355S. doi: 10.1093/jn/132.8.2350S
- Glover, T. W., Berger, C., Coyle, J., and Echo, B. (1984). DNA polymerase- α inhibition by aphidicolin induces gaps and breaks at common fragile sites in human-chromosomes. *Hum. Genet.* 67, 136–142. doi: 10.1007/Bf00272988
- Green, R., and Miller, J. W. (1999). Folate deficiency beyond megaloblastic anemia: hyperhomocysteinemia and other manifestations of dysfunctional folate status. *Semin. Hematol.* 36, 47–64.
- Karolchik, D., Hinrichs, A. S., Furey, T. S., Roskin, K. M., Sugnet, C. W., Haussler, D., et al. (2004). The UCSC table browser data retrieval tool. *Nucleic Acids Res.* 32, D493–D496. doi: 10.1093/nar/gkh103
- Lamm, N., Maoz, K., Bester, A. C., Im, M. M., Shewach, D. S., Karni, R., et al. (2015). Folate levels modulate oncogene-induced replication stress and tumorigenicity. *EMBO Mol. Med.* 7, 1138–1152. doi: 10.15252/emmm.201404824
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. doi: 10.1038/35057062
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Larsson, S. C., Hakansson, N., Giovannucci, E., and Wolk, A. (2006). Folate intake and pancreatic cancer incidence: a prospective study of Swedish women and men. *J. Natl. Cancer Inst.* 98, 407–413. doi: 10.1093/jnci/djj094
- Lerdrup, M., Johansen, J. V., Agrawal-Singh, S., and Hansen, K. (2016). An interactive environment for agile analysis and visualization of ChIP-sequencing data. *Nat. Struct. Mol. Biol.* 23, 349–357. doi: 10.1038/nsmb.3180
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [Preprint]* arXiv:1303.3997v1.
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Long, H. K., King, H. W., Patient, R. K., Odom, D. T., and Klose, R. J. (2016). Protection of CpG islands from DNA methylation is DNA-encoded and evolutionarily conserved. *Nucleic Acids Res.* 44, 6693–6706. doi: 10.1093/nar/gkw258
- Mastracci, T. L., Shadco, A., Colby, S. M., Tuck, A. B., O'Malley, F. P., Bull, S. B., et al. (2006). Genomic alterations in lobular neoplasia: a microarray comparative genomic hybridization signature for early neoplastic proliferation in the breast. *Genes Chromosomes Cancer* 45, 1007–1017. doi: 10.1002/gcc.20368
- Mensink, G. B., Fletcher, R., Gurinovic, M., Huybrechts, I., Lafay, L., Serra-Majem, L., et al. (2013). Mapping low intake of micronutrients across Europe. *Br. J. Nutr.* 110, 755–773. doi: 10.1017/S000711451200565X
- Min, J., Wright, W. E., and Shay, J. W. (2017). Alternative lengthening of telomeres mediated by mitotic DNA synthesis engages break-induced replication processes. *Mol. Cell. Biol.* 37:e00226-17. doi: 10.1128/MCB.00226-17
- Minocherhomji, S., Ying, S., Bjerregaard, V. A., Bursomanno, S., Aleliunaite, A., Wu, W., et al. (2015). Replication stress activates DNA repair synthesis in mitosis. *Nature* 528, 286–290. doi: 10.1038/nature16139
- Nakashima, H., Nishikawa, K., and Ooi, T. (1997). Differences in dinucleotide frequencies of human, yeast, and *Escherichia coli* genes. *DNA Res.* 4, 185–192. doi: 10.1093/dnares/4.3.185
- Neph, S., Kuehn, M. S., Reynolds, A. P., Haugen, E., Thurman, R. E., Johnson, A. K., et al. (2012). BEDOPS: high-performance genomic feature operations. *Bioinformatics* 28, 1919–1920. doi: 10.1093/bioinformatics/bts277
- Ni, J., Lu, L., Fenech, M., and Wang, X. (2010). Folate deficiency in human peripheral blood lymphocytes induces chromosome 8 aneuploidy but this effect is not modified by riboflavin. *Environ. Mol. Mutagen.* 51, 15–22. doi: 10.1002/em.20502
- No authors listed (1991). Prevention of neural tube defects: results of the medical research council vitamin study. MRC vitamin study research group. *Lancet* 338, 131–137. doi: 10.1016/0140-6736(91)90133-a
- Oh, E. K., Kim, Y. W., Kim, I. W., Liu, H. B., Lee, K. H., Chun, H. J., et al. (2012). Differential DNA copy number aberrations in the progression of cervical lesions to invasive cervical carcinoma. *Int. J. Oncol.* 41, 2038–2046. doi: 10.3892/ijo.2012.1644
- Ozer, O., Bhowmick, R., Liu, Y., and Hickson, I. D. (2018). Human cancer cells utilize mitotic DNA synthesis to resist replication stress at telomeres regardless of their telomere maintenance mechanism. *Oncotarget* 9, 15836–15846. doi: 10.18632/oncotarget.24745
- Pieretti, M., Zhang, F. P., Fu, Y. H., Warren, S. T., Oostra, B. A., Caskey, C. T., et al. (1991). Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell* 66, 817–822. doi: 10.1016/0092-8674(91)90125-i
- Piovesan, A., Pelleri, M. C., Antonaros, F., Strippoli, P., Caracausi, M., and Vitale, L. (2019). On the length, weight and GC content of the human genome. *BMC Res. Notes* 12:106. doi: 10.1186/s13104-019-4137-z
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033
- Roman Vinas, B., Ribas Barba, L., Ngo, J., Gurinovic, M., Novakovic, R., Cavelaars, A., et al. (2011). Projected prevalence of inadequate nutrient intakes in Europe. *Ann. Nutr. Metab.* 59, 84–95. doi: 10.1159/000332762
- Rosenbloom, K. R., Armstrong, J., Barber, G. P., Casper, J., Clawson, H., Diekhans, M., et al. (2015). The UCSC genome browser database: 2015 update. *Nucleic Acids Res.* 43, D670–D681. doi: 10.1093/nar/gku1177
- Sansone, M., Sansone, A., Romano, M., Seraceno, S., Di Luigi, L., and Romanelli, F. (2018). Folate: a possible role in erectile dysfunction? *Aging Male* 21, 116–120. doi: 10.1080/13685538.2017.1404022
- Santoro, M. R., Bray, S. M., and Warren, S. T. (2012). Molecular mechanisms of fragile X syndrome: a twenty-year perspective. *Annu. Rev. Pathol.* 7, 219–245. doi: 10.1146/annurev-pathol-011811-132457
- Sedlazeck, F. J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A., et al. (2018). Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods* 15, 461–468. doi: 10.1038/s41592-018-0001-7
- Seneca, S., Lissens, W., Endels, K., Caljon, B., Bonduelle, M., Keymolen, K., et al. (2012). Reliable and sensitive detection of fragile X (expanded) alleles in clinical prenatal DNA samples with a fast turnaround time. *J. Mol. Diagn.* 14, 560–568. doi: 10.1016/j.jmoldx.2012.05.003
- Senti, F. R., and Pilch, S. M. (1985). Analysis of folate data from the second national health and nutrition examination survey (NHANES II). *J. Nutr.* 115, 1398–1402. doi: 10.1093/jn/115.11.1398
- Tomso, D. J., and Bell, D. A. (2003). Sequence context at human single nucleotide polymorphisms: overrepresentation of CpG dinucleotide at polymorphic sites and suppression of variation in CpG islands. *J. Mol. Biol.* 327, 303–308. doi: 10.1016/s0022-2836(03)00120-7
- Udin, K., and Woodford, K. J. (1995). CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis in vitro. *Nucleic Acids Res.* 23, 4202–4209. doi: 10.1093/nar/23.20.4202
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., et al. (2001). The sequence of the human genome. *Science* 291, 1304–1351. doi: 10.1126/science.1058040
- Vincent, A., Heitz, D., Petit, C., Kretz, C., Oberle, I., and Mandel, J. L. (1991). Abnormal pattern detected in fragile-X patients by pulsed-field gel electrophoresis. *Nature* 349, 624–626. doi: 10.1038/349624a0
- Vinogradov, A. E. (2003). DNA helix: the importance of being GC-rich. *Nucleic Acids Res.* 31, 1838–1844. doi: 10.1093/nar/gkg296
- Voineagu, I., Surka, C. F., Shishkin, A. A., Krasilnikova, M. M., and Mirkin, S. M. (2009). Replisome stalling and stabilization at CGG repeats, which are responsible for chromosomal fragility. *Nat. Struct. Mol. Biol.* 16, 226–228. doi: 10.1038/nsmb.1527
- Walsh, C. P., and Xu, G. L. (2006). Cytosine methylation and DNA repair. *Curr. Top. Microbiol. Immunol.* 301, 283–315. doi: 10.1007/3-540-31390-7_11
- Wang, Y. H., Gellibolian, R., Shimizu, M., Wells, R. D., and Griffith, J. (1996). Long CCG triplet repeat blocks exclude nucleosomes: a possible mechanism

- for the nature of fragile sites in chromosomes. *J. Mol. Biol.* 263, 511–516. doi: 10.1006/jmbi.1996.0593
- Wilson, P. M., Danenberg, P. V., Johnston, P. G., Lenz, H. J., and Ladner, R. D. (2014). Standing the test of time: targeting thymidylate biosynthesis in cancer therapy. *Nat. Rev. Clin. Oncol.* 11, 282–298. doi: 10.1038/nrclinonc.2014.51
- Wingett, S. W., and Andrews, S. (2018). FastQ screen: a tool for multi-genome mapping and quality control. *F1000Res.* 7:1338. doi: 10.12688/f1000research.15931.2
- Yrigollen, C. M., Martorell, L., Durbin-Johnson, B., Naudo, M., Genoves, J., Murgia, A., et al. (2014). AGG interruptions and maternal age affect FMR1 CGG repeat allele stability during transmission. *J. Neurodev. Disord.* 6:24. doi: 10.1186/1866-1955-6-24

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Garribba, Vogel, Lerdrup, Gonçalves Dinis, Ren and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



OPEN ACCESS

Edited by:

Stefano Gnan,
Institut Curie, France

Reviewed by:

Lora Boteva,
University of Edinburgh,
United Kingdom
Victoria Alexandra Bjerregaard,
Danish Cancer Society Research
Center, Denmark
Valeria Naim,
UMR 8200 Stabilité génétique et
oncogénèse, France

***Correspondence:**

Vibe H. Oestergaard
Vibe@bio.ku.dk

[†] These authors have contributed
equally to this work

***Present address:**

Sebastian H. N. Munk,
DNA Replication and Cancer Group,
Genome Integrity Unit, Danish Cancer
Society Research Center,
Copenhagen, Denmark

Specialty section:

This article was submitted to
Human and Medical Genomics,
a section of the journal
Frontiers in Genetics

Received: 14 April 2021

Accepted: 28 June 2021

Published: 20 July 2021

Citation:

Munk SHN, Voutsinos V and
Oestergaard VH (2021) Large Intronic
Deletion of the Fragile Site Gene
PRKN Dramatically Lowers Its Fragility
Without Impacting Gene Expression.
Front. Genet. 12:695172.
doi: 10.3389/fgene.2021.695172

Large Intronic Deletion of the Fragile Site Gene *PRKN* Dramatically Lowers Its Fragility Without Impacting Gene Expression

Sebastian H. N. Munk^{†‡}, Vasileios Voutsinos[†] and Vibe H. Oestergaard^{*}

Department of Biology, University of Copenhagen, Copenhagen, Denmark

Common chromosomal fragile sites (CFSs) are genomic regions prone to form breaks and gaps on metaphase chromosomes during conditions of replication stress. Moreover, CFSs are hotspots for deletions and amplifications in cancer genomes. Fragility at CFSs is caused by transcription of extremely large genes, which contributes to replication problems. These extremely large genes do not encode large proteins, but the extreme sizes of the genes originate from vast introns. Intriguingly, the intron sizes of extremely large genes are conserved between mammals and birds. Here, we have used reverse genetics to address the function and significance of the largest intron in the extremely large gene *PRKN*, which is highly fragile in our model system. Specifically, we have introduced an 80-kilobase deletion in intron 7 of *PRKN*. We find that gene expression of *PRKN* is largely unaffected by this intronic deletion. Strikingly, the intronic deletion, which leads to a 12% reduction of the overall size of the *PRKN* gene body, results in an almost twofold reduction of the *PRKN* fragility. Our results stress that while the large intron clearly contributes to the fragility of *PRKN*, it does not play an important role for *PRKN* expression. Taken together, our findings further add to the mystery concerning conservation of the seemingly non-functional but troublesome large introns in *PRKN*.

Keywords: common chromosomal fragile sites, large genes, *PRKN*, parkin, genomic instability, genome editing

INTRODUCTION

CFSs are specific regions of the genome that often fail to replicate before mitosis, which results in chromosome breakage and high mutation rates (Debatisse et al., 2012). Several pan-cancer genome analyses have also revealed that CFSs are hotspots for structural variants in cancer genomes (Beroukheim et al., 2010; Bignell et al., 2010) with large deletions at the center of CFSs and insertions at the borders of CFSs (Li et al., 2020). In addition, DNA double-strand breaks are remarkably recurrent at CFSs in neuronal progenitor cells (Schwer et al., 2016; Wei et al., 2016).

It has become apparent that breakage and mutations at CFSs are due to replication problems caused by transcription of extremely large genes located at CFSs (Le Tallec et al., 2013; Wilson et al., 2015; Pentzold et al., 2018). To understand how transcription of large genes perturb replication, it is important to keep in mind that eukaryotic replication is initiated bidirectionally from origin of replication complexes (ORCs) scattered across the genome. The distance between these complexes thus determines the minimum distance that two opposing replication forks have to travel to complete replication of the region. Most origins of replication are not used during a normal cell cycle, but during replication stress, excess origins of replication are engaged to ensure complete replication of the genome. The process of transcription repositions the ORCs and thereby clears active intragenic regions of replication origins (Gros et al., 2015; Macheret and Halazonetis, 2018). Hence, transcription of extremely large genes clears vast genomic regions of ORCs and in that way suppresses the firing of backup replication origins in these regions, thus impeding genome replication in an indirect manner. Moreover, clashes between transcription and replication machineries may directly challenge replication of CFSs (Helmrich et al., 2011; Oestergaard and Lisby, 2016; Hamperl et al., 2017). Finally, AT-dinucleotide rich regions capable of forming secondary structures can further perturb replication of certain regions within CFSs (Kaushal and Freudenreich, 2019).

One of the most fragile regions of the human genome is called FRA6E. Here, transcription of the 1.4 Mb *PRKN* underlies its fragility (Glover et al., 2017). Intriguingly, the mature *PRKN* mRNA is only 4 kb despite the fact that the RNA polymerase has to synthesize 1.4 Mb of pre-mRNA. This is because *PRKN* as well as other extremely large genes mainly consist of introns (Voutsinos et al., 2018). Despite their unstable nature and scarce coding information, we recently showed that the size of *PRKN* and other extremely large genes at CFSs are conserved in vertebrates, suggesting that the large introns of these genes possess currently unknown biological functions (Pentzold et al., 2018).

The *PRKN* gene product, parkin, is an E3 ubiquitin-protein ligase that plays a key role in removal of damaged mitochondria through mitophagy (Frank et al., 2012). This process prevents excessive production of reactive oxygen species from dysfunctional mitochondria. Inherited mutations in *PRKN* are the most common cause of autosomal recessive juvenile form of Parkinson's disease, thus emphasizing the neuroprotective importance of *PRKN* (Klein and Westenberger, 2012). Numerous studies suggest that impaired mitophagy is involved in Parkinson's disease etiology (Frank et al., 2012; Guo, 2012). Additionally, loss or down-regulation of *PRKN* has been associated with various types of cancer (Gupta et al., 2017), and its loss has been shown to result in a switch to aerobic glycolysis, known as the Warburg effect, which is a characteristic of many cancer types (Zhang et al., 2011).

To investigate the functional significance of extremely large introns, we deleted 80 kb of intron 7 in *PRKN* in our model system, the avian cell line DT40. We previously showed that

PRKN is transcribed and fragile in this cell line (Pentzold et al., 2018). Here, we find that the deletion does not affect *PRKN* expression levels but leads to a drastic reduction in *PRKN* fragility.

METHODS

Generation of Constructs

All constructs generated in this study are listed in **Supplementary Table 1** and all primers plus other DNA oligos used in this study are listed in **Supplementary Table 2**. The *PRKN* homology arms for C-terminal fluorescent tagging were amplified with the primer pairs VV5/VV6 and VV7/VV8 for the 5' arm or the 3' arm, respectively. The Venus-YFP (2YFP) was amplified using VV47 and VV49. All primers were designed to facilitate directional cloning and they were synthesized by TAG Copenhagen. The amplified products were cloned into pCR2.1-TOPO (Invitrogen) and confirmed by Sanger sequencing (performed by Eurofins Genomics).

The fragments for the *PRKN* 2YFP-tagging construct were then assembled in pBluescript (SK+). Specifically, the 5' homology arm was inserted as a *KpnI* *SalI* fragment, the 3' homology arm was inserted as a *BamHI* *NotI* fragment, and a resistance cassette (BSR or PURO) was inserted as a *BamHI* fragment. Finally, the 2YFP fragment was inserted as a *XhoI* *SalI* fragment into the *SalI* site. Correct orientation was confirmed by restriction digest. The resulting constructs were named pVV6 and pVV15 encoding puromycin (PURO) or blasticidin (BSR) resistance, respectively.

To construct the repair template for *PRKN* intron-7 deletion, genomic regions flanking gRNA Target Site 1 (TS1) and gRNA TS2 were first amplified by PCR templated by genomic DNA (gDNA) from DT40 cells to obtain homology arms.

To create the 5' homology arm extending 2 kb 5' of gRNA TS1, PCR was conducted on gDNA with the primers 5' fwd and 5' rev adapted with *ApaI* and *BamHI* restriction sites, respectively. Similarly, PCR was conducted on gDNA with primers 3' fwd and 3' rev adapted with *BamHI* and *XbaI* restriction sites, respectively, to create the 3' homology arm extending 3' of gRNA TS2. The homology arm fragments were subcloned into TOPO TA vectors (Invitrogen; according to manufacturer's protocol) and sequenced (Eurofins Genomics).

Then, the 5' homology arm was subcloned from the TOPO TA vector into pBluescript (SK+) as an *ApaI*-*BamHI* fragment. Subsequently, the 3' homology arm was subcloned from the TOPO TA vector into the 5' homology arm-pBluescript as a *XbaI*-*BamHI* fragment. Finally, the BSR cassette fragment was cloned in as a *BamHI* fragment. The final construct was sequenced to confirm correct assembly (Eurofins Genomics).

Cas9/gRNA Constructs

The backbone for the Cas9/gRNA constructs was pX458 (Addgene). The expression of specific target gRNAs was obtained by annealing the oligonucleotides listed in **Supplementary Table 2** and integrating them into pX458 at the *BbsI* site. Correct integration was confirmed by sequencing (Eurofins

Genomics). The constructs were named pX458 *PRKN* TS1 and pX458 *PRKN* TS2.

Cell Culture and Transfection

All DT40 cell lines used in this study are listed in **Supplementary Table 3**. DT40 cells were cultured in RPMI 1640 medium GlutaMAX (Thermo Fisher Scientific) supplemented with 2% chicken serum (Sigma-Aldrich), 8% fetal bovine serum (Thermo Fisher Scientific), 50 U/ml penicillin, 50 µg/ml streptomycin, and 50 µM β-mercaptoethanol at 39°C with 5% CO₂.

Transfections for targeted integration were performed by electroporation with Gene Pulser Xcell™ (BioRad) with the settings 25 µF and 0.6 kV. Approximately 35 µg of linearized plasmid DNA was used for transfection with the 2YFP targeting construct. For deletion of *PRKN* intron 7, 20 million cells were transfected with 50 µg linearized repair template and 30 µg of each of the two Cas9/gRNA expression vectors (110 µg DNA in total). Transfection with Cas9/gRNA was transient.

For transient expression of the Cre recombinase, 3.5 million cells were transfected with 15 µg plasmid DNA using the nucleofector system developed by Amaxa Biosystems GmbH (Franklin and Sale, 2006).

Image Cytometry

For quantification of fluorescently tagged protein levels, the Xcyto® 10 image cytometer (ChemoMetec A/S) was used. Cells were stained by Vybrant Ruby Stain (5 µM, V10309, Thermo Fisher Scientific) to exclude dead cells based on their DNA content. Only cells meeting the following criteria were included in the analysis: not in aggregate, circularity > 0.6, and with DNA content of viable cells.

Reverse Transcription Quantitative PCR (RT-qPCR)

Total RNA was isolated using the GeneJET RNA Purification Kit (Thermo Fisher Scientific). For RNA samples used for analysis of *PACRG* mRNA levels, 1 µg RNA was pre-treated with DNase I (Fermentas) to remove gDNA according to manufacturer's instructions in the GeneJET RNA Purification Kit (Thermo Fisher Scientific). cDNA was made using RevertAid Premium Reverse Transcriptase (Thermo Fisher Scientific) with random hexamers and oligo(dT) primers.

Each qPCR reaction was performed in triplicates. All the primer pairs used for qPCR are shown in **Supplementary Table 2**. All the primer pair efficiencies were close to 1 (100%) and within the acceptable range according to the Minimum Information for Publication of Quantitative Real-Time PCR Experiments guidelines (Bustin et al., 2009). qPCR was performed with Maxima SYBR Green/ROX qPCR Master Mix (Thermo Scientific) according to manufacturer's instructions for three-step RT-qPCR cycling protocol on CFX96 Real-Time PCR Detection System (BioRad). Fold changes were calculated using the $2^{-\Delta\Delta C_t}$ method (Livak and Schmittgen, 2001). In all cases *GAPDH* was used as reference gene.

Fluorescence *in situ* Hybridization (FISH)

Metaphase spreads were prepared as previously described (Smith et al., 1990; Pentzold et al., 2018). Briefly, cells were treated with DMSO or aphidicolin (APH) (0.3 µM) (Sigma-Aldrich) over 16 h followed by a 3-h treatment of 0.1 µg/ml colcemid (Life Technologies). Next, cells were swelled in 8 ml hypotonic buffer [20% FBS (vol/vol), 15 mM KCl] for 15 min. Subsequently, 1 ml fixation buffer (25% acetone, 75% methanol) was gradually added. Cells were harvested by centrifugation and then resuspended in 10 ml fixation buffer. Cells were stored at −20°C O/N. Finally, the cells were splatted onto the slides to spread the metaphase chromosomes (Pentzold et al., 2018).

FISH was carried out as previously described with minor modifications (El Achkar et al., 2005; Pentzold et al., 2018). Briefly, the probe used for *PRKN* detection was made with the BAC CH261-119N16 from the CHORI library, and the probes for intron 7 detection were made by amplifying ≈10 kb fractions of intron 7 with the primers listed in **Supplementary Table 2**. Probes were labeled either by biotin or by digoxigenin by using BioPrime DNA Labeling system (Invitrogen). Metaphase spreads had been treated with RNase H (Thermo Fisher Scientific) before they were incubated with probes.

Metaphase spreads were incubated with blocking reagent and Streptavidin-Cy3 (Alexa 594) (1:200), Biotinylated rabbit anti-streptavidin (Rockland) (1:266) for biotin-labeled probe detection and mouse anti-digoxigenin FITC (Interchim) (1:50) and goat anti-mouse (Alexa 488) for digoxigenin-labeled probe detection.

Slides with metaphase spreads were mounted with coverslips using 15 µl of mounting medium containing DAPI (4% n-propyl gallate, 80% glycerol, 1 µg/ml DAPI). Metaphase chromosomes were visualized on a widefield microscope (AxioImager Z1; Carl Zeiss) equipped with a 100× objective lens (Plan Apochromat, NA 1.4; Carl Zeiss), a cooled CCD camera (Orca-ER; Hamamatsu Photonics), differential interference contrast (DIC), and an illumination source (HXP120C; Carl Zeiss).

RESULTS

Establishing *PRKN* Intron 7-Deleted Cell Lines

To investigate the role of large introns in genes coinciding with CFSs, we chose to study the *PRKN* gene, which is highly fragile in our model cell line (Pentzold et al., 2018). To enable live-cell detection of parkin protein levels, we first generated a DT40 cell line with *PRKN* endogenously tagged with a Venus-YFP (2YFP) tandem tag in a background where the non-fragile gene *TOPBP1* was endogenously tagged with TFP on one of its three alleles. The resulting cell line thus has the genotype *PRKN*^{WT/2YFP} *TOPBP1*^{WT/WT/TFP} and is referred to as “P2Y-TT”. Following tagging of *PRKN*, the allele remained fragile in response to replication stress and the tagged gene product was expressed at full length (**Supplementary Figures 1A–C**). This cell line, P2Y-TT, was

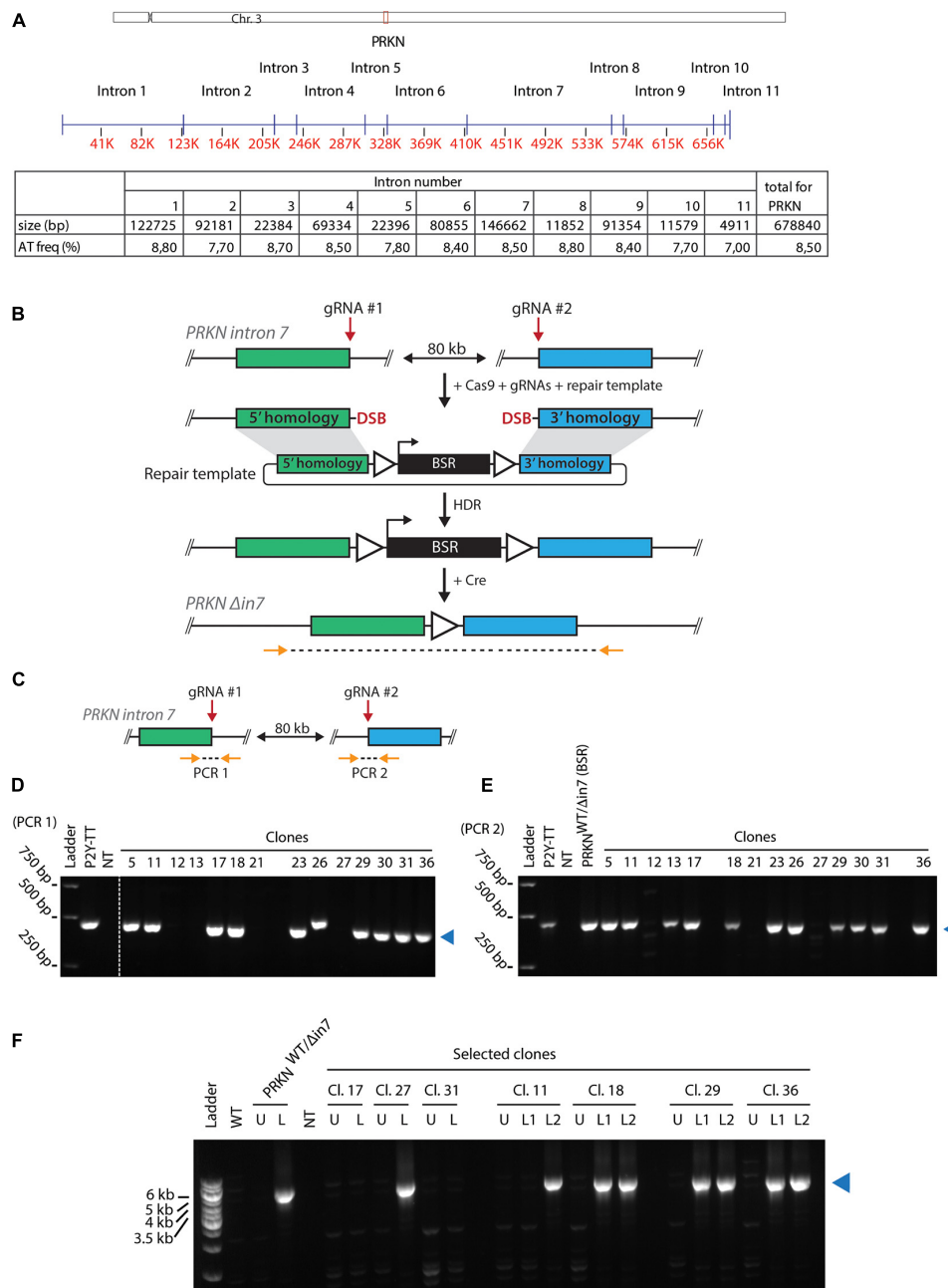


FIGURE 1 | Generating cell lines with large intronic deletion in the fragile site gene *PRKN*. **(A)** Upper panel, schematic representation of chicken chromosome 3 drawn to scale. Positions of the centromere (x) and *PRKN* (red box) are indicated. Middle panel, schematic representation of *PRKN* drawn to scale. Intron numbers are indicated. Exons are represented as vertical pins on a horizontal line. Ruler indicates position along the gene (kb). Lower panel, table showing sizes of introns given in basepairs (bp) as well as percentage of AT-dinucleotide frequency (AT freq). **(B)** Outline of the strategy for targeted deletion of 80 kb in *PRKN* intron 7. To delete 80 kb in *PRKN* intron 7, cells were transfected with Cas9 and guide RNA (gRNA) expression vectors along with a repair template. The Cas9 nuclease was directed to induce double-strand breaks (DSB) at two target sites (gRNA #1 and gRNA #2) 80 kb apart in *PRKN* intron 7. The DSBs can be repaired by homology-directed repair (HDR) using the repair template containing a selectable blasticidin resistance (BSR) cassette flanked by *loxP* sites (triangles) and homology regions (green and blue rectangles). After Cre-mediated removal of the floxed BSR cassette, PCR was used to amplify across the 80-kb deletion with primers (horizontal arrows) binding at the indicated positions outside of the homology regions. **(C)** Outline of the PCR strategy used to amplify across each of the two gRNA target sites (gRNA #1 and gRNA #2) denoted PCR 1 and PCR 2, respectively. Primers are shown as orange arrows. **(D,E)** PCR amplification across gRNA #1 **(D)** and gRNA #2 **(E)**. Analysis of PCR products from the indicated clones and the parental P2Y-TT cell line (positive control). In E, an additional positive control (*PRKN*^{WT/Δin7(BSR)}) was included. A no-template control (NT) was included in both analyses. A blue triangle on the right of each gel indicates the product of the predicted size. **(F)** PCR amplification across the 80-kb deletion (as shown in B). Analysis of PCR products from WT cells and selected clones before (U) or after (L) removal of the BSR cassette. If two clonal populations were tested after loxing, this is indicated with L1 and L2. A positive control with the indicated genotype was also included in the analysis. NT denotes the no-template control. The product of the predicted size is indicated by a blue triangle on the right of the gel.

used as background for all further genetic manipulations unless otherwise stated.

PRKN, which is located on the long arm of gallus chromosome 3, contains 11 introns of varying sizes (Figure 1A). *PRKN* is not enriched for repetitive sequences and replicates in the middle of the S phase (Shang et al., 2013; Pentzold et al., 2018). The total AT-dinucleotide percentage of *PRKN* is 8.5 (Figure 1A, lower panel). We generated cell lines deleted for most of intron 7, which is the largest intron in *PRKN* and has a representative AT-dinucleotide frequency (Figure 1A). Deletion of this intron was achieved by combining a selectable targeting construct, with homology to each side of the region targeted for deletion, with CRISPR/Cas9-mediated cleavage at two target sites flanking the desired 80-kb deletion as outlined in Figure 1B. Successful targeting yields clones with 80 kb of *PRKN* intron 7 replaced with a blasticidin resistance (BSR) cassette. Flanking loxP sites enabled subsequent removal of the BSR cassette. Initial PCR screening suggested that some clones potentially had the 80 kb region replaced by the cassette (Supplementary Figures 2A–E). Subsequently, we tested whether the clones still contained a wild-type *PRKN* allele with PCR analyses of the two guide RNA target sites (Figure 1C). Most clones retained a wild-type allele, but three clones appeared to have lost both wild-type alleles (Figures 1D,E). Then, one potential homozygote and six potential heterozygotes for the 80-kb deletion were transiently transfected with the Cre recombinase to mediate removal of the BSR cassette followed by isolation of single clones. The resulting clones were analyzed with PCR across the region targeted for deletion, and an amplicon of the expected size confirmed successful deletion of 80 kb in *PRKN* intron 7 in a subset of the clones (Figure 1F). The successful homozygote (clone 27) and heterozygotes (clone 11, 18, 29, and 36) are referred to as *PRKN*^{Δ in7/Δ in7} and *PRKN*^{WT/Δ in7}, respectively.

Interestingly, in a previous attempt to generate cell lines with the 80-kb deletion, we isolated two clones with the deletion that both turned out to be trisomic for chromosome 3, which is the chromosome that contains the *PRKN* gene (not shown). Thus, we examined the karyotype of the clones derived from this transfection for aneuploidy. Here, we found that 2 out of 14 clones were trisomic for chromosome 3 (Supplementary Figure 2F). This suggests that there is a high risk of chromosome mis-segregation associated with targeting of the *PRKN* gene.

Taken together, the fact that we were able to isolate a homozygous *PRKN* intron 7-deleted DT40 cell line demonstrates that this part of the genome does not contain functional elements essential for cell viability.

The 80-kb Deletion in *PRKN* Intron 7 Does Not Significantly Change *PRKN* Expression

Although intron 7 is not essential for cell viability, it may contain regulatory elements that influence *PRKN* expression. We thus asked if deletion of the intron has an effect on parkin levels in the cell. First, we used fluorescence image cytometry to evaluate how *PRKN*-2YFP expression was affected by biallelic intron 7 deletion (Figure 2A). The levels of parkin-2YFP were similar in the intron

7-deleted clone and the parental cell line. Because cells with this genotype must contain the deletion in the 2YFP-tagged *PRKN* allele this indicates that the 80-kb deletion in *PRKN* intron 7 does not alter *PRKN* expression.

To evaluate the effect of intron 7 deletion on *PRKN* transcript levels, we performed reverse transcription quantitative PCR (RT-qPCR) with two different primer sets: One set binding specifically to transcripts from the 2YFP-tagged allele (*PRKN*^{2YFP}) and one set binding to transcripts from both the 2YFP-tagged and untagged (*PRKN*^{WT}) allele (referred to as “total *PRKN* transcripts”) (Supplementary Figure 3). Only the clone with homozygous intron 7 deletion was included in the analysis of *PRKN*^{2YFP} transcript levels while the homozygote and two heterozygotes were included in the analysis of total *PRKN* transcript levels. No change in either *PRKN*^{2YFP} transcript levels (Figure 2B) or total *PRKN* transcript levels (Figure 2C) were detected in any of the clones, supporting that the intron 7 deletion does not alter *PRKN* expression.

We further investigated whether the intron 7 deletion induced changes in the promoter activity of *PRKN*. Specifically, we exploited that *PRKN* shares its promoter (marked by high GC content in Figure 2D, upper panel) with the gene *PACRG* (parkin coregulated) (West et al., 2003), which is transcribed in the opposite direction of *PRKN*. Thus, we would expect changes in the promoter activity to affect both genes, and we therefore extended our RT-qPCR investigations to include *PACRG* transcription (Figure 2D). While a significant decrease in *PACRG* transcript levels were detected in *PRKN*^{WT/Δ in7} clone 18 compared to the parental cell line, no significant differences were detected in any of the other clones including the homozygote for intron 7 deletion, suggesting that the difference seen in one of the clones is due to clonal variation.

Altogether, this indicates that the 80-kb deletion in *PRKN* intron 7 does not markedly alter *PRKN* or *PACRG* expression.

Truncation of *PRKN* Significantly Reduces Its Fragility

To test the hypothesis that the large introns in *PRKN* are underlying its fragility, we performed FISH on metaphase spreads from the two *PRKN*^{WT/Δ in7} clones after inducing replication stress by treatment with the DNA polymerase inhibitor aphidicolin (APH). These clones enabled us to use the full-length wild-type allele of *PRKN* as an internal control. For FISH, we used a probe that binds *PRKN* outside of intron 7, which detects both full-length and intron 7-deleted *PRKN*, as well as a probe that binds the deleted region of intron 7 and therefore only detects the full-length *PRKN* (Figures 3A,B). While the full-length *PRKN* locus in the *PRKN*^{WT/Δ in7} clones was as fragile as the full-length *PRKN* in the parental cell line, the intron 7-deleted *PRKN* allele was significantly less fragile in both *PRKN*^{WT/Δ in7} clones (Figure 3C). Notably, the deletion, which is equivalent to approximately 12% of the length of *PRKN*, resulted in an approximately 50% reduction of the fragility of the gene. Thus, the 80 kb region in intron 7 clearly contributes to *PRKN* fragility even though this region does not have a clear role in *PRKN* expression or cell viability.

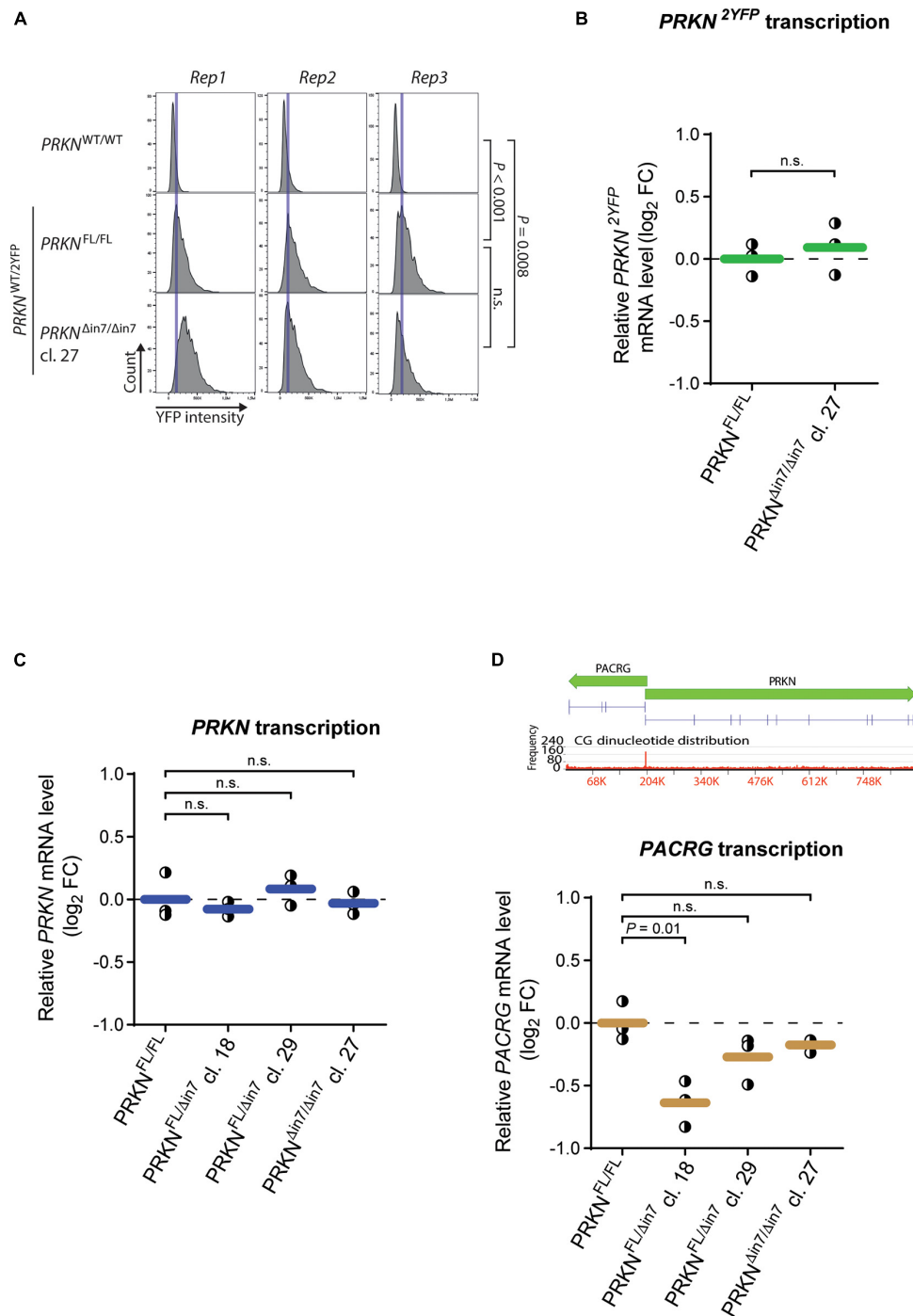


FIGURE 2 | 80-kb deletion of intron 7 does not influence *prkn* expression levels. **(A)** Histograms of parkin-2YFP fluorescence levels measured by fluorescence image cytometry in WT cells ($PRKN^{WT/WT}$) and the P2Y-TT cell line with full-length *PRKN* ($PRKN^{FL/FL}$) or intron 7-deleted *PRKN* ($PRKN^{\Delta in7/\Delta in7}$). Rep1-3 denote three individual experiments. Blue vertical lines indicate the histogram peak for P2Y-TT cells in each replicate. Between 900 and 3,800 cells were analyzed per cell line per replicate. *P*-values were calculated using the mean YFP intensities from all replicates with Student's *t*-test (n.s. = not significant). **(B,C)** RT-qPCR analysis of *PRKN*-2YFP **(B)** and *PRKN* **(C)** mRNA levels in P2Y-TT cells with full-length *PRKN* ($PRKN^{FL/FL}$) or intron 7 deletion in one ($PRKN^{FL/\Delta in7}$) or both alleles ($PRKN^{\Delta in7/\Delta in7}$). Y-axis shows \log_2 fold change (\log_2 FC) in mRNA levels relative to $PRKN^{FL/FL}$. Dots represent individual experiments ($n = 3$), and horizontal lines indicate the mean. Dashed line denotes \log_2 FC = 0. *P*-values were calculated with Student's *t*-test (n.s. = not significant). **(D)** Upper panel, Schematic representation of *PACRG* and *PRKN* including their shared promoter. Gene sizes and orientations are indicated with green arrows. Exons are represented as vertical pins on a horizontal line. Ruler indicates position along the genes (kb). Bar chart showing CG dinucleotide frequency is shown below the genes (2,930-bp windows). Lower panel, RT-qPCR analysis of *PACRG* mRNA levels as in **(B,C)**.

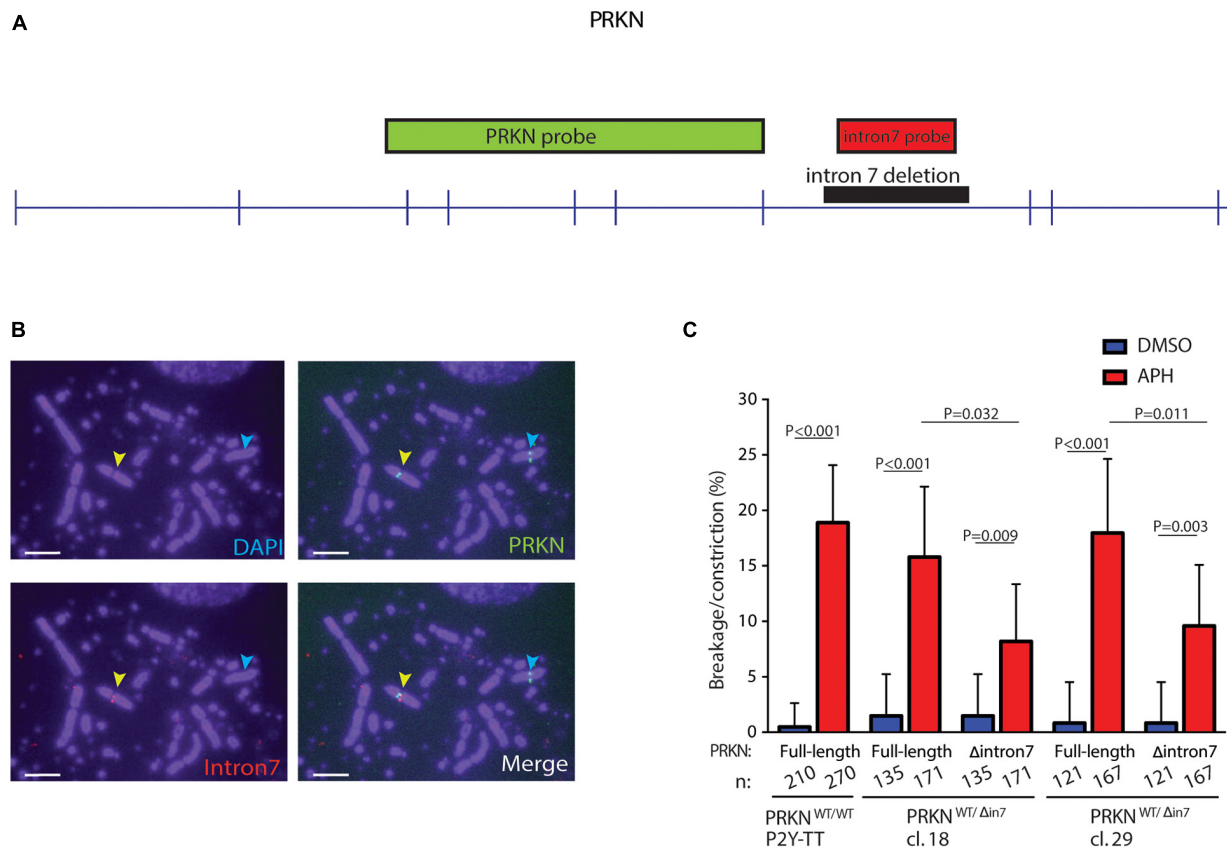


FIGURE 3 | 80 kb deletion in intron 7 significantly affects *prkn* fragility. **(A)** Scaled schematic representation of *PRKN*. Positions of the FISH probes for *PRKN* are indicated with green and red boxes. The deleted region of intron 7 is indicated by a black box. Exons are represented as vertical pins on a horizontal line. **(B)** Representative images of metaphase spreads with FISH probes against *PRKN* (green) and *PRKN* intron 7 (red) on DAPI-stained metaphase spreads. Cells were treated with 0.3 μ M of APH for 16 h before harvesting. Yellow arrows point to a break/constriction at the full-length *PRKN* allele and the cyan arrows point to the intron 7-deleted *PRKN* allele only recognized by the *PRKN* FISH probe. Scale bars are 5 μ m. **(C)** Quantification of breakage/constriction at *PRKN* in P2Y-TT cells and two independently derived clones of *PRKN*^{WT/ Δ in7}. Cells were treated with DMSO or 0.3 μ M APH for 16 h. n denotes the total number of the indicated *PRKN* allele that was quantified for each clone. The data were analyzed by Fisher's exact test. Error bars indicate 95% confidence intervals.

DISCUSSION

Extremely large genes arose in an early vertebrate ancestor due to intron expansions (Voutsinos et al., 2018). Furthermore, the size of extremely large genes seems to be conserved during evolution even though they pose a threat to genome integrity (Pentzold et al., 2018; Voutsinos et al., 2018). In this paper, we have investigated the cellular role of the largest intron in the *PRKN* gene, which is located in a highly fragile CFS (Wilson et al., 2015; Okamoto et al., 2018; Pentzold et al., 2018; Voutsinos et al., 2018). This is to our knowledge the first controlled experiment addressing the function of an extremely large intron. We were able to generate a cell line homozygous for an 80 kb deletion in *PRKN* intron 7, clearly demonstrating that the deleted region is not essential for cell viability. Moreover, we find that this intron 7 truncation does not have any significant effect on *PRKN* gene expression. Yet, the 80-kb intronic deletion leads to an almost 50% reduction of *PRKN* fragility although only shortening the gene length by 12%, which does not appear to be a consequence of altered transcriptional activity. Thus, these 80 kb

of intronic sequence with no apparent function are significantly contributing to *PRKN* fragility, most likely reflecting that extreme gene size is a trigger for chromosomal fragility, which suggests a disproportional significance of gene length on chromosomal fragility. The reason for the reduction of fragility upon intron deletion might be that conflicts between transcription and replication are avoided due to the shorter traveling time for the RNA polymerase. However, given the importance of replication timing for fragility, the reduced fragility may result from change in replication timing, which we expect to occur because the transcription unit is shortened and thereby the distance between replication origins at each side of the gene will be reduced. However, further studies are needed to experimentally determine the effect of intron deletion on replication timing. We note that the size of the gene or elements within the intron may play a functional role in certain tissues. It may even be possible that genomic instability at CFSs play a physiological role for instance in neurons where it might serve to generate genetic diversity (Schwer et al., 2016; Wei et al., 2016; Voutsinos et al., 2018). Alternatively, the replication difficulties induced by long introns

may provoke epigenetic diversification as shown for replication problems induced by G4 quadruplex forming DNA sequences (Schiavone et al., 2014).

Here, we find that intronic truncation does not lead to changes in gene expression, thus adding to the mystery regarding the conservation of large introns. Therefore, further studies are needed to unravel the functional significance of large introns in genes at CFSs that clearly cause problems for dividing cells.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

Ethical review and approval was not required for the animal study because the DT40 parental bursal lymphoblast cell line is commercially available from ATCC.

AUTHOR CONTRIBUTIONS

SM performed the 80-kb deletion of PRKN intron 7 and most of the experiments that relate to this. VV constructed the cell

lines expressing tagged parkin and generated the intron 7 FISH probe, and also assisted with supervision and data analysis. VO conceived, coordinated, and supervised the project. VO, SM, and VV wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This project was supported by the Villum Foundation (00011407) and the Novo Nordisk Foundation (NNF18OC0052089) to VO. SM was supported by a Novo Scholarship.

ACKNOWLEDGMENTS

We are grateful to Michael Lisby for sharing reagents. We would also like to thank Chemometec A/S for providing the Xcyto image cytometer and training for the instrument.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.695172/full#supplementary-material>

REFERENCES

- Beroukhi, R., Mermel, C. H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., et al. (2010). The landscape of somatic copy-number alteration across human cancers. *Nature* 463, 899–905. doi: 10.1038/nature08822
- Bignell, G. R., Greenman, C. D., Davies, H., Butler, A. P., Edkins, S., Andrews, J. M., et al. (2010). Signatures of mutation and selection in the cancer genome. *Nature* 463, 893–898. doi: 10.1038/nature08768
- Bustin, S. A., Benes, V., Garson, J. A., Hellems, J., Huggett, J., Kubista, M., et al. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* 55:611. doi: 10.1373/clinchem.2008.112797
- Debatisse, M., Le Tallec, B., Letessier, A., Dutrillaux, B., and Brison, O. (2012). Common fragile sites: mechanisms of instability revisited. *Trends Genet.* 28, 22–32. doi: 10.1016/j.tig.2011.10.003
- El Achkar, E., Gerbault-Seureau, M., Muleris, M., Dutrillaux, B., and Debatisse, M. (2005). Premature condensation induces breaks at the interface of early and late replicating chromosome bands bearing common fragile sites. *Proc. Natl. Acad. Sci. U. S. A.* 102, 18069–18074. doi: 10.1073/pnas.0506497102
- Frank, M., Duvezin-Caubet, S., Koob, S., Occhipinti, A., Jagasia, R., Petcherski, A., et al. (2012). Mitophagy is triggered by mild oxidative stress in a mitochondrial fission dependent manner. *Biochim. Biophys. Acta* 1823, 2297–2310. doi: 10.1016/j.bbamcr.2012.08.007
- Franklin, R., and Sale, J. E. (2006). Transient transfection of DT40. *Subcell. Biochem.* 40, 379–382. doi: 10.1007/978-1-4020-4896-8_29
- Glover, T. W., Wilson, T. E., and Arlt, M. F. (2017). Fragile sites in cancer: more than meets the eye. *Nat. Rev. Cancer* 17, 489–501. doi: 10.1038/nrc.2017.52
- Gros, J., Kumar, C., Lynch, G., Yadav, T., Whitehouse, I., and Remus, D. (2015). Post-licensing Specification of Eukaryotic Replication Origins by Facilitated Mcm2-7 Sliding along DNA. *Mol. Cell* 60, 797–807. doi: 10.1016/j.molcel.2015.10.022
- Guo, M. (2012). Drosophila as a model to study mitochondrial dysfunction in Parkinson's disease. *Cold Spring Harb. Perspect. Med.* 2:a009944. doi: 10.1101/cshperspect.a009944
- Gupta, A., Anjomani-Virmouni, S., Koundouros, N., and Poulogiannis, G. (2017). PARK2 loss promotes cancer progression via redox-mediated inactivation of PTEN. *Mol. Cell Oncol.* 4:e1329692. doi: 10.1080/23723556.2017.1329692
- Hamperl, S., Bocek, M. J., Saldivar, J. C., Swigut, T., and Cimprich, K. A. (2017). Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. *Cell* 170, 774–786e719. doi: 10.1016/j.cell.2017.07.043
- Helmrich, A., Ballarino, M., and Tora, L. (2011). Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol. Cell* 44, 966–977. doi: 10.1016/j.molcel.2011.10.013
- Kaushal, S., and Freudenreich, C. H. (2019). The role of fork stalling and DNA structures in causing chromosome fragility. *Genes Chromosomes Cancer* 58, 270–283. doi: 10.1002/gcc.22721
- Klein, C., and Westenberger, A. (2012). Genetics of Parkinson's disease. *Cold Spring Harb. Perspect. Med.* 2:a008888. doi: 10.1101/cshperspect.a008888
- Le Tallec, B., Millot, G. A., Blin, M. E., Brison, O., Dutrillaux, B., and Debatisse, M. (2013). Common fragile site profiling in epithelial and erythroid cells reveals that most recurrent cancer deletions lie in fragile sites hosting large genes. *Cell Rep.* 4, 420–428. doi: 10.1016/j.celrep.2013.07.003
- Li, Y., Roberts, N. D., Wala, J. A., Shapira, O., Schumacher, S. E., Kumar, K., et al. (2020). Patterns of somatic structural variation in human cancer genomes. *Nature* 578, 112–121. doi: 10.1038/s41586-019-1913-9
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(Delta Delta C(T)) Method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Macheret, M., and Halazonetis, T. D. (2018). Intragenic origins due to short G1 phases underlie oncogene-induced DNA replication stress. *Nature* 555, 112–116. doi: 10.1038/nature25507
- Oestergaard, V. H., and Lisby, M. (2016). Transcription-replication conflicts at chromosomal fragile sites—consequences in M phase and beyond. *Chromosoma* 126, 213–222. doi: 10.1007/s00412-016-0617-2
- Okamoto, Y., Iwasaki, W. M., Kugou, K., Takahashi, K. K., Oda, A., Sato, K., et al. (2018). Replication stress induces accumulation of FANCD2 at central region of large fragile genes. *Nucleic Acids Res.* 46, 2932–2944. doi: 10.1093/nar/gky058

- Pentzold, C., Shah, S. A., Hansen, N. R., Le Tallec, B., Seguin-Orlando, A., Debatisse, M., et al. (2018). FANCD2 binding identifies conserved fragile sites at large transcribed genes in avian cells. *Nucleic Acids Res.* 46, 1280–1294. doi: 10.1093/nar/gkx1260
- Schiavone, D., Guilbaud, G., Murat, P., Papadopoulou, C., Sarkies, P., Prioleau, M. N., et al. (2014). Determinants of G quadruplex-induced epigenetic instability in REV1-deficient cells. *EMBO J.* 33, 2507–2520. doi: 10.15252/embj.201488398
- Schwer, B., Wei, P. C., Chang, A. N., Kao, J., Du, Z., Meyers, R. M., et al. (2016). Transcription-associated processes cause DNA double-strand breaks and translocations in neural stem/progenitor cells. *Proc. Natl. Acad. Sci. U. S. A.* 113, 2258–2263. doi: 10.1073/pnas.1525564113
- Shang, W. H., Hori, T., Martins, N. M., Toyoda, A., Misu, S., Monma, N., et al. (2013). Chromosome engineering allows the efficient isolation of vertebrate neocentromeres. *Dev. Cell* 24, 635–648. doi: 10.1016/j.devcel.2013.02.009
- Smith, K. A., Gorman, P. A., Stark, M. B., Groves, R. P., and Stark, G. R. (1990). Distinctive chromosomal structures are formed very early in the amplification of CAD genes in Syrian hamster cells. *Cell* 63, 1219–1227. doi: 10.1016/0092-8674(90)90417-d
- Voutsinos, V., Munk, S. H. N., and Oestergaard, V. H. (2018). Common Chromosomal Fragile Sites-Conserved Failure Stories. *Genes* 9:580. doi: 10.3390/genes9120580
- Wei, P. C., Chang, A. N., Kao, J., Du, Z., Meyers, R. M., Alt, F. W., et al. (2016). Long Neural Genes Harbor Recurrent DNA Break Clusters in Neural Stem/Progenitor Cells. *Cell* 164, 644–655. doi: 10.1016/j.cell.2015.12.039
- West, A. B., Lockhart, P. J., O'Farrell, C., and Farrer, M. J. (2003). Identification of a novel gene linked to parkin via a bi-directional promoter. *J. Mol. Biol.* 326, 11–19. doi: 10.1016/s0022-2836(02)01376-1
- Wilson, T. E., Arlt, M. F., Park, S. H., Rajendran, S., Paulsen, M., Ljungman, M., et al. (2015). Large transcription units unify copy number variants and common fragile sites arising under replication stress. *Genome Res.* 25, 189–200. doi: 10.1101/gr.177121.114
- Zhang, C., Lin, M., Wu, R., Wang, X., Yang, B., Levine, A. J., et al. (2011). Parkin, a p53 target gene, mediates the role of p53 in glucose metabolism and the Warburg effect. *Proc. Natl. Acad. Sci. U. S. A.* 108, 16259–16264. doi: 10.1073/pnas.1113884108

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer VB declared a shared affiliation with one of the authors SM to the handling editor at the time of review.

Copyright © 2021 Munk, Voutsinos and Oestergaard. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Common Threads: Aphidicolin-Inducible and Folate-Sensitive Fragile Sites in the Human Genome

Rachel Adihe Lokanga¹, Daman Kumari² and Karen Usdin^{2*}

¹Cancer Genetics Branch, National Cancer Institute, Bethesda, MD, United States, ²Laboratory of Cell and Molecular Biology, National Institute of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, United States

OPEN ACCESS

Edited by:

Advaita Madireddy,
Rutgers, The State University of New
Jersey, United States

Reviewed by:

Jeannine Gerhardt,
Cornell University, United States
Emanuela Volpi,
University of Westminster,
United Kingdom
Batsheva Kerem,
Hebrew University of Jerusalem,
Israel

*Correspondence:

Karen Usdin
karenu@nih.gov

Specialty section:

This article was submitted to
Human and Medical Genomics,
a section of the journal
Frontiers in Genetics

Received: 12 May 2021

Accepted: 28 July 2021

Published: 08 September 2021

Citation:

Lokanga RA, Kumari D and
Usdin K (2021) Common Threads:
Aphidicolin-Inducible and Folate-
Sensitive Fragile Sites in the
Human Genome.
Front. Genet. 12:708860.
doi: 10.3389/fgene.2021.708860

The human genome has many chromosomal regions that are fragile, demonstrating chromatin breaks, gaps, or constrictions on exposure to replication stress. Common fragile sites (CFSs) are found widely distributed in the population, with the largest subset of these sites being induced by aphidicolin (APH). Other fragile sites are only found in a subset of the population. One group of these so-called rare fragile sites (RFSs) is induced by folate stress. APH-inducible CFSs are generally located in large transcriptionally active genes that are A + T rich and often enriched for tracts of AT-dinucleotide repeats. In contrast, all the folate-sensitive sites mapped to date consist of transcriptionally silenced CGG microsatellites. Thus, all the folate-sensitive fragile sites may have a very similar molecular basis that differs in key ways from that of the APH CFSs. The folate-sensitive FSs include FRAXA that is associated with Fragile X syndrome (FXS), the most common heritable form of intellectual disability. Both CFSs and RFSs can cause chromosomal abnormalities. Recent work suggests that both APH-inducible fragile sites and FRAXA undergo Mitotic DNA synthesis (MiDAS) when exposed to APH or folate stress, respectively. Interestingly, blocking MiDAS in both cases prevents chromosome fragility but increases the risk of chromosome mis-segregation. MiDAS of both APH-inducible and FRAXA involves conservative DNA replication and POLD3, an accessory subunit of the replicative polymerase Pol δ that is essential for break-induced replication (BIR). Thus, MiDAS is thought to proceed *via* some form of BIR-like process. This review will discuss the recent work that highlights the similarities and differences between these two groups of fragile sites and the growing evidence for the presence of many more novel fragile sites in the human genome.

Keywords: break-induced DNA replication, mitotic DNA synthesis, SLX1-SLX4, MUS81/EME1, replication fork blockage, R-loops, origins of replication, secondary DNA structures

INTRODUCTION

Fragile sites are apparent as chromatin gaps, constrictions, or breaks in cells exposed to replication stress (Sutherland, 1991). These sites are typically classified based on the reagent that induces their expression most effectively. They are also classified as common or rare, depending on their frequency in the population (Feng and Chakraborty, 2017). The largest known group of fragile

sites are most efficiently induced by aphidicolin (APH), an inhibitor of DNA polymerases α , δ , and ϵ . FRA3B and FRA16D are among the best known APH inducible CFSs. FRA3B is associated with the fragile histidine triad (*FHIT*) gene, a tumor suppressor gene located on chromosome 3p14.2 and FRA16D is associated with the WW domain-containing oxidoreductase (*WWOX*) gene, a tumor suppressor gene located on chromosome 16 (Bednarek et al., 2000). Another group of fragile sites are referred to as being folate-sensitive since they are induced by either too much or too little folate, with both situations resulting in nucleotide pool imbalances (Glover, 1981; James et al., 1993). Perhaps, the best known of the folate-sensitive fragile sites is the rare fragile site, FRAXA, a site on the X chromosome that is seen in individuals with fragile X syndrome (FXS), the most common heritable cause of intellectual disability and autism spectrum disorder (Lozano et al., 2014). Other fragile sites are induced by agents, such as 5-azacytidine, 5-bromo-2-deoxyuridine, or distamycin A that can be incorporated or intercalated into DNA (Schmid et al., 1980; Sutherland et al., 1985; Hori et al., 1988). Interestingly, FRA16B and FRA10B, two rare distamycin-inducible fragile sites, are AT-rich minisatellites (Yu et al., 1997; Hewett et al., 1998) that are expansions of the AT microsatellites normally present in the CFSs FRA16C and FRA10E, respectively (Zlotorynski et al., 2003). As such, they may share common features with the APH-inducible sites. While most fragile sites replicate late in the cell cycle, early replicating fragile sites (ERFSs) have also been identified that are readily induced by hydroxyurea, a reagent that causes depletion of deoxynucleotide pools (Barlow et al., 2013).

Fragile sites are all thought to be regions of the genome that for some reason are slow to complete replication, and their presence is associated with a variety of chromosome abnormalities. Genome instability at CFSs is thought to be a driving force for tumorigenesis with APH-CFSs being associated with copy number variations, including a variety of recurrent cancer deletions (Le Tallec et al., 2013; Wilson et al., 2015; Zheglo et al., 2019). Some CFS-associated CNVs are also associated with neurological disorders (Denison et al., 2003; Ambroziak et al., 2015; Zheglo et al., 2019). CFSs are also frequent sites of viral integration associated with cancer (Thorland et al., 2000; Yu et al., 2005). In contrast to the CNVs associated with CFSs, ERFSs are associated with recurrent chromosomal rearrangements during lymphomagenesis (Barlow et al., 2013). The RFS FRAXA is associated with a high frequency loss of the affected X chromosome *in vitro* in response to folate stress (Bjerregaard et al., 2018) and *in vivo* (Dobkin et al., 2009), and many cases of Jacobsen (11q-) syndrome, a chromosomal deletion disorder affecting chromosome 11, have been attributed to the presence of folate-sensitive fragile sites on that chromosome (Jones et al., 1994).

THE MOLECULAR BASIS OF THE REPLICATION PROBLEMS AT CFSs AND FOLATE-SENSITIVE FSs

Unlike ERFSs which are located in early replicating G + C-rich, gene-dense regions with high numbers of activated origins of

replication (ORIs) (Barlow et al., 2013), many APH-inducible CFSs are located in active, A + T-rich genes that are >300kb in size, replicate late, and are frequently ORI-poor (Glover et al., 2017; Debatisse and Rosselli, 2019). CFSs have been reported to be located at topologically associated domains (TADs) in some studies (Sarni et al., 2020), but not others (Ji et al., 2020). Some CFSs are associated with the expression of different oncogenes that can modulate replication stress (Miron et al., 2015). Transcription is required for CFS expression (Helmrich et al., 2011; Park et al., 2021), although higher transcription rates are associated with reduced fragility, perhaps due to the associated shift of the locus to replication earlier in the cell cycle (Blin et al., 2019). The relationship to transcription likely explains the reported tissue specificity of CFS expression.

Many different models have been proposed to account for the replication difficulties of CFSs, including those invoking replication-transcription collisions that promote R-loop formation and ultimately the stalling of the replication fork (Helmrich et al., 2011) and/or structural blocks to replication fork progression resulting from hairpin or cruciform formation by the AT-dinucleotide-rich regions embedded within many CFSs (Zlotorynski et al., 2003; Ozeri-Galai et al., 2011; Irony-Tur Sinai et al., 2019; Van Wietmarschen et al., 2020). In addition, TAD boundaries located between different replication timing zones are known to be prone to replication fork stalling (Lombardi and Tarsounas, 2020). Since ORIs are only licensed in G1 and bound pre-replication complexes can be displaced by RNA Pol II, at least in yeast (Snyder et al., 1988), it has also been suggested that transcription of long genes results in a paucity of active ORIs within the gene body that delays the completion of replication (Brison et al., 2019). Parenthetically, while a paucity of ORIs is associated with replication stress at CFSs, it has been suggested that increased ORI initiation at ERFSs also causes replication stress, perhaps by prematurely depleting nucleotide pools or by increasing replication-transcription collisions (Barlow et al., 2013).

However, while some studies support a role of R-loops in replication stress at fragile sites, including FRA3B (Helmrich et al., 2011), others do not (Park et al., 2021). Furthermore, while molecular combing has demonstrated replication stalling at FRA16C (Ozeri-Galai et al., 2011) and at FRA16D and FRA6E in *FANCD2*^{-/-} cells (Madireddy et al., 2016), combing studies of FRA3B and FRA6E showed no evidence of abnormal fork speed or replication fork stalling in normal APH-treated cells (Palumbo et al., 2010; Letessier et al., 2011). The lack of stalling at FRA3B together with the fact that transcription inhibition in S phase did not affect fragile site expression would be consistent with the idea that stalled replication forks and/or replication-transcription collisions are not a major source of replication stress at all CFSs (Brison et al., 2019). In addition, while delayed replication and their presence within large, transcriptionally active genes are consistent features of CFSs, these features are not sufficient for fragility, since a number of active, large genes that replicate late are not fragile (Wilson et al., 2015; Sarni et al., 2020; Park et al., 2021). Thus, the precise nature of the replication problem or problems at CFSs remains enigmatic and current thinking is that a combination

of different factors may contribute to replication stress at different loci.

Unlike CFSs, many of the RFSs involve a much shorter region of DNA, usually 0.6–5 kb. Of the 10 folate-sensitive RFSs characterized to date, all consist of a single tract of >200 CGG repeats (Table 1). In most cases, the repeat is in the 5' UTR of a gene that is epigenetically silenced (Lukusa and Fryns, 2008). Thus, fragility of these sites is likely to have a similar molecular basis. These sites are often associated with human disease, most commonly intellectual disability and autism spectrum disorder. However, it is not the fragile site itself that is responsible for this pathology, but rather the silencing-associated loss of the affected gene product. In the case of FRAXA and its associated disorder, FXS, the CGG repeat tract is located in the 5' UTR of the X-linked *FMR1* gene. The CGG repeat tract is prone to two forms of instability, the tendency to gain repeats with time, a hallmark of the repeat expansion diseases (Paulson, 2018) and the propensity to show fragility and sex chromosome aneuploidy (Dobkin et al., 2009). Both CGG repeats and the complementary CCG repeat form secondary structures, including hairpins and either G4 quadruplexes or i-motif structures [reviewed in Mirkin (2006)]. *In vitro* the CGG repeats show a K⁺ specific block to DNA synthesis consistent with the underlying problem being the formation of a G4 structure (Usdin and Woodford, 1995). The repeats also stall DNA synthesis in mammalian model systems (Voineagu et al., 2009) and in the endogenous *FMR1* locus (Gerhardt et al., 2014). In contrast to APH-inducible sites, the expression of FRAXA requires transcriptional silencing since those rare FXS alleles that escape silencing are not fragile (Yudkin et al., 2014). DNA methylation associated with gene silencing could increase the stability of fork blocking structures (Hardin et al., 1993; Lin et al., 2013). However, it is probable that silencing *per se* is not the trigger for fragility, but rather the delayed replication associated with silencing; transcribed *FMR1* alleles replicate late in the cell cycle, in S4 or G2 (Hansen et al., 2010), with silenced FXS alleles replicating even later (Webb, 1992; Hansen et al., 1997). Folate stress

would delay this even further. FXS alleles lack the association with a TAD boundary that is seen in normal alleles (Sun et al., 2018). Thus, while APH CFSs and FRAXA share some common features, the underlying problems responsible for these different classes of fragile sites are likely to also be different.

THE DOWNSTREAM CONSEQUENCES OF THE REPLICATION PROBLEMS AT FRAGILE SITES

Both APH-inducible CFSs and the FRAXA locus undergo mitotic DNA synthesis (MiDAS; Minocherhomji et al., 2015; Bhowmick et al., 2016; Garribba et al., 2020), a salvage pathway that ensures that regions of the genome that have not completed replication by the start of mitosis are successfully duplicated before the cell divides (Minocherhomji et al., 2015). Given that folate-sensitive fragile sites are all comprised of long CGG microsatellites, it is reasonable to think that other folate-sensitive fragile sites undergo MiDAS as well. MiDAS at both APH-inducible CFSs and FRAXA shares some common features. Both proceed *via* conservative DNA replication, in which DNA synthesis is confined to just one of the sister chromatids. They both also require POLD3 (Bhowmick et al., 2016; Garribba et al., 2020). POLD3, an accessory subunit of the replicative polymerase Pol δ , is not required for normal chromosomal replication, but is required for break-induced replication (BIR) (Costantino et al., 2014), a form of homologous recombination involved in the repair of one-sided double-strand breaks (DSBs) arising at collapsed replication forks. Thus, MiDAS has many of the hallmarks of a BIR-related process. BIR usually proceeds *via* the cleavage of the leading strand template by one of the structure-selective endonucleases: XPF-ERCC1, MUS81-EME1, or SLX1-SLX4. Cleavage results in a free 3' DNA tail that can strand-invade the sister chromatid to create a D-loop thus allowing POLD3-dependent DNA synthesis to proceed using the sister chromatid as a template. Successful completion of BIR at fragile site loci results in completely replicated chromatids that can be properly segregated into daughter cells in anaphase. Inhibition of BIR, on the other hand, results in the reduced expression of both CFSs and FRAXA, consistent with the idea that fragility is an active process resulting from MiDAS that has not been completed by the time normal chromatin condensation begins (Minocherhomji et al., 2015; Garribba et al., 2020). BIR frequently involves multiple rounds of strand invasion, DNA synthesis, and dissociation (Smith et al., 2007). Dissociation at one interspersed repeat and reinvasion into a different one could produce deletions, if the second repeat was downstream of the first one, and duplications if upstream. This could contribute to the high incidence of CNVs associated with fragile sites. Repeated mispriming within the repeat tract during BIR could also account for the tendency of CGG repeat tracts to expand (Kononenko et al., 2018). However, in the case of CGG repeats at the *FMR1* locus expansions occur in cells like ova that do not replicate (Yrigollen et al., 2014; Zhao and Usdin, 2018) and, in contrast to fragility, expansion at

TABLE 1 | Folate-sensitive rare fragile sites known to be associated with CGG microsatellites.

Fragile site/disorder	Gene	References
FRA2A ID	<i>AFF3</i>	Metsu et al., 2014b
FRA7A autism spectrum disorder	<i>ZNF713</i>	Metsu et al., 2014a
FRA10A*	<i>FRA10AC1</i>	Sarafidou et al., 2004
FRA11A*	<i>C11orf80</i>	Debacker et al., 2007
FRA11B [§]	<i>CBL2</i>	Jones et al., 1994
FRA12A ID	<i>DIP2B</i>	Winnepenninckx et al., 2007
FRA16A*	<i>XYLT1</i>	Nancarrow et al., 1994
FRAXA ID/ <i>FMR1</i> disorders	<i>FMR1</i>	Verkerk et al., 1991
FRAXE ID	<i>FMR2/AFF2</i>	Knight et al., 1993
FRAXF*	<i>FAM11A</i>	Parrish et al., 1994; Shaw et al., 2002

*not associated with disease. [§]responsible for some cases of Jacobsen syndrome, a chromosome deletion syndrome.

this locus requires transcription (Lokanga et al., 2014). Thus, the trigger for fragility and expansion of CGG repeats may differ.

While both the expression of CFSs and FRAXA likely involve some form of BIR, the process at these sites differs with respect to some of the proteins involved as illustrated in **Figure 1**. Specifically, initiation of BIR at APH-inducible sites involves cleavage of the replication intermediates by Mus81-EME1 (Ying et al., 2013) acting in conjunction with the scaffolding protein, SLX4 (Minocherhomji et al., 2015). Processing of the cleavage product requires Rad52 (Ying et al., 2013). In contrast, BIR at FRAXA requires the RAD51 recombinase and the SLX1-SLX4 endonuclease (Garribba et al., 2020). It has been suggested that the DNA secondary structures formed by the CGG repeat tract result in an atypical stalled fork that is a poor substrate for MUS81-EME1 (Garribba et al., 2020), a complex that specifically nicks duplex DNA on the 5'-side of a single-stranded/double-stranded DNA branch point (Wyatt et al., 2013). In contrast, the SLX1 endonuclease, which is activated by binding to SLX4, has a wider range of possible substrates and can incise duplex or single-stranded DNA on either the 5'- or 3'-sides of the branch point, thus allowing SLX1-SLX4 to nick either the leading or lagging strand template (Wyatt et al., 2013).

This difference in substrates may account for the involvement of different enzymes for processing stalled replication forks in the case of CFSs and FRAXA (Garribba et al., 2020). Interestingly, in a tissue culture reporter system, siRNA knockdown of either RAD51 or RAD52 significantly reduced BIR-associated mutagenesis of the region flanking a CGG repeat tract (Kononenko et al., 2018). Whether this reflects two different BIR subpathways operating in these cells or some sort of hybrid process is unclear.

While initiation of MiDAS is required for cytogenetic expression of the fragile site, failure to initiate MiDAS at both sets of loci leads to increased formation of ultrafine bridges (UFBs) in anaphase (Minocherhomji et al., 2015; Bhowmick et al., 2016; Garribba et al., 2020). These UFBs are anaphase bridges that do not stain with conventional DNA stains like Hoechst or DAPI and are not associated with histones (Chan et al., 2009). Failure to resolve these UFBs results the formation of micronuclei and chromosome mis-segregation (Fernandez-Casas and Chan, 2018). Perhaps not surprisingly given the differences in the underlying cause of replication fork stalling, CFSs and the FRAXA locus also differ in the nature of the UFBs that are formed when MiDAS does not occur. The UFBs formed at APH-inducible sites are coated with PICH, a DNA

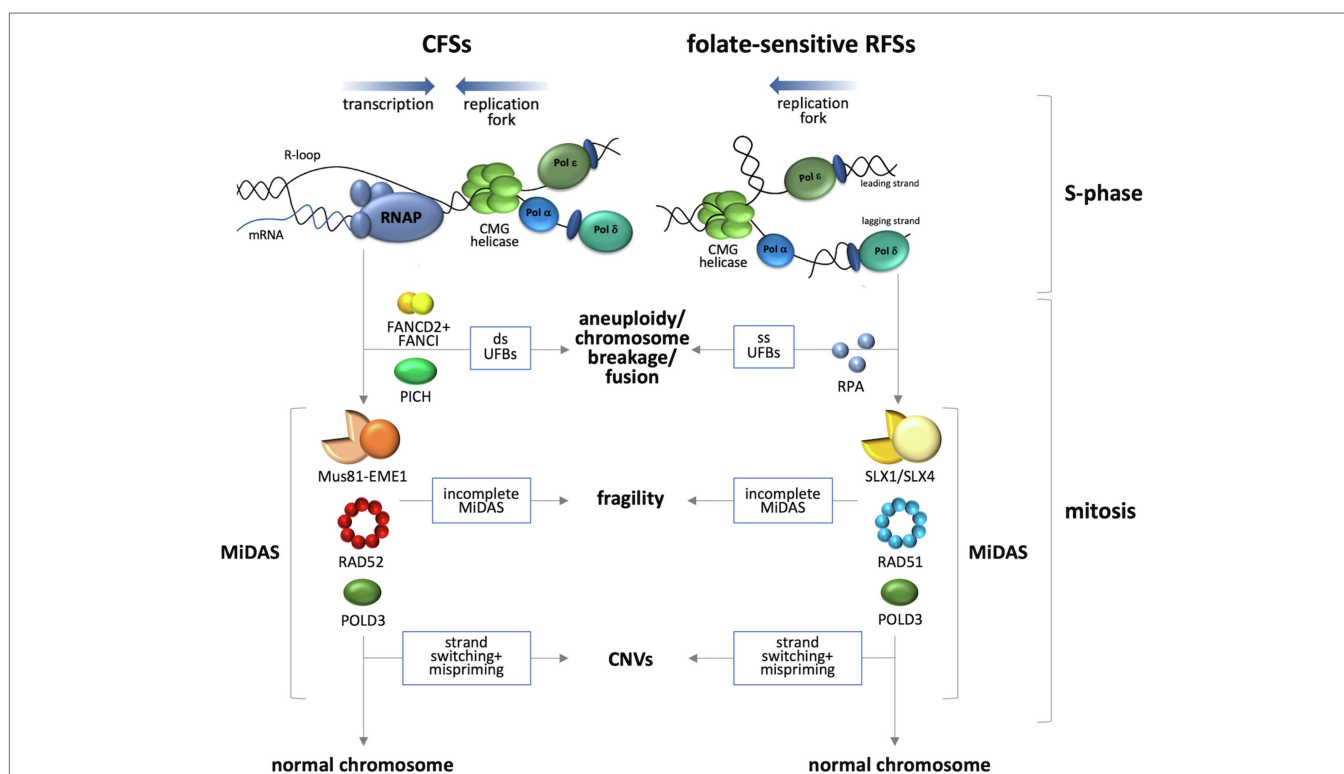


FIGURE 1 | Current models for events occurring at loci containing APH-inducible and folate-sensitive FSs. APH-inducible sites are difficult to replicate, a situation that may be exacerbated by collisions between the replication fork and the transcription complex. A head-on collision is depicted here since it is associated with elevated levels of DSBs as well as the formation of R-loops (Hamperl et al., 2017) suggested to be important for fragility of these sites (Helmrich et al., 2011). Folate-sensitive RFSs are associated with replication problems resulting from the formation of a fork blocking lesion at a locus that is transcriptionally silent. Rescue of the stalled replication forks occurs by MiDAS that involves two different BIR subpathways (Wyatt et al., 2013; Ying et al., 2013; Minocherhomji et al., 2015; Bhowmick et al., 2016; Garribba et al., 2020). Completion of MiDAS produces a normal chromosome, while failure to do so results in chromosome fragility. Strand switching during BIR that results in mispriming can result in CNVs. Failure to initiate MiDAS results in either double-stranded UFBs in the case of the APH sites (Chan et al., 2007) or single-stranded UFBs in the case of FRAXA (Bjerregaard et al., 2018).

translocase, are double-stranded (Liu et al., 2014), and are bounded by FANCD2/FANCI foci (Chan et al., 2009). The absence of an effect of topoisomerase II inhibition on the frequency of these UFBs suggests that they reflect the presence of under-replicated DNA or unresolved replication intermediates rather than dsDNA catenanes (Chan et al., 2009). The UFBs associated with FRAXA on the other hand are not associated with FANCD2, FANCI, or PICH. Instead, they are coated with RPA (Bjerregaard et al., 2018) and are thus likely to be single-stranded, consistent with unresolved HR intermediates (Fernandez-Casas and Chan, 2018).

CONCLUDING REMARKS

Lessons learnt from these two groups of fragile sites have allowed many more potential fragile sites to be identified. For example, genome-wide mapping of loci that undergo MiDAS in the presence of APH has identified hundreds of potential new CFSs (Ji et al., 2020; Macheret et al., 2020). Genome-wide studies of loci showing a fragility signature consisting of a TAD boundary that overlaps a highly transcribed, large gene with APH-induced replication delay, also suggest the presence of additional sites (Sarni et al., 2020). In addition, folate stress induces MiDAS or γ -H2AX foci, a marker of DSBs, at many genomic loci in normal human cells (Kumari et al., 2009; Garribba et al., 2020), suggesting that there are also several common folate-sensitive fragile sites that are as yet uncharacterized. Furthermore, a recent study of epigenetic variation in the human genome suggests the existence of at least 19 rare, long, and silenced CGG repeat tracts that could well also be fragile (Garg et al., 2020).

In addition to the CGG repeat diseases associated with folate-sensitive RFSs, many other repeat expansion diseases are known (Paulson, 2018). The CTG repeats responsible for a subset of these disorders, block replication (Samadashwily et al., 1997; Pelletier et al., 2003), induce BIR (Kim et al., 2017), and cause chromosome fragility in yeast (Freudenreich et al., 1998; Freudenreich and Lahiri, 2004). They also block replication in human cells (Liu et al., 2012). Furthermore, when cells containing a reporter construct with (CTG)₁₀₀ repeats were treated with hydroxyurea replication-dependent DSBs were seen close to the replication fork (Gadgil et al., 2020). Increased fragility as evidenced by the loss of an adjacent fluorescent reporter was also seen along with evidence of BIR. Interestingly, unlike BIR at the FRAXA locus, this BIR was dependent on MUS81 (Gadgil et al., 2020). CTG and CAG repeats form

hairpins (reviewed in Mirkin, 2006), like the CGG and CCG repeats responsible for FRAXA. However, they do not form G4 or i-motif structures. The MUS81 requirement for fragility may reflect this difference. Since some of the CTG/CAG expansion disorders can involve thousands of repeats (Fu et al., 1992; Mahadevan et al., 1992; Van Kuilenburg et al., 2019), they may also be fragile. However, no mitotic fragility has been reported for individuals with one such disorder, myotonic dystrophy type 1 (DM1; Jalal et al., 1993; Wenger et al., 1996). This may reflect the fact that, according to the ENCODE dataset, DMPK, the affected gene, replicates in the G1 or G1b phase of the cell cycle (Thurman et al., 2007; Hansen et al., 2010). Similarly, the GAA repeat tract responsible for Friedreich ataxia has key hallmarks of a mammalian fragile site: It blocks DNA synthesis (Krasilnikova and Mirkin, 2004; Gerhardt et al., 2016; Murat et al., 2020), is fragile in yeast (Kim et al., 2008), and is prone to chromosomal duplications in culture (Kumari et al., 2015). It is also associated with a high frequency of *de novo* mutations in the flanking regions (Bidichandani et al., 1999), a hallmark of BIR. However, as with the DM1 repeats, these loci are not apparent as gaps or constrictions in the chromatin in metaphase spreads, perhaps because they too replicate early in S phase (Kumari et al., 2015).

Thus, the repeats responsible for the repeat expansion diseases, may represent an unappreciated double threat to the human genome: the first threat being mediated *via* the deleterious effects of having large numbers of repeats in the DNA, RNA, and/or protein encoded by the affected loci (Paulson, 2018) and the second posed by the difficulty of replicating the repeats, with downstream effects on genome integrity, including aneuploidy, translocations, and CNVs. In addition to the repeats currently known to be associated with pathology, many thousands of other microsatellites with potential to stall DNA replication are known to be present in the human genome (Murat et al., 2020). Thus, the number of potentially fragile sites in the human genome could well be much higher than currently appreciated.

AUTHOR CONTRIBUTIONS

All authors contributed equally to this manuscript.

FUNDING

Funding for this work is provided by a grant from the Intramural Program of NIDDK, NIH to KU (1ZIADK057808).

REFERENCES

- Ambroziak, W., Koziorowski, D., Duszyc, K., Gorka-Skoczylas, P., Potulska-Chromik, A., Slawek, J., et al. (2015). Genomic instability in the PARK2 locus is associated with Parkinson's disease. *J. Appl. Genet.* 56, 451–461. doi: 10.1007/s13353-015-0282-9
- Barlow, J. H., Faryabi, R. B., Callen, E., Wong, N., Malhowski, A., Chen, H. T., et al. (2013). Identification of early replicating fragile sites that contribute to genome instability. *Cell* 152, 620–632. doi: 10.1016/j.cell.2013.01.006
- Bednarek, A. K., Laflin, K. J., Daniel, R. L., Liao, Q., Hawkins, K. A., and Aldaz, C. M. (2000). WWOX, a novel WW domain-containing protein mapping to human chromosome 16q23.3-24.1, a region frequently affected in breast cancer. *Cancer Res.* 60, 2140–2145.
- Bhowmick, R., Minocherhomji, S., and Hickson, I. D. (2016). RAD52 facilitates mitotic DNA synthesis following replication stress. *Mol. Cell* 64, 1117–1126. doi: 10.1016/j.molcel.2016.10.037

- Bidichandani, S. I., Purandare, S. M., Taylor, E. E., Gumin, G., Machkhas, H., and Harati, Y. (1999). Somatic sequence variation at the Friedreich ataxia locus includes complete contraction of the expanded GAA triplet repeat, significant length variation in serially passaged lymphoblasts and enhanced mutagenesis in the flanking sequence. *Hum. Mol. Genet.* 8, 2425–2436. doi: 10.1093/hmg/8.13.2425
- Bjerregaard, V. A., Garribba, L., McMurray, C. T., Hickson, I. D., and Liu, Y. (2018). Folate deficiency drives mitotic missegregation of the human FRAXA locus. *Proc. Natl. Acad. Sci.* 115, 13003–13008. doi: 10.1073/pnas.1808377115
- Blin, M., Le Tallec, B., Nahse, V., Schmidt, M., Brossas, C., Millot, G. A., et al. (2019). Transcription-dependent regulation of replication dynamics modulates genome stability. *Nat. Struct. Mol. Biol.* 26, 58–66. doi: 10.1038/s41594-018-0170-1
- Brisson, O., El-Hilali, S., Azar, D., Koundrioukoff, S., Schmidt, M., Nahse, V., et al. (2019). Transcription-mediated organization of the replication initiation program across large genes sets common fragile sites genome-wide. *Nat. Commun.* 10:5693. doi: 10.1038/s41467-019-13674-5
- Chan, K. L., North, P. S., and Hickson, I. D. (2007). BLM is required for faithful chromosome segregation and its localization defines a class of ultrafine anaphase bridges. *EMBO J.* 26, 3397–3409. doi: 10.1038/sj.emboj.7601777
- Chan, K. L., Palmai-Pallag, T., Ying, S., and Hickson, I. D. (2009). Replication stress induces sister-chromatid bridging at fragile site loci in mitosis. *Nat. Cell Biol.* 11, 753–760. doi: 10.1038/ncb1882
- Costantino, L., Sotiriou, S. K., Rantala, J. K., Magin, S., Mladenov, E., Helleday, T., et al. (2014). Break-induced replication repair of damaged forks induces genomic duplications in human cells. *Science* 343, 88–91. doi: 10.1126/science.1243211
- Debacker, K., Winnepeninckx, B., Longman, C., Colgan, J., Tolmie, J., Murray, R., et al. (2007). The molecular basis of the folate-sensitive fragile site FRA11A at 11q13. *Cytogenet. Genome Res.* 119, 9–14. doi: 10.1159/000109612
- Debatisse, M., and Rosselli, F. (2019). A journey with common fragile sites: From S phase to telophase. *Genes Chromosomes Cancer* 58, 305–316. doi: 10.1002/gcc.22704
- Denison, S. R., Callahan, G., Becker, N. A., Phillips, L. A., and Smith, D. I. (2003). Characterization of FRA6E and its potential role in autosomal recessive juvenile parkinsonism and ovarian cancer. *Genes Chromosomes Cancer* 38, 40–52. doi: 10.1002/gcc.10236
- Dobkin, C., Radu, G., Ding, X. H., Brown, W. T., and Nolin, S. L. (2009). Fragile X prenatal analyses show full mutation females at high risk for mosaic turner syndrome: fragile X leads to chromosome loss. *Am. J. Med. Genet. A* 149A, 2152–2157. doi: 10.1002/ajmg.a.33011
- Feng, W., and Chakraborty, A. (2017). Fragility extraordinaire: unsolved mysteries of chromosome fragile sites. *Adv. Exp. Med. Biol.* 1042, 489–526. doi: 10.1007/978-981-10-6955-0_21
- Fernandez-Casas, M., and Chan, K. L. (2018). The unresolved problem of DNA bridging. *Genes* 9:623. doi: 10.3390/genes9120623
- Freudenreich, C. H., Kantrow, S. M., and Zakian, V. A. (1998). Expansion and length-dependent fragility of CTG repeats in yeast. *Science* 279, 853–856. doi: 10.1126/science.279.5352.853
- Freudenreich, C. H., and Lahiri, M. (2004). Structure-forming CAG/CTG repeat sequences are sensitive to breakage in the absence of Mrc1 checkpoint function and S-phase checkpoint signaling: implications for trinucleotide repeat expansion diseases. *Cell Cycle* 3, 1370–1374. doi: 10.4161/cc.3.11.1246
- Fu, Y. H., Pizzuti, A., Fenwick, R. G., King, J., Rajnarayan, S., Dunne, P. W., et al. (1992). An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* 255, 1256–1258. doi: 10.1126/science.1546326
- Gadgil, R. Y., Romer, E. J., Goodman, C. C., Rider, S. D., Damewood, F. J., Barthelemy, J. R., et al. (2020). Replication stress at microsatellites causes DNA double-strand breaks and break-induced replication. *J. Biol. Chem.* 295, 15378–15397. doi: 10.1074/jbc.RA120.013495
- Garg, P., Jadhav, B., Rodriguez, O. L., Patel, N., Martin-Trujillo, A., Jain, M., et al. (2020). A survey of rare epigenetic variation in 23,116 human genomes identifies disease-relevant Epivariations and CGG expansions. *Am. J. Hum. Genet.* 107, 654–669. doi: 10.1016/j.ajhg.2020.08.019
- Garribba, L., Bjerregaard, V. A., Goncalves Dinis, M. M., Ozer, O., Wu, W., Sakellariou, D., et al. (2020). Folate stress induces SLX1- and RAD51-dependent mitotic DNA synthesis at the fragile X locus in human cells. *Proc. Natl. Acad. Sci.* 117, 16527–16536. doi: 10.1073/pnas.1921219117
- Gerhardt, J., Bhalla, A. D., Butler, J. S., Puckett, J. W., Dervan, P. B., Rosenwaks, Z., et al. (2016). Stalled DNA replication forks at the endogenous GAA repeats drive repeat expansion in Friedreich's ataxia cells. *Cell Rep.* 16, 1218–1227. doi: 10.1016/j.celrep.2016.06.075
- Gerhardt, J., Tomishima, M. J., Zaninovic, N., Colak, D., Yan, Z., Zhan, Q., et al. (2014). The DNA replication program is altered at the FMR1 locus in fragile X embryonic stem cells. *Mol. Cell* 53, 19–31. doi: 10.1016/j.molcel.2013.10.029
- Glover, T. W. (1981). FUDR induction of the X chromosome fragile site: evidence for the mechanism of folic acid and thymidine inhibition. *Am. J. Hum. Genet.* 33, 234–242.
- Glover, T. W., Wilson, T. E., and Arlt, M. F. (2017). Fragile sites in cancer: more than meets the eye. *Nat. Rev. Cancer* 17, 489–501. doi: 10.1038/nrc.2017.52
- Hamperl, S., Bocek, M. J., Saldivar, J. C., Swigut, T., and Cimprich, K. A. (2017). Transcription-replication conflict orientation modulates R-loop levels and activates distinct DNA damage responses. *Cell* 170:e719. doi: 10.1016/j.cell.2017.07.043
- Hansen, R. S., Canfield, T. K., Fjeld, A. D., Mumm, S., Laird, C. D., and Gartler, S. M. (1997). A variable domain of delayed replication in FRAXA fragile X chromosomes: X inactivation-like spread of late replication. *Proc. Natl. Acad. Sci.* 94, 4587–4592.
- Hansen, R. S., Thomas, S., Sandstrom, R., Canfield, T. K., Thurman, R. E., Weaver, M., et al. (2010). Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl. Acad. Sci.* 107, 139–144. doi: 10.1073/pnas.0912402107
- Hardin, C. C., Corregan, M., Brown, B. A. 2nd, and Frederick, L. N. (1993). Cytosine-cytosine+ base pairing stabilizes DNA quadruplexes and cytosine methylation greatly enhances the effect. *Biochemistry* 32, 5870–5880. doi: 10.1021/bi00073a021
- Helmrich, A., Ballarino, M., and Tora, L. (2011). Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol. Cell* 44, 966–977. doi: 10.1016/j.molcel.2011.10.013
- Hewett, D. R., Handt, O., Hobson, L., Mangelsdorf, M., Eyre, H. J., Baker, E., et al. (1998). FRA10B structure reveals common elements in repeat expansion and chromosomal fragile site genesis. *Mol. Cell* 1, 773–781. doi: 10.1016/S1097-2765(00)80077-5
- Hori, T., Takahashi, E., and Murata, M. (1988). Nature of distamycin A-inducible fragile sites. *Cancer Genet. Cytogenet.* 34, 189–194. doi: 10.1016/0165-4608(88)90258-0
- Irony-Tur Sinai, M., Salamon, A., Stanleigh, N., Goldberg, T., Weiss, A., Wang, Y. H., et al. (2019). AT-dinucleotide rich sequences drive fragile site formation. *Nucleic Acids Res.* 47, 9685–9695. doi: 10.1093/nar/gkz689
- Jalal, S. M., Lindor, N. M., Michels, V. V., Buckley, D. D., Hoppe, D. A., Sarkar, G., et al. (1993). Absence of chromosome fragility at 19q13.3 in patients with myotonic dystrophy. *Am. J. Med. Genet.* 46, 441–443. doi: 10.1002/ajmg.1320460419
- James, S. J., Miller, B. J., Cross, D. R., McGarrity, L. J., and Morris, S. M. (1993). The essentiality of folate for the maintenance of deoxynucleotide precursor pools, DNA synthesis, and cell cycle progression in PHA-stimulated lymphocytes. *Environ. Health Perspect.* 101(Suppl 5), 173–178.
- Ji, F., Liao, H., Pan, S., Ouyang, L., Jia, F., Fu, Z., et al. (2020). Genome-wide high-resolution mapping of mitotic DNA synthesis sites and common fragile sites by direct sequencing. *Cell Res.* 30, 1009–1023. doi: 10.1038/s41422-020-0357-y
- Jones, C., Slijepcevic, P., Marsh, S., Baker, E., Langdon, W. Y., Richards, R. I., et al. (1994). Physical linkage of the fragile site FRA11B and a Jacobsen syndrome chromosome deletion breakpoint in 11q23.3. *Hum. Mol. Genet.* 3, 2123–2130. doi: 10.1093/hmg/3.12.2123
- Kim, H. M., Narayanan, V., Mieczkowski, P. A., Petes, T. D., Krasilnikova, M. M., Mirkin, S. M., et al. (2008). Chromosome fragility at GAA tracts in yeast depends on repeat orientation and requires mismatch repair. *EMBO J.* 27, 2896–2906. doi: 10.1038/emboj.2008.205
- Kim, J. C., Harris, S. T., Dinter, T., Shah, K. A., and Mirkin, S. M. (2017). The role of break-induced replication in large-scale expansions of (CAG)_n/(CTG)_n repeats. *Nat. Struct. Mol. Biol.* 24, 55–60. doi: 10.1038/nsmb.3334
- Knight, S. J., Flannery, A. V., Hirst, M. C., Campbell, L., Christodoulou, Z., Phelps, S. R., et al. (1993). Trinucleotide repeat amplification and

- hypermethylation of a CpG island in FRAXE mental retardation. *Cell* 74, 127–134. doi: 10.1016/0092-8674(93)90300-F
- Kononenko, A. V., Ebersole, T., Vasquez, K. M., and Mirkin, S. M. (2018). Mechanisms of genetic instability caused by (CGG)_n repeats in an experimental mammalian system. *Nat. Struct. Mol. Biol.* 25, 669–676. doi: 10.1038/s41594-018-0094-9
- Krasilnikova, M. M., and Mirkin, S. M. (2004). Replication stalling at Friedreich's ataxia (GAA)_n repeats in vivo. *Mol. Cell. Biol.* 24, 2286–2295. doi: 10.1128/MCB.24.6.2286-2295.2004
- Kumari, D., Hayward, B., Nakamura, A. J., Bonner, W. M., and Usdin, K. (2015). Evidence for chromosome fragility at the frataxin locus in Friedreich ataxia. *Mutat. Res.* 781, 14–21. doi: 10.1016/j.mrfmmm.2015.08.007
- Kumari, D., Somma, V., Nakamura, A. J., Bonner, W. M., D'ambrosio, E., and Usdin, K. (2009). The role of DNA damage response pathways in chromosome fragility in fragile X syndrome. *Nucleic Acids Res.* 37, 4385–4392. doi: 10.1093/nar/gkp391
- Le Tallec, B., Millot, G. A., Blin, M. E., Brison, O., Dutrillaux, B., and Debatisse, M. (2013). Common fragile site profiling in epithelial and erythroid cells reveals that most recurrent cancer deletions lie in fragile sites hosting large genes. *Cell Rep.* 4, 420–428. doi: 10.1016/j.celrep.2013.07.003
- Letessier, A., Millot, G. A., Koundrioukoff, S., Lachages, A. M., Vogt, N., Hansen, R. S., et al. (2011). Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. *Nature* 470, 120–123. doi: 10.1038/nature09745
- Lin, J., Hou, J. Q., Xiang, H. D., Yan, Y. Y., Gu, Y. C., Tan, J. H., et al. (2013). Stabilization of G-quadruplex DNA by C-5-methyl-cytosine in bcl-2 promoter: implications for epigenetic regulation. *Biochem. Biophys. Res. Commun.* 433, 368–373. doi: 10.1016/j.bbrc.2012.12.040
- Liu, G., Chen, X., Gao, Y., Lewis, T., Barthelemy, J., and Leffak, M. (2012). Altered replication in human cells promotes DMPK (CTG)_n. (CAG)_n repeat instability. *Mol. Cell. Biol.* 32, 1618–1632. doi: 10.1128/MCB.06727-11
- Liu, Y., Nielsen, C. F., Yao, Q., and Hickson, I. D. (2014). The origins and processing of ultra fine anaphase DNA bridges. *Curr. Opin. Genet. Dev.* 26, 1–5. doi: 10.1016/j.gde.2014.03.003
- Lokanga, A. R., Zhao, X. N., Entezam, A., and Usdin, K. (2014). X inactivation plays a major role in the gender bias in somatic expansion in a mouse model of the fragile X-related disorders: implications for the mechanism of repeat expansion. *Hum. Mol. Genet.* 23, 4985–4994. doi: 10.1093/hmg/ddu213
- Lombardi, E. P., and Tarsounas, M. (2020). Topologically associating domain boundaries are enriched in early firing origins and restrict replication fork progression. *bioRxiv* [Preprint]. doi: 10.1101/2020.10.21.348946
- Lozano, R., Rosero, C. A., and Hagerman, R. J. (2014). Fragile X spectrum disorders. *Intractable Rare Dis Res* 3, 134–146. doi: 10.5582/iridr.2014.01022
- Lukusa, T. and Fryns, J. P., 2008. Human chromosome fragility. *Biochim Biophys Acta*, 1779: 3–16.
- Macheret, M., Bhowmick, R., Sobkowiak, K., Padayachy, L., Mailler, J., Hickson, I. D., et al. (2020). High-resolution mapping of mitotic DNA synthesis regions and common fragile sites in the human genome through direct sequencing. *Cell Res.* 30, 997–1008. doi: 10.1038/s41422-020-0358-x
- Madireddy, A., Kosiyatrakul, S. T., Boisvert, R. A., Herrera-Moyano, E., Garcia-Rubio, M. L., Gerhardt, J., et al. (2016). FANCD2 facilitates replication through common fragile sites. *Mol. Cell* 64, 388–404. doi: 10.1016/j.molcel.2016.09.017
- Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., et al. (1992). Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science* 255, 1253–1255. doi: 10.1126/science.1546325
- Metsu, S., Rainger, J. K., Debacker, K., Bernhard, B., Rooms, L., Grafodatskaya, D., et al. (2014a). A CGG-repeat expansion mutation in ZNF713 causes FRA7A: association with autistic spectrum disorder in two families. *Hum. Mutat.* 35, 1295–1300. doi: 10.1002/humu.22683
- Metsu, S., Rooms, L., Rainger, J., Taylor, M. S., Bengani, H., Wilson, D. I., et al. (2014b). FRA2A is a CGG repeat expansion associated with silencing of AFF3. *PLoS Genet.* 10:e1004242. doi: 10.1371/journal.pgen.1004242
- Minocherhomji, S., Ying, S., Bjerregaard, V. A., Bursomanno, S., Aleliunaitė, A., Wu, W., et al. (2015). Replication stress activates DNA repair synthesis in mitosis. *Nature* 528, 286–290. doi: 10.1038/nature16139
- Mirkin, S. M. (2006). DNA structures, repeat expansions and human hereditary disorders. *Curr. Opin. Struct. Biol.* 16, 351–358. doi: 10.1016/j.sbi.2006.05.004
- Miron, K., Golan-Lev, T., Dvir, R., Ben-David, E., and Kerem, B. (2015). Oncogenes create a unique landscape of fragile sites. *Nat. Commun.* 6:7094. doi: 10.1038/ncomms8094
- Murat, P., Guilbaud, G., and Sale, J. E. (2020). DNA polymerase stalling at structured DNA constrains the expansion of short tandem repeats. *Genome Biol.* 21:209. doi: 10.1186/s13059-020-02124-x
- Naim, V., Wilhelm, T., Debatisse, M., and Rosselli, F. (2013). ERCC1 and MUS81-EME1 promote sister chromatid separation by processing late replication intermediates at common fragile sites during mitosis. *Nat. Cell Biol.* 15, 1008–1015. doi: 10.1038/ncb2793
- Nancarrow, J. K., Kremer, E., Holman, K., Eyre, H., Doggett, N. A., Le Paslier, D., et al. (1994). Implications of FRA16A structure for the mechanism of chromosomal fragile site genesis. *Science* 264, 1938–1941. doi: 10.1126/science.8009225
- Ozeri-Galai, E., Lebofsky, R., Rahat, A., Bester, A. C., Bensimon, A., and Kerem, B. (2011). Failure of origin activation in response to fork stalling leads to chromosomal instability at fragile sites. *Mol. Cell* 43, 122–131. doi: 10.1016/j.molcel.2011.05.019
- Palumbo, E., Matricardi, L., Tosoni, E., Bensimon, A., and Russo, A. (2010). Replication dynamics at common fragile site FRA6E. *Chromosoma* 119, 575–587. doi: 10.1007/s00412-010-0279-4
- Park, S. H., Bennett-Baker, P., Ahmed, S., Arlt, M. F., Ljungman, M., Glover, T. W., et al. (2021). Locus-specific transcription silencing at the FHIT gene suppresses replication stress-induced copy number variant formation and associated replication delay. *Nucleic Acids Res.* 49, 7507–7524. doi: 10.1093/nar/gkab559
- Parrish, J. E., Oostra, B. A., Verkerk, A. J., Richards, C. S., Reynolds, J., Spikes, A. S., et al. (1994). Isolation of a GCC repeat showing expansion in FRAXE, a fragile site distal to FRAXA and FRAXE. *Nat. Genet.* 8, 229–235. doi: 10.1038/ng1194-229
- Paulson, H. (2018). Repeat expansion diseases. *Handb. Clin. Neurol.* 147, 105–123. doi: 10.1016/B978-0-444-63233-3.00009-9
- Pelletier, R., Krasilnikova, M. M., Samadashwily, G. M., Lahue, R., and Mirkin, S. M. (2003). Replication and expansion of trinucleotide repeats in yeast. *Mol. Cell. Biol.* 23, 1349–1357. doi: 10.1128/MCB.23.4.1349-1357.2003
- Samadashwily, G. M., Raca, G., and Mirkin, S. M. (1997). Trinucleotide repeats affect DNA replication in vivo. *Nat. Genet.* 17, 298–304. doi: 10.1038/ng1197-298
- Sarafidou, T., Kahl, C., Martinez-Garay, I., Mangelsdorf, M., Gesk, S., Baker, E., et al. (2004). Folate-sensitive fragile site FRA10A is due to an expansion of a CGG repeat in a novel gene, FRA10AC1, encoding a nuclear protein. *Genomics* 84, 69–81. doi: 10.1016/j.ygeno.2003.12.017
- Sarni, D., Sasaki, T., Irony Tur-Sinai, M., Miron, K., Rivera-Mulia, J. C., Magnuson, B., et al. (2020). 3D genome organization contributes to genome instability at fragile sites. *Nat. Commun.* 11:3613. doi: 10.1038/s41467-020-17448-2
- Schmid, M., Klett, C., and Niederhofer, A. (1980). Demonstration of a heritable fragile site in human chromosome 16 with distamycin A. *Cytogenet. Cell Genet.* 28, 87–94. doi: 10.1159/000131516
- Shaw, M. A., Chiurazzi, P., Romain, D. R., Neri, G., and Gecz, J. (2002). A novel gene, FAM11A, associated with the FRAXE CpG island is transcriptionally silent in FRAXE full mutation. *Eur. J. Hum. Genet.* 10, 767–772. doi: 10.1038/sj.ejhg.5200881
- Smith, C. E., Llorente, B., and Symington, L. S. (2007). Template switching during break-induced replication. *Nature* 447, 102–105. doi: 10.1038/nature05723
- Snyder, M., Sapolsky, R. J., and Davis, R. W. (1988). Transcription interferes with elements important for chromosome maintenance in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 8, 2184–2194. doi: 10.1128/mcb.8.5.2184-2194.1988
- Sun, J. H., Zhou, L., Emerson, D. J., Phyo, S. A., Titus, K. R., Gong, W., et al. (2018). Disease-associated short tandem repeats co-localize with chromatin domain boundaries. *Cell* 175:e215. doi: 10.1016/j.cell.2018.08.005
- Sutherland, G. R. (1991). Chromosomal fragile sites. *Genet. Anal. Tech. Appl.* 8, 161–166.
- Sutherland, G. R., Parslow, M. I., and Baker, E. (1985). New classes of common fragile sites induced by 5-azacytidine and bromodeoxyuridine. *Hum. Genet.* 69, 233–237. doi: 10.1007/BF00293031
- Thorland, E. C., Myers, S. L., Persing, D. H., Sarkar, G., McGovern, R. M., Gostout, B. S., et al. (2000). Human papillomavirus type 16 integrations in

- cervical tumors frequently occur in common fragile sites. *Cancer Res.* 60, 5916–5921.
- Thurman, R. E., Day, N., Noble, W. S., and Stamatoyannopoulos, J. A. (2007). Identification of higher-order functional domains in the human ENCODE regions. *Genome Res.* 17, 917–927. doi: 10.1101/gr.6081407
- Usdin, K., and Woodford, K. J. (1995). CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis in vitro. *Nucleic Acids Res.* 23, 4202–4209. doi: 10.1093/nar/23.10.4202
- Van Kuilenburg, A. B. P., Tarailo-Graovac, M., Richmond, P. A., Drogemoller, B. I., Pouladi, M. A., Leen, R., et al. (2019). Glutaminase deficiency caused by short tandem repeat expansion in GLS. *N. Engl. J. Med.* 380, 1433–1441. doi: 10.1056/NEJMoa1806627
- Van Wietmarschen, N., Sridharan, S., Nathan, W. J., Tubbs, A., Chan, E. M., Callen, E., et al. (2020). Repeat expansions confer WRN dependence in microsatellite-unstable cancers. *Nature* 586, 292–298. doi: 10.1038/s41586-020-2769-8
- Verkerk, A. J., Pieretti, M., Sutcliffe, J. S., Fu, Y. H., Kuhl, D. P., Pizzuti, A., et al. (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* 65, 905–914. doi: 10.1016/0092-8674(91)90397-H
- Voineagu, I., Surka, C. F., Shishkin, A. A., Krasilnikova, M. M., and Mirkin, S. M. (2009). Replisome stalling and stabilization at CGG repeats, which are responsible for chromosomal fragility. *Nat. Struct. Mol. Biol.* 16, 226–228. doi: 10.1038/nsmb.1527
- Webb, T. (1992). Delayed replication of Xq27 in individuals with the fragile X syndrome. *Am. J. Med. Genet.* 43, 1057–1062. doi: 10.1002/ajmg.1320430633
- Wenger, S. L., Giangreco, C. A., Tarleton, J., and Wessel, H. B. (1996). Inability to induce fragile sites at CTG repeats in congenital myotonic dystrophy. *Am. J. Med. Genet.* 66, 60–63. doi: 10.1002/(SICI)1096-8628(19961202)66:1<60::AID-AJMG13>3.0.CO;2-O
- Wilson, T. E., Arlt, M. F., Park, S. H., Rajendran, S., Paulsen, M., Ljungman, M., et al. (2015). Large transcription units unify copy number variants and common fragile sites arising under replication stress. *Genome Res.* 25, 189–200. doi: 10.1101/gr.177121.114
- Winnepenninckx, B., Debacker, K., Ramsay, J., Smeets, D., Smits, A., Fitzpatrick, D. R., et al. (2007). CGG-repeat expansion in the DIP2B gene is associated with the fragile site FRA12A on chromosome 12q13.1. *Am. J. Hum. Genet.* 80, 221–231. doi: 10.1086/510800
- Wyatt, H. D., Sarbajna, S., Matos, J., and West, S. C. (2013). Coordinated actions of SLX1-SLX4 and MUS81-EME1 for Holliday junction resolution in human cells. *Mol. Cell* 52, 234–247. doi: 10.1016/j.molcel.2013.08.035
- Ying, S., Minocherhomji, S., Chan, K. L., Palmai-Pallag, T., Chu, W. K., Wass, T., et al. (2013). MUS81 promotes common fragile site expression. *Nat. Cell Biol.* 15, 1001–1007. doi: 10.1038/ncb2773
- Yrigollen, C. M., Martorell, L., Durbin-Johnson, B., Naudo, M., Genoves, J., Murgia, A., et al. (2014). AGG interruptions and maternal age affect FMR1 CGG repeat allele stability during transmission. *J. Neurodev. Disord.* 6:24. doi: 10.1186/1866-1955-6-24
- Yu, S., Mangelsdorf, M., Hewett, D., Hobson, L., Baker, E., Eyre, H. J., et al. (1997). Human chromosomal fragile site FRA16B is an amplified AT-rich minisatellite repeat. *Cell* 88, 367–374. doi: 10.1016/S0092-8674(00)81875-9
- Yu, T., Ferber, M. J., Cheung, T. H., Chung, T. K., Wong, Y. F., and Smith, D. I. (2005). The role of viral integration in the development of cervical cancer. *Cancer Genet. Cytogenet.* 158, 27–34. doi: 10.1016/j.cancergencyto.2004.08.021
- Yudkin, D., Hayward, B. E., Aladjem, M. I., Kumari, D., and Usdin, K. (2014). Chromosome fragility and the abnormal replication of the FMR1 locus in fragile X syndrome. *Hum. Mol. Genet.* 23, 2940–2952. doi: 10.1093/hmg/ddu006
- Zhao, X. N., and Usdin, K. (2018). Timing of expansion of fragile X Premutation alleles During intergenerational transmission in a mouse model of the fragile X-related disorders. *Front. Genet.* 9:314. doi: 10.3389/fgene.2018.00314
- Zheglo, D., Brueckner, L. M., Sepman, O., Wecht, E. M., Kuligina, E., Susptsin, E., et al. (2019). The FRA14B common fragile site maps to a region prone to somatic and germline rearrangements within the large GPHN gene. *Genes Chromosomes Cancer* 58, 284–294. doi: 10.1002/gcc.22706
- Zlotorynski, E., Rahat, A., Skaug, J., Ben-Porat, N., Ozeri, E., Hershsberg, R., et al. (2003). Molecular basis for expression of common and rare fragile sites. *Mol. Cell. Biol.* 23, 7143–7151. doi: 10.1128/MCB.23.20.7143-7151.2003

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Lokanga, Kumari and Usdin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



RNF4 Regulates the BLM Helicase in Recovery From Replication Fork Collapse

Nathan Ellis^{1*}, Jianmei Zhu², Mary K Yagle¹, Wei-Chih Yang², Jing Huang³, Alexander Kwako¹, Michael M. Seidman³ and Michael J. Matunis^{2*}

¹University of Arizona Cancer Center, University of Arizona, Tucson, AZ, United States, ²Department of Biochemistry and Molecular Biology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States, ³Laboratory of Molecular Gerontology, National Institute on Aging, Baltimore, MD, United States

OPEN ACCESS

Edited by:

Qing Hu,
University of Texas Southwestern
Medical Center, United States

Reviewed by:

Samuel Bunting,
Rutgers, The State University of New
Jersey - Busch Campus, United States
Sudha Sharma,
Howard University, United States

*Correspondence:

Michael J. Matunis
mmatuni1@jhu.edu
Nathan Ellis
naellis@email.arizona.edu

Specialty section:

This article was submitted to
Human and Medical Genomics,
a section of the journal
Frontiers in Genetics

Received: 05 August 2021

Accepted: 25 October 2021

Published: 12 November 2021

Citation:

Ellis N, Zhu J, Yagle MK, Yang W-C,
Huang J, Kwako A, Seidman MM and
Matunis MJ (2021) RNF4 Regulates
the BLM Helicase in Recovery From
Replication Fork Collapse.
Front. Genet. 12:753535.
doi: 10.3389/fgene.2021.753535

Sumoylation is an important enhancer of responses to DNA replication stress and the SUMO-targeted ubiquitin E3 ligase RNF4 regulates these responses by ubiquitylation of sumoylated DNA damage response factors. The specific targets and functional consequences of RNF4 regulation in response to replication stress, however, have not been fully characterized. Here we demonstrated that RNF4 is required for the restart of DNA replication following prolonged hydroxyurea (HU)-induced replication stress. Contrary to its role in repair of γ -irradiation-induced DNA double-strand breaks (DSBs), our analysis revealed that RNF4 does not significantly impact recognition or repair of replication stress-associated DSBs. Rather, using DNA fiber assays, we found that the firing of new DNA replication origins, which is required for replication restart following prolonged stress, was inhibited in cells depleted of RNF4. We also provided evidence that RNF4 recognizes and ubiquitylates sumoylated Bloom syndrome DNA helicase BLM and thereby promotes its proteasome-mediated turnover at damaged DNA replication forks. Consistent with it being a functionally important RNF4 substrate, co-depletion of BLM rescued defects in the firing of new replication origins observed in cells depleted of RNF4 alone. We concluded that RNF4 acts to remove sumoylated BLM from collapsed DNA replication forks, which is required to facilitate normal resumption of DNA synthesis after prolonged replication fork stalling and collapse.

Keywords: Bloom syndrome, BLM, DNA repair, dormant origins, fork collapse, homologous recombination, hydroxyurea, RAD51

INTRODUCTION

Accurate DNA replication is essential to maintenance of genome integrity. When the replicative polymerase encounters DNA damage such as chemical modifications of bases, the polymerase stalls at the site of the DNA lesion and the CDC45-MCM2-7-GINS (CMG) helicase uncouples from the polymerase and continues to unwind downstream duplex to expose single-stranded DNA (ssDNA) (Cortez, 2019). ssDNA binding protein (RPA) binds to ssDNA, and the complex activates the ATR kinase, which is required for the recruitment of factors from the homologous recombination (HR) pathway (Dungrawala et al., 2015). These factors stabilize and protect the replication fork from nascent-strand degradation (Schlachter et al., 2011). Fork protection and stabilization are dependent on the RAD51 recombinase (Zellweger et al., 2015; Mijic et al., 2017), which is thought to engage with

other factors, including BRCA2 and SMARCA1, that reverse the fork to form a structure resembling a Holliday junction and prevent nascent strand degradation due to exonucleolytic attack (Hashimoto et al., 2010; Kolinjivadi et al., 2017).

Nucleotide deprivation using the ribonucleotide reductase inhibitor hydroxyurea (HU) has provided a powerful probe to examine the mechanisms of fork stabilization and protection because no specific DNA lesion is generated (Vesela et al., 2017). Uncoupling occurs due to DNA polymerases stalling on the template following a >50% decrease in purine deoxynucleotide concentration (Skoog and Bjursell, 1974). In this context, activation of ATR and the recruitment of HR factors, such as RAD51 and BRCA2, are replication-specific (Petermann et al., 2010; Zellweger et al., 2015). Early studies conducted with radiolabeled thymidine tracer demonstrated that ribonucleotide reductase inhibition achieved a cessation of DNA replication and, upon removal of HU, DNA replication resumed (Bianchi et al., 1986). Analysis of DNA replication dynamics based on single molecule DNA fiber assay has demonstrated that HU-stalled replication forks resume synthesis at the sites where they stalled (Petermann et al., 2010; Sidorova et al., 2013; Thangavel et al., 2015). However, prolonged treatment with HU (≥ 16 h) leads to irreversible fork collapse. Cells treated with HU ≥ 16 h have two cell populations, namely, cells that were in S phase when drug was added to the medium and cells that transited the cell cycle then entered and were stalled at the beginning of S phase (Karnani and Dutta, 2011). Because collapsed forks cannot be restarted, successful genome duplication in cells depends on the firing of dormant origins (Woodward et al., 2006; Ge and Blow, 2010). The majority of dormant origins fired in these cells are >100 kb away from the collapsed forks.

Multiple proteins that are associated with replication fork stability and HR are sumoylated (Xiao et al., 2015). Sumoylation is in turn regulated by recruitment of SUMO-targeted ubiquitin E3 ligases (STUbLs). STUbLs contain tandem SUMO interaction motifs (SIMs), which bind poly-sumoylated proteins, and RING domains that mediate ubiquitylation (Prudden et al., 2007). The prototypical mammalian STUbL, RNF4, was first identified as a transcriptional co-regulator of hormone receptors (Moilanen et al., 1998). The *S. cerevisiae* ortholog of mammalian RNF4, the Slx5/Slx8 heterodimer, was discovered in a screen for genes synthetically lethal with the BLM ortholog Sgs1 (Mullen et al., 2001). Yeast strains lacking *Slx5* or *Slx8* are hypersensitive to chronic DNA replication stress and exhibit elevated levels of gross chromosome rearrangements and spontaneous mutations (Mullen et al., 2001; Zhang et al., 2006; Prudden et al., 2007); *Slx5* or *Slx8* regulate the levels of numerous sumoylated HR proteins (Burgess et al., 2007). In mammalian cells, RNF4 has been found to operate in a variety of DNA repair functions. It controls the formation of double-strand breaks (DSB) in ATR-deficient cells undergoing replication stress (Ragland et al., 2013). It mediates the recruitment of BRCA1 to DSB sites through the generation of SUMO-ubiquitin hybrid chains (Guzzo et al., 2012). It is recruited to sites of DNA damage via sumoylated MDC1 and is required for exonucleolytic processing of DSBs preceding HR-

mediated repair (Galanty et al., 2012; Luo et al., 2012; Yin et al., 2012). RNF4-mediated ubiquitylation facilitates the extraction of proteins from DNA repair sites through recruitment of the Cdc48/p97 segregase (Nie et al., 2012), it regulates FANCI/FANCD2 turnover at stalled forks (Gibbs-Seymour et al., 2015), and it mediates the release of FAAP20 from sumoylated FANCA during interstrand crosslink repair (Xie et al., 2015). Although the function of RNF4 in DSB repair has been studied by many laboratories, its role in responding to fork collapse has not been well characterized in mammalian cells.

The BLM helicase has been implicated in replication fork stability as BLM-deficient cells exhibit multiple defects in DNA replication, including accumulation of abnormal DNA replication intermediates (Lonn et al., 1990), slower replication fork velocity (Rao et al., 2007), and excessive firing of dormant origins (Davies et al., 2007). BLM-deficient cells exhibit increased levels of chromatid breakage and HR, in particular, elevated sister chromatid exchanges (SCEs) (Chaganti et al., 1974). BLM interacts directly with both RAD51 (Wu et al., 2001; Bugreev et al., 2007; Patel et al., 2017) and RPA (Brosh et al., 2000; Doherty et al., 2005; Xu et al., 2008; Shorrocks et al., 2021). We have shown previously that BLM's function in DNA replication is regulated by sumoylation (Ouyang et al., 2009). Cells that express a sumoylation-deficient BLM (SD-BLM) accumulate lower levels of RAD51 at collapsed forks but higher levels of RPA (Eladad et al., 2005; Ouyang et al., 2009; Ouyang et al., 2013). Study of an RPA-binding-deficient BLM showed that BLM's capacity to bind RPA is required for its role in fork protection (Shorrocks et al., 2021), but the mechanism is unclear because RPA can block BLM's DNA unwinding activity *in vitro* (Xue et al., 2019). Levels of SCE are normal in both untreated SD-BLM cells (Eladad et al., 2005) and untreated RPA-binding-deficient BLM cells (Shorrocks et al., 2021); however, in SD-BLM cells subject to prolonged HU treatment, SCEs are not induced but instead DSBs accumulate, associating the failure to recruit RAD51 to collapsed forks with failure to repair DSBs that are generated there. Our recent work with mutations of *NSMCE2*, the SUMO E3 ligase responsible for BLM sumoylation at collapsed forks, suggests that at least some of these DSBs occur subsequent to dormant origin firing when converging forks merge with collapsed forks (Pond et al., 2019).

Because BLM's functions are regulated by sumoylation, we hypothesized that RNF4 promotes turnover of BLM at collapsed replication forks, possibly to facilitate downstream DSB processing events. We show here that sumoylated BLM is indeed an RNF4 substrate, but RNF4 is not required for DSB repair at collapsed forks. Instead, we found that deficiency in BLM turnover at collapsed forks resulted in defects in dormant origin firing.

MATERIALS AND METHODS

Cell Culture and Transfection

For knockdown experiments, cells were cultured to 30–50% confluence and transfected with Lipofectamine RNAiMAX (Invitrogen). NC1 negative control siRNA was obtained from

IDT, Inc., RNF4 siRNA #1 (5'-GACTCACAATGACTCTGT TGTGATT-3') from Invitrogen, and RNF4 siRNA #2 (5'-GAA UGGACGUCUCAUCGUUUU-3') from Dharmacon. SENP6 siRNA (5'-AAGAAAGTGAAGGAGATACAGUU-3') was obtained from Qiagen, Inc. and BLM siRNA (5'-UCCCGG GAUACUGCUCUCA-3') was from IDT, Inc. siRNAs were used at a final concentration of 20 nM.

Antibodies

Antibodies were obtained from the following sources: anti-RPA/p34 (Neomarkers MS-691-P0), anti- γ -H2AX (Millipore 05-636), rabbit anti-BLM (Eladad et al., 2005), rabbit anti-RNF4 (a gift from Dr. Jorma Palvimo), rabbit anti-CHK1 (Cell Signaling Technology 2345) at 1:400, rabbit anti-phospho-CHK1 (ser317) (Cell Signaling Technology 2344S) at 1:400, rat anti-HSC70 (Assay Design) at 1:45,000, anti-tubulin (Sigma T9026), anti-SUMO2 (Zhang et al., 2008), anti-Myc (Cell Signaling Technology 2276S), and anti-SENP6 (a gift from Dr. Mary Dasso). AlexaFluor-labeled secondary antibodies (A11029; A11035), were obtained from Invitrogen. HRP-labeled secondary antibodies were anti-mouse IgG (Cell Signaling Technology 7076S), anti-rat IgG (Jackson Labs), and anti-rabbit IgG (GE Healthcare NA934V). For DNA fiber assays, antibodies for detection of 5-iodo-2'-deoxyuridine (IdU) were mouse anti-IdU (BD) and for detection of 5-chloro-2'-deoxyuridine (CldU) were rat anti-CldU (Abcam); the secondary antibodies were anti-mouse Dylight 488 and anti-rat Dylight 649 (Jackson ImmunoResearch).

Clonogenic Survival Assay

U2OS or HeLa cells were transfected with control or RNF4 siRNAs. 48 h after transfection, cells were incubated with varying concentrations of HU for 72 h or of camptothecin (CPT) for 3 h. Cells were trypsinized after treatment and counted with a hemocytometer. Then, 200, 400, and 800 cells were seeded in duplicate into six-well or 60 mm plates in normal medium. After one to two weeks, clones with >50 cells were scored. Clonogenic survival was calculated as the average of number of clones over the number of cells seeded for all scorable wells or dishes. The results were normalized to the clonogenic survival of untreated, negative control condition. The experiment was repeated three times. The average of experiments and standard deviations were calculated.

Immunofluorescence Microscopy

For analysis of RNF4 localization, U2OS cells were seeded in 35-mm glass-bottom culture dishes. For HU-induced DNA damage, cells were incubated in the presence of 2 mM HU for 16 h. Laser-induced irradiation was performed as described (Muniandy et al., 2009). Irradiated cells were allowed to recover for 2 h before antibody staining. Cells were fixed and stained as described (Zhang et al., 2008). Images were collected using Zeiss Observer Z1 fluorescence microscope with an Apotome VH optical sectioning grid and processed using AxioVision Software Release 4.8.1.

For analysis of DNA repair foci numbers and focal areas, 50,000 HeLa cells were seeded on cover slips, transfected with siRNAs, and treated with 2 mM HU for 16 h. As a positive control for a known RNF4-dependent phenotype, we also treated cells with 10 μ M etoposide for 4 h, which is known to generate DSBs in all cell cycle phases. Following treatment, cells were pre-processed as described (Dimitrova and Gilbert, 2000) then fixed and stained as described (Zhang et al., 2008). Images were collected on a Zeiss LSM 710 Meta Confocal microscope using Zeiss LSM4.2 software. Foci numbers and areas were acquired in ImageJ/FIJI. The DAPI image was used to generate the nuclear regions of interest. For each experiment an image threshold of the red and green channels was determined using the brightest conditions (e.g., HU-treated control), and this threshold was used for all images acquired. Particles >2 pixels were counted in the resulting binary images using the Analyze Particles function. Quantification of numbers of foci was carried out using CellProfiler (version 2.0). 30 to 70 cells were counted in each experiment. Three experiments were performed and the three experiments were combined and box and whiskers plots prepared from the merged data.

Pulsed-Field Gel Electrophoresis to Measure Double-Strand Breaks

One million HeLa or U2OS cells were seeded into 35 mm dishes. The next day cells were transfected with negative control or RNF4 siRNAs. 48 h after transfection of siRNAs, the cells were treated with varying concentrations of CPT for 3 h. The cells were harvested, counted, and 250,000–3,000,000 cells were embedded in agarose plugs in agarose insert buffer (10 mM Tris-HCl pH 7.5, 20 mM NaCl, 50 mM EDTA). Plug preparation, cell lysis, and agarose gel electrophoresis were carried as described (Ouyang et al., 2009; Pond and Ellis, 2019). Gels were stained with SYBR Gold (1 part in 10,000 parts water) and the UV transilluminator image was analyzed using ImageJ Gel Analyzer. Arbitrary fluorescence units were normalized to untreated, untransfected controls. Experiments were repeated three times. The average of experiments and standard deviations were calculated.

Flow Cytometric Analysis

300,000 U2OS or HeLa cells or 150,000 HCT116 cells were seeded into 40 mm dishes and forward transfected the following day with negative control or RNF4-specific siRNAs. 24–30 h after transfection, the cells were treated or not with 2 mM HU for 16 h. The cells were released from the HU block and incubated in media containing 20 μ M BrdU for 20, 30, 40, 60, or 120 min prior to harvest. 20 min of BrdU labeling was used for each time point prior to harvest. Processing of cells for flow cytometry was carried out using the APC BrdU Flow kit (BD Pharmingen) according to the manufacturer's instructions. The fixed and stained cells were analyzed on a Beckman Coulter Cyan ADP. Data was analyzed using Summit 4.3 software from Beckman Coulter. After gating, percent of cells in each cell-cycle phase (G1, S,

and G2/M) was calculated. Experiments were repeated three or more times. The average of experiments and standard deviations were calculated.

DNA Fiber Analysis

U2OS cells were transfected with control or RNF4 siRNAs. 48 h after transfection, cells were exposed to 20 μ M IdU for 20 min. Cells were incubated or not with 2 mM HU for 2 or 16 h. Cells were washed and then exposed to 100 μ M CldU for 30 min. DNA fibers were prepared and visualized as described (Davies et al., 2007). Microscopy was carried out using a Zeiss Observer Z1 fluorescence microscope.

Analysis of BLM Sumoylation and RNF4-Mediated Ubiquitylation

U2OS cells stably transfected with His-SUMO-1 or His-SUMO-2 were a gift from Dr. Mary Dasso (NIH). His-SUMO conjugates were purified as described (Jaffray and Hay, 2006). For *in vitro* sumoylation and ubiquitylation reactions, recombinant GST-tagged BLM (amino acids 1-431), SUMO E1 (Aos1/Uba2), E2 (Ubc9), and SUMO proteins were expressed and purified from *E. coli* as previously described (Zhu et al., 2008). Recombinant ubiquitin E1 (Uba1), E2 (UbcH5c), RNF4, and ubiquitin were kindly provided by Dr. Cynthia Wolberger (Johns Hopkins University). Sumoylated BLM on GST beads was produced as previously described (Zhu et al., 2008). The GST-BLM-SUMO beads were washed and incubated with 1 μ M ubiquitin E1, 25 μ M UbcH5a and 1 mM ubiquitin with or without 50 μ M RNF4 in reaction buffer [1 mM ATP, 20 U/ml creatine phosphokinase, 5 mM phosphocreatine, 0.6 mg/ml inorganic pyrophosphatase in 20 mM HEPES-KOH (pH 7.3), 110 mM potassium acetate, 2 mM magnesium acetate and 1 mM DTT] at 37°C for 2 h. After five washes with 500 mM NaCl in PBS, proteins were eluted with 2X SDS sample buffer and analyzed by immunoblot analysis.

Analysis of BLM Stability

Cells were transfected with negative control or RNF4 siRNA oligos on two sequential days. 24 h after the second transfection, cells were treated with or without 100 ng/ml cycloheximide for different times and lysed directly in sample buffer at the end of treatment. Proteins were analyzed by immunoblot analysis.

Mitotic Index Analysis

Mitotic index was measure by seeding 50,000 transfected HeLa or HCT116 cells onto cover slips and treating them the next day with 2 mM HU for 16 h. Cells were released from the HU block by replacement with fresh medium, and the cover slips were processed for examination of mitoses at 1-h time points 9–15 h after release. For processing, cells were fixed in 4% paraformaldehyde diluted into PBS at room temperature for 20 min, washed in PBS, permeabilized in 0.5% Triton X-100 for 10 min, washed 3 times in PBS (the middle wash containing 0.1 M glycine), and mounted in Prolong Gold with DAPI (Invitrogen). Cells were examined at 40x magnification with

a UV filter on a Nikon Eclipse E600 controlled by NIS-Elements BR 3.0 software. Fields were imaged and mitotic cells were identified as cells with chromosomes undergoing condensation or cells containing condensed chromosomes. The percent of mitotic cells was calculated as the total number of mitotic cells divided by the total number of cells. A small percentage of cells (<0.2%) that exhibited nuclear blebbing or other nuclear changes indicative of apoptosis, were not included in this total. Experiments were repeated three times. The average of experiments and standard deviations were calculated.

Sister Chromatid Exchange

165,000 HeLa cells were seeded into T25 flasks overnight and were subsequently transfected with siRNAs on two sequential days. On the third day, cells were incubated in medium containing 20 μ M bromodeoxyuridine (BrdU) for 42 h (untreated samples) or cells were incubated in 20 μ M BrdU for 24 h, in 20 μ M BrdU and 2 mM HU for 16 h, and finally in 20 μ M BrdU for 14 h (HU-treated samples). 45 min prior to harvest, colcemid was added to achieve a concentration of 0.15 μ g/ml. For HU-treated cultures, 3 h prior to harvest freshly prepared caffeine was added at a concentration of 1 mM, otherwise mitoses were not obtained in RNF4-depleted cells. The failure to obtain mitoses in RNF4-depleted cells was specific to colcemid treatment, because RNF4-depleted cells underwent mitosis and cell division in the absence of colcemid. Metaphases were prepared and stained using the fluorescence plus Geimsa method as described (Ouyang et al., 2009). Metaphases were examined at 100x magnification under oil with a Nikon Eclipse E600 controlled by NIS-Elements BR 3.0 software. SCEs in each metaphase and the numbers of chromosomes were counted, and the levels of SCEs expressed as the number of exchanges per 46 chromosomes over all scorable metaphases. Experiments were repeated twice. The average of experiments and standard deviations were calculated.

Statistical Analyses

T-tests were performed to compare the effects of RNF4 depletion vs control on proliferation; on focal accumulations of RAD51, γ -H2AX, BLM, and RPA; on SCEs; and, on replication dynamics exhibited by the DNA fiber analysis. To compare flow cytometry cell cycle profiles, the difference in the mean values for RNF4 versus NC1 were tested for percentage phase of the cell cycle (% G1, % S, and % G2) within each cell line using two sample *t*-tests. The interaction between caffeine and RNF4 (versus NC1) was initially tested using linear regression. Since it was not statistically significant for any conditions, it was removed from the model. Subsequently the difference between the caffeine treatment in the no HU and HU conditions was assessed using two sample *t*-tests. To analyze recovery from replication arrest, the difference in the percent of cells between siRNF4 and control was assessed using linear regression models. The initial model assessed whether the profile of the change across time differed between the two conditions (i.e., was there interaction). If the profile was not significantly different, the potential for a shift between the two conditions was assessed.

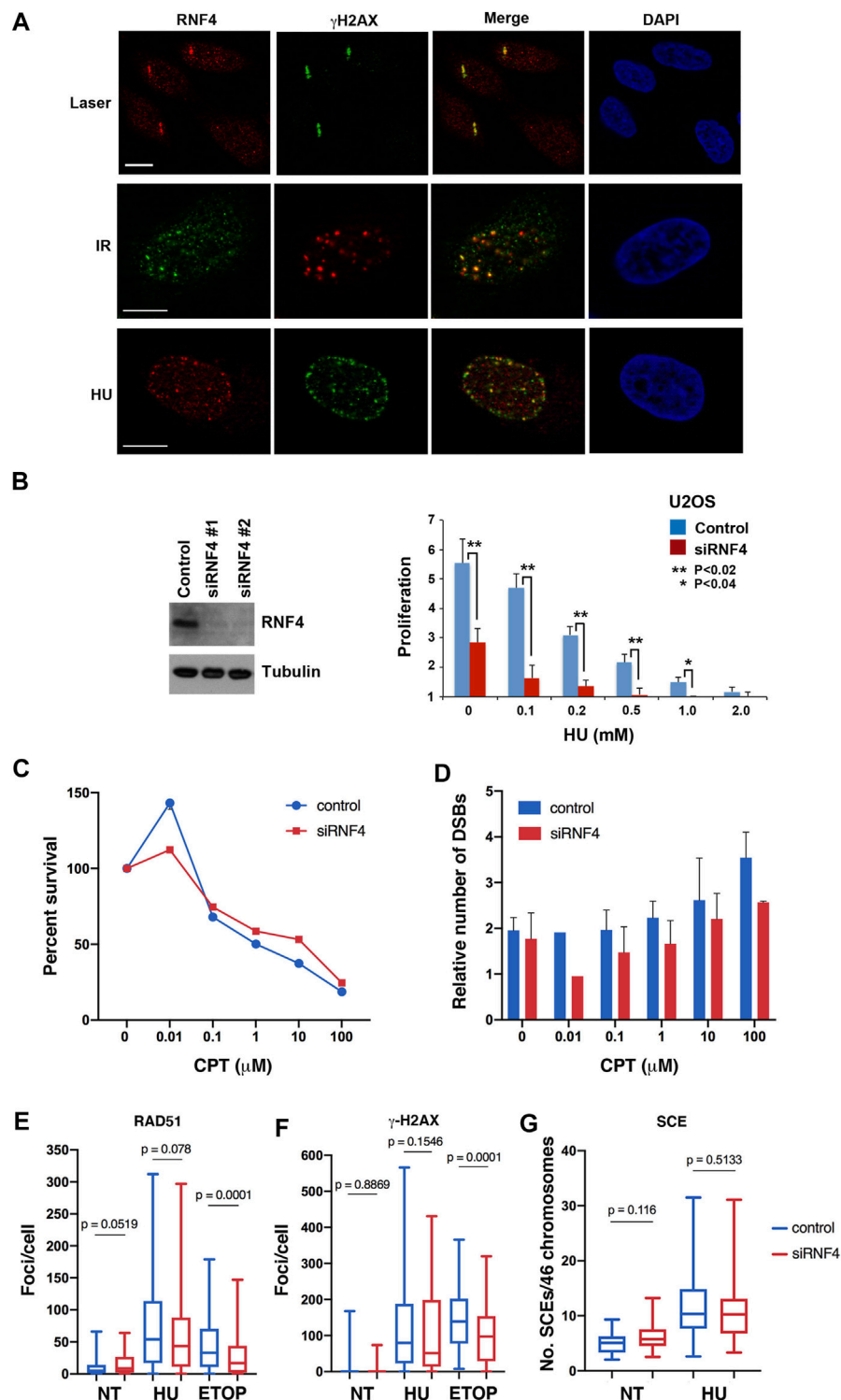


FIGURE 1 | RNF4 was not required for HR repair of replication-associated DSBs. **(A)** RNF4 co-localized with phosphorylated histone H2AX (γ -H2AX) in U2OS cells treated with a micro-laser, ionizing radiation (IR), or hydroxyurea (HU). DNA was detected with DAPI. Bars = 10 μ m. **(B)** RNF4 depletion in U2OS cells by two unique RNF4-specific siRNAs was evaluated by immunoblot analysis. Proliferation defect in RNF4-depleted U2OS cells exposed to varying concentrations of HU for 72 h. Error bars represent standard deviations from three biological replicates. p values were determined using a Student's t -test. **(C)** Line graph showing results of clonogenic survival assays on RNF4-depleted and control-depleted HeLa cells exposed to varying concentrations of camptothecin (CPT) for 3 h. Results from three independent assays were averaged and standard deviations shown. **(D)** Bar graphs showing the relative levels of DSBs in RNF4-depleted and control-depleted HeLa cells exposed to

(Continued)

FIGURE 1 | varying concentrations of CPT for 3 h, as determined by pulsed field gel electrophoresis. Induced DSBs were normalized to DSBs in untransfected and untreated cells. Results from three independent experiments were averaged and standard deviations shown. **(E,F)** Box and whiskers plots showing the enumerations of focal accumulations of RAD51 and γ -H2AX, detected by indirect immunofluorescence, in RNF4-depleted and control-depleted HeLa cells untreated or treated 2 mM HU for 16 h. Results from three independent experiments were combined. As a positive control, cells were treated with 10 μ M etoposide for 4 h. **(G)** Box and whiskers plot showing levels of sister chromatid exchange (SCE) in RNF4-depleted and control-depleted HeLa cells untreated or treated 2 mM HU. Cells were labeled with BrdU for one cell division, blocked in HU for 16 h, released from the HU block into medium containing BrdU, collected in metaphase with colcemid and processed for assaying SCEs. Results from two independent experiments were combined. NT = not treated.

RESULTS

RNF4 Exhibited a Unique Role in Response to Prolonged Replication Stress

Although known to have important roles in cellular response to DNA replication stress (Ragland et al., 2013; Kumar and Sabapathy, 2019), the exact functions of RNF4 at stalled and collapsed replication forks are not fully understood. To explore these functions, we first used indirect immunofluorescence confocal microscopy to investigate RNF4 localization in U2OS cells treated with 2 mM HU for 16 h, as localization under these conditions has not previously been reported. This analysis revealed that RNF4 co-localized with γ -H2AX, a marker for sites of collapsed replication forks and the modicum of DSBs induced by this duration of HU treatment (Petermann et al., 2010). As positive controls, we confirmed previous findings that RNF4 co-localizes with γ -H2AX at DSBs generated by microlaser- and γ -irradiation (Galanty et al., 2012; Yin et al., 2012; Vyas et al., 2013) (**Figure 1A**).

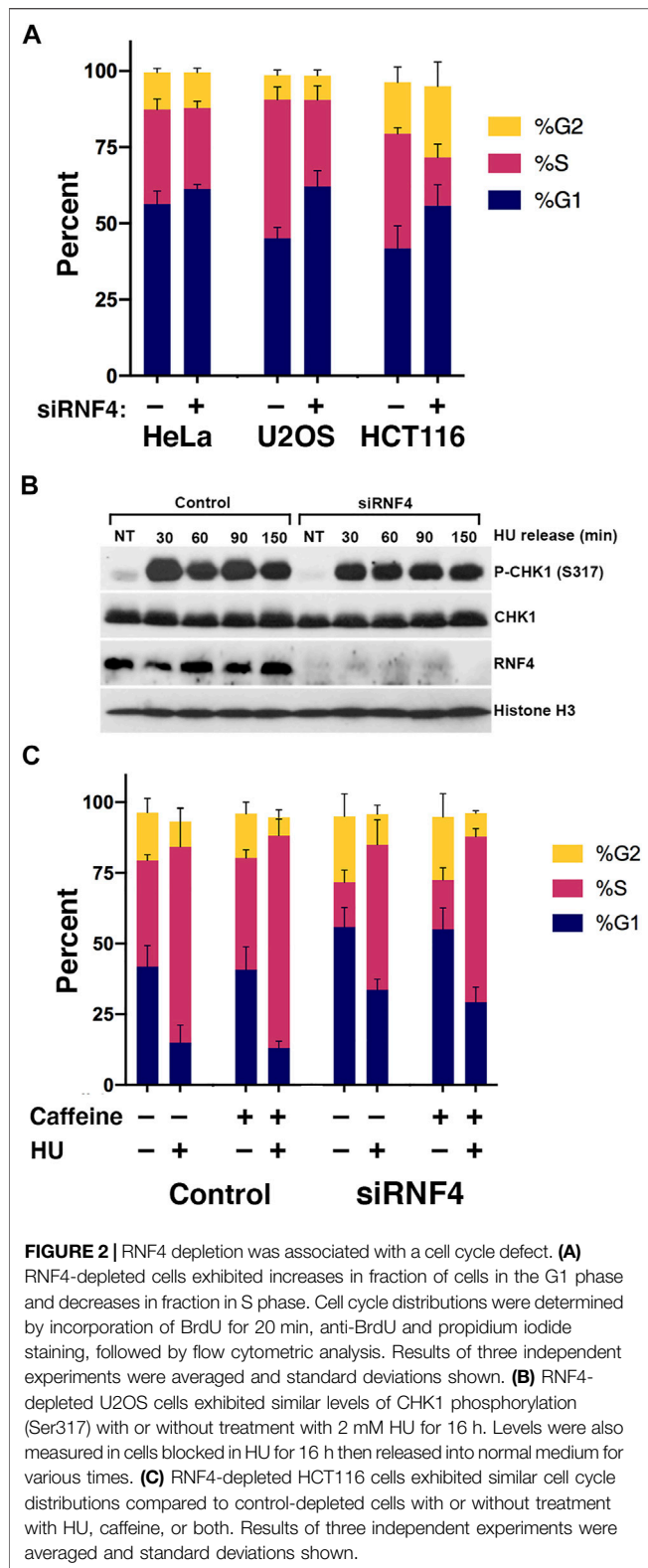
To further explore the functions of RNF4 in the cellular response to DNA replication stress, we used two independent siRNA oligos to effectively deplete expression levels by >90% in human U2OS cells (**Figure 1B**). Consistent with previous studies (Galanty et al., 2012; Yin et al., 2012; Vyas et al., 2013), we found that RNF4 depletion enhanced sensitivity to chronic HU-induced replication stress (**Figure 1B**; **Supplementary Figure S1A**). We also examined the sensitivity of RNF4-depleted HeLa and U2OS cells to CPT, a topoisomerase I (TopI) inhibitor that stabilizes the normally transient TopI cleavage complex. CPT treatment also induces DNA replication stress in S phase, and the collision of replication forks with CPT-TopI-DNA complexes induces formation of DSBs (Vesela et al., 2017). We found that RNF4-depletion had no effect on the sensitivity to CPT (**Figure 1C**; **Supplementary Figure S1B**). Because CPT toxicity is linked to generation of DSBs, we measured the number of CPT-induced DSBs in HeLa and U2OS cells by pulsed-field gel electrophoresis. This analysis revealed that DSB formation was also unaffected in RNF4-depleted cells compared to control cells (**Figure 1D**; **Supplementary Figure S1C**). These observations suggested that RNF4 plays a unique role in the response to DNA replication stress, and that this role involves functions at sites of collapsed replication forks that may be independent of DSB repair.

RNF4 functions in the repair of DSBs generated by γ -irradiation in part by facilitating the recruitment of RAD51

to sites of DNA damage (Galanty et al., 2012; Yin et al., 2012; Vyas et al., 2013). We therefore examined whether RNF4 depletion in HeLa cells impaired RAD51 accumulation at collapsed replication forks and DSBs generated by prolonged HU treatment. Unexpectedly, the numbers of RAD51 and γ -H2AX foci, as well as their focal areas, were similar in RNF4-depleted cells with or without HU treatment compared to control-depleted cells (**Figures 1E,F**; **Supplementary Figure S1D**), suggesting normal responses to collapsed replication forks. In contrast, and as expected, the number of RAD51 foci were reduced in RNF4-depleted HeLa cells treated with etoposide, an inhibitor of Topoisomerase II that induces DSBs directly in all phases of the cell cycle (Vesela et al., 2017). Contrary to expectation, γ -H2AX foci were also reduced, suggesting that RNF4 depletion leads to slower proteolytic processing of the TopII-DNA cleavage complex or of the tyrosyl-DNA moiety left over after proteolysis (Sciascia et al., 2020). The tyrosyl-DNA moiety is processed by tyrosyl-DNA phosphodiesterase 2, which is a target of sumoylation and regulated by RNF4 (Sun et al., 2020). We also measured rates of SCE in HU-treated HeLa cells, which serves as a readout for DSB repair through RAD51-dependent HR. In agreement with the normal RAD51 recruitment to DSBs, the levels of SCE were similar in RNF4-depleted compared to control-depleted cells with or without HU treatment (**Figure 1G**). These findings provided further evidence that RNF4 plays a unique role in the response to HU-induced replication stress that is distinct from DSB recognition and repair.

RNF4 was Required for Normal Recovery of DNA Synthesis After Prolonged Replication Stress

The lack of a correlation between DSBs and HU sensitivity left unanswered the question of how RNF4 protects cells from prolonged DNA replication stress. To further investigate the roles of RNF4 in DNA synthesis under normal and replication-stress conditions, we examined the cell cycle profiles of control and RNF4-depleted HeLa, U2OS, and HCT116 cells by flow cytometry. For these experiments, we included the colon cancer cell line HCT116, because it has defects in the expression of the MRE11-RAD50-NBS1 complex and is thus more sensitive to replication stress. In the absence of DNA replication stress, RNF4-depleted cells exhibited reproducible perturbations of the G1 and S cell cycle phases in U2OS and HCT116 cells, showing a higher fraction of cells in G1 and a lower fraction in S phase compared to control-depleted cells (**Figure 2A**; **Supplementary Table S1A**). RNF4-depleted U2OS



cells exhibited 17% more cells in G1 and 15% less in S phase, and for HCT116, it was 14% more in G1 and 22% less in S phase. For HeLa cells, it was 5% more cells in G1 and 4% less in S phase, but

the results were not significant. These results show cell type-specific sensitivities to RNF4 depletion with HeLa cells being relatively resistant and U2OS and HCT116 cells being sensitive.

To test whether the excess accumulation of cells in G1 in the absence of HU was due to a checkpoint response, we measured the phosphorylation of CHK1 (Ser317) in untreated cells and found no evidence of more or less DNA damage checkpoint activation compared to control-depleted U2OS cells (**Figure 2B**). Consistent with a lack of DNA damage checkpoint activation, treating HCT116 cells with the general checkpoint inhibitor caffeine had negligible effects on cell cycle distributions (**Figure 2C**; **Supplementary Table S1B**). Thus, the effect of RNF4 depletion on the cell cycle in untreated cells is not caused by activation of ATR or ATM by a DNA damage signal.

We next investigated how RNF4 depletion affects the progression of cells through S phase by examining the kinetics of DNA synthesis resumption following release from prolonged HU treatment. We treated HeLa, U2OS, and HCT116 cells with 2 mM HU for 16 h and released them into normal medium containing BrdU. Cells were then analyzed by flow cytometry at different time points following HU release. This analysis revealed that RNF4-depleted cells treated with HU exhibited a delay in the incorporation of BrdU compared to control-depleted cells (**Figure 3A**), with HCT116 cells exhibiting the most severe defect. An effect of replication stress was also evidenced by a reduction in the total percent of RNF4-depleted cells that incorporated BrdU after HU release (6% reduction in HeLa cells, 13% reduction in U2OS cells, and 25% reduction in HCT116 cells at 40 min after HU release).

We also examined levels of CHK1 phosphorylation following release from overnight HU treatment. In both control- and RNF4-depleted cells, CHK1 phosphorylation remained elevated at least 150 min after release (**Figure 2B**), suggesting that dormant origin firing (see below) does not depend on reduction of CHK1 phosphorylation after HU release.

As an independent measure of the effect of RNF4 on S phase progression and exit, we treated HeLa and HCT116 cells with HU for 16 h, released the cells into normal medium, and then measured the mitotic index at different times after release. RNF4-depleted HeLa cells exhibited a delay in transit to metaphase following HU release compared to control-depleted cells, but the results were not significant (**Figure 3B**). HCT116 cells exhibited a more severe delay (>2 h), and a majority of cells appeared to fail to complete S phase. These results were consistent with the extreme hypersensitivity of HCT116 cells to RNF4 depletion (**Supplementary Figure S2A**). Thus, RNF4 is required for efficient resumption of DNA replication and completion of S phase following prolonged HU-induced DNA replication stress.

RNF4 was Required for Activation of Dormant Origins After Replication Fork Collapse

The delay in resumption of DNA synthesis after prolonged HU treatment suggested that RNF4 may be required for the activation of dormant origins in proximity to collapsed replication forks.

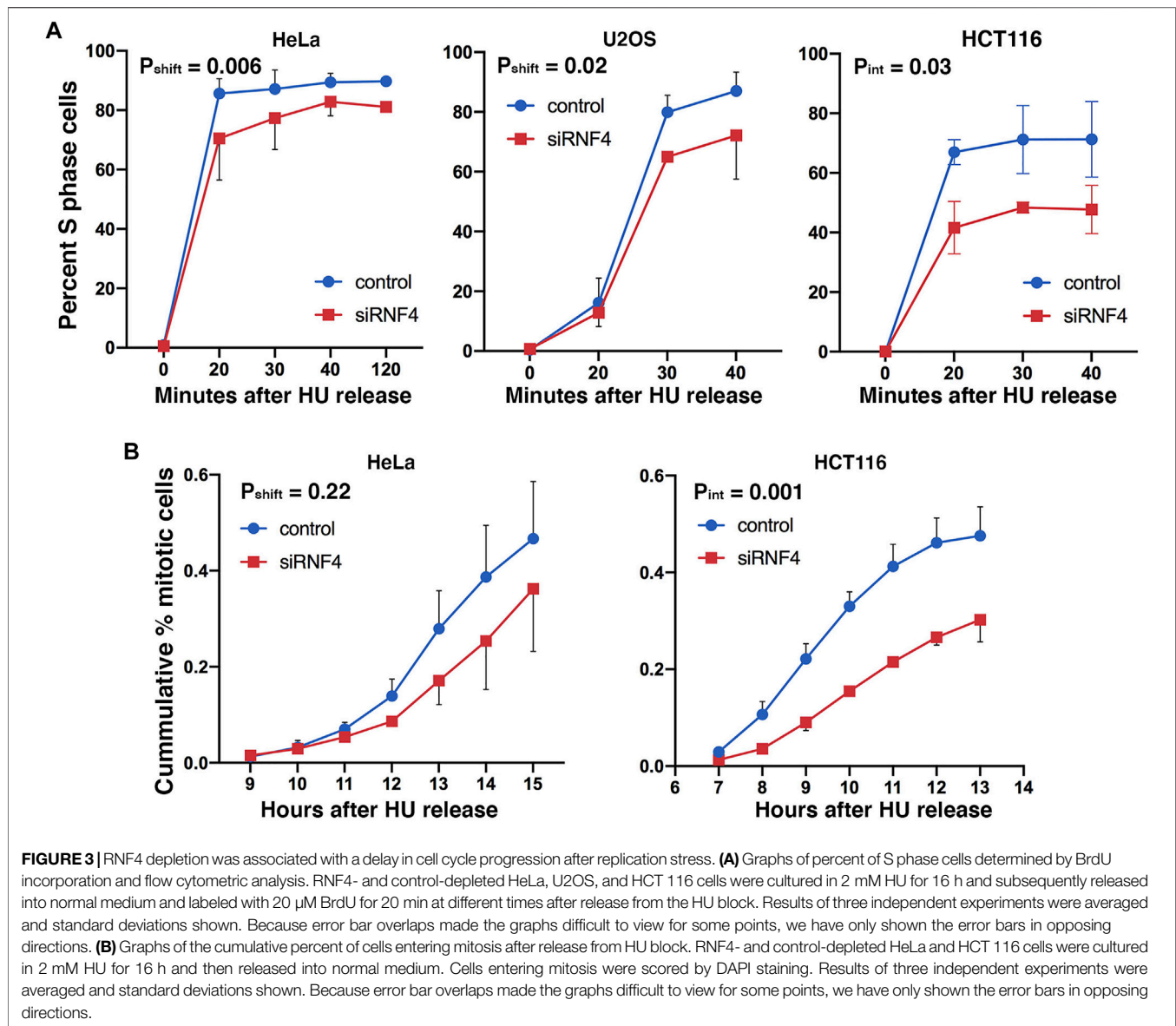
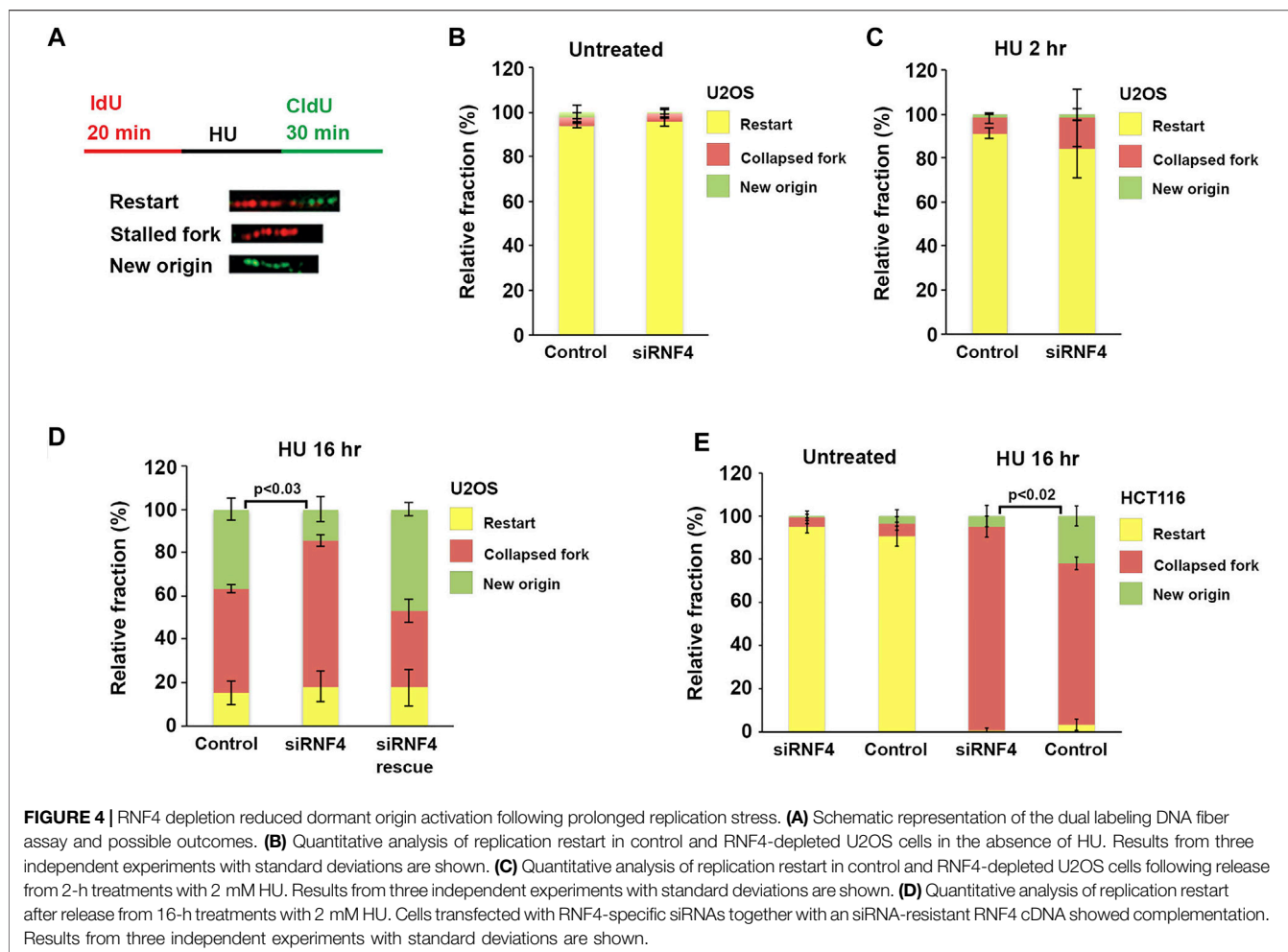


FIGURE 3 | RNF4 depletion was associated with a delay in cell cycle progression after replication stress. **(A)** Graphs of percent of S phase cells determined by BrdU incorporation and flow cytometric analysis. RNF4- and control-depleted HeLa, U2OS, and HCT 116 cells were cultured in 2 mM HU for 16 h and subsequently released into normal medium and labeled with 20 μ M BrdU for 20 min at different times after release from the HU block. Results of three independent experiments were averaged and standard deviations shown. Because error bar overlaps made the graphs difficult to view for some points, we have only shown the error bars in opposing directions. **(B)** Graphs of the cumulative percent of cells entering mitosis after release from HU block. RNF4- and control-depleted HeLa and HCT 116 cells were cultured in 2 mM HU for 16 h and then released into normal medium. Cells entering mitosis were scored by DAPI staining. Results of three independent experiments were averaged and standard deviations shown. Because error bar overlaps made the graphs difficult to view for some points, we have only shown the error bars in opposing directions.

We therefore measured the effects of RNF4 depletion on replication dynamics using the DNA fiber assay. To carry out this assay, RNF4-depleted and control-depleted U2OS cells were incubated in medium supplemented with 20 μ M IdU for 20 min. The IdU-containing medium was then replaced with medium containing 2 mM HU for 2 or 16 h, followed by release from HU and incubation in medium supplemented with 100 μ M CldU for 30 min (Figure 4A). Cells were then processed for single molecule stretching and immunofluorescence detection of halogenated nucleotide incorporation. We then calculated the percentage of DNA molecules labeled with both IdU and CldU (representing replication fork restart), IdU only (representing irreversible fork collapse and termination of replication), or CldU only (representing replication from newly fired origins). In the absence of HU, >95% of labeled DNA molecules in both RNF4-depleted and control-depleted cells contained IdU and

CldU labels (Figure 4B), indicating that RNF4 is not required for ongoing replication in the absence of replication stress.

In cells treated with HU for 2 hours, the increase in collapsed replication forks in RNF4-depleted cells compared to control-depleted cells was not significant (Figure 4C). Following treatment with HU for 16 h, RNF4-depleted and control-depleted cells also exhibited minimal differences in the percentage of replication forks undergoing restart. In contrast, however, RNF4-depleted cells exhibited a significantly higher percentage of collapsed forks and a lower percentage of forks starting at new origins compared to control-depleted cells (Figure 4D). This observed defect in new origin firing was rescued by the ectopic expression of siRNA-resistant *RNF4* mRNA. These results provided further evidence that RNF4 is required for efficient resumption of DNA synthesis following replication stress and pointed to a



role in the activation of dormant origins following fork collapse.

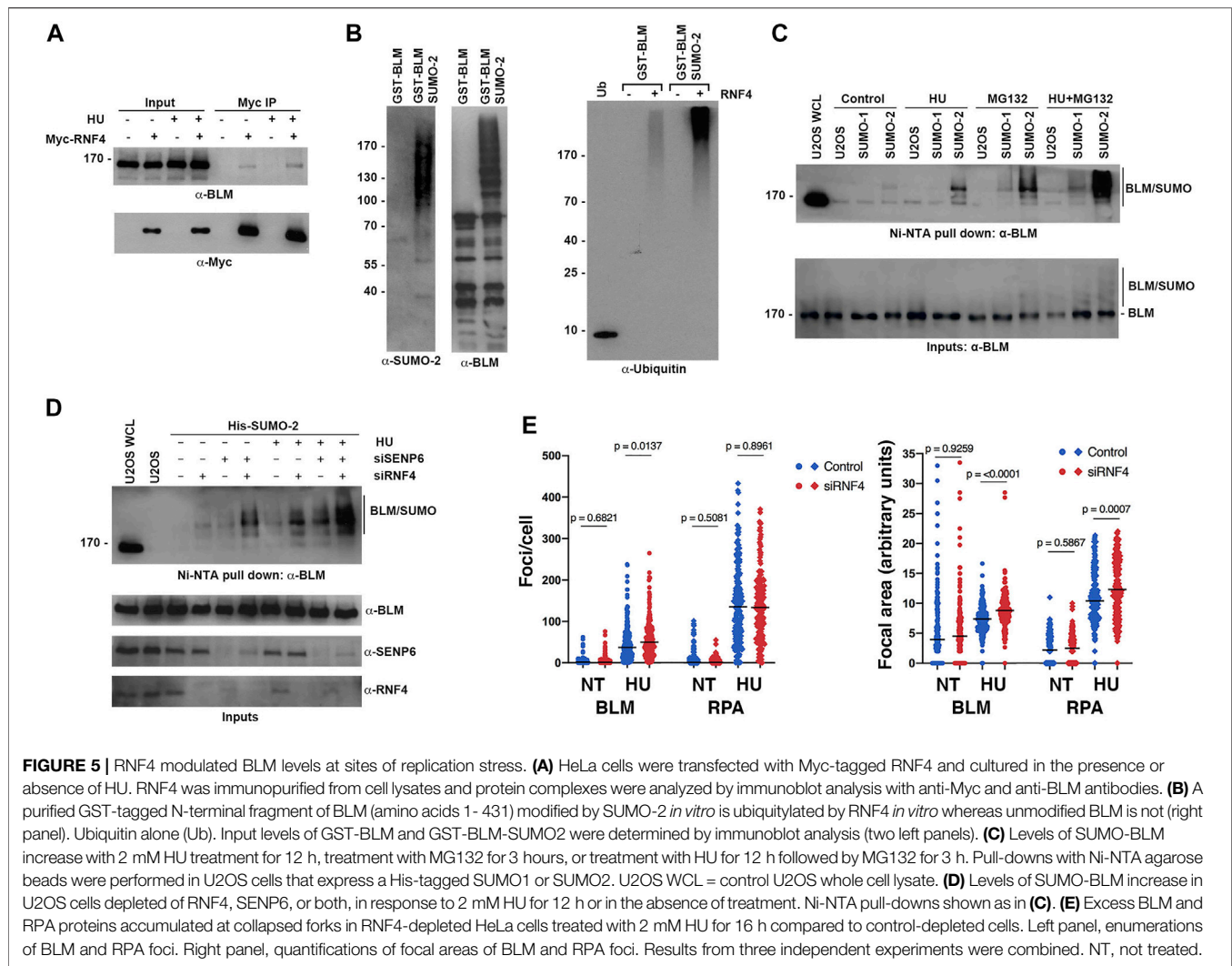
BLM was Found to be Regulated by RNF4 at Sites of Replication Stress

BLM accumulates at stalled and collapsed forks, is sumoylated in response to replication stress (Ouyang et al., 2009; Xiao et al., 2015), and it has been reported to be a substrate of RNF4 in proteomic studies (Kumar et al., 2017). We therefore hypothesized that RNF4 may regulate ubiquitin-mediated turnover of BLM at collapsed forks, and that an accumulation of excess BLM in the absence of RNF4 may inhibit the normal resumption of DNA synthesis after prolonged HU treatment.

To begin to test this hypothesis, we first investigated whether RNF4 interacts with BLM and whether this interaction is regulated by HU-induced replication stress. HeLa cells were transfected with a Myc-RNF4 expression construct and Myc-RNF4 was immunopurified from cell lysates prepared from control and HU-treated cells. Immunoblot analysis with BLM antibodies revealed an interaction in the absence of HU that increased following HU treatment (Figure 5A). The predominant

form of BLM detected in the pulldown was unsumoylated BLM, suggesting a possible direct interaction between BLM and RNF4. Next, we tested whether sumoylated BLM is ubiquitinated by RNF4 by performing *in vitro* conjugation assays using purified recombinant proteins. Using an N-terminal fragment of GST-tagged BLM (BLM 1-431) that is readily sumoylated *in vitro* (Zhu et al., 2008), we found that sumoylated BLM was robustly ubiquitinated in comparison to unmodified BLM (Figure 5B).

To further investigate whether sumoylation targets BLM for ubiquitin-mediated turnover *in vivo*, we analyzed levels of sumoylated BLM in U2OS cell lines that stably expressed either a His-tagged SUMO-1 or His-tagged SUMO-2 using nickel-NTA bead affinity pull down and immunoblot analysis. As anticipated from previous studies (Eladad et al., 2005; Zhu et al., 2008; Ouyang et al., 2009), BLM was preferentially modified by SUMO-2 at low levels under control conditions, and these levels increased in response to HU treatment (Figure 5C). Consistent with sumoylation functioning as a signal for proteasome-mediated turnover, levels of sumoylated BLM were greater in cells treated with MG132 compared to untreated cells, and sumoylated BLM levels were further increased in cells treated with both HU and MG132 (Figure 5C). We note that the ratio of sumoylated to



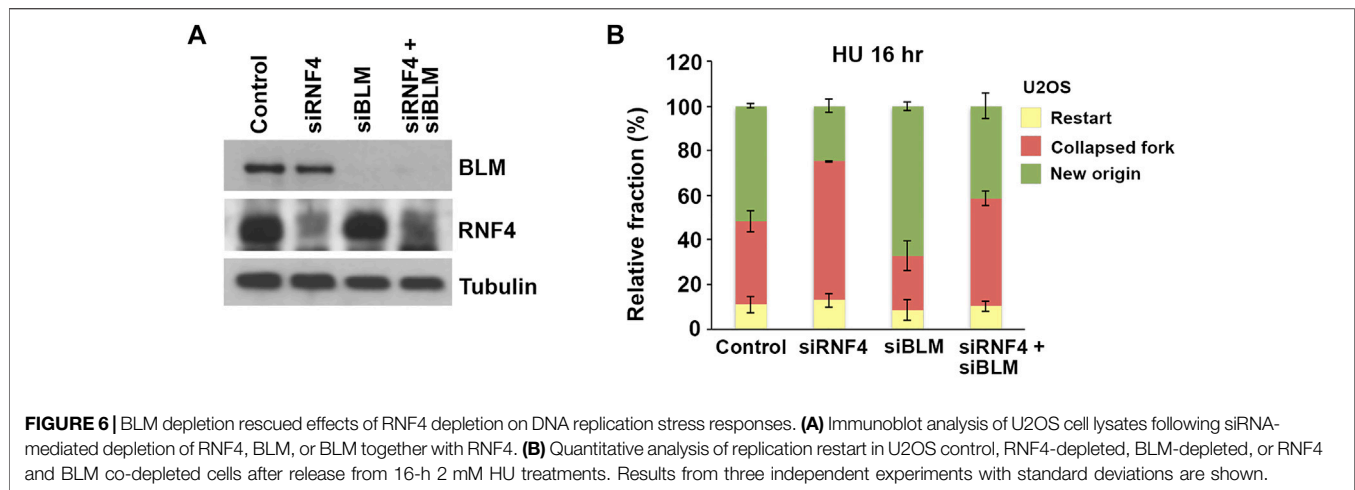
unsumoylated BLM was low, even under conditions of HU and MG132 treatment (**Figure 5C**).

To test whether RNF4 regulates the turnover of sumoylated BLM, we next measured sumoylated BLM levels in RNF4-depleted cells. We also tested cells depleted for SENP6, a chain-editing SUMO isopeptidase capable of limiting poly-sumoylation and thereby RNF4 recognition (Mukhopadhyay et al., 2010). In untreated cells, RNF4 depletion had minimal effect on sumoylated BLM levels, whereas levels were significantly increased when RNF4 was depleted in HU-treated cells (**Figure 5D**). Similarly, SENP6 depletion alone had minimal effect on sumoylated BLM levels in untreated cells, whereas levels were increased in combination with HU treatment. In comparison to the single knockdowns, an increase in sumoylated BLM levels was observed in cells co-depleted of both SENP6 and RNF4, and the highest levels of sumoylated BLM were detected in co-depleted cells treated with HU (**Figure 5D**).

Direct investigation of kinetics of sumoylated BLM turnover in control and RNF4-depleted cells was complicated by the low ratio

of modified to unmodified BLM. In addition, studies using inhibition of new protein synthesis with cycloheximide were not possible because RNF4 is itself turned over within 2 h of cycloheximide addition (**Supplementary Figure S3A**).

Finally, to investigate the effect of RNF4 on BLM accumulation at sites of collapsed DNA replication forks, we quantified the number and area of BLM foci in control and RNF4-depleted cells by indirect immunofluorescence confocal microscopy (**Figure 5E**; **Supplementary Figure S3B**). In the absence of HU, we found that the number and area of BLM foci were similar in RNF4- and control-depleted cells. In contrast, following HU treatment for 16 h, both the number of BLM foci and their focal areas were significantly greater in RNF4-depleted cells compared to control-depleted cells. Because RPA is also sumoylated and accumulates at collapsed replication forks (Dou et al., 2010), we quantified RPA foci and focal areas in both untreated and treated cells. Although the numbers of RPA foci remained the same, the areas of RPA foci were significantly greater in RNF4-depleted cells compared to control-depleted cells following HU treatment. Altogether, these findings



demonstrated that sumoylated BLM is a substrate for RNF4-mediated ubiquitylation and proteasome-mediated turnover, and that BLM accumulates at sites of collapsed DNA replication forks in the absence of RNF4.

BLM Contributed to Suppression of Dormant Origin Firing in RNF4-Depleted Cells

To investigate whether the accumulation of BLM at collapsed replication forks in RNF4-depleted cells contributed to defects in replication restart, we next asked whether co-depletion of BLM could rescue the observed defects in dormant origin firing. We again used DNA fiber assays to analyze replication dynamics in U2OS cells co-depleted of BLM and RNF4 (**Figure 6A**). Consistent with previous findings (Davies et al., 2007), we observed an increase in the proportion of DNA replication restart events from new origins in BLM-depleted cells after release from treatment with HU for 16 h compared to control-depleted cells (**Figure 6B**). In addition, we again observed a decrease in replication restart from new origins in RNF4-depleted cells. In contrast, replication restart from new origins in cells co-depleted of RNF4 and BLM was similar to control-depleted cells, which demonstrated that BLM depletion rescued the defect in dormant origin firing observed in RNF4-depleted cells. These findings were consistent with our hypothesis that an accumulation of excess BLM at collapsed replication forks inhibits the normal resumption of DNA replication following prolonged HU treatment.

DISCUSSION

The experiments presented here have implicated RNF4 function in recovery from fork collapse. Brief exposure of RNF4-depleted cells to HU elicited little effect on replication fork stability, whereas RNF4-depleted cells exposed to prolonged HU treatment—a treatment that causes widespread fork collapse—exhibited a delay in the resumption of DNA

synthesis after removal of drug and an increase in the percentage of cells that permanently left the cell cycle, which was demonstrated in both flow cytometric and DNA fiber assays (**Figures 3, 4**). Extending the results of earlier proteomic studies (Kumar et al., 2017), we showed that BLM interacts with RNF4, sumoylated BLM is ubiquitylated by RNF4 *in vitro*, and RNF4 depletion led to a substantial increase in SUMO-BLM in HU treated cells (**Figure 5**). Moreover, depletion of RNF4 led to excessive accumulation of BLM and RPA at collapsed forks as evidenced by increased focal areas. The observation that the number of RPA foci per cell remained the same whilst the number of BLM foci per cells increased could be an effect of better detection, as increased BLM protein accumulation at collapsed forks may have brought more sites of collapsed forks at which BLM localized above the detection threshold. The generation of excess SUMO-BLM in RNF4-depleted, HU-treated cells could explain the observed increases in BLM focal areas due to higher-order interactions, because BLM contains both multiple SUMO acceptors sites and a SIM and BLM can form a multimer *in vitro* (Karow et al., 1999; Xu et al., 2012). Earlier studies showed that the SUMO E2 conjugating enzyme UBC9 and E3 ligase NSMCE2 are targets for RNF4 that enforce a negative regulatory loop (Kumar et al., 2017). *In vitro*, UBC9 is sufficient for poly-sumoylation or multi-mono-sumoylation of BLM (Eladad et al., 2005) and we showed recently that NSMCE2 is required for sumoylation of BLM and for accumulation of BLM at collapsed replication forks in response to HU treatment (Pond et al., 2019). Although we cannot rule out the possibility that BLM accumulates excessively in HU-treated, RNF4-depleted cells due to damage at the replication fork, we favor the hypothesis that the excessive accumulation of BLM has pathological effects on recovery of DNA synthesis after prolonged HU exposure and that, by controlling the levels of sumoylated BLM, RNF4 facilitates resumption of DNA synthesis after widespread fork collapse. We previously showed that SD-BLM accumulates excessively at stressed replication forks, suggesting this accumulation is an upstream event. We found here that the delay in resumption of DNA synthesis in RNF4-depleted cells was rescued by co-depletion of the BLM protein (**Figure 6**).

The mechanism by which RNF4 contributes to immediate resumption of DNA synthesis is not known. One possibility is that the excess SUMO-BLM that accumulates at collapsed forks in RNF4-depleted cells ties up large amounts of the unsumoylated BLM and that BLM needs to be released and recycled in order to achieve efficient dormant origin firing or resolution of fork impediments that prevent rapid replisome take off at already-fired dormant origins. We do note, however, that the ratio of sumoylated to unsumoylated BLM is low even under conditions that limit its turnover (RNF4 and SENP6 knockdown and MG132 treatment). These low levels are nonetheless consistent with functionally relevant levels of sumoylation observed with other SUMO substrates (Hay, 2005). Another possibility is that RNF4-mediated degradation of SUMO-BLM at collapsed forks drives the disassembly of multiple factors in repair foci that are needed for dormant origin firing or replisome take off. Moreover, we do not know that BLM is the only sumoylated protein whose depletion would result in rescue. Because multiple damage-response factors are sumoylated, it is possible that depletion of other sumoylated proteins at sites of replication stress could also rescue the DNA synthesis resumption defect. In particular, interesting candidates would be proteins that contain both SUMO acceptor sites and SIMs, such as SLX4, which could mediate higher-order interactions and focus formation (Bursomanno et al., 2015; Hendriks et al., 2015; Gonzalez-Prieto et al., 2021).

Short-term treatment of *BLM*-deficient cells with HU induces excessive fork collapse and dormant-origin firing relative to normal cells (Davies et al., 2007; Patel et al., 2017). This evidence argues that BLM itself does not play a direct role in activating dormant origin firing, nor has it a known activity in the firing of origins of replication in early or late periods of S phase from *in vitro* studies (On et al., 2014; Kurat et al., 2017). For these reasons, we do not favor a hypothesis that places SUMO-BLM in an inhibitory role in firing of dormant origins.

Because RNF4 depletion promotes an increase in the size and numbers of PML bodies (Lallemant-Breitenbach et al., 2008; Tatham et al., 2008), it was formally possible that the flux of trafficking of damage response proteins that normally accumulate in PML bodies due to sumoylation or SIMs could be delayed by RNF4 depletion. In untreated RNF4-depleted cells, there is an increase in BLM foci approximately corresponding to the increase in PML bodies (Bohm et al., 2015). BLM's localization to PML bodies relies primarily on its SIM (Eladad et al., 2005; Zhu et al., 2008); however, we did not find evidence that BLM recruitment to sites of replication stress was less efficient, because BLM substantially co-localized with RPA foci after HU treatment and RPA is not a PML-associated nuclear protein.

RNF4 is important in the repair of DSBs, because RNF4-depleted cells are hypersensitive to γ -radiation and have a defect in the recruitment of RAD51 to DSBs (Galanty et al., 2012; Yin et al., 2012; Vyas et al., 2013). Our investigation began with the question concerning the hypersensitivity of RNF4-depleted cells to HU treatment and with the hypothesis that RNF4 would be important for the repair of replication-associated DSBs (Figure 1). However, we did not observe a role for RNF4 in recruitment of RAD51 to collapsed replication forks caused by

prolonged HU treatment. RNF4 depletion did not impair ATR- or ATM-dependent checkpoint signaling in response to HU, as indicated by normal γ -H2AX levels, nor were levels of phosphorylated CHK1 affected by RNF4 depletion with or without HU treatment (Figure 2B). Levels of SCE were similar in RNF4-depleted compared to control-depleted cells with or without HU treatment (Figure 1F). To our surprise, RNF4-depleted cells were not hypersensitive to CPT, and levels of CPT-induced DSBs were unaffected by RNF4 depletion (Figures 1C,D), despite the role of RNF4 in degradation of topoisomerase I-DNA cleavage complexes (Sun et al., 2020). Treatment with CPT generates a predominance of single-ended, replication-associated DSBs, and our evidence indicates that the repair of these breaks is not affected by RNF4 depletion. Because excess BLM accumulated at collapsed forks without affecting the rate of SCEs, our results also excluded a hypothesis in which BLM or other RNF4-regulated HR factors must be extracted from collapsed forks in order for HR repair to proceed. These data highlight the importance of damage context. Previous evidence has shown that the requirements for recruitment of RAD51 to two-ended DSBs and stalled forks are different (Chaudhuri et al., 2016). Moreover, previous results have shown that cells held in HU for up to 24 h do not accumulate many DSBs (Petermann et al., 2010); instead, DSBs accumulate after release from HU blockade (Pond et al., 2019) or after longer treatments with HU. Some fraction of the DSBs that accumulate after release from HU occur in late S phase, indicating that breaks occur when active forks converge on collapsed forks. The SCE results shown here indicate that repair of these DSBs was normal in RNF4-depleted cells. Our results did not address the question whether RNF4 was required for the repair of breaks generated during mitosis or at cytokinesis, where two-ended DSBs are thought to be generated (Spies et al., 2019).

In mouse knockout studies, RNF4 was found to be essential for embryogenesis and *Rnf4*^{-/-} mouse embryonic fibroblast lines could not be obtained in at least one study (Hu et al., 2010; Vyas et al., 2013). *RNF4*^{-/-} chicken DT40 cells are viable, but they have limited proliferation capacity due to chromosomal loss (Yin et al., 2012; Hirota et al., 2014); *RNF4* knock-out human cell lines generated using CRISPR technology have been reported (Maure et al., 2016; Sun et al., 2020), which suggests that RNF4 is not cell essential. However, colony survival assays have consistently shown a ~50% reduction in RNF4-depleted cells relative to control, indicating that RNF4 has a role in cell viability. Similarly, *BLM*-deficient cells proliferate less robustly than normal cells. These observations are consistent with the synthetic lethality of the yeast orthologs of *BLM* and *RNF4*, namely, *SGS1* and the *SLX5-SLX8* complex (Mullen et al., 2001). We found that RNF4-depleted cells exhibited an increase in the fraction of cells in the G1 phase and showed evidence that this increase was not a result of a DNA damage signal (Figure 2). Although it has been proposed that the essential function of RNF4 is due to its role in maintaining genomic integrity, RNF4 also plays important roles in gene transcription (Moilanen et al., 1998), global DNA methylation levels (Hu et al., 2010), chromatin structure (Hendriks et al., 2015), and regulation of oncogenes (Thomas et al., 2016), and it

could be the combination of all these roles that leads to loss of viability in untreated cells.

The present study showed that the process of replication fork collapse and dormant origin firing are connected through the action of RNF4. RNF4 is required for the clearance of BLM from collapsed forks and the failure to release BLM from collapsed forks affected the recovery of cells from prolonged replication stress. With the varied roles that RNF4 plays in DNA damage responses, further investigation into its efficacy as a potential target in cancer treatments seems warranted. As a cancer target, the function of RNF4 in turnover of sumoylated BLM and other HR proteins could perhaps be utilized to slow recovery in replication stressed cells.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

Conception of research: NE and MM. Experimentation: JZ, MY, W-CY, JH, AK. Resources: MS. Writing and revision: NE and MM. Funding: NE, MS, and MM.

FUNDING

This work was supported by grants from the National Institutes of Health (R01 GM060980 to MM and R01 CA140804 to NE, and the Cancer Center Support Grant P30 CA023074) and by the Intramural Research Program of the National Institutes of Health, National Institute on Aging (Z01 AG000746-08 to MS). The content of this article is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health.

REFERENCES

- Bianchi, V., Pontis, E., and Reichard, P. (1986). Changes of Deoxyribonucleoside Triphosphate Pools Induced by Hydroxyurea and Their Relation to DNA Synthesis. *J. Biol. Chem.* 261, 16037–16042. doi:10.1016/s0021-9258(18)66672-4
- Böhm, S., Mihalevic, M. J., Casal, M. A., and Bernstein, K. A. (2015). Disruption of SUMO-Targeted Ubiquitin Ligases Slx5-Slx8/RNF4 Alters RecQ-like Helicase Sgs1/BLM Localization in Yeast and Human Cells. *DNA Repair*. 26, 1–14. doi:10.1016/j.dnarep.2014.12.004
- Brosh, R. M., Jr., Li, J.-L., Kenny, M. K., Karow, J. K., Cooper, M. P., Kurekattil, R. P., et al. (2000). Replication Protein A Physically Interacts with the Bloom's Syndrome Protein and Stimulates its Helicase Activity. *J. Biol. Chem.* 275, 23500–23508. doi:10.1074/jbc.m001557200
- Bugreev, D. V., Yu, X., Egelman, E. H., and Mazin, A. V. (2007). Novel Pro- and Anti-recombination Activities of the Bloom's Syndrome Helicase. *Genes Dev.* 21, 3085–3094. doi:10.1101/gad.1609007
- Burgess, R. C., Rahman, S., Lisby, M., Rothstein, R., and Zhao, X. (2007). The Slx5-Slx8 Complex Affects Sumoylation of DNA Repair Proteins and Negatively

ACKNOWLEDGMENTS

Thank Ke Ma in the University of Illinois at Chicago Confocal Microscopy Facility for technical assistance. We are indebted to Mary Dasso and Jorma Palvimo for sharing antibodies to SENP6 and RNF4, respectively.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.753535/full#supplementary-material>

Supplementary Figure S1 | (A) Bar graphs showing results of clonogenic survival assays on RNF4-depleted and control-depleted U2OS cells exposed to varying concentrations of HU. Results from three independent assays were averaged and standard deviations shown. **(B)** Bar graphs showing results of clonogenic survival assays on RNF4-depleted and control-depleted U2OS cells exposed to varying concentrations of camptothecin (CPT) for three hours. Results from three independent assays were averaged and standard deviations shown. **(C)** Bar graphs showing the relative levels of DSBs in RNF4-depleted and control-depleted HeLa cells exposed to varying concentrations of CPT for three hours, as determined by pulsed field gel electrophoresis. Induced DSBs were normalized to DSBs in untransfected and untreated cells. Results from three independent experiments were averaged and standard deviations shown. **(D)** Box and whiskers plot showing the quantifications of focal areas of RAD51 focal accumulations, in RNF4-depleted and control-depleted cells with and without treatment with 2 mM HU for 16 h. RAD51 localized to PML nuclear bodies in untreated cells, and PML bodies have been noted to enlarge in RNF4-depleted cells. Results from three independent experiments were combined. NT, not treated.

Supplementary Figure S2 | Proliferation defect in RNF4-depleted HCT116 cells exposed to 2 mM HU for 16 h. Results from three independent experiments with standard deviations are shown.

Supplementary Figure S3 | (A) BLM stability was unaffected by RNF4 depletion in HeLa cells treated with cycloheximide. Cycloheximide was added at time zero and treated cells examined hourly for six hours. Untreated control examined at six hours is indicated by an asterisk. Note that RNF4 levels were reduced by more than half after 2 h. **(B)** Representative indirect immunofluorescence image of BLM and RPA foci detected in HeLa cells transfected with negative control and RNF4 siRNAs and treated with 2 mM HU for 16 h.

Regulates Recombination. *Mol. Cell Biol.* 27, 6153–6162. doi:10.1128/mcb.00787-07

- Bursomanno, S., Beli, P., Khan, A. M., Minocherhomji, S., Wagner, S. A., Bekker-Jensen, S., et al. (2015). Proteome-wide Analysis of SUMO2 Targets in Response to Pathological DNA Replication Stress in Human Cells. *DNA Repair*. 25, 84–96. doi:10.1016/j.dnarep.2014.10.011
- Chaganti, R. S. K., Schonberg, S., and German, J. (1974). A Manyfold Increase in Sister Chromatid Exchanges in Bloom's Syndrome Lymphocytes. *Proc. Natl. Acad. Sci.* 71, 4508–4512. doi:10.1073/pnas.71.11.4508
- Chaudhuri, A. R., Callen, E., Ding, X., Gogola, E., Duarte, A. A., Lee, J.-E., et al. (2016). Erratum: Replication fork Stability Confers Chemoresistance in BRCA-Deficient Cells. *Nature* 539, 456. doi:10.1038/nature19826
- Cortez, D. (2019). Replication-Coupled DNA Repair. *Mol. Cell*. 74, 866–876. doi:10.1016/j.molcel.2019.04.027
- Davies, S. L., North, P. S., and Hickson, I. D. (2007). Role for BLM in Replication-fork Restart and Suppression of Origin Firing after Replicative Stress. *Nat. Struct. Mol. Biol.* 14, 677–679. doi:10.1038/nsmb1267
- Dimitrova, D. S., and Gilbert, D. M. (2000). Temporally Coordinated Assembly and Disassembly of Replication Factories in the Absence of DNA Synthesis. *Nat. Cell Biol.* 2, 686–694. doi:10.1038/35036309

- Doherty, K. M., Sommers, J. A., Gray, M. D., Lee, J. W., von Kobbe, C., Thoma, N. H., et al. (2005). Physical and Functional Mapping of the Replication Protein A Interaction Domain of the Werner and Bloom Syndrome Helicases. *J. Biol. Chem.* 280, 29494–29505. doi:10.1074/jbc.m500653200
- Dou, H., Huang, C., Singh, M., Carpenter, P. B., and Yeh, E. T. H. (2010). Regulation of DNA Repair through deSUMOylation and SUMOylation of Replication Protein A Complex. *Mol. Cell.* 39, 333–345. doi:10.1016/j.molcel.2010.07.021
- Dungrawala, H., Rose, K. L., Bhat, K. P., Mohni, K. N., Glick, G. G., Couch, F. B., et al. (2015). The Replication Checkpoint Prevents Two Types of Fork Collapse without Regulating Replisome Stability. *Mol. Cell.* 59, 998–1010. doi:10.1016/j.molcel.2015.07.030
- Eladad, S., Ye, T.-Z., Hu, P., Leversha, M., Beresten, S., Matunis, M. J., et al. (2005). Intra-nuclear Trafficking of the BLM Helicase to DNA Damage-Induced Foci Is Regulated by SUMO Modification. *Hum. Mol. Genet.* 14, 1351–1365. doi:10.1093/hmg/ddi145
- Galanty, Y., Belotserkovskaya, R., Coates, J., and Jackson, S. P. (2012). RNF4, a SUMO-Targeted Ubiquitin E3 Ligase, Promotes DNA Double-Strand Break Repair. *Genes Dev.* 26, 1179–1195. doi:10.1101/gad.188284.112
- Ge, X. Q., and Blow, J. J. (2010). Chk1 Inhibits Replication Factory Activation but Allows Dormant Origin Firing in Existing Factories. *J. Cell Biol.* 191, 1285–1297. doi:10.1083/jcb.201007074
- Gibbs-Seymour, I., Oka, Y., Rajendra, E., Weinert, B. T., Passmore, L. A., Patel, K. J., et al. (2015). Ubiquitin-SUMO Circuitry Controls Activated Fanconi Anemia ID Complex Dosage in Response to DNA Damage. *Mol. Cell.* 57, 150–164. doi:10.1016/j.molcel.2014.12.001
- González-Prieto, R., Eifler-Olivi, K., Claessens, L. A., Willemstein, E., Xiao, Z., Talavera Ormeno, C. M. P., et al. (2021). Global Non-covalent SUMO Interaction Networks Reveal SUMO-dependent Stabilization of the Non-homologous End Joining Complex. *Cell Rep.* 34, 108691. doi:10.1016/j.celrep.2021.108691
- Guzzo, C. M., Berendsen, C. E., Zhu, J., Gupta, V., Datta, A., Greenberg, R. A., et al. (2012). RNF4-dependent Hybrid SUMO-Ubiquitin Chains Are Signals for RAP80 and Thereby Mediate the Recruitment of BRCA1 to Sites of DNA Damage. *Sci. Signal.* 5, ra88. doi:10.1126/scisignal.2003485
- Hashimoto, Y., Ray Chaudhuri, A., Lopes, M., and Costanzo, V. (2010). Rad51 Protects Nascent DNA from Mre11-dependent Degradation and Promotes Continuous DNA Synthesis. *Nat. Struct. Mol. Biol.* 17, 1305–1311. doi:10.1038/nsmb.1927
- Hay, R. T. (2005). Sumo. *Mol. Cell.* 18, 1–12. doi:10.1016/j.molcel.2005.03.012
- Hendriks, I. A., Treffers, L. W., Verlaan-de Vries, M., Olsen, J. V., and Vertegaal, A. C. O. (2015). SUMO-2 Orchestrates Chromatin Modifiers in Response to DNA Damage. *Cell Rep.* 10, 1778–1791. doi:10.1016/j.celrep.2015.02.033
- Hirota, K., Tsuda, M., Murai, J., Takagi, T., Keka, I. S., Narita, T., et al. (2014). SUMO-targeted Ubiquitin Ligase RNF4 Plays a Critical Role in Preventing Chromosome Loss. *Genes Cells.* 19, 743–754. doi:10.1111/gtc.12173
- Hu, X. V., Rodrigues, T. M. A., Tao, H., Baker, R. K., Miraglia, L., Orth, A. P., et al. (2010). Identification of RING finger Protein 4 (RNF4) as a Modulator of DNA Demethylation through a Functional Genomics Screen. *Proc. Natl. Acad. Sci.* 107, 15087–15092. doi:10.1073/pnas.1009025107
- Jaffray, E. G., and Hay, R. T. (2006). Detection of Modification by Ubiquitin-like Proteins. *Methods.* 38, 35–38. doi:10.1016/j.jmeth.2005.07.020
- Karnani, N., and Dutta, A. (2011). The Effect of the Intra-S-phase Checkpoint on Origins of Replication in Human Cells. *Genes Dev.* 25, 621–633. doi:10.1101/gad.2029711
- Karow, J. K., Newman, R. H., Freemont, P. S., and Hickson, I. D. (1999). Oligomeric Ring Structure of the Bloom's Syndrome Helicase. *Curr. Biol.* 9, 597–600. doi:10.1016/s0960-9822(99)80264-4
- Kolinjivadi, A. M., Sannino, V., De Antoni, A., Zadorozhny, K., Kilkenny, M., Técher, H., et al. (2017). Smarcal1-Mediated Fork Reversal Triggers Mre11-dependent Degradation of Nascent DNA in the Absence of Brca2 and Stable Rad51 Nucleofilaments. *Mol. Cell.* 67, 867–881. doi:10.1016/j.molcel.2017.07.001
- Kumar, R., and Sabapathy, K. (2019). RNF4-A Paradigm for SUMOylation-Mediated Ubiquitination. *Proteomics* 19, e1900185. doi:10.1002/pmic.201900185
- Kumar, R., González-Prieto, R., Xiao, Z., Verlaan-de Vries, M., and Vertegaal, A. C. O. (2017). The STUB1 RNF4 Regulates Protein Group SUMOylation by Targeting the SUMO Conjugation Machinery. *Nat. Commun.* 8, 1809. doi:10.1038/s41467-017-01900-x
- Kurat, C. F., Yeeles, J. T. P., Patel, H., Early, A., and Diffley, J. F. X. (2017). Chromatin Controls DNA Replication Origin Selection, Lagging-Strand Synthesis, and Replication Fork Rates. *Mol. Cell.* 65, 117–130. doi:10.1016/j.molcel.2016.11.016
- Lallemant-Breitenbach, V., Jeanne, M., Benhenda, S., Nasr, R., Lei, M., Peres, L., et al. (2008). Arsenic Degrades PML or PML-Rara through a SUMO-Triggered RNF4/ubiquitin-Mediated Pathway. *Nat. Cell Biol.* 10, 547–555. doi:10.1038/ncb1717
- Lönn, U., Lönn, S., Nylen, U., Winblad, G., and German, J. (1990). An Abnormal Profile of DNA Replication Intermediates in Bloom's Syndrome. *Cancer Res.* 50, 3141–3145.
- Luo, K., Zhang, H., Wang, L., Yuan, J., and Lou, Z. (2012). Sumoylation of MDC1 Is Important for Proper DNA Damage Response. *EMBO J.* 31, 3008–3019. doi:10.1038/emboj.2012.158
- Maure, J.-F., Moser, S. C., Jaffray, E. G., Alpi, A. F., and Hay, R. T. (2016). Loss of Ubiquitin E2 Ube2w Rescues Hypersensitivity of Rnf4 Mutant Cells to DNA Damage. *Sci. Rep.* 6, 26178. doi:10.1038/srep26178
- Mijic, S., Zellweger, R., Chappidi, N., Berti, M., Jacobs, K., Mutreja, K., et al. (2017). Replication fork Reversal Triggers fork Degradation in BRCA2-Defective Cells. *Nat. Commun.* 8, 859. doi:10.1038/s41467-017-01164-5
- Moilanen, A.-M., Poukka, H., Karvonen, U., Häkli, M., Jänne, O. A., and Palvimo, J. J. (1998). Identification of a Novel RING finger Protein as a Coregulator in Steroid Receptor-Mediated Gene Transcription. *Mol. Cell Biol.* 18, 5128–5139. doi:10.1128/mcb.18.9.5128
- Mukhopadhyay, D., Arnaoutov, A., and Dasso, M. (2010). The SUMO Protease SENP6 Is Essential for Inner Kinetochore Assembly. *J. Cell Biol.* 188, 681–692. doi:10.1083/jcb.200909008
- Mullen, J. R., Kaliraman, V., Ibrahim, S. S., and Brill, S. J. (2001). Requirement for Three Novel Protein Complexes in the Absence of the Sgs1 DNA Helicase in *Saccharomyces cerevisiae*. *Genetics* 157, 103–118. doi:10.1093/genetics/157.1.103
- Muniandy, P. A., Thapa, D., Thazhathveetil, A. K., Liu, S.-t., and Seidman, M. M. (2009). Repair of Laser-Localized DNA Interstrand Cross-Links in G1 Phase Mammalian Cells. *J. Biol. Chem.* 284, 27908–27917. doi:10.1074/jbc.m109.029025
- Nie, M., Aslanian, A., Prudden, J., Heideker, J., Vashisht, A. A., Wohlschlegel, J. A., et al. (2012). Dual Recruitment of Cdc48 (P97)-Ufd1-Npl4 Ubiquitin-Selective Segregase by Small Ubiquitin-like Modifier Protein (SUMO) and Ubiquitin in SUMO-Targeted Ubiquitin Ligase-Mediated Genome Stability Functions. *J. Biol. Chem.* 287, 29610–29619. doi:10.1074/jbc.m112.379768
- On, K. F., Beuron, F., Frith, D., Snijders, A. P., Morris, E. P., and Diffley, J. F. X. (2014). Prereplicative Complexes Assembled *In Vitro* Support Origin-dependent and Independent DNA Replication. *EMBO J.* 33, 605–620. doi:10.1002/emboj.201387369
- Ouyang, K. J., Woo, L. L., Zhu, J., Huo, D., Matunis, M. J., and Ellis, N. A. (2009). SUMO Modification Regulates BLM and RAD51 Interaction at Damaged Replication forks. *Plos Biol.* 7, e1000252. doi:10.1371/journal.pbio.1000252
- Ouyang, K. J., Yagle, M. K., Matunis, M. J., and Ellis, N. A. (2013). BLM SUMOylation Regulates ssDNA Accumulation at Stalled Replication forks. *Front. Genet.* 4, 167. doi:10.3389/fgene.2013.00167
- Patel, D. S., Misenko, S. M., Her, J., and Bunting, S. F. (2017). BLM Helicase Regulates DNA Repair by Counteracting RAD51 Loading at DNA Double-Strand Break Sites. *J. Cell Biol.* 216, 3521–3534. doi:10.1083/jcb.201703144
- Petermann, E., Orta, M. L., Issaeva, N., Schultz, N., and Helleday, T. (2010). Hydroxyurea-stalled Replication forks Become Progressively Inactivated and Require Two Different RAD51-Mediated Pathways for Restart and Repair. *Mol. Cell.* 37, 492–502. doi:10.1016/j.molcel.2010.01.021
- Pond, K. W., de Renty, C., Yagle, M. K., and Ellis, N. A. (2019). Rescue of Collapsed Replication forks Is Dependent on NSMCE2 to Prevent Mitotic DNA Damage. *Plos Genet.* 15, e1007942. doi:10.1371/journal.pgen.1007942
- Pond, K. W., and Ellis, N. A. (2019). Quantification of Double-Strand Breaks in Mammalian Cells Using Pulsed-Field Gel Electrophoresis. *Methods Mol. Biol.* 1999, 75–85. doi:10.1007/978-1-4939-9500-4_4
- Prudden, J., Pebernard, S., Raffa, G., Slavin, D. A., Perry, J. J. P., Tainer, J. A., et al. (2007). SUMO-targeted Ubiquitin Ligases in Genome Stability. *EMBO J.* 26, 4089–4101. doi:10.1038/sj.emboj.7601838

- Ragland, R. L., Patel, S., Rivard, R. S., Smith, K., Peters, A. A., Bielinsky, A.-K., et al. (2013). RNF4 and PLK1 Are Required for Replication fork Collapse in ATR-Deficient Cells. *Genes Dev.* 27, 2259–2273. doi:10.1101/gad.223180.113
- Rao, V. A., Conti, C., Guirouilh-Barbat, J., Nakamura, A., Miao, Z.-H., Davies, S. L., et al. (2007). Endogenous γ -H2AX-ATM-Chk2 Checkpoint Activation in Bloom's Syndrome Helicase-Deficient Cells Is Related to DNA Replication Arrested Forks. *Mol. Cancer Res.* 5, 713–724. doi:10.1158/1541-7786.mcr-07-0028
- Schlacher, K., Christ, N., Siaud, N., Egashira, A., Wu, H., and Jasini, M. (2011). Double-strand Break Repair-independent Role for BRCA2 in Blocking Stalled Replication fork Degradation by MRE11. *Cell* 145, 529–542. doi:10.1016/j.cell.2011.03.041
- Sciascia, N., Wu, W., Zong, D., Sun, Y., Wong, N., John, S., et al. (2020). Suppressing Proteasome Mediated Processing of Topoisomerase II DNA-Protein Complexes Preserves Genome Integrity. *Elife*. 9, e53447. doi:10.7554/eLife.53447
- Shorrock, A.-M. K., Jones, S. E., Tsukada, K., Morrow, C. A., Belblidia, Z., Shen, J., et al. (2021). The Bloom Syndrome Complex Senses RPA-Coated Single-Stranded DNA to Restart Stalled Replication forks. *Nat. Commun.* 12, 585. doi:10.1038/s41467-020-20818-5
- Sidorova, J. M., Kehrl, K., Mao, F., and Monnat, R., Jr. (2013). Distinct Functions of Human RECQ Helicases WRN and BLM in Replication fork Recovery and Progression after Hydroxyurea-Induced Stalling. *DNA Repair*. 12, 128–139. doi:10.1016/j.dnarep.2012.11.005
- Skoog, L., and Bjursell, G. (1974). Nuclear and Cytoplasmic Pools of Deoxyribonucleoside Triphosphates in Chinese Hamster Ovary Cells. *J. Biol. Chem.* 249, 6434–6438. doi:10.1016/s0021-9258(19)42175-3
- Spies, J., Lukas, C., Somyajit, K., Rask, M.-B., Lukas, J., and Neelsen, K. J. (2019). 53BP1 Nuclear Bodies Enforce Replication Timing at Under-replicated DNA to Limit Heritable DNA Damage. *Nat. Cell Biol.* 21, 487–497. doi:10.1038/s41556-019-0293-6
- Sun, Y., Miller Jenkins, L. M., Su, Y. P., Nitiss, K. C., Nitiss, J. L., and Pommier, Y. (2020). A Conserved SUMO Pathway Repairs Topoisomerase DNA-Protein Cross-Links by Engaging Ubiquitin-Mediated Proteasomal Degradation. *Sci. Adv.* 6, eaba6290. doi:10.1126/sciadv.aba6290
- Tatham, M. H., Geoffroy, M.-C., Shen, L., Plechanovova, A., Hattersley, N., Jaffray, E. G., et al. (2008). RNF4 Is a Poly-SUMO-specific E3 Ubiquitin Ligase Required for Arsenic-Induced PML Degradation. *Nat. Cell Biol.* 10, 538–546. doi:10.1038/ncb1716
- Thangavel, S., Berti, M., Levikova, M., Pinto, C., Gomathinayagam, S., Vujanovic, M., et al. (2015). DNA2 Drives Processing and Restart of Reversed Replication forks in Human Cells. *J. Cell Biol.* 208, 545–562. doi:10.1083/jcb.201406100
- Thomas, J. J., Abed, M., Heuberger, J., Novak, R., Zohar, Y., Beltran Lopez, A. P., et al. (2016). RNF4-Dependent Oncogene Activation by Protein Stabilization. *Cel Rep.* 16, 3388–3400. doi:10.1016/j.celrep.2016.08.024
- Vesela, E., Chroma, K., Turi, Z., and Mistrik, M. (2017). Common Chemical Inducers of Replication Stress: Focus on Cell-Based Studies. *Biomolecules* 7, 19. doi:10.3390/biom7010019
- Vyas, R., Kumar, R., Clermont, F., Helfricht, A., Kalev, P., Sotiropoulou, P., et al. (2013). RNF4 Is Required for DNA Double-Strand Break Repair *In Vivo*. *Cell Death Differ* 20, 490–502. doi:10.1038/cdd.2012.145
- Woodward, A. M., Göhler, T., Luciani, M. G., Oehlmann, M., Ge, X., Gartner, A., et al. (2006). Excess Mcm2-7 License Dormant Origins of Replication that Can Be Used under Conditions of Replicative Stress. *J. Cell Biol.* 173, 673–683. doi:10.1083/jcb.200602108
- Wu, L., Davies, S. L., Levitt, N. C., and Hickson, I. D. (2001). Potential Role for the BLM Helicase in Recombinational Repair via a Conserved Interaction with RAD51. *J. Biol. Chem.* 276, 19375–19381. doi:10.1074/jbc.m009471200
- Xiao, Z., Chang, J.-G., Hendriks, I. A., Sigurdsson, J. O., Olsen, J. V., and Vertegaal, A. C. O. (2015). System-wide Analysis of SUMOylation Dynamics in Response to Replication Stress Reveals Novel Small Ubiquitin-like Modified Target Proteins and Acceptor Lysines Relevant for Genome Stability. *Mol. Cell Proteomics*. 14, 1419–1434. doi:10.1074/mcp.o114.044792
- Xie, J., Kim, H., Moreau, L. A., Puhalla, S., Garber, J., Al Abo, M., et al. (2015). RNF4-mediated Polyubiquitination Regulates the Fanconi Anemia/BRCA Pathway. *J. Clin. Invest.* 125, 1523–1532. doi:10.1172/jci79325
- Xu, D., Guo, R., Soback, A., Bachrati, C. Z., Yang, J., Enomoto, T., et al. (2008). RMI, a New OB-fold Complex Essential for Bloom Syndrome Protein to Maintain Genome Stability. *Genes Dev.* 22, 2843–2855. doi:10.1101/gad.1708608
- Xu, Y.-N., Bazeille, N., Ding, X.-Y., Lu, X.-M., Wang, P.-Y., Bugnard, E., et al. (2012). Multimeric BLM Is Dissociated upon ATP Hydrolysis and Functions as Monomers in Resolving DNA Structures. *Nucleic Acids Res.* 40, 9802–9814. doi:10.1093/nar/gks728
- Xue, C., Daley, J. M., Xue, X., Steinfeld, J., Kwon, Y., Sung, P., et al. (2019). Single-molecule Visualization of Human BLM Helicase as it Acts upon Double- and Single-Stranded DNA Substrates. *Nucleic Acids Res.* 47, 11225–11237. doi:10.1093/nar/gkz810
- Yin, Y., Seifert, A., Chua, J. S., Maure, J.-F., Golebiowski, F., and Hay, R. T. (2012). SUMO-targeted Ubiquitin E3 Ligase RNF4 Is Required for the Response of Human Cells to DNA Damage. *Genes Dev.* 26, 1196–1208. doi:10.1101/gad.189274.112
- Zellweger, R., Dalcher, D., Mutreja, K., Berti, M., Schmid, J. A., Herrador, R., et al. (2015). Rad51-mediated Replication fork Reversal Is a Global Response to Genotoxic Treatments in Human Cells. *J. Cell Biol.* 208, 563–579. doi:10.1083/jcb.201406099
- Zhang, C., Roberts, T. M., Yang, J., Desai, R., and Brown, G. W. (2006). Suppression of Genomic Instability by SLX5 and SLX8 in *Saccharomyces cerevisiae*. *DNA Repair*. 5, 336–346. doi:10.1016/j.dnarep.2005.10.010
- Zhang, X.-D., Goeres, J., Zhang, H., Yen, T. J., Porter, A. C. G., and Matunis, M. J. (2008). SUMO-2/3 Modification and Binding Regulate the Association of CENP-E with Kinetochores and Progression through Mitosis. *Mol. Cell*. 29, 729–741. doi:10.1016/j.molcel.2008.01.013
- Zhu, J., Zhu, S., Guzzo, C. M., Ellis, N. A., Sung, K. S., Choi, C. Y., et al. (2008). Small Ubiquitin-Related Modifier (SUMO) Binding Determines Substrate Recognition and Paralog-Selective SUMO Modification. *J. Biol. Chem.* 283, 29405–29415. doi:10.1074/jbc.m803632000

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Ellis, Zhu, Yagle, Yang, Huang, Kwako, Seidman and Matunis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Transcription-Replication Collisions and Chromosome Fragility

Wei Wu¹, Jing Na He², Mengjiao Lan¹, Pumin Zhang¹ and Wai Kit Chu^{2*}

¹Zhejiang Provincial Key Laboratory of Pancreatic Diseases, The First Affiliated Hospital of Zhejiang University, Hangzhou, China,

²Department of Ophthalmology and Visual Sciences, The Chinese University of Hong Kong, Hong Kong, Hong Kong, China

Accurate replication of the entire genome is critical for cell division and propagation. Certain regions in the genome, such as fragile sites (common fragile sites, rare fragile sites, early replicating fragile sites), rDNA and telomeres, are intrinsically difficult to replicate, especially in the presence of replication stress caused by, for example, oncogene activation during tumor development. Therefore, these regions are particularly prone to deletions and chromosome rearrangements during tumorigenesis, rendering chromosome fragility. Although, the mechanism underlying their “difficult-to-replicate” nature and genomic instability is still not fully understood, accumulating evidence suggests transcription might be a major source of endogenous replication stress (RS) leading to chromosome fragility. Here, we provide an updated overview of how transcription affects chromosome fragility. Furthermore, we will use the well characterized common fragile sites (CFSs) as a model to discuss pathways involved in offsetting transcription-induced RS at these loci with a focus on the recently discovered atypical DNA synthesis repair pathway Mitotic DNA Synthesis (MiDAS).

Keywords: transcription, replication, mitotic DNA synthesis (MiDAS), replication stress, fragile sites

OPEN ACCESS

Edited by:

Advaitha Madireddy,
Rutgers, The State University of New
Jersey, United States

Reviewed by:

Kirill Lobachev,
Georgia Institute of Technology,
United States

*Correspondence:

Wai Kit Chu
waikit@cuhk.edu.hk

Specialty section:

This article was submitted to
Human and Medical Genomics,
a section of the journal
Frontiers in Genetics

Received: 29 October 2021

Accepted: 29 November 2021

Published: 10 December 2021

Citation:

Wu W, He JN, Lan M, Zhang P and
Chu WK (2021) Transcription-
Replication Collisions and
Chromosome Fragility.
Front. Genet. 12:804547.
doi: 10.3389/fgene.2021.804547

INTRODUCTION

To proliferate, a cell needs to go through a division cycle, where it duplicates its chromosomes in S phase. The replicated chromosomes are then separated and segregated into daughter cells during mitosis. Due to the large size of human genome, cells start DNA replication from multiple origins and up to thousands of replication forks are established and coordinated to replicate the genome in a very short time period (S phase). The untimely duplication of the whole genome in S phase can lead to cells entering mitosis with under replicated DNA (URD). URD can affect sister chromatids separation and genome stability, which is a hallmark of cancer (Hanahan and Weinberg, 2011).

There are certain regions in the human genome that are inherently hard-to-replicate, posing a great challenge for the timely duplication of the entire genome. Such regions including fragile sites, rDNA and telomeres have been well documented and are the major cause of chromosome fragility (Özer and Hickson, 2018). Intriguingly, many of these regions are either characterized as containing very large genes or harboring highly transcribed genes, indicating transcription might play an important role in determining their fragility. Indeed, for very long genes that take more than one cell cycle for them to be transcribed, collisions of transcription and replication machineries are inevitable (Helmrich et al., 2011). In this article, we will first review the transcription-replication conflicts and the associated DNA:RNA hybrid called R-loop. R-loop can trigger RS leading to chromosome fragility. We will then discuss the strategies employed by cells to counteract this transcription associated RS to maintain chromosome stability.

TRANSCRIPTION-MEDIATED REPLICATION OBSTACLES

During transcription, RNA polymerase (RNAP) together with multiple transcription elongation and RNA processing factors form a large complex, which is tightly bound to DNA. In addition, eukaryotic transcription needs to be coupled with other downstream events, like RNA splicing to generate mature mRNA. Thus, transcription machineries together with RNA processing factors could be obstacles for an advancing replication fork. Since DNA replication and transcription compete for the same DNA template, it is inevitable that on some occasions there will be a collision between the two machineries. Indeed, it has long been known that transcription can trigger replication-dependent genome instability that would lead to recombination and mutations (Keil and Roeder, 1984; Voelkel-Meiman et al., 1987; Chávez and Aguilera, 1997; Prado and Aguilera, 2005; Kim and Jinks-Robertson, 2012).

Depending on how transcription and replication machineries approach each other, they can either collide co-directionally or in a head-on manner. When transcription uses the leading strand of DNA replication as a template, co-directional collisions might happen; when transcription uses the lagging strand as the substrate, head-on collisions might take place. In general, it is thought that a co-directional encounter is less toxic to cells than the head-on collision. Consistent with this notion, an *in vitro* study suggested that a reconstituted *Escherichia coli* replisome can remove or bypass a co-directional RNAP and use the newly synthesized mRNA as a primer to carry on DNA synthesis (Liu et al., 1993; Pomerantz and O'Donnell, 2008). Another fact is that highly transcribed regions of rRNA, tRNA and some other essential genes are almost exclusively co-directional to fork progression in most of the studied bacteria (Rocha and Danchin, 2003; Guy and Roten, 2004). This is probably because of the evolutionary pressure for the proper replication fork progression at these regions, which is critical for genome stability and cell survival. Indeed, inverting the highly transcribed ribosomal RNA (rRNA) operons can severely affect replication fork progression at these regions and induce DNA damage responses and cell death (Srivatsan et al., 2010). Intriguingly, there is also a bias towards co-directional collision between transcription and replication in the human genome (Petryk et al., 2016). On the other hand, co-directional encounters of transcription and replication machineries can also cause DNA damages. For example, a co-directional encounter can induce replication restart with the help from helicases and replication restart proteins in *Bacillus subtilis* (Merrikh et al., 2011). In budding yeast, DNA polymerases tend to accumulate at highly transcribed genes in an orientation independent manner (Azvolinsky et al., 2009). Furthermore, by generating an episomal system in human cells, Hamperl et al. discovered that co-directional collisions can provoke ATM-dependent DNA damage responses (Hamperl et al., 2017).

Multiple experimental data has indicated that head-on collisions are more deleterious to fork progression than co-directional collisions and are commonly associated with DNA damage formation (Deshpande and Newlon, 1996; Prado and

Aguilera, 2005; Merrikh et al., 2012). First, the tightly bound large RNAP complex can be a physical barrier that is difficult for the replisome to bypass. Second, when transcription and replication machineries move towards each other, the accumulated positive supercoiled DNA structures would slow down the progression of replication fork (Bermejo et al., 2012). Consistent with this, torsional stress relieving factors: DNA topoisomerases I and II, travel with replication forks and are required for the avoidance of transcription and replication conflicts (Tuduri et al., 2009; Bermejo et al., 2012). Lastly, head-on collisions are always associated with the formation of stable pathological nucleic acid structure: R-loop. R-loop is a three-stranded nucleic acid structure containing one DNA-RNA hybrid and one single strand of DNA. It is much more stable than double stranded DNA, and therefore can directly interfere with DNA replication leading to fork stalling or collapse (Aguilera and García-Muse, 2012; Santos-Pereira and Aguilera, 2015). Consistent with this idea, R-loops are significantly enriched at head-on regions in human cells, and introducing a head-on collision on a plasmid causes plasmid loss in a R-loop dependent manner (Hamperl et al., 2017).

TRANSCRIPTION AT CHROMOSOMAL FRAGILE LOCI

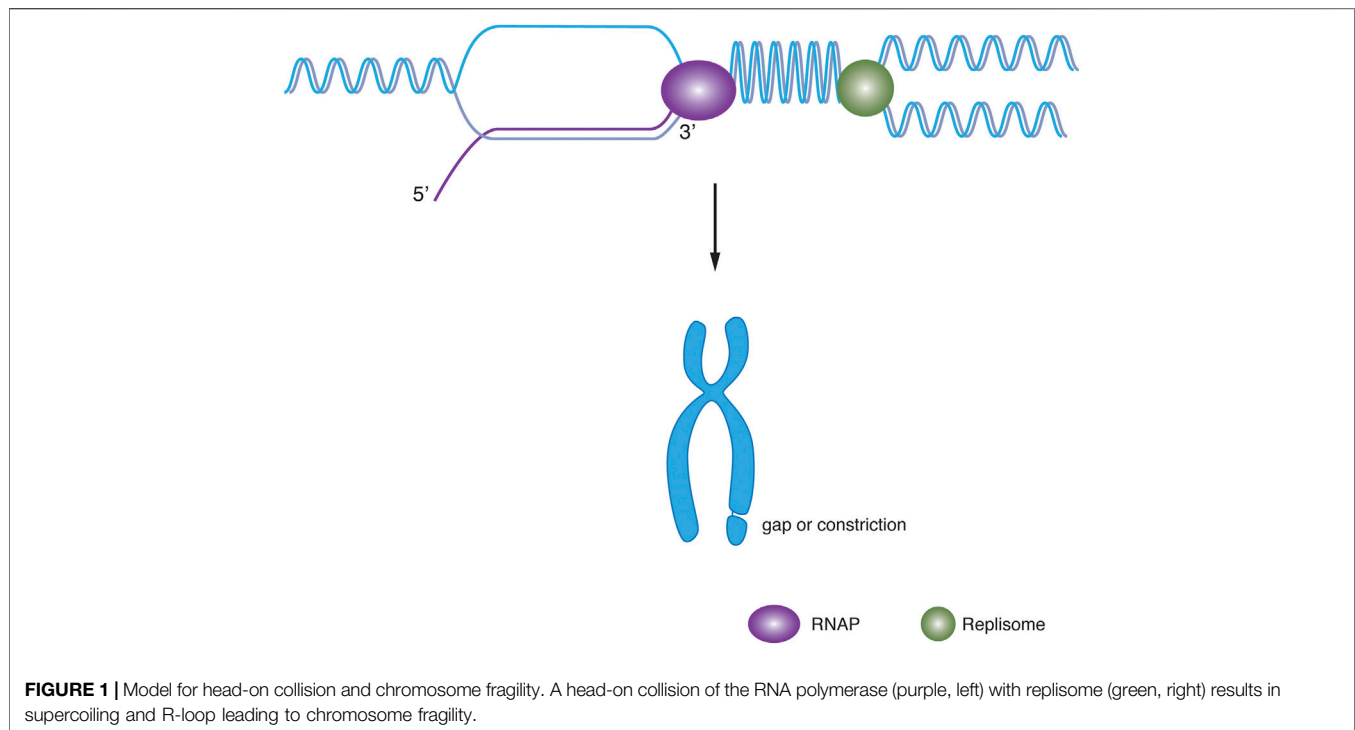
As mentioned above, certain regions in the human genome are intrinsically difficult to replicate and are prone to recombination and viral integration (Thorland et al., 2003; Gao et al., 2017). These regions are called chromosomal fragile loci. Accumulating evidence suggests that transcription and replication conflict is an important factor contributing to chromosomal fragility. Some of these fragile regions are discussed below.

Fragile Sites

Fragile sites are specific regions that tend to form gaps or breaks on metaphase chromosomes in the presence of RS (Sutherland, 1979). Fragile sites could potentially comprise genome stability and are highly associated with the development of various diseases, including cancer and neurological diseases. Currently, there are three known types of fragile sites:

Rare fragile sites (RFSs) are only found in less than 5% individuals in the general population, and their fragility is associated with the expansion of dinucleotide or trinucleotide repeats, which can form secondary DNA structures, affecting DNA replication and transcription (Magenis et al., 1970). Furthermore, R-loop has been found to be at RFS:FRAXA, and is associated with its silencing (Kumari et al., 2015; Bhowmick et al., 2016).

Common fragile sites (CFSs) are present in all individuals. Some of the characteristics of CFSs might explain their fragility. First, CFSs are late replicating, which might leave insufficient time for cells to clear DNA replication obstacles or to repair damaged DNA following replication perturbations (Le Beau, 1998; Hellman et al., 2000). Second, CFSs are short of replication origins (Letessier et al., 2011; Sugimoto et al., 2018; Pruitt et al., 2017). Thus, to completely duplicate CFSs, two replication forks need to travel a long distance to converge,



which is challenging to replication forks, especially when cells experience RS. Recent evidence suggests CFSs fragility is tightly associated with transcription. Indeed, CFSs harboring actively transcribed long genes tend to be hotspots for copy number variants (CNVs) (Wilson et al., 2015). Interestingly, more than 80% of human CFSs and all mouse embryonic fibroblast CFSs overlap with long genes of more than 300 kb (et al., 2013). Due to their large sizes, these regions can take the entire cell cycle for them to be completely transcribed, therefore the collisions between replication and transcription are inevitable, creating stable R-loops and consequent CFSs fragility (Helmrich et al., 2011) (**Figure 1**). Consistent with this notion, the depletion of R-loop by RNase H1 overexpression can reduce CFSs fragility (Helmrich et al., 2011). Also, low doses of aphidicolin (APH, commonly used as a CFSs fragility inducer) induced the formation of FANCD2 protein foci, which decorate the location of CFS loci, were found to bind to the central regions of large genes in human and chicken DT40 cells (Pentzold et al., 2018). In addition, FANCD2 has also been shown to be required to promote DNA replication at CFSs by preventing R-loop formation (García-Rubio et al., 2015; Schwab et al., 2015; Madireddy et al., 2016). Our recent study has identified RTEL1 as a key factor in suppressing CFSs fragility by resolving transcription-replication conflicts caused by G-quadruplex and R-loop structures. Moreover, through DNA-RNA immunoprecipitation (DRIP) sequencing, we found both RTEL1 and low dose APH can induce genome-wide R-loop formation. Importantly, around 70% of well characterized CFSs contains at least one region with high signal intensity of R-loop, which are significantly enriched following RTEL1 depletion or APH treatment (Wu et al.,

2020). Therefore, transcription and replication collisions and their associated R-loops might underline CFSs fragility.

Early replicating fragile sites (ERFSs) are newly discovered fragile sites that reside in highly transcribed, early replicating genes. Besides, ERFSs are very close to replication origins, which further increases the incidence of transcription and replication collisions (Barlow et al., 2013; Mortusewicz et al., 2013). Recent data suggested that transcription and replication conflicts might explain the mechanism of oncogene-induced replication stress. It has been proposed that oncogene activation forces cells exiting G1 phase and entering S phase prematurely, where transcription is not able to suppress replication origins within the body of such genes, resulting in transcription and replication conflicts. Of note, ERFSs have been mapped as hotspots for sites of oncogene-induced replication and transcription collisions (Halazonetis et al., 2008; Macheret and Halazonetis, 2018).

rDNA

Ribosomal DNA (rDNA) are regions clustered with long tandem repeats, where rRNA are transcribed. rDNA is highly transcribed by RNA polymerase I to meet the demand for the translation activity in cells. Each copy of a repeat harbors a potential replication origin (Linskens and Huberman, 1988). In yeast, it is estimated that each active rDNA gene can accommodate up to 150 RNA polymerases (French et al., 2003), which can form a physical barrier for the replication machinery (Berger et al., 1997; Coffman et al., 2005). Therefore, there are very high chances of transcription and replication collisions at rDNA (Deshpande and Newlon, 1996; Takeuchi et al., 2003; Kobayashi, 2014). In agreement with this, transcription and replication conflicts associated with R-loops are commonly found at rDNA

(Chen et al., 2017; Lindström, 2018). Immunofluorescence staining analysis using a specific antibody recognizing R-loops showed the strongest signal in the nucleoli. Nuclear regions containing lots of rDNA, implying stable R-loops formation at rDNA loci. In addition, proteins involved in R-loops prevention or resolution, such as RNase H1, PIF1, and Top I are found to accumulate at rDNA region and are essential for their integrity (El Hage et al., 2010; El Hage et al., 2014; Shen et al., 2017; Tran et al., 2017). Moreover, the RNAi (RNA interference) component, Dicer, was found to be required to limit transcription and replication collisions at rDNA through promoting transcription termination (Castel et al., 2014).

Telomere

Telomeres are regions composed of thousands of TTAGGG repeats that lie at the ends of chromosomes. Usually, telomeres are bound by the shelterin complex. These nucleoprotein structures can prevent telomeres from being recognized as DNA double strand breaks (DSBs), which is critical for genome stability and cell survival (Palm and de Lange, 2008; de Lange, 2009). Although, telomeres are heterochromatic, transcription still happens from the subtelomeric regions towards the end of chromosomes. Currently, two types of telomere associated RNA have been reported, telomeric repeat-containing G-rich RNA (TERRA) and the complementary C-rich transcripts (ARIA) (Kwapisz and Morillon, 2020). Although the biological functions of ARIA remain to be explored, TERRA has been shown to invade into telomeres and induce the formation of stable R-loop structures (Azzalin et al., 2007). TERRA R-loops can be found in all eukaryotes (Rippe and Luke, 2015). As mentioned above, stable R-loops can be a challenging obstacle for replication forks. Furthermore, the displaced G-rich single strand DNA during TERRA R-loop formation promotes the generation of another noncanonical secondary DNA structure: G-quadruplex (G4). G4 is also known to obstruct DNA replication resulting in RS and genome instability (Wang et al., 2019). Therefore, TERRA R-loop may be a major source for telomere fragility. Interestingly, the level of TERRA transcription is negatively correlated with DNA replication process, with the highest being at the G1-S transition and the lowest at the S-G2 transition (Flynn et al., 2015), which can reduce the chance of replication machineries colliding with TERRA R-loops. Besides, several R-loops processing factors have been found to be essential for telomere stability. Specifically, RNaseH1 can

regulate TERRA-telomeric DNA hybrids to maintain telomere stability in cells employing non-telomerase approach to maintain the length of their telomers (Arora et al., 2014). FEN1 and UPF1 (Up-frameshift 1) can promote telomeric leading stand DNA synthesis by processing TERRA R-loops through the flap endonuclease activity and 5'-3' helicase activity, respectively, (Azzalin et al., 2007; Teasley et al., 2015).

PATHWAYS IN DEALING WITH TRANSCRIPTION ASSOCIATED RS AT CFSs

Despite the high incidence of transcription and replication collisions and the prevalence of R-loops at chromosome fragile loci, cells seem to proliferate with a fairly stable genome. Therefore, it is conceivable that cells have evolved surveillance pathways to avoid or minimize the deleterious effects cause by these transcription and replication collision events. Indeed, multiple strategies or pathways have been identified that can prevent or resolve transcription and replication collisions and R-loops, such as spatial and temporal separation of transcription and replication (Wei et al., 1998), the coupling of transcription with mRNA processing (Aguilera and García-Muse, 2012; Huertas and Aguilera, 2003; Li and Manley, 2005), RNAP anti-backtracking and clearance (Castel et al., 2014; Saponaro et al., 2014; Urban et al., 2016; Cheung and Cramer, 2011; Thomas et al., 1998; Roghanian et al., 2015; Ganesan et al., 2012; Epshtein et al., 2014; Zaratiegui et al., 2011; Zatreanu et al., 2019), Fanconi anemia pathway (Madireddy et al., 2016; Schwab et al., 2015; García-Rubio, 2015a), ATR-mediated DNA damage response and the expression of specific R-loop processing factors including RNase H1 (Lockhart et al., 2019) and SETX (García-Muse et al., 2019). Some of the key proteins involved in these pathways are summarized in **Table 1**. Interestingly, deficiency in some of these proteins such as FANCD2, ATR, RNase H1, and Dicer, have been linked to CFSs fragility (Casper et al., 2002; Madireddy et al., 2016; Di Marco et al., 2017; Fragkos et al., 2019), indicating cells employ these pathways to counteract transcription induced obstacles or RS to maintain CFS stability. Of note, most of these pathways seem to largely operate in S phase of the cell cycle.

Recently, an unscheduled mitotic DNA synthesis (MiDAS) pathway functioning in early mitosis has been found to be

TABLE 1 | Key proteins involved in preventing or resolving transcription associated RS.

Pathways	Proteins
Transcription and mRNA processing coupling	THO complex (Gómez-González, (2011), Sin3A (Salas-Armenteros, (2017), Splicing factors (ASF/SF2) (Li and Manley, (2005)
RNAP anti-backtracking	TFIIS Zatreanu et al. (2019), RECQ5 (Saponaro et al., (2014))
RNAP clearance	TC-NER factors (XPF, XPG and CSB) (Sollier et al., (2014), Def1 (Wilson et al., (2013), Dicer (Castel et al., (2014)
Fanconi anemia pathway	FANCD2 (García-Rubio, (2015b)
ATR-mediated DNA damage responses	ATR (Hodroj et al., (2017); Matos et al., (2020)
R-loop processing factors	RNase H1 and RNase H2 (Lockhart et al., (2019), SETX Cohen et al., (2018), DXH9 Cristini et al., (2018), DDX39B Pérez-Calero, (2020)
Mitotic DNA synthesis pathway	POLD3, RAD52, SLX4, RTEL1 (Minocherhomji et al., (2015); Bhowmick et al., (2016); Wu et al., (2020)

essential for CFS stability and is recognized as a “salvage pathway” to alleviate RS accumulated at CFSs (Minocherhomji et al., 2015). In response to RS, CFSs can remain under-replicated from interphase till mitosis without necessarily activating G2/M checkpoint. In early mitosis, under-replicated DNA (URD) can activate MiDAS to complete the under-replicated regions. Failure in filling up URD can cause DNA bridges (chromatin bridges and ultra-fine DNA bridges), which can affect sister chromatids separation and lead to genome instability or mitotic catastrophe (Ying et al., 2013). MiDAS is thought to be a break-induced replication (BIR) like pathway. Consistently, more than half of MiDAS foci, defined by visualizing nascent DNA synthesis in mitosis, were found to be on only one of the two sister chromatids, which is consistent with the conserved DNA synthesis feature in BIR (Özer and Hickson, 2018). Until now, many proteins have been identified to be essential for MiDAS. TRAIIP is an E3 ubiquitin ligase required for replisome unloading at G2/M boundary, which is a prerequisite for SLX4-dependent endonucleases to access and cleave incomplete DNA structures. After that, RAD52 and POLD3 dependent DNA synthesis can take place. Our recent data has linked R-loop to MiDAS (Wu et al., 2020). In the presence of RS, RNase H1 depletion can stimulate robust MiDAS, while overexpression of RNase H1 can reduce MiDAS. The simplest and more understandable interpretation to these results could be envisaged as R-loop induced RS cannot be completely resolved in interphase, making the cells enter mitosis with incompletely replicated DNA, and consequently triggering the MiDAS pathway. However, there is another possibility that R-loop might directly take part in MiDAS pathway and regulate MiDAS. Indeed, R-loop can stimulate RAD52-POLD3-mediated BIR at ROS-induced telomeric DNA breaks (Tan et al., 2020). Which of the two possibilities is closer to what truly happens in cells warrants further investigation.

REFERENCES

- Aguilera, A., and García-Muse, T. (2012). R Loops: from Transcription Byproducts to Threats to Genome Stability. *Mol. Cell* 46 (2), 115–124. doi:10.1016/j.molcel.2012.04.009
- Arora, R., Lee, Y., Wischniewski, H., Brun, C. M., Schwarz, T., and Azzalin, C. M. (2014). RNaseH1 Regulates TERRA-Telomeric DNA Hybrids and Telomere Maintenance in ALT Tumour Cells. *Nat. Commun.* 5, 5220. doi:10.1038/ncomms6220
- Azvolinsky, A., Giresi, P. G., Lieb, J. D., and Zakian, V. A. (2009). Highly Transcribed RNA Polymerase II Genes Are Impediments to Replication Fork Progression in *Saccharomyces cerevisiae*. *Mol. Cell* 34 (6), 722–734. doi:10.1016/j.molcel.2009.05.022
- Azzalin, C. M., Reichenbach, P., Khoraiuli, L., Giulotto, E., and Lingner, J. (2007). Telomeric Repeat-Containing RNA and RNA Surveillance Factors at Mammalian Chromosome Ends. *Science* 318 (5851), 798–801. doi:10.1126/science.1147182
- Barlow, J. H., Faryabi, R. B., Callén, E., Wong, N., Malhowski, A., Chen, H. T., et al. (2013). Identification of Early Replicating Fragile Sites that Contribute to Genome Instability. *Cell* 152 (3), 620–632. doi:10.1016/j.cell.2013.01.006
- Berger, C., Horlebein, A., Gögel, E., and Grummt, F. (1997). Temporal Order of Replication of Mouse Ribosomal RNA Genes during the Cell Cycle. *Chromosoma* 106 (8), 479–484. doi:10.1007/s004120050269
- Bermejo, R., Lai, M. S., and Foiani, M. (2012). Preventing Replication Stress to Maintain Genome Stability: Resolving Conflicts between Replication and Transcription. *Mol. Cell* 45 (6), 710–718. doi:10.1016/j.molcel.2012.03.001

CONCLUSION

Tumorigenesis is a process involving mutations in tumor suppressor genes and the activation of oncogenes. In the early stage of tumor, oncogene activation induces RS and drives genomic instability. Many oncogenes act as growth factors to support proliferation by upregulating transcription factors which in turn can stimulate RNA synthesis (Kotsantis et al., 2016). The increased level of transcription might enhance the conflicts between transcription and replication machineries and its associated R-loops, giving rise to transcription associated RS in S phase. Indeed, accumulating evidence suggests transcription mediated RS is one of the major sources of genome instability, especially at difficult-to-replicate loci mentioned above. Of note, these loci are preferably targeted by RS generated during cancer development (Tsantoulis et al., 2008; Barlow et al., 2013), indicating these loci might be constantly being challenged and mutated during tumorigenesis. Understanding the mechanisms underlying their fragility and identifying molecular pathways in preventing or resolving transcription associated RS will help us understand the molecular basis of cancer development and identify more cancer specific druggable targets.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This work was supported by National Natural Science Foundation of China (Project 82103232, to WW).

- Bhowmick, R., Minocherhomji, S., and Hickson, I. D. (2016). RAD52 Facilitates Mitotic DNA Synthesis Following Replication Stress. *Mol. Cell* 64 (6), 1117–1126. doi:10.1016/j.molcel.2016.10.037
- Casper, A. M., Nghiem, P., Arlt, M. F., and Glover, T. W. (2002). ATR Regulates Fragile Site Stability. *Cell* 111 (6), 779–789. doi:10.1016/s0092-8674(02)01113-3
- Castel, S. E., Ren, J., Bhattacharjee, S., Chang, A.-Y., Sánchez, M., Valbuena, A., et al. (2014). Dicer Promotes Transcription Termination at Sites of Replication Stress to Maintain Genome Stability. *Cell* 159 (3), 572–583. doi:10.1016/j.cell.2014.09.031
- Chávez, S., and Aguilera, A. (1997). The Yeast HPR1 Gene Has a Functional Role in Transcriptional Elongation that Uncovers a Novel Source of Genome Instability. *Genes Dev.* 11 (24), 3459–3470.
- Chen, L., Chen, J.-Y., Zhang, X., Gu, Y., Xiao, R., Shao, C., et al. (2017). R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-Loops with Transcriptional Pausing at Gene Promoters. *Mol. Cell* 68 (4), 745–757. doi:10.1016/j.molcel.2017.10.008
- Cheung, A. C. M., and Cramer, P. (2011). Structural Basis of RNA Polymerase II Backtracking, Arrest and Reactivation. *Nature* 471 (7337), 249–253. doi:10.1038/nature09785
- Coffman, F. D., He, M., Diaz, M.-L., and Cohen, S. (2005). DNA Replication Initiates at Different Sites in Early and Late S Phase within Human Ribosomal RNA Genes. *Cell Cycle* 4 (9), 1223–1226. doi:10.4161/cc.4.9.1961
- Cohen, S., Puget, N., Lin, Y.-L., Clouaire, T., Aguirrebengoa, M., Rocher, V., et al. (2018). Senataxin Resolves RNA:DNA Hybrids Forming at DNA Double-Strand Breaks to Prevent Translocations. *Nat. Commun.* 9 (1), 533. doi:10.1038/s41467-018-02894-w

- Cristini, A., Groh, M., Kristiansen, M. S., and Gromak, N. (2018). RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage. *Cel Rep.* 23 (6), 1891–1905. doi:10.1016/j.celrep.2018.04.025
- de Lange, T. (2009). How Telomeres Solve the End-protection Problem. *Science* 326 (5955), 948–952. doi:10.1126/science.1170633
- Deshpande, A. M., and Newlon, C. S. (1996). DNA Replication fork Pause Sites Dependent on Transcription. *Science* 272 (5264), 1030–1033. doi:10.1126/science.272.5264.1030
- Di Marco, S., Hasanova, Z., Kanagaraj, R., Chappidi, N., Altmannova, V., Menon, S., et al. (2017). RECQ5 Helicase Cooperates with MUS81 Endonuclease in Processing Stalled Replication Forks at Common Fragile Sites during Mitosis. *Mol. Cel* 66 (5), 658–671. doi:10.1016/j.molcel.2017.05.006
- El Hage, A., French, S. L., Beyer, A. L., and Tollervey, D. (2010). Loss of Topoisomerase I Leads to R-Loop-Mediated Transcriptional Blocks during Ribosomal RNA Synthesis. *Genes Dev.* 24 (14), 1546–1558. doi:10.1101/gad.573310
- El Hage, A., Webb, S., Kerr, A., and Tollervey, D. (2014). Genome-wide Distribution of RNA-DNA Hybrids Identifies RNase H Targets in tRNA Genes, Retrotransposons and Mitochondria. *Plos Genet.* 10 (10), e1004716. doi:10.1371/journal.pgen.1004716
- Epshtein, V., Kamarthapu, V., McGary, K., Svetlov, V., Ueberheide, B., Proshkin, S., et al. (2014). UvrD Facilitates DNA Repair by Pulling RNA Polymerase Backwards. *Nature* 505 (7483), 372–377. doi:10.1038/nature12928
- , et al. (2013). Common Fragile Site Profiling in Epithelial and Erythroid Cells Reveals that Most Recurrent Cancer Deletions Lie in Fragile Sites Hosting Large Genes. *Cell Rep* 4 (3), 420–428. doi:10.1016/j.celrep.2013.07.003
- Flynn, R. L., Cox, K. E., Jeitany, M., Wakimoto, H., Bryll, A. R., Ganem, N. J., et al. (2015). Alternative Lengthening of Telomeres Renders Cancer Cells Hypersensitive to ATR Inhibitors. *Science* 347 (6219), 273–277. doi:10.1126/science.1257216
- Fragkos, M., Barra, V., Egger, T., Bordignon, B., Lemacon, D., Naim, V., et al. (2019). Dicer Prevents Genome Instability in Response to Replication Stress. *Oncotarget* 10 (43), 4407–4423. doi:10.18632/oncotarget.27034
- French, S. L., Osheim, Y. N., Cioci, F., Nomura, M., and Beyer, A. L. (2003). In Exponentially Growing *Saccharomyces cerevisiae* Cells, rRNA Synthesis Is Determined by the Summed RNA Polymerase I Loading Rate rather Than by the Number of Active Genes. *Mol. Cel Biol* 23 (5), 1558–1568. doi:10.1128/mcb.23.5.1558-1568.2003
- Gómez-González, B. (2012). *Genome-wide Function of THO/TREX in Active Genes Prevents R-loop-dependent Replication Obstacles*. 30(15). 3106–3119.
- Ganesan, A., Spivak, G., and Hanawalt, P. C. (2012). “Transcription-Coupled DNA Repair in Prokaryotes,” in *Progress in Molecular Biology and Translational Science*. Editor P. W. Doetsch (Academic Press), 25–40. doi:10.1016/b978-0-12-387665-2.00002-x
- Gao, G., Johnson, S. H., Vasmataz, G., Pauley, C. E., Tombers, N. M., Kasperbauer, J. L., et al. (2017). Common Fragile Sites (CFS) and Extremely Large CFS Genes Are Targets for Human Papillomavirus Integrations and Chromosome Rearrangements in Oropharyngeal Squamous Cell Carcinoma. *Genes Chromosomes Cancer* 56 (1), 59–74. doi:10.1002/gcc.22415
- García-Rubio, M. L. (2015b). *The Fanconi Anemia Pathway Protects Genome Integrity from R-Loops*. 11(11): e1005674.
- García-Muse, T., Aguilera, A., and Loops, R. (2019). From Physiological to Pathological Roles. *Cell* 179 (3), 604–618.
- García-Rubio, M. L., Pérez-Calero, C., Barroso, S. I., Tumini, E., Herrera-Moyano, E., Rosado, I. V., et al. (2015). The Fanconi Anemia Pathway Protects Genome Integrity from R-Loops. *Plos Genet.* 11 (11), e1005674. doi:10.1371/journal.pgen.1005674
- García-Rubio, M. L. (2015). The Fanconi Anemia Pathway Protects Genome Integrity from R-Loops. *PLOS Genet.* 11 (11), e1005674.
- Guy, L., and Roten, C.-A. H. (2004). Genometric Analyses of the Organization of Circular Chromosomes: a Universal Pressure Determines the Direction of Ribosomal RNA Genes Transcription Relative to Chromosome Replication. *Gene* 340 (1), 45–52. doi:10.1016/j.gene.2004.06.056
- Halazonetis, T. D., Gorgoulis, V. G., and Bartek, J. (2008). An Oncogene-Induced DNA Damage Model for Cancer Development. *Science* 319 (5868), 1352–1355. doi:10.1126/science.1140735
- Hamperl, S., Bocek, M. J., Saldivar, J. C., Swigut, T., and Cimprich, K. A. (2017). Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. *Cell* 170 (4), 774–786. doi:10.1016/j.cell.2017.07.043
- Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of Cancer: the Next Generation. *Cell* 144 (5), 646–674. doi:10.1016/j.cell.2011.02.013
- Hellman, A., Rahat, A., Scherer, S. W., Darvasi, A., Tsui, L.-C., and Kerem, B. (2000). Replication Delay along FRA7H, a Common Fragile Site on Human Chromosome 7, Leads to Chromosomal Instability. *Mol. Cel Biol* 20 (12), 4420–4427. doi:10.1128/mcb.20.12.4420-4427.2000
- Helmrich, A., Ballarino, M., and Tora, L. (2011). Collisions between Replication and Transcription Complexes Cause Common Fragile Site Instability at the Longest Human Genes. *Mol. Cel* 44 (6), 966–977. doi:10.1016/j.molcel.2011.10.013
- Hodroj, D., Recolin, B., Serhal, K., Martinez, S., Tsanov, N., Abou Merhi, R., et al. (2017). An ATR-dependent Function for the Ddx19 RNA Helicase in Nuclear R-loop Metabolism. *Embo J.* 36 (9), 1182–1198. doi:10.15252/emboj.201695131
- Huertas, P., and Aguilera, A. (2003). Cotranscriptionally Formed DNA:RNA Hybrids Mediate Transcription Elongation Impairment and Transcription-Associated Recombination. *Mol. Cel* 12 (3), 711–721. doi:10.1016/j.molcel.2003.08.010
- Keil, R. L., and Roeder, G. S. (1984). Cis-acting, Recombination-Stimulating Activity in a Fragment of the Ribosomal DNA of *S. cerevisiae*. *Cell* 39 (2 Pt 1), 377–386. doi:10.1016/0092-8674(84)90016-3
- Kim, N., and Jinks-Robertson, S. (2012). Transcription as a Source of Genome Instability. *Nat. Rev. Genet.* 13 (3), 204–214. doi:10.1038/nrg3152
- Kobayashi, T. (2014). Ribosomal RNA Gene Repeats, Their Stability and Cellular Senescence. *Proc. Jpn. Acad. Ser. B: Phys. Biol. Sci.* 90 (4), 119–129. doi:10.2183/pjab.90.119
- Kotsantis, P., Silva, L. M., Irmscher, S., Jones, R. M., Folkes, L., Gromak, N., et al. (2016). Increased Global Transcription Activity as a Mechanism of Replication Stress in Cancer. *Nat. Commun.* 7 (1), 13087. doi:10.1038/ncomms13087
- Kumari, D., Hayward, B., Nakamura, A. J., Bonner, W. M., and Usdin, K. (2015). Evidence for Chromosome Fragility at the Frataxin Locus in Friedreich Ataxia. *Mutat. Research/Fundamental Mol. Mech. Mutagenesis* 781, 14–21. doi:10.1016/j.mrfmmm.2015.08.007
- Kwapisz, M., and Morillon, A. (2020). Subtelomeric Transcription and its Regulation. *J. Mol. Biol.* 432 (15), 4199–4219. doi:10.1016/j.jmb.2020.01.026
- Le Beau, M. (1998). Replication of a Common Fragile Site, FRA3B, Occurs Late in S Phase and Is Delayed Further upon Induction: Implications for the Mechanism of Fragile Site Induction. 7(4). 755–761. doi:10.1093/hmg/7.4.755
- Letessier, A., Millot, G. A., Koundrioukoff, S., Lachagès, A.-M., Vogt, N., Hansen, R. S., et al. (2011). Cell-type-specific Replication Initiation Programs Set Fragility of the FRA3B Fragile Site. *Nature* 470 (7332), 120–123. doi:10.1038/nature09745
- Li, X., and Manley, J. L. (2005). Inactivation of the SR Protein Splicing Factor ASF/SF2 Results in Genomic Instability. *Cell* 122 (3), 365–378. doi:10.1016/j.cell.2005.06.008
- Lindström, M. S. (2018). Nucleolus as an Emerging Hub in Maintenance of Genome Stability and Cancer Pathogenesis. *Oncogene* 37 (18), 2351–2366.
- Linskens, M. H., and Huberman, J. A. (1988). Organization of Replication of Ribosomal DNA in *Saccharomyces cerevisiae*. *Mol. Cel Biol* 8 (11), 4927–4935. doi:10.1128/mcb.8.11.4927-4935.1988
- Liu, B., Lie Wong, M., Tinker, R. L., Geiduschek, E. P., and Alberts, B. M. (1993). The DNA Replication fork Can Pass RNA Polymerase without Displacing the Nascent Transcript. *Nature* 366 (6450), 33–39. doi:10.1038/366033a0
- Lockhart, A., Pires, V. B., Bento, F., Kellner, V., Luke-Glaser, S., Yakoub, G., et al. (2019). RNase H1 and H2 Are Differentially Regulated to Process RNA-DNA Hybrids. *Cel Rep.* 29 (9), 2890–2900. doi:10.1016/j.celrep.2019.10.108
- Macheret, M., and Halazonetis, T. D. (2018). Intragenic Origins Due to Short G1 Phases Underlie Oncogene-Induced DNA Replication Stress. *Nature* 555 (7694), 112–116. doi:10.1038/nature25507
- Madireddy, A., Kosiyatrakul, S. T., Boisvert, R. A., Herrera-Moyano, E., García-Rubio, M. L., Gerhardt, J., et al. (2016). FANCD2 Facilitates Replication through Common Fragile Sites. *Mol. Cel* 64 (2), 388–404. doi:10.1016/j.molcel.2016.09.017

- Magenis, R. E., Hecht, F., and Lovrien, E. W. (1970). Heritable Fragile Site on Chromosome 16: Probable Localization of Haptoglobin Locus in Man. *Science* 170 (3953), 85–87. doi:10.1126/science.170.3953.85
- Matos, D. A., Zhang, J.-M., Ouyang, J., Nguyen, H. D., Genois, M.-M., and Zou, L. (2020). ATR Protects the Genome against R Loops through a MUS81-triggered Feedback Loop. *Mol. Cell* 77 (3), 514–527. doi:10.1016/j.molcel.2019.10.010
- Merrikh, H., Machón, C., Grainger, W. H., Grossman, A. D., and Soultanas, P. (2011). Co-directional Replication-Transcription Conflicts lead to Replication Restart. *Nature* 470 (7335), 554–557. doi:10.1038/nature09758
- Merrikh, H., Zhang, Y., Grossman, A. D., and Wang, J. D. (2012). Replication-transcription Conflicts in Bacteria. *Nat. Rev. Microbiol.* 10 (7), 449–458. doi:10.1038/nrmicro2800
- Minocherhomji, S., Ying, S., Bjerregaard, V. A., Bursomanno, S., Aleliunaite, A., Wu, W., et al. (2015). Replication Stress Activates DNA Repair Synthesis in Mitosis. *Nature* 528 (7581), 286–290. doi:10.1038/nature16139
- Mortusewicz, O., Herr, P., and Helleday, T. (2013). Early Replication Fragile Sites: where Replication-Transcription Collisions Cause Genetic Instability. *Embo J.* 32 (4), 493–495. doi:10.1038/emboj.2013.20
- Özer, Ö., and Hickson, I. D. (2018). Pathways for Maintenance of Telomeres and Common Fragile Sites during DNA Replication Stress. *Open Biol.* 8 (4), 180018.
- Pérez-Calero, C. (2020). *UAP56/DDX39B Is a Major Cotranscriptional RNA-DNA Helicase that Unwinds Harmful R Loops Genome-wide.* 34(13-14). 898.
- Palm, W., and de Lange, T. (2008). How Shelterin Protects Mammalian Telomeres. *Annu. Rev. Genet.* 42, 301–334. doi:10.1146/annurev.genet.41.110306.130350
- Pentzold, C., Shah, S. A., Hansen, N. R., Le Tallec, B., Seguin-Orlando, A., Debatisse, M., et al. (2018). FANCD2 Binding Identifies Conserved Fragile Sites at Large Transcribed Genes in Avian Cells. *Nucleic Acids Res.* 46 (3), 1280–1294. doi:10.1093/nar/gkx1260
- Petryk, N., Kahli, M., d'Aubenton-Carafa, Y., Jaszczyszyn, Y., Shen, Y., Silvain, M., et al. (2016). Replication Landscape of the Human Genome. *Nat. Commun.* 7, 10208. doi:10.1038/ncomms10208
- Pomerantz, R. T., and O'Donnell, M. (2008). The Replisome Uses mRNA as a Primer after Colliding with RNA Polymerase. *Nature* 456 (7223), 762–766. doi:10.1038/nature07527
- Prado, F., and Aguilera, A. (2005). Impairment of Replication fork Progression Mediates RNA polII Transcription-Associated Recombination. *Embo j* 24 (6), 1267–1276. doi:10.1038/sj.emboj.7600602
- Pruitt, S. C., Qin, M., Wang, J., Kunnev, D., and Freeland, A. (2017). A Signature of Genomic Instability Resulting from Deficient Replication Licensing. *Plos Genet.* 13 (1), e1006547. doi:10.1371/journal.pgen.1006547
- Rippe, K., and Luke, B. (2015). TERRA and the State of the Telomere. *Nat. Struct. Mol. Biol.* 22 (11), 853–858. doi:10.1038/nsmb.3078
- Rocha, E. P. C., and Danchin, A. (2003). Essentiality, Not Expressiveness, Drives Gene-Strand Bias in Bacteria. *Nat. Genet.* 34 (4), 377–378. doi:10.1038/ng1209
- Roghani, M., Zenkin, N., and Yuzenkova, Y. (2015). Bacterial Global Regulators DksA/ppGpp Increase Fidelity of Transcription. *Nucleic Acids Res.* 43 (3), 1529–1536. doi:10.1093/nar/gkv003
- Salas-Armenteros, I. (2017). *Human THO-Sin3A Interaction Reveals New Mechanisms to Prevent R-loops that Cause Genome Instability.* 36(23). 3532–3547.
- Santos-Pereira, J. M., and Aguilera, A. (2015). R Loops: New Modulators of Genome Dynamics and Function. *Nat. Rev. Genet.* 16 (10), 583–597. doi:10.1038/nrg3961
- Saponaro, M., Kantidakis, T., Mitter, R., Kelly, G. P., Heron, M., Williams, H., et al. (2014). RECQL5 Controls Transcript Elongation and Suppresses Genome Instability Associated with Transcription Stress. *Cell* 157 (5), 1037–1049. doi:10.1016/j.cell.2014.03.048
- Schwab, R. A., Nieminszczy, J., Shah, F., Langton, J., Lopez Martinez, D., Liang, C.-C., et al. (2015). The Fanconi Anemia Pathway Maintains Genome Stability by Coordinating Replication and Transcription. *Mol. Cell* 60 (3), 351–361. doi:10.1016/j.molcel.2015.09.012
- Shen, W., Sun, H., De Hoyos, C. L., Bailey, J. K., Liang, X.-h., and Crooke, S. T. (2017). Dynamic Nucleoplasmic and Nucleolar Localization of Mammalian RNase H1 in Response to RNAP I Transcriptional R-Loops. *Nucleic Acids Res.* 45 (18), 10672–10692. doi:10.1093/nar/gkx710
- Sollier, J., Stork, C. T., García-Rubio, M. L., Paulsen, R. D., Aguilera, A., and Cimprich, K. A. (2014). Transcription-coupled Nucleotide Excision Repair Factors Promote R-Loop-Induced Genome Instability. *Mol. Cell* 56 (6), 777–785. doi:10.1016/j.molcel.2014.10.020
- Srivatsan, A., Tehrani, A., MacAlpine, D. M., and Wang, J. D. (2010). Co-Orientation of Replication and Transcription Preserves Genome Integrity. *Plos Genet.* 6 (1), e1000810. doi:10.1371/journal.pgen.1000810
- Sugimoto, N., Maehara, K., Yoshida, K., Ohkawa, Y., and Fujita, M. (2018). *Genome-wide Analysis of the Spatiotemporal Regulation of Firing and Dormant Replication Origins in Human Cells.* 46(13). 6683–6696. doi:10.1093/nar/gky476
- Sutherland, G. R. (1979). Heritable Fragile Sites on Human Chromosomes I. Factors Affecting Expression in Lymphocyte Culture. *Am. J. Hum. Genet.* 31 (2), 125–135.
- Takeuchi, Y., Horiuchi, T., and Kobayashi, T. (2003). Transcription-dependent Recombination and the Role of fork Collision in Yeast rDNA. *Genes Dev.* 17 (12), 1497–1506. doi:10.1101/gad.1085403
- Tan, J., Duan, M., Yadav, T., Phoon, L., Wang, X., Zhang, J.-M., et al. (2020). An R-Loop-Initiated CSB-RAD52-POLD3 Pathway Suppresses ROS-Induced Telomeric DNA Breaks. *Nucleic Acids Res.* 48 (3), 1285–1300. doi:10.1093/nar/gkz1114
- Teasley, D. C., Parajuli, S., Nguyen, M., Moore, H. R., Alspach, E., Lock, Y. J., et al. (2015). Flap Endonuclease 1 Limits Telomere Fragility on the Leading Strand. *J. Biol. Chem.* 290 (24), 15133–15145. doi:10.1074/jbc.m115.647388
- Thomas, M. J., Platas, A. A., and Hawley, D. K. (1998). Transcriptional Fidelity and Proofreading by RNA Polymerase II. *Cell* 93 (4), 627–637. doi:10.1016/s0092-8674(00)81191-5
- Thorland, E. C., Myers, S. L., Gostout, B. S., and Smith, D. I. (2003). Common Fragile Sites Are Preferential Targets for HPV16 Integrations in Cervical Tumors. *Oncogene* 22 (8), 1225–1237. doi:10.1038/sj.onc.1206170
- Tran, P. L. T., Pohl, T. J., Chen, C.-F., Chan, A., Pott, S., and Zakian, V. A. (2017). PIF1 Family DNA Helicases Suppress R-Loop Mediated Genome Instability at tRNA Genes. *Nat. Commun.* 8, 15025. doi:10.1038/ncomms15025
- Tsantoulis, P. K., Kotsinas, A., Sfrikakis, P. P., Evangelou, K., Sideridou, M., Levy, B., et al. (2008). Oncogene-induced Replication Stress Preferentially Targets Common Fragile Sites in Preneoplastic Lesions. A Genome-wide Study. *Oncogene* 27 (23), 3256–3264. doi:10.1038/sj.onc.1210989
- Tuduri, S., Crabbé, L., Conti, C., Tourrière, H., Holtgreve-Grez, H., Jauch, A., et al. (2009). Topoisomerase I Suppresses Genomic Instability by Preventing Interference between Replication and Transcription. *Nat. Cell Biol.* 11 (11), 1315–1324. doi:10.1038/ncb1984
- Urban, V., Dobrovolska, J., Hühner, D., Fryzelkova, J., Bartek, J., and Janscak, P. (2016). RECQ5 Helicase Promotes Resolution of Conflicts between Replication and Transcription in Human Cells. *J. Cell Biol.* 214 (4), 401–415. doi:10.1083/jcb.201507099
- Voelkel-Meiman, K., Keil, R. L., and Roeder, G. S. (1987). Recombination-stimulating Sequences in Yeast Ribosomal DNA Correspond to Sequences Regulating Transcription by RNA Polymerase I. *Cell* 48 (6), 1071–1079. doi:10.1016/0092-8674(87)90714-8
- Wang, Y., Yang, J., Wild, A. T., Wu, W. H., Shah, R., Danussi, C., et al. (2019). G-quadruplex DNA Drives Genomic Instability and Represents a Targetable Molecular Abnormality in ATRX-Deficient Malignant Glioma. *Nat. Commun.* 10 (1), 943. doi:10.1038/s41467-019-08905-8
- Wei, X., Samarabandu, J., Devdhar, R. S., Siegel, A. J., Acharya, R., and Berezney, R. (1998). Segregation of Transcription and Replication Sites into Higher Order Domains. *Science* 281 (5382), 1502–1505. doi:10.1126/science.281.5382.1502
- Wilson, M. D., Harreman, M., Taschner, M., Reid, J., Walker, J., Erdjument-Bromage, H., et al. (2013). Proteasome-mediated Processing of Def1, a Critical Step in the Cellular Response to Transcription Stress. *Cell* 154 (5), 983–995. doi:10.1016/j.cell.2013.07.028
- Wilson, T. E., Arlt, M. F., Park, S. H., Rajendran, S., Paulsen, M., Ljungman, M., et al. (2015). Large Transcription Units Unify Copy Number Variants and Common Fragile Sites Arising under Replication Stress. *Genome Res.* 25 (2), 189–200. doi:10.1101/gr.177121.114
- Wu, W., Bhowmick, R., Vogel, I., Özer, Ö., Ghisays, F., Thakur, R. S., et al. (2020). RTEL1 Suppresses G-Quadruplex-Associated R-Loops at Difficult-To-Replicate Loci in the Human Genome. *Nat. Struct. Mol. Biol.* 27 (5), 424–437. doi:10.1038/s41594-020-0408-6
- Ying, S., Minocherhomji, S., Chan, K. L., Palmai-Pallag, T., Chu, W. K., Wass, T., et al. (2013). MUS81 Promotes Common Fragile Site Expression. *Nat. Cell Biol.* 15 (8), 1001–1007. doi:10.1038/ncb2773

- Zaratiegui, M., Castel, S. E., Irvine, D. V., Kloc, A., Ren, J., Li, F., et al. (2011). RNAi Promotes Heterochromatic Silencing through Replication-Coupled Release of RNA Pol II. *Nature* 479 (7371), 135–138. doi:10.1038/nature10501
- Zatreanu, D., Han, Z., Mitter, R., Tumini, E., Williams, H., Gregersen, L., et al. (2019). Elongation Factor TFIIS Prevents Transcription Stress and R-Loop Accumulation to Maintain Genome Stability. *Mol. Cell* 76 (1), 57–69. doi:10.1016/j.molcel.2019.07.037

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Wu, He, Lan, Zhang and Chu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Characterization of Chromosomal Instability in Glioblastoma

Elisa Balzano^{1,2}, Elena Di Tommaso^{1,2}, Antonio Antoccia³, Franca Pelliccia^{1*} and Simona Giunta^{2*}

¹Laboratory of Molecular Cytogenetics, Dipartimento di Biologia e Biotechnologie "Charles Darwin", Sapienza Università di Roma, Roma, Italy, ²Laboratory of Genome Evolution, Dipartimento di Biologia e Biotechnologie "Charles Darwin", Sapienza Università di Roma, Roma, Italy, ³Laboratory of Genetics and Cytogenetics, Dipartimento di Scienze, Università Degli Studi Roma Tre, Roma, Italy

OPEN ACCESS

Edited by:

Qing Hu,
University of Texas Southwestern
Medical Center, United States

Reviewed by:

Arvind Panday,
Harvard Medical School,
United States
Vibe Oestergaard,
University of Copenhagen, Denmark

*Correspondence:

Franca Pelliccia
franca.pelliccia@uniroma1.it
Simona Giunta
simona.giunta@uniroma1.it

Specialty section:

This article was submitted to
Human and Medical Genomics,
a section of the journal
Frontiers in Genetics

Received: 07 November 2021

Accepted: 23 December 2021

Published: 28 January 2022

Citation:

Balzano E, Di Tommaso E, Antoccia A,
Pelliccia F and Giunta S (2022)
Characterization of Chromosomal
Instability in Glioblastoma.
Front. Genet. 12:810793.
doi: 10.3389/fgene.2021.810793

Glioblastoma multiforme (GBM) is a malignant tumor of the central nervous system (CNS). The poor prognosis of GBM due to resistance to therapy has been associated with high chromosomal instability (CIN). Replication stress is a major cause of CIN that manifests as chromosome rearrangements, fragility, and breaks, including those cytologically expressed within specific chromosome regions named common fragile sites (CFSs). In this work, we characterized the expression of human CFSs in the glioblastoma U-251 MG cell line upon treatment with the inhibitor of DNA polymerase alpha aphidicolin (APH). We observed 52 gaps/breaks located within previously characterized CFSs. We found 17 to be CFSs in GBM cells upon treatment with APH, showing a frequency equal to at least 1% of the total gaps/breaks. We report that two CFSs localized to regions FRA2E (2p13/p12) and FRA2F (2q22) were only found in U-251 MG cells, but not lymphocytes or fibroblasts, after APH treatment. Notably, these glioblastoma-specific CFSs had a relatively high expression compared to the other CFSs with breakage frequency between ~7 and 9%. Presence of long genes, incomplete replication, and delayed DNA synthesis during mitosis (MiDAS) after APH treatment suggest that an impaired replication process may contribute to this loci-specific fragility in U-251 MG cells. Altogether, our work offers a characterization of common fragile site expression in glioblastoma U-251 MG cells that may be further exploited for cytogenetic and clinical studies to advance our understanding of this incurable cancer.

Keywords: glioblastoma, cancer, chromosome instability, replicative stress, fragile sites

INTRODUCTION

Fragile sites are defined as gaps/breaks induced by replication stress that are visible on metaphase chromosomes. The Human Genome database currently reports 120 chromosomal regions to be fragile sites of which 30 are classified as rare fragile sites (RFSs) and 90 as common fragile sites (CFSs) (Feng and Chakraborty, 2018). To be considered fragile sites, these loci must exhibit as a chromosomal gap/break in at least 1% out of all gaps/breaks upon replication stress induced by treatments such as the DNA polymerase alpha inhibitor, aphidicolin (APH) (Le Tallec et al., 2011; Le Tallec et al., 2013; Arlt et al., 2003). The expression of common fragile sites as chromosome gaps/breaks may not be triggered by a single factor but rather by a combination of different mechanisms. Several studies in the last decade have added to our understanding of which CFS molecular features contribute to their fragility, especially regarding the convergence of replication and transcription

machineries within these loci (Helmrich et al., 2011). Growing evidence shows that CFS instability can vary between different cell types in response to replicative stress conditions, partly due to tissue-specific expression of genes located within each CFS (Debatisse et al., 2012). Similarly, transcription of non-coding RNAs can also trigger chromosomal fragility within CFSs as has been shown for other fragile regions of the genome such as centromeres (Balzano et al., 2021; Balzano and Giunta, 2020; Giunta, 2018), leading to rearrangements and aneuploidy (Giunta et al., 2021; Giunta and Funabiki, 2017). Indeed, many of the common fragile sites recorded in lymphocytes harbor genes longer than 650 kb, called very long genes (VLGs) (Smith et al., 2006; Bosco et al., 2010) and long-non-coding RNAs (lncRNAs) as we recently characterized in fibroblasts (Maccaroni et al., 2020). A connection between the level of transcription and the frequency of instability of the corresponding region has been reported (Helmrich et al., 2011; Brison et al., 2019). However, APH delays replication timing (RT) of large genes in either fragile or non-fragile loci, highlighting that delayed replication timing and transcription processes alone are insufficient to drive CFS fragility (Sarni et al., 2020). Recent evidence shows that CFSs are chromatin regions with a defective condensin loading due to an under-replicated state which persists until mitosis (Boteva et al., 2020). These faulty chromatin-folding regions have been detected as sites of mitotic DNA synthesis (MiDAS) (Ji et al., 2020; Macheret et al., 2020), implying that chromatin conformation and delayed replication further underline their fragility (Miotto et al., 2016).

Fragile sites represent threats to genome stability as breaks on metaphase chromosomes but also as regions that are hot spots for deletions of tumor suppressor genes (TSGs) in cancer cells (Casper et al., 2002; Smith et al., 2007; Sarni and Kerem, 2016). The degree of a cause-effect relationship between the expression of fragile sites and TSG-driven transformation is unclear. However, mapping these fragile regions throughout the human genome in several tissues might identify specific tumor-associated stress sites that can contribute to malignancy.

To this end, we used the glioblastoma multiforme U-251 MG cell line as a genetic model in which we characterized common fragile site expression under mild replication stress using the DNA polymerase alpha inhibitor, aphidicolin (APH). We utilized low doses of APH to analyze its effects on metaphases and interphase nuclei and obtained a detailed quantification of CFS expression by scoring for breakage frequency across all known human CFS chromosomal regions. As a control, we compared GBM cells against lymphocytes taken from peripheral blood of healthy individuals and fibroblasts previously used in the study by Maccaroni et al., 2020 to analyze the tissue-specific responses to APH treatment. We found two glioblastoma-specific CFSs along with several CFSs that are expressed in GBM cells at a higher frequency than in primary tissues. We observed presence of long genes and delayed replication within some of these fragile regions, likely contributing to their breakage and expression as CFSs. Future studies of these CFSs can offer a window of opportunity to better understand CIN in GBM cells and may inform novel cancer treatments, such as use of transcription or DNA damage response

(DDR) inhibitors to prevent clone selection, evolution, progression, and resistance of this deadly tumor.

MATERIALS AND METHODS

Human Cell Cultures

The human glioblastoma U-251 MG cell line (Astrocytoma IV WHO grade) was purchased from Banca Biologica and Cell Factory (Banca Biologica and Cell Factory, Genoa, Italy) and was provided by A.A.; U-251 MG cells were grown in Eagle's Minimum Essential Medium (EMEM; Euroclone) supplemented with 10% fetal bovine serum (FBS) (Corning), 1% penicillin (Thermo Fisher Scientific), and 1% L-glutamine (Thermo Fisher Scientific) at 37°C with 5% CO₂. Lymphocyte cultures were prepared from human peripheral whole blood of healthy individuals, collected with heparin; cells were grown in RPMI medium (Corning) supplemented with 10% FBS, 1% penicillin, and 1% L-glutamine at 37°C with 5% CO₂. For stimulation of T lymphocyte proliferation, phytohemagglutinin (PHA, GIBCO, 3%) was added to the culture medium for 72 h. To induce common fragile sites, aphidicolin (0.4 µM) was added to the medium of both lymphocytes and U-251 MG cells for 22–24 h. To collect the mitotic cells, colchicine (Sigma-Aldrich) was supplemented to the medium of lymphocytes (1 mM) and U-251 MG cells (5 µM) for 2 and 4 h, respectively. For replication timing analysis, 5-Bromo-2'-deoxyuridine (BrdU; 10 µM) was added 20 min prior to harvesting the cells.

Metaphase Spreads Preparation

Upon colchicine treatment, lymphocytes and U-251 MG cells were harvested for metaphase spreads preparation. U-251 MG cells were trypsinized (trypsin 0.1% EDTA, Corning) and centrifuged prior to the addition of hypotonic solution (KCl 0.075 M) for 20 min at 37°C. The hypotonic treatment was performed for 8 min in lymphocytes; after centrifugation, Ibraimov's solution (3% methanol and 5% acetic acid in dH₂O) was used to remove the erythrocytes. Swollen cells were centrifuged twice and resuspended in cold fixative solution (methanol:acetic acid at a ratio of 3:1) and then stored overnight at –20°C. The metaphase spreads and interphasic nuclei were dropped onto clean glass slides and air-dried. The slides were stored at 4°C until subsequent analysis.

Cytogenetic Observation and Analysis

The slides were stained with 4% Giemsa (Carlo Erba) to detect chromosome aberrations and then with Chromomycin A3 (R-banding) to localize gaps/breaks, according to ISNC recommendation (Maccaroni et al., 2020). The karyotype of this glioblastoma multiforme U-251 MG clone using mFISH (multicolor-FISH) was performed by Antocchia Laboratory as can be seen in the study by Berardinelli et al., 2018. The quantification of chromosome aberrations was done according to the OECD guideline (OECD Test No, 2016).

BAC Extraction and Labeling by Nick Translation

The bacterial artificial chromosomes (BACs) were chosen from NCBI GenBank, 2121 (<https://www.ncbi.nlm.nih.gov/genbank/>) for chromosome 1 (RP11-316C12, chr1: 71,385,313-71,476,945), chromosome 3 (RP11-324H4 chr3: 116,954,325-117,125,019), and chromosome 7 (RP11 321C7 chr7: 67,705,408 - 67,771,498). Bacteria were grown in 10 ml of Luria-Bertani (LB) medium and selected with chloramphenicol (20 µg/ml). BACs were extracted by alkaline lysis and subsequently labeled by Nick Translation with bio-16-dUTP (biotin-16-deoxy-Uridine Triphosphate) and/or dig-16-dUTP (digoxigenin-16-deoxy-Uridine Triphosphate). The labeled probes were used for fluorescent *in situ* hybridization (FISH) experiments on interphase nuclei.

DNA Fluorescence *in situ* Hybridization

The slides were treated with RNase (100 µg/ml in 2× saline sodium citrate SSC solution) for 1 h at 37°C and dehydrated by washing for 5 min in 70, 90, and 100% ethanol. After air-drying, the slides were aged at 65°C for 60 min and denatured at 80°C for 2 min with 70% formamide (Sigma) in 2× SSC. The denaturation was stopped with cold 70% ethanol for 5 min, and the slides were dehydrated again with 90 and 100% ethanol and air-dried prior to hybridization using the denatured probes (200 ng). The used BACs were RP11-316C12 (1p31.1), RP11-324H4 (3q13.3), and RP11 321C7 (7q11.2). Sequentially, the overnight BAC probe incubation at 37°C and 3 × 5-min post-hybridization washes with 1× SSC were performed at 60°C. The slides were then incubated for 30 min with anti-digoxigenin-rhodamine antibody (1:20, Roche). Three washes with 0.1% Tween20 in 2× SSC were performed. The slides were counterstained with DAPI (4',6'-diamidino-2-phenylindole hydrochloride, Sigma; 1 µg/ml), diluted 1:300 in VECTASHIELD Antifade Mounting Medium (Vector Laboratories).

Immunofluorescence on Interphase Nuclei

Immunofluorescence (IF) against BrdU was performed to distinguish the different stages of the S-phase. The slides were incubated for 1 h at room temperature with the anti-mouse BrdU monoclonal antibody (MoBU-1, Thermo Fisher Scientific), diluted 1:300 in 5% FBS in 1X PBS, pH 7.4. After 3 × 5-min washes in 1X PBS, the slides were incubated with FITC-conjugated anti-mouse IgG H&L antibody (Abcam) (1:1000 in 1X PBS, pH 7.4) for 1 h at room temperature. After 4 × 5-min washes in 1X PBS, a DAPI:VECTASHIELD Antifade solution (1:300) was used to mount the slides.

Acquisition and Processing of Sample Observations

Metaphase spreads and nuclei were observed at a magnification of 100X using an epifluorescence microscope (Zeiss Axioplan) equipped with a CCD camera (charge-coupled device). The images of metaphase spreads and nuclei were taken using

RSImage software and then processed using Photoshop (Adobe) software.

Fragile Sites Sequence Analysis

In addition to fragile site expression, we analyzed the sequence and gene composition for the region using three Human Genome Resources (NCBI Genome Data Viewer, 2021: <https://www.ncbi.nlm.nih.gov/genome/gdv/> [Release Data May 16, 2021]; Ensembl, 2021: http://www.ensembl.org/Homo_sapiens/Location/Genome [Human GRCh38p13]; GeneCards, 2021: <https://www.genecards.org/>). The characterization is visible in **Supplementary Table S1**.

The elements within the fragile sites characterized include location, length, and gene expression. **Figure 6** and **Supplementary Figure S2** represent ideograms for the chosen fragile sites with representations of the cytogenetic band and the genes expressed in brain tissue.

Statistical Analysis

Statistical paired *t*-tests were calculated on Prism (TablePad software). Individual *p* values are indicated in the figures. *p* values: ns (not significant) *p* > 0.05, **p* ≤ 0.05, ***p* < 0.01, ****p* < 0.001, and *****p* < 0.0001.

RESULTS

Replication Stress Induces Chromosomal Instability in the U-251 MG Glioblastoma Cell Line

A complete and accurate DNA replication process is essential for chromosome integrity. In cancer, DNA synthesis can be compromised by a lack of checkpoint control leading to mitotic arrest and/or genome instability.

To understand how replication stress affects GBM cells, we evaluated the mitotic index of U-251 MG glioblastoma cells compared with primary lymphocytes and MRC-5 fibroblasts upon treatment with a low dose of APH (0.4 µM) for 24 h. We quantified the ratio of mitotic cells on the total of 500 cells and found that aphidicolin-treated cells show a significant reduction in the mitotic index (M.I.) of U-251 MG glioblastoma cells similarly to lymphocytes (**Figure 1**).

We further assessed a variety of cellular phenotypes in both metaphase spreads and in interphase nuclei such as blebbing, cytoplasmic bridges, and micronuclei (**Supplementary Figures S1A–C**). Comparing the response under the APH stress condition in all three cell types, we found a trend of increased cytoplasmic bridges (**Supplementary Figure S1B**). Instead, the frequency of cellular blebbing and micronuclei remained unaffected between untreated and treated conditions in glioblastoma, lymphocytes, and fibroblasts (**Supplementary Figures S1A,C**). These results suggest that cells are likely arrested before entry into mitosis or after chromosome segregation in G1 after replication stress or that low-dose APH triggers only very mild phenotypes.

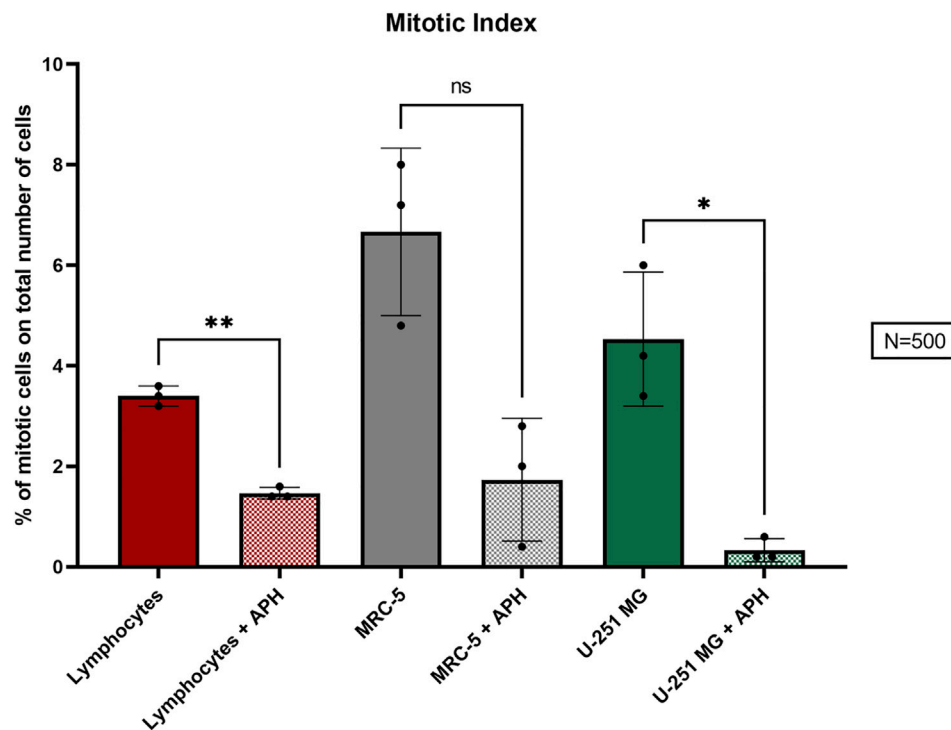


FIGURE 1 | Percentage of the M.I. (mitotic index) in the three cell types, lymphocytes (red), MRC-5 fibroblasts (gray), and U-251 MG GBM cells (green) in control and APH-treated conditions. The color coding of the three cell types is maintained throughout the figures. The error bars represent the standard deviation (SD) determined from 3 independent experiments ($N = 500$ cells for each replicate). Paired t -test was used to calculate the p values, where $p > 0.05$ ns, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, and $****p < 0.0001$.

The effect of APH in mitotic cells was evaluated on 100 metaphases, where we scored chromosome aberrations (Figure 2A) such as biradials, double minutes (DMs), fragments, extra-chromatin, fragile chromatin, and dicentric chromosomes. As expected, we found several of these phenotypes to be present in untreated glioblastoma cells that were absent in lymphocytes. Interestingly, APH treatment caused an increase in the total amount of chromosome aberrations only in glioblastoma but not in lymphocytes, including biradials, fragile chromatin, and dicentric chromosomes (Figure 2B, graph). Conversely, some aberrations decreased upon APH treatment, such as DMs, DNA fragments, and extra chromatin, suggesting potential activation of checkpoints or repair upon replication stress under these conditions in U-251 MG glioblastoma cells.

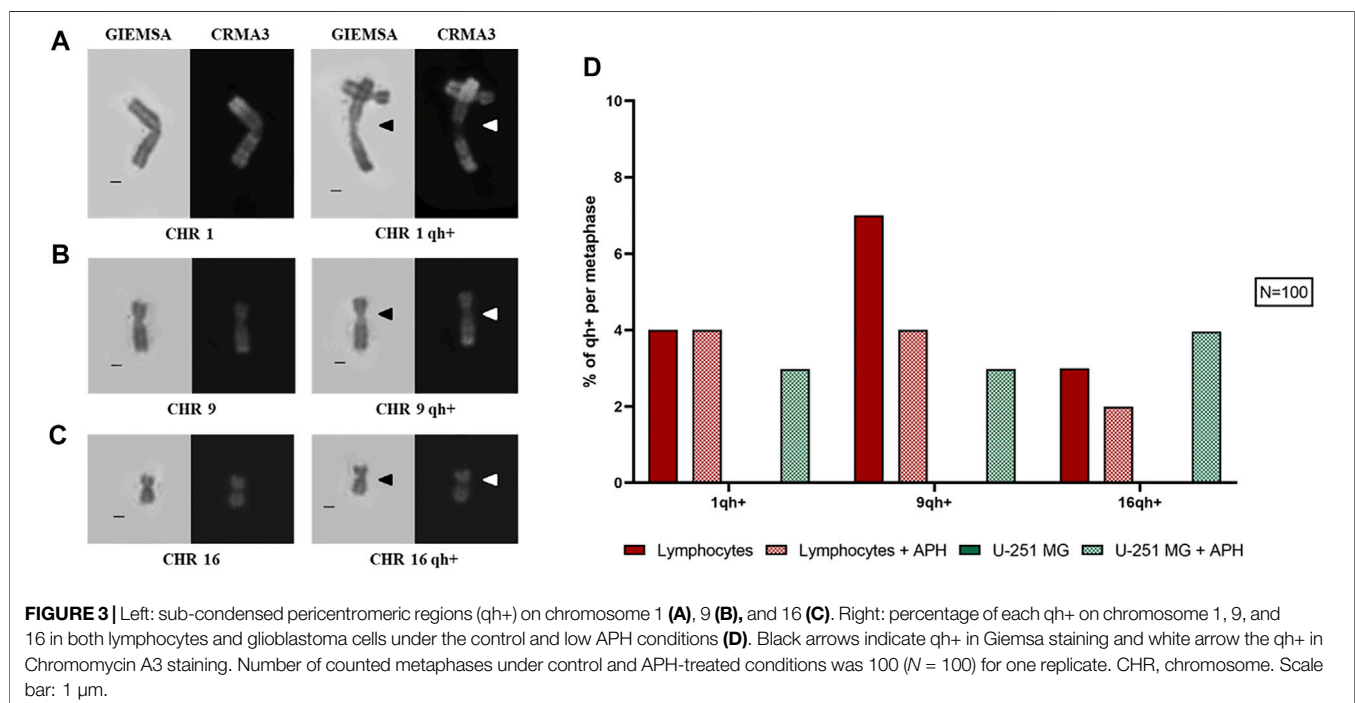
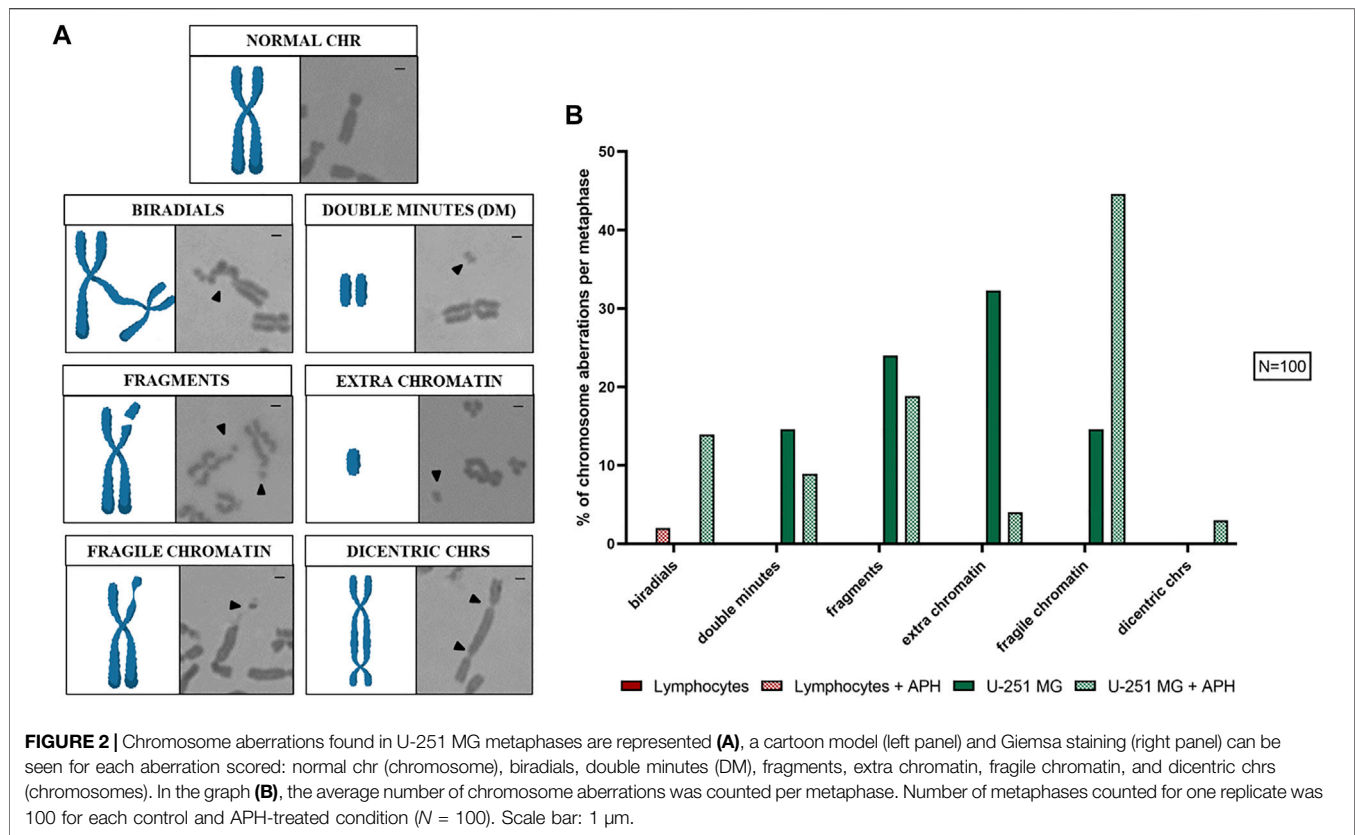
On 100 metaphase spreads, we also analyzed the pericentromeric heterochromatin for chromosomes 1, 9, and 16 being present as a morphological variant in a sub-condensed state named “qh+,” visible similarly to a secondary constriction (Figures 3A–C). In the human karyotype, these are the only chromosomes, including the Y chromosome, that exhibit this structural peculiarity (Purandare et al., 2011; Sipek Jr et al., 2014). In the case of glioblastoma cells, under-condensed pericentromeric chromatin was highly increased only upon treatment with APH. In lymphocytes, instead, we observed a

similar qh+ expression under both untreated and APH conditions, except for chromosome 9 (Figure 3D, graph).

Altogether, our data show that GBM cells are highly affected by replication stressors even low-dose APH, leading to a decrease in overall number of mitoses and specific fragile chromosome phenotypes implying chromatin fragility.

Glioblastoma-Specific Expression of Common Fragile Sites After Replication Stress

Next, we investigated the incidence of DNA gaps/breaks under normal and replication stress conditions, and we compared the occurrence of common fragile sites. We used two stains, Giemsa and CRMA3 (Chromomycin A3), on the same metaphase to recognize both expression of the site and the specific cytogenetic band involved in the gap/break (Figure 4A). Notably, gaps and breaks were mainly detected after APH treatment within fragile sites in both U-251 MG cells and in lymphocytes (Figure 4B, graph). Cytogenetic observation of over 100 metaphases enabled us to detect and map gaps/breaks to CFSs on glioblastoma metaphase chromosomes. We scored all known human CFSs across primary cells (lymphocytes and MRC-5 fibroblasts) and GBM cells (Figure 5F, graph). We found 52 CFSs expressed as gaps/breaks in GBM cells; among them, only 17 CFSs showed fragility with a



frequency equal to at least 1% on the total of gaps/breaks upon APH treatment. Importantly, two of these CFSs localized to regions FRA2E (2p13/p12) and FRA2F (2q22) appeared to be

glioblastoma-specific since the breaks/gaps induced by APH were only seen in U-251 MG cells (Figure 5B) and were not found in lymphocytes or fibroblasts under these experimental conditions.

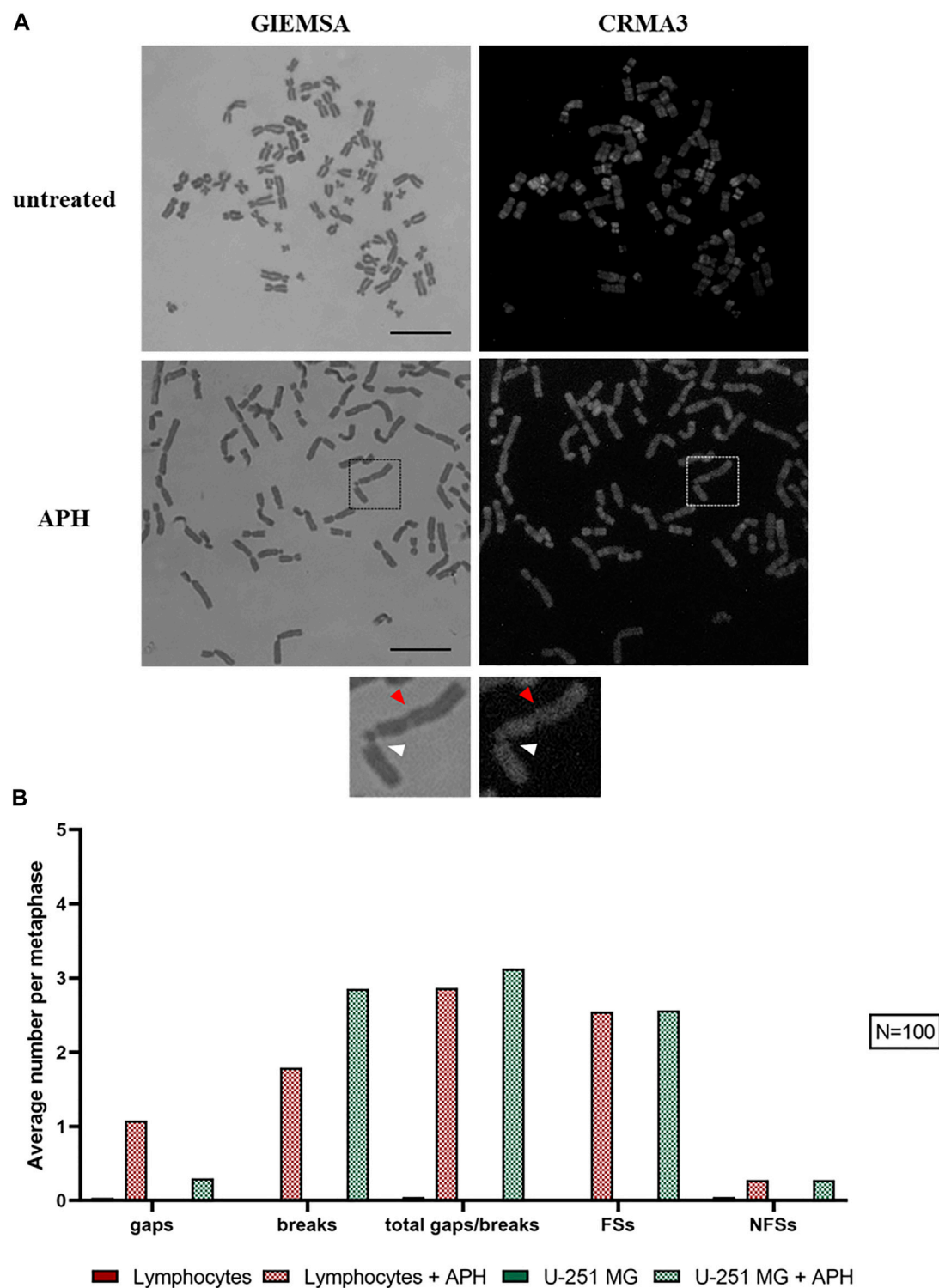
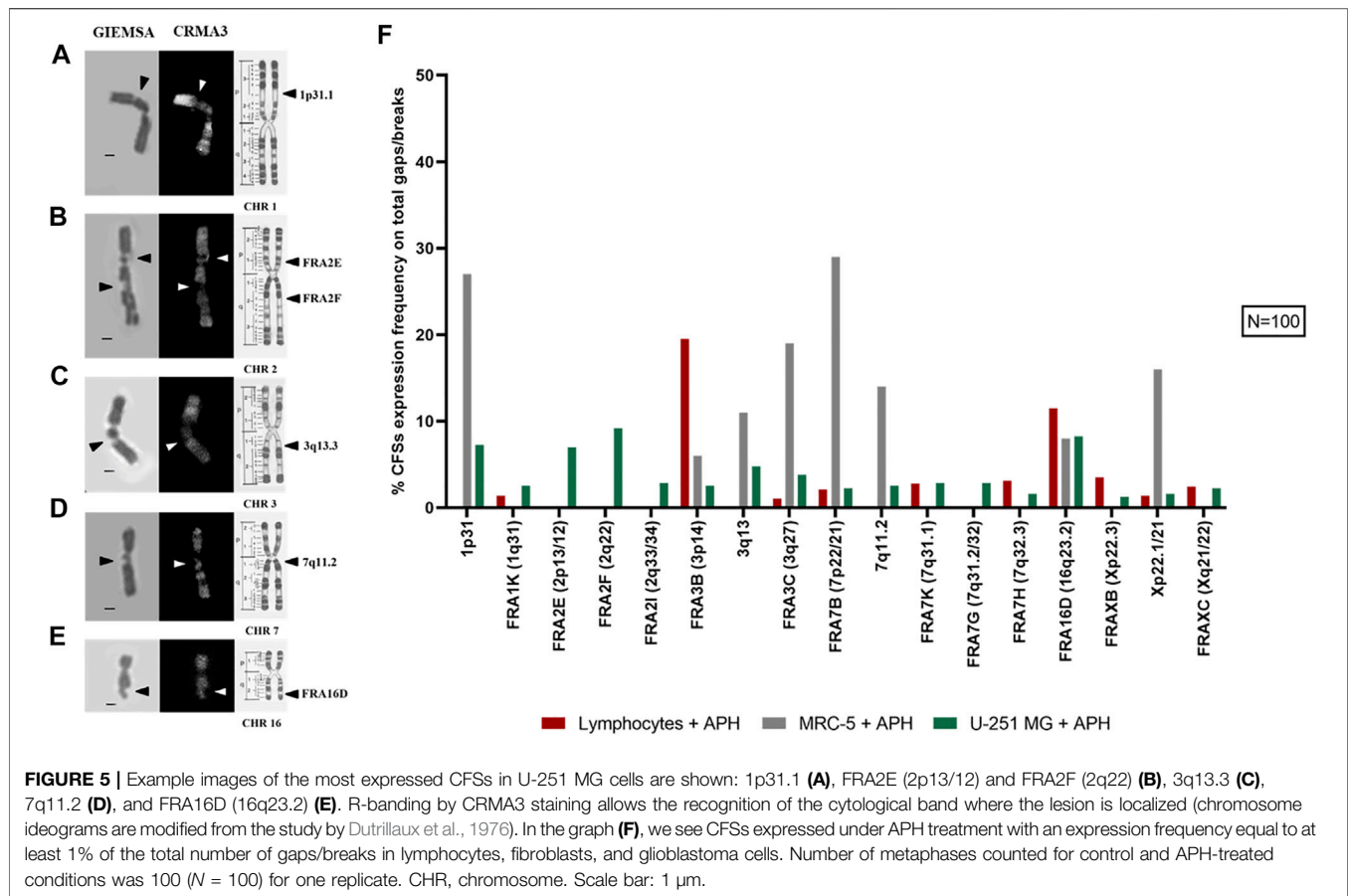


FIGURE 4 | U-251 MG metaphases stained with Giemsa and CRMA3 in control and APH-treated conditions; the brackets highlight chromosome 2, where white arrows indicate the break within FRA2E (2p12/13) and red arrows the gap within FRA2F (2q22) (**A**). The graph shows the average number of gaps/breaks per metaphase under both conditions and how many of these lesions were fragile sites (FSs) or non-fragile sites (NFSs) (**B**). Number of counted metaphases under control and APH-treated conditions was 100 ($N = 100$) for one replicate. Scale bar: 10 μ m.



In line with previous evidence, our data show the striking heterogeneity in expressing specific CFSs at different frequencies across the three cell types. For all cells, the fragile sites did not occur without replicative stress. After APH is supplemented to the cell culture medium, high expression of specific fragile sites such as FRA2E and FRA2F (Figure 5B) were found in the glioblastoma cell line.

The effect of APH in promoting expression varies in frequency between CFSs observed in lymphocytes and in U-251 MG cells. For instance, the FRA3B site in lymphocytes has an expression greater than 19%, while in GBM cells it is less than 3% (Figure 5F, graph). Conversely, glioblastoma-specific breaks localized to regions FRA2E (2p13/p12) and FRA2F (2q22) were only seen in U-251 MG cells and were not observed in either lymphocytes or fibroblasts.

FRA16D (Figure 5E) appears as a CFS in all three cell types, potentially underscoring a different mechanism of fragility of this region that is not tissue-specific. We also identified several characteristic fibroblasts sites, such as 1p31.1 (Figure 5A), 3q13.3 (Figure 5C), and 7q11.2 (Figure 5D) as previously shown (Murano et al., 1989; Le Tallec et al., 2011; Maccaroni et al., 2020), which were also expressed as gaps and breaks in GBM (Figure 5F, graph).

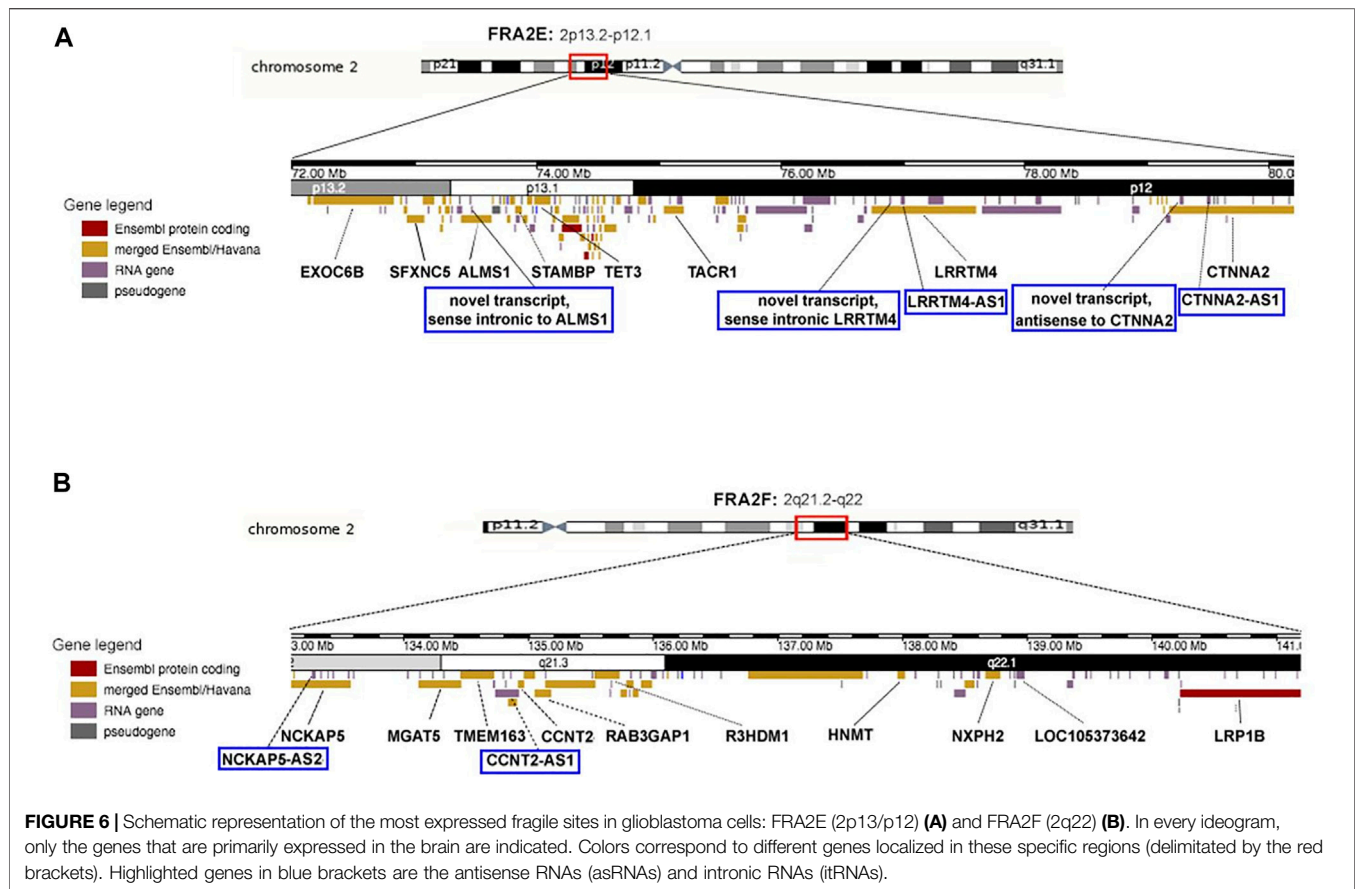
Altogether, our data show that CFSs have a very different expression frequency in different cell types underscoring the tissue-specific expression of CFSs (Le Tallec et al., 2011;

Maccaroni et al., 2020) and pointing to potential glioblastoma-specific vulnerability upon replication stress within CFSs FRA2E (2p13/p12) and FRA2F (2q22) seen in U-251 MG cells.

Given the specificity of the CFS FRA2E and FRA2F expression in GBM, we analyzed the sequence composition and the presence of transcripts of these regions. A detailed characterization was obtained by using different databases (NCBI, Ensembl, and GeneCards) (Supplementary Table S1). We found evidence of long genes that are frequently expressed in brain tissue (for FRA2E and FRA2F see Figure 6; for 1p31.1, 3q13.3, and 7q11.2, see Supplementary Figure S2). Interestingly, in addition to several long transcriptionally active genes in the brain, we also observed many antisense and intronic RNA transcripts that we hypothesize may also contribute to promoting glioblastoma-specific chromosomal fragility.

CFSs Replication Timing Analysis

We wondered whether replication stress affected the dynamics of replication during the S-phase within fragile regions expressed in U-251 MG cells. To this end, we analyzed the replication timing of three CFSs (1p31.1, 3q13.3, and 7q11.2) previously studied in MRC-5 fibroblasts (Maccaroni et al., 2020) by using a specific probe for each fragile region to compare their behavior upon mild replicative stress. These replication data were also compared with the replication analysis in lymphocytes. More than 100 nuclei for each FISH probe were observed both in the absence and the



presence of APH. The spots indicate the replication status of the CFS: one spot for unreplicated allele (**Figure 7D**, yellow arrows) and double spots for replicated alleles (**Figure 7D**, white arrows). Immunofluorescence (IF) against BrdU on the same nuclei revealed each specific substage of the S-phase (early, mid, and late; **Figure 7D**). The combined FISH-IF allowed us to monitor the CFS replication status throughout the S-phase.

In the control condition, the site 1p31.1 showed only 50% of replicated alleles in MRC-5 fibroblasts in the late S-phase; under APH treatment, we observed an increase in the total amount of replicated alleles (until ~75%) in the late S-phase, as seen in APH-treated U-251 MG cells and lymphocytes (**Figure 7A**, graphs).

For 3q13.3, we did not observe any difference between untreated and treated cells in spite of being a highly expressed CFS in GBM (**Figure 5F**, graph). In U-251 MG cells and lymphocytes under APH stress, we observed a ~5% reduction of replicated alleles compared to the control (**Figure 7B**, graphs).

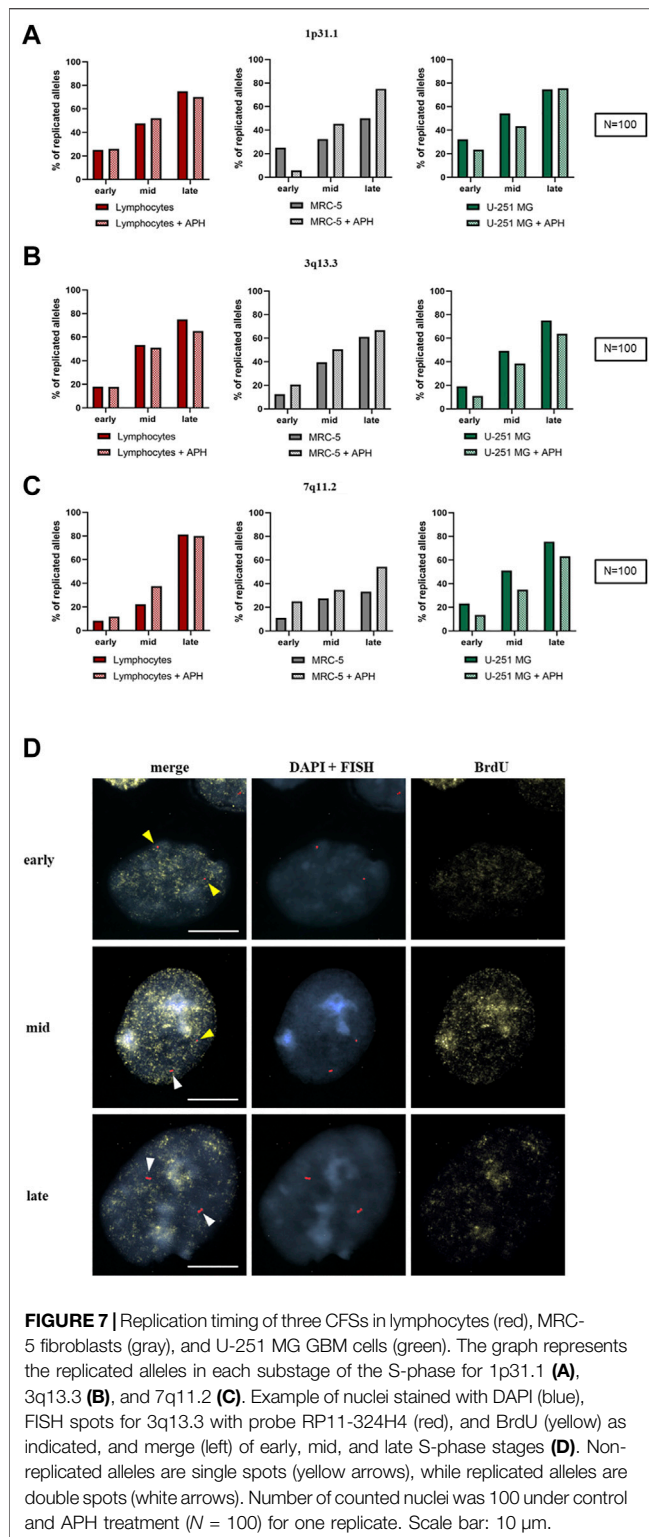
Regarding the fragile site 7q11.2, in lymphocytes, it reached the same percentages of replicated alleles in the late S-phase under both normal and replicative stress conditions. Under APH, U-251 MG cells showed a slight decrease in replicated alleles, while MRC-5 fibroblasts showed a ~10% increase in replicated alleles before the end of the S-phase (**Figure 7C**, graphs). However, for all the analyzed samples, we did not detect complete replications for these alleles, highlighting a possible delay in replication after the late S-phase/G2. Collectively, our

data show no significant difference in replication dynamics within these regions, whether they were non-fragile (lymphocytes) or expressed as CFSs at different frequencies (U-251 MG and MRC-5 fibroblasts).

Replication in Interphase and Mitotic Cells

Our data suggest that upon exposure to a low dose of replication stress, GBM cells suffer insults that result in a lower level of mitotic cells (**Figure 1**) and chromosome instability (**Figure 2** and **Figure 4**), including expression of specific CFSs (**Figure 5**). Some of these CFSs fail to duplicate within the timeframe of the S-phase (**Figure 7**), potentially entering mitosis unreplicated. To better understand the nature of these CFSs' expression, we assessed the cell cycle distribution using 5-Bromo-2'-deoxyuridine (BrdU) incorporation under untreated and APH conditions in U-251 MG cells. We scored either replicating cells (BrdU-positive nuclei, white arrows in **Figure 8A** and **Supplementary Figure S3**) representing different stages of the S-phase (as in **Figure 7D**) or non-replicating cells (BrdU-negative, red arrows in **Figure 8A**), likely representing G1 or G2.

We found a shift of glioblastoma cells exposed to replication stress from G1/G2 BrdU negative to a replicative state with a significant increase in BrdU-positive cells in APH-treated cells (**Figure 8B**, graph). The higher proportion of cells residing in the S-phase may imply the replication struggle induced by APH and activation of the intra-S-phase checkpoint and justify the



decreased number of cells that make it into mitosis (Figure 1). We hypothesized that unfinished replication may cause MiDAS in U-251 MG cells, a phenomenon where DNA synthesis continues into mitosis. Thus, we assessed persistent BrdU incorporation on metaphase spreads (Supplementary Figure

S4). Indeed, we observed the presence of BrdU signals on 1–4 chromosomes per metaphase spread only under the APH condition (Supplementary Figure S4), indicating that defective replication in U-251 MG cells is not dealt with in G2 but persists into mitosis, likely contributing to the chromosome defects as observed in Figures 2 and Figure 3 and resulting in gaps and breaks at specific CFSs (Figures 4 and Figure 5), given the small number of BrdU-positive loci seen in U-251 MG cells only after APH treatment.

Collectively, our data suggest that these chromosomal regions show glioblastoma-specific CFSs likely due to multiple converging features including presence of long genes actively transcribed in the tissue of origin, slower/impaired replication, and evidence of MiDAS to attempt completing DNA synthesis at these regions ahead of chromosome segregation in mitosis (Figure 9).

DISCUSSION

Our work presents a characterization of CFS expression in a cell line derived from the malignant brain tumor GBM. Here, we assessed the expression of every CFS found in the human genome in U-251 MG cells under conditions of mild replicative stress using APH. Out of all known CFSs, we identified 52 CFSs expressed as gaps/breaks only after APH treatment in GBM cells and 17 CFSs that showed fragility with frequency equal to at least 1% on the total gaps/breaks. We also identified two CFSs that appear to be glioblastoma-specific localized to regions FRA2E (2p13/p12) and FRA2F (2q22), only seen as breaks in U-251 MG cells after APH treatment. Within the CFSs highly expressed in GBM after APH treatment, we observed presence of long genes, incomplete replication, and delayed DNA synthesis that persisted into mitosis (MiDAS). Given our data showing similar replication dynamics during the S-phase for untreated U-251 MG or treated with APH, we suggest that the replication issues at these sites partly activate a replication checkpoint as we observed upon scoring BrdU-positive and -negative cells. Likely, however, several cells continue into mitosis with under-replicated DNA. Presence of long genes, especially expressed in brain tissue, may further enhance the fragility of a specific region that manifests as gaps or breaks within mitotic chromosomes. We also noted fragile and uncondensed chromatin and other chromosome phenotypes, indicating that issues generated by replication stressors such as a low dose of APH result in a variety of phenotypes, including but not limited to fragile site expression that can readily compromise overall genome stability.

In cancer, DNA synthesis may be compromised by the lack of basic replication-components, leading to mitotic arrest (Miron et al., 2015). After APH treatment, we observed a specific pattern of CFSs in glioblastoma cells that is not detectable in lymphocytes and fibroblasts. This different response of the three cell types can be explained by the fact that cancer cells as glioblastoma could bypass DNA damage and the cell cycle checkpoint, including intra-S-phase ones, and proceed through the cell cycle despite the persistent damage and/or unfinished replication within the

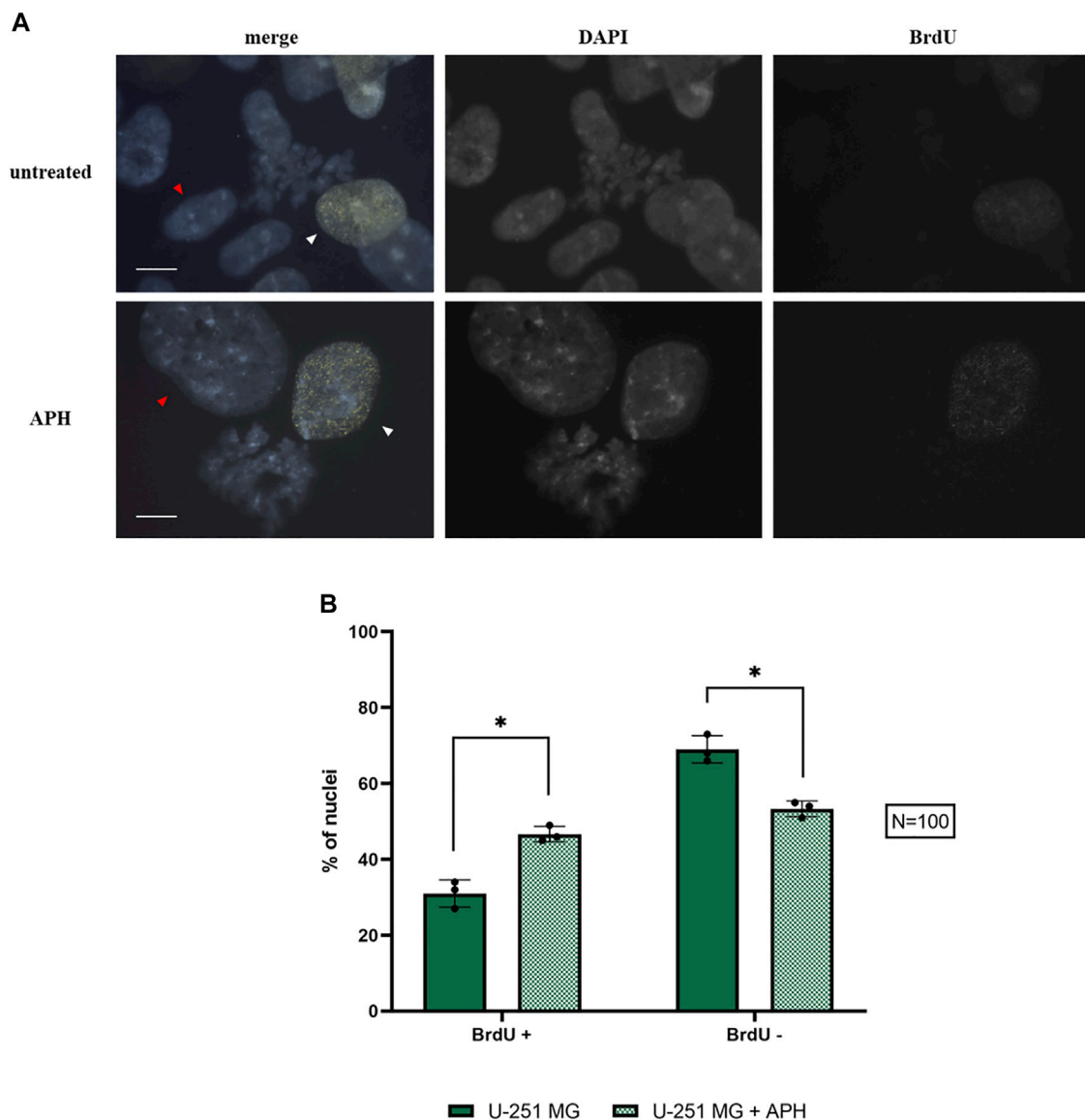


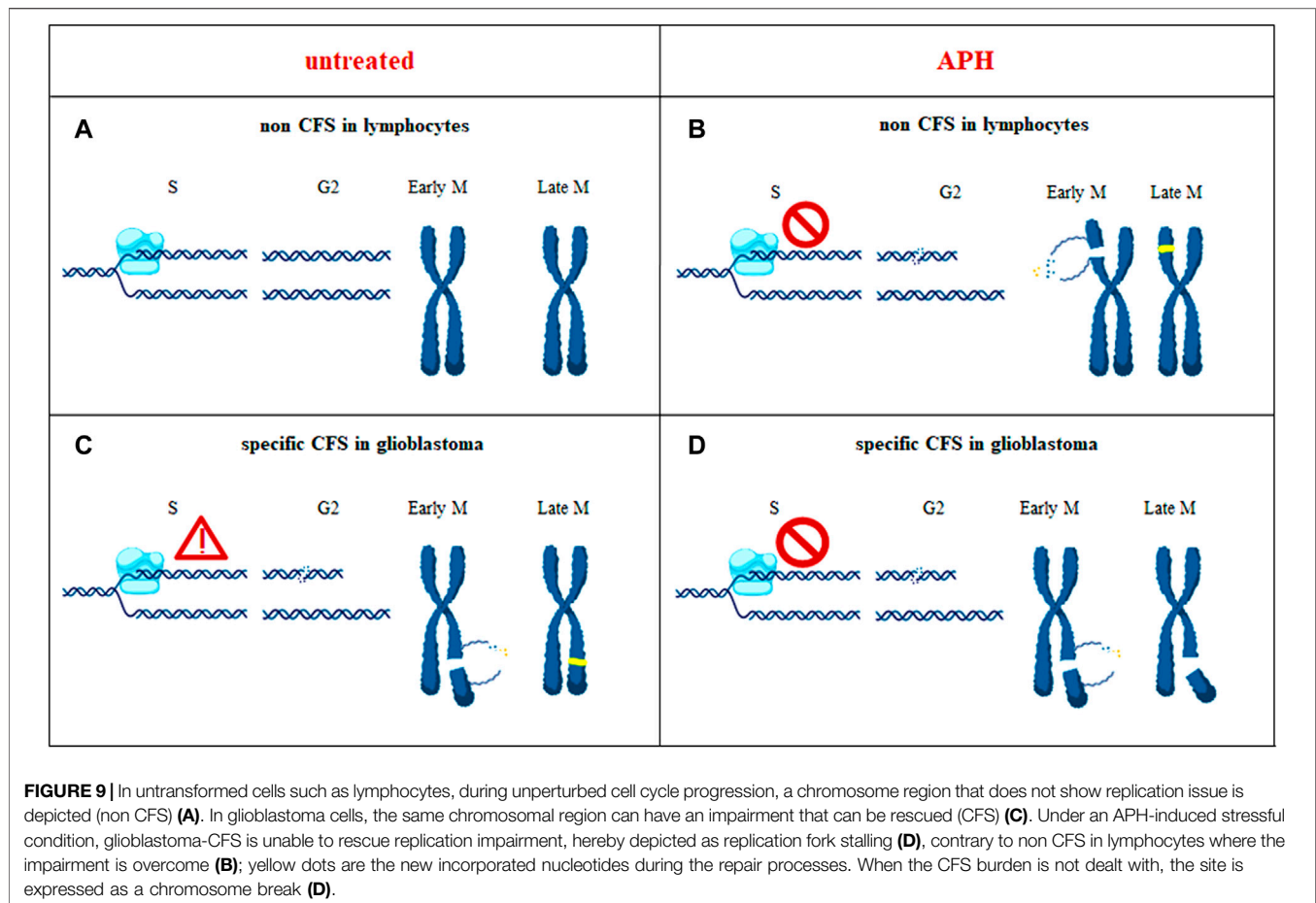
FIGURE 8 | Replicating nuclei under control and APH-treated conditions with BrdU foci (yellow) and DAPI staining (blue) (**A**); the graph indicates the amount of BrdU-negative (non-replicating cells, red arrows) and BrdU-positive (replicating cells, white arrows) glioblastoma cells under control and APH-treated conditions (**B**). Scale bar: 10 μ m. The error bars represent standard deviation (SD) determined from 3 independent experiments ($N = 100$ nuclei for each replicate). Paired t -test was used to calculate the p values, where $p > 0.05$ ns, $*p \leq 0.05$, $**p < 0.01$, $***p < 0.001$, and $****p < 0.0001$.

timeframe of the S-phase or G2, with the subsequent effect of aberrant mitoses and the potential for rearrangements and accumulation of chromosome aberrations.

Indeed, it can be noted that glioblastoma cells express several more CFSs than lymphocytes, implying that CFSs in primary cells such as lymphocytes may express breaks in specific regions after mild replication stress due to intrinsic and tissue-specific features; in glioblastoma, on the other hand, to these intrinsic features get added additional factors likely related to malignant transformation and decreased proficiency of checkpoints and DDR. This is in line with the notion that

glioblastoma often displays high CIN. Some of the CFSs we have identified have been specifically characterized in fibroblasts (i.e., 1p31, 3q13, and 7q11; refer to Murano et al., 1989; Le Tallec et al., 2011; Maccaroni et al., 2020), suggesting that the chromosome fragility phenotype of GBM appears under our experimental conditions to be more similar to fibroblasts than lymphocytes.

Our work offers an initial overview of common fragile site expression in glioblastoma that may be further exploited for future cytogenetic, molecular, and clinical studies to advance our understanding of this incurable cancer.



DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**; further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

EB, FP, and SG contributed to data conception and interpretation and manuscript writing; all authors contributed to data analysis; EB and EDT contributed to cell culture; EB performed the cytogenetic observations and FISH and IF experiments and redefined all the figures and graphs; EDT contributed to the counting of the mitotic index and nuclei aberrations and also to the figure and graph design and **Figure 9**; AA contributed the cell line and experimental advice; FP contributed to the conceptualization, data assembly, recognition of CFSs, and the design of all figures and **Supplementary Table S1**; SG contributed to the conceptual idea, validation of formal analysis, project supervision, and manuscript writing and editing.

FUNDING

Research was supported by a grant by La Sapienza University grants 2020/2021 to FP, an AIRC grant (AIRC Start-Up 2020 25189) to SG and Rita Levi Montalcini Award by the Italian Ministry of University and Research to SG.

ACKNOWLEDGMENTS

The authors are grateful to all the members of the Giunta and Pelliccia Labs for useful discussions and advice, especially Edna Patricia Cigarroa for her unwavering support during this work. We thank Evelyne Tassone for critical reading of the manuscript and Luca Corda for advice with the figures' layout. EB was the recipient of a PhD fellowship from the Sapienza University of Rome.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.810793/full#supplementary-material>

REFERENCES

- Arlt, M. F., Casper, A. M., and Glover, T. W. (2003). Common Fragile Sites. *Cytogenet. Genome Res.* 100, 92–100. doi:10.1159/000072843
- Balzano, E., and Giunta, S. (2020). Centromeres under Pressure: Evolutionary Innovation in Conflict with Conserved Function. *Genes (Basel)* 11. doi:10.3390/genes11080912
- Balzano, E., Pelliccia, F., and Giunta, S. (2021). Genome (In)stability at Tandem Repeats. *Semin. Cel Dev. Biol.* 113. doi:10.1016/j.semcdb.2020.10.003
- Berardinelli, F., Tanori, M., Muoio, D., Buccarelli, M., di Masi, A., Leone, S., et al. (2018). G-quadruplex Ligand RHPS4 Radiosensitizes Glioblastoma Xenograft *In Vivo* through a Differential Targeting of Bulky Differentiated- and Stem-Cancer Cells. *J. Exp. Clin. Cancer Res.* 38, 311. doi:10.1186/s13046-019-1293-x
- Bosco, N., Pelliccia, F., and Rocchi, A. (2010). Characterization of FRA7B, a Human Common Fragile Site Mapped at the 7p Chromosome Terminal Region. *Cancer Genet. Cytogenet.* 202, 47–52. doi:10.1016/j.cancergencyto.2010.06.008
- Boteva, L., Nozawa, R. S., Naughton, C., Samejima, K., Earnshaw, W. C., and Gilbert, N. (2020). Common Fragile Sites Are Characterized by Faulty Condensin Loading after Replication Stress. *Cell Rep* 32, 108177. doi:10.1016/j.celrep.2020.108177
- Brisson, O., El-Hilali, S., Azar, D., Koundrioukoff, S., Schmidt, M., Nähse, V., et al. (2019). Transcription-mediated Organization of the Replication Initiation Program across Large Genes Sets Common Fragile Sites Genome-wide. *Nat. Commun.* 10, 5693. doi:10.1038/s41467-019-13674-5
- Casper, A. M., Nghiem, P., Arlt, M. F., and Glover, T. W. (2002). ATR Regulates Fragile Site Stability. *Cell* 111, 779–789. doi:10.1016/s0092-8674(02)01113-3
- Debatisse, M., Letessier, B., Dutrillaux, A., Brisson, B., and au, O. (2012). Common Fragile Sites: Mechanisms of Instability Revisited. *Trends Genet.* 28, 22–32. doi:10.1016/j.tig.2011.10.003
- Dutrillaux, B., Couturier, J., Richer, C. L., and Viegas-Péquignot, E. (1976). Sequence of DNA Replication in 277 R- and Q-Bands of Human Chromosomes Using a BrdU Treatment. *Chromosoma* 58, 51–61. doi:10.1007/BF00293440
- Ensembl (2021). Available at: http://www.ensembl.org/Homo_sapiens/Location/Genome (GRCh38p13 May, 2021).
- Feng, W., and Chakraborty, A. (2018). Fragility Extraordinaire: Unsolved Mysteries of Chromosome Fragile Sites. *Adv. Exp. Med. Biol.*, 489–526. doi:10.1007/978-981-10-6955-0_21
- GeneCards (2021). The Human Gene Database. Available at: <https://www.genecards.org/> (Release version 56October 27, 2021).
- Giunta, S. (2018). Centromere Chromosome Orientation Fluorescent *In Situ* Hybridization (Cen-CO-FISH) Detects Sister Chromatid Exchange at the Centromere in Human Cells. *Bio Protoc.* 8, e2792. doi:10.21769/BioProtoc.2792
- Giunta, S., and Funabiki, H. (2017). Integrity of the Human Centromere DNA Repeats Is Protected by CENP-A, CENP-C, and CENP-T. *Proc. Natl. Acad. Sci. U S A.* 114, 1928–1933. doi:10.1073/pnas.1615133114
- Giunta, S., Hervé, S., White, R. R., Wilhelm, T., Dumont, M., Scelfo, A., et al. (2021). CENP-A Chromatin Prevents Replication Stress at Centromeres to Avoid Structural Aneuploidy. *Proc. Natl. Acad. Sci.* 118. doi:10.1073/pnas.2015634118
- Helmrich, A., Ballarino, M., and Tora, L. (2011). Collisions between Replication and Transcription Complexes Cause Common Fragile Site Instability at the Longest Human Genes. *Mol. Cell* 44, 966–977. doi:10.1016/j.molcel.2011.10.013
- Ji, F., Liao, H., Pan, S., Ouyang, L., Jia, F., Fu, Z., et al. (2020). Genome-wide High-Resolution Mapping of Mitotic DNA Synthesis Sites and Common Fragile Sites by Direct Sequencing. *Cell Res* 30, 1009–1023. doi:10.1038/s41422-020-0357-y
- LeTallec, B., Armel Millot, G., Esther Blin, M., Brisson, O., Dutrillaux, B., and Debatisse, M. (2013). Common Fragile Site Profiling in Epithelial and Erythroid Cells Reveals that Most Recurrent Cancer Deletions Lie in Fragile Sites Hosting Large Genes. *Cel Rep* 4, 420–428. doi:10.1016/j.celrep.2013.07.003
- LeTallec, B., Dutrillaux, B., Lachages, A.-M., Armel Millot, G., Brisson, O., and Debatisse, M. (2011). Molecular Profiling of Common Fragile Sites in Human Fibroblasts. *Nat. Struct. Mol. Biol.* 18, 1421–1423. doi:10.1038/nsmb.2155
- Maccaroni, K., Balzano, E., Mirimao, F., Giunta, S., and Pelliccia, F. (2020). Impaired Replication Timing Promotes Tissue-specific Expression of Common Fragile Sites. *Genes (Basel)* 11. doi:10.3390/genes11030326
- Macheret, M., Bhowmick, R., Sobkowiak, K., Padayachy, L., Mailler, J., Hickson, I. D., et al. (2020). High-resolution Mapping of Mitotic DNA Synthesis Regions and Common Fragile Sites in the Human Genome through Direct Sequencing. *Cel Res* 30, 997–1008. doi:10.1038/s41422-020-0358-x
- Miotto, B., Ji, Z., and Struhl, K. (2016). Selectivity of ORC Binding Sites and the Relation to Replication Timing, Fragile Sites, and Deletions in Cancers. *Proc. Natl. Acad. Sci. U S A.* 113, E4810–E4819. doi:10.1073/pnas.1609060113
- Miron, K., Golan-Lev, T., Dvir, R., Ben-David, E., and Kerem, B. (2015). Oncogenes Create a Unique Landscape of Fragile Sites. *Nat. Commun.* 6, 7094. doi:10.1038/ncomms8094
- Murano, I., Kuwano, A., and Kajii, T. (1989). Fibroblast-specific Common Fragile Sites Induced by Aphidicolin. *Hum. Genet.* 83, 45–48. doi:10.1007/BF00274145
- NCBI GenBank (2021). Available at: <https://www.ncbi.nlm.nih.gov/genbank/> (Release Data April 26, 2021).
- NCBI Genome Data Viewer (2021). Available at: <https://www.ncbi.nlm.nih.gov/genome/gdv/> (Release Data May 16, 2021).
- OECD Test No (2016). Section 4 *In Vitro Mammalian Chromosomal Aberration Test, OECD Guidelines for the Testing of Chemicals* (Paris: OECD Publishing), 473.
- Purandare, H., Fernandes, N., Deshmukh, S., and Chavan, S. (2011). Heterochromatic Variations and Pregnancy Losses in Humans. *Int. J. Hum. Genet.* 11, 167–175. doi:10.1080/09723757.2011.11886139
- Sarni, D., and Kerem, B. (2016). The Complex Nature of Fragile Site Plasticity and its Importance in Cancer. *Curr. Opin. Cel Biol* 40, 131–136. doi:10.1016/j.jceb.2016.03.017
- Sarni, D., Sasaki, T., Irony Tur-Sinai, M., Miron, K., Rivera-Mulia, J. C., Magnuson, B., et al. (2020). 3D Genome Organization Contributes to Genome Instability at Fragile Sites. *Nat. Commun.* 11, 3613. doi:10.1038/s41467-020-17448-2
- Sipek, A., Jr, Mihalová, R., Panczak, A., Hrčková, L., Janashia, M., Kaspříková, N., et al. (2014). Heterochromatin Variants in Human Karyotypes: a Possible Association with Reproductive Failure. *Reprod. Biomed. Online* 29 (2), 245–250. doi:10.1016/j.rbmo.2014.04.021
- Smith, D. I., McAvoy, S., Zhu, Y., and Perez, D. S. (2007). Large Common Fragile Site Genes and Cancer. *Semin. Cancer Biol.* 17, 31–41. doi:10.1016/j.semcancer.2006.10.003
- Smith, D. I., Zhu, Y., McAvoy, S., and Kuhn, R. (2006). Common Fragile Sites, Extremely Large Genes, Neural Development and Cancer. *Cancer Lett.* 232, 48–57. doi:10.1016/j.canlet.2005.06.049

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Balzano, Di Tommaso, Antoccia, Pelliccia and Giunta. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



OPEN ACCESS

EDITED BY

Qing Hu,
University of Texas Southwestern
Medical Center, United States

REVIEWED BY

Wai Kit Chu,
The Chinese University of Hong Kong,
China
Philipp Oberdoerffer,
Johns Hopkins University, United States
Huiming Lu,
University of Texas Southwestern
Medical Center, United States

*CORRESPONDENCE

Wenyi Feng,
fengw@upstate.edu

†PRESENT ADDRESS

Arijita Chakraborty,
Tessera Therapeutics, Somerville,
Massachusetts, MN, United States

SPECIALTY SECTION

This article was submitted to Human
and Medical Genomics,
a section of the journal
Frontiers in Genetics

RECEIVED 29 March 2022

ACCEPTED 07 November 2022

PUBLISHED 24 November 2022

CITATION

Kodali S, Meyer-Nava S, Landry S,
Chakraborty A, Rivera-Mulia JC and
Feng W (2022), Epigenomic signatures
associated with spontaneous and
replication stress-induced DNA double
strand breaks.
Front. Genet. 13:907547.
doi: 10.3389/fgene.2022.907547

COPYRIGHT

© 2022 Kodali, Meyer-Nava, Landry,
Chakraborty, Rivera-Mulia and Feng.
This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License](#)
(CC BY). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Epigenomic signatures associated with spontaneous and replication stress-induced DNA double strand breaks

Sravan Kodali¹, Silvia Meyer-Nava², Stephen Landry¹,
Arijita Chakraborty^{1†}, Juan Carlos Rivera-Mulia² and
Wenyi Feng^{1*}

¹Department of Biochemistry and Molecular Biology, Upstate Medical University, Syracuse, NY, United States, ²Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, MN, United States

Common fragile sites (CFSs) are specific regions of all individuals' genome that are predisposed to DNA double strand breaks (DSBs) and undergo subsequent rearrangements. CFS formation can be induced *in vitro* by mild level of DNA replication stress, such as DNA polymerase inhibition or nucleotide pool disturbance. The mechanisms of CFS formation have been linked to DNA replication timing control, transcription activities, as well as chromatin organization. However, it is unclear what specific cis- or trans-factors regulate the interplay between replication and transcription that determine CFS formation. We recently reported genome-wide mapping of DNA DSBs under replication stress induced by aphidicolin in human lymphoblastoids for the first time. Here, we systematically compared these DSBs with regards to nearby epigenomic features mapped in the same cell line from published studies. We demonstrate that aphidicolin-induced DSBs are strongly correlated with histone 3 lysine 36 trimethylation, a marker for active transcription. We further demonstrate that this DSB signature is a composite effect by the dual treatment of aphidicolin and its solvent, dimethylsulfoxide, the latter of which potentially induces transcription on its own. We also present complementing evidence for the association between DSBs and 3D chromosome architectural domains with high density gene cluster and active transcription. Additionally, we show that while DSBs were detected at all but one of the fourteen finely mapped CFSs, they were not enriched in the CFS core sequences and rather demarcated the CFS core region. Related to this point, DSB density was not higher in large genes of greater than 300 kb, contrary to reported enrichment of CFS sites at these large genes. Finally, replication timing analyses demonstrate that the CFS core region contain initiation events, suggesting that altered replication dynamics are responsible for CFS formation in relatively higher level of replication stress.

KEYWORDS

DNA double strand break (DSB), common fragile site (CFS), DNA replication stress, histone H3K36 trimethylation, histone H3K27 trimethylation, CTCF, topologically associated domain (TAD)

Introduction

CFSs are genomic regions that are prone to DNA strand breakage, observable as gaps or other abnormalities on the metaphase chromosomes. The manifestation, or expression, of CFSs is induced by mild level of DNA replication stress such as DNA polymerase inhibition or nucleotide pool limitation, as reviewed in (Feng and Chakraborty 2017). There are two major mechanisms proposed to underlie CFS formation: defective DNA initiation/progression and replication-transcription conflict (Le Tallec et al., 2014; Ozeri-Galai et al., 2014; Sarni and Kerem 2016). These theories are predicated on the observations that 1) with noted exceptions (El Achkar et al., 2005; Barlow et al., 2013; Handt et al., 2014), CFSs are generally characterized by late replication timing in an unperturbed S phase and experience persistent delay under replication stress (Le Beau et al., 1998; Wang et al., 1999; Hellman et al., 2000; Palakodeti et al., 2004; Pelliccia et al., 2008; Handt et al., 2014); and 2) CFSs tend to nest in large transcribed genes. It is thought that the persistently under-replicated regions, presumably as a result of replication fork breakdown, become unstable and induce genomic rearrangements. Additionally, it is thought that transcription suppresses initiation of DNA replication within these genes, thus contributing to the persistent replication delay (Brisson et al., 2019). Finally, direct collisions between the replication and transcription machineries, particularly at sequence locations prone to form R-loops, is thought to cause DNA strand breakage at the CFSs (Helmrich et al., 2011).

CFSs are an intrinsic feature of the human genome and are hot spots for large scale amplification, deletion, and rearrangements, which are thought to underlie genome instability that are prevalent in cancer as well as neurological disorders (Arlt et al., 2003; Glover et al., 2005; Arlt et al., 2006; Ozeri-Galai et al., 2012; Alt et al., 2017; Alt and Schwer 2018; Palumbo and Russo 2019). Therefore, normal cells are arguably the most important model for CFS mapping in order to understand mechanisms of disease onset (Palumbo and Russo 2019). We recently applied the Break-seq method to map DSBs, both spontaneous or chemically induced, in the GM06990 cell line (Chakraborty et al., 2020). These endeavors led to the first high resolution map of DSBs under conditions used to induce CFSs in human lymphoblastoids. The salient points from this study are as follows. First, the vehicle control (DMSO) potently induce DSBs and APH further enhances DSBs; thus APH-induced DSB formation is necessarily a composite effect of the two chemicals. For simplicity we will refer to DSBs induced by both chemicals as APH-induced. Second, both DMSO- and APH-induced DSBs are predominantly located in late-replicating regions, consistent with the noted feature of CFSs. Third, while neither spontaneous nor APH-induced DSBs are enriched for R-loop forming sequences (RLFSs), the DMSO-induced DSBs are enriched for RLFSs, suggesting that transcription induction in the DMSO-treated cells played a major role in DSB formation. Fourth, these DSBs did not show significant correlation with the core sequences of 76 CFSs previously described in lymphoblasts (Le Tallec et al., 2013; Savelyeva and Brueckner 2014). Related to this final point, because CFS cores

have been shown to have strong association with large genes of greater than 300 kb (Smith et al., 2006; Le Tallec et al., 2013), our results would suggest that DSBs were not enriched in large genes. Therefore, in this study we investigated the association, or the lack thereof, between DSBs and large genes harboring CFS cores and asked what cis- or trans-factors determine DSB formation.

We systematically examined the relationship between DSBs and select key genomic features mapped by published studies, the majority of which were curated by the Encyclopedia of DNA Elements (ENCODE) project. We specifically focused on data sets generated from the same cell line (GM06990) as our data were, to minimize confounding genetic factors. These data sets included histone modification sites—specifically histone 3 lysine 4 trimethylation (H3K4me3), H3K36me3, and H3K27me3—DNaseI hypersensitive sites (DNaseI HSS), and topologically associated domain (TAD) architectural protein CTCF binding sites. Each of these elements has been implicated in replication and/or transcription regulation (Kimura 2013; Zhang et al., 2015). Of note, it has been suggested that spontaneous DSB sites are correlated with epigenetic markers for chromatin accessibility, including DNaseI HSSs, H3K4me3, and CTCF binding sites (Mourad et al., 2018). We have also found that APH-induced CFSs are associated with TAD boundaries enriched for CTCF binding sites (Sarni et al., 2020). Finally, CTCF binding sites have been shown to be susceptible to DSBs induced by a topoisomerase inhibitor (Canela et al., 2017). Therefore, these studies provided compelling evidence for a connection between DNA strand breakage and 3D genome organization. Additionally, in our previous study (Chakraborty et al., 2020) we analyzed our DSBs for replication timing using Repli-seq data (Hansen et al., 2010); here we also compared the DSB locations to origins of replication mapped by Bubble-seq (Mesner et al., 2013). Our analysis demonstrated a correlation between DMSO- and APH-induced DSBs with H3K36me3 at two locations: first at the TSS, where CTCF bindings sites are also enriched; second within gene bodies downstream from the TSS; origins of replication are broadly distributed at both locales. These observations are consistent with a model where replication stress-induced DSBs are correlated with active transcription and are enriched at TAD boundaries.

Materials and methods

Downloaded data sets

DNaseI HSS_1: bigWig

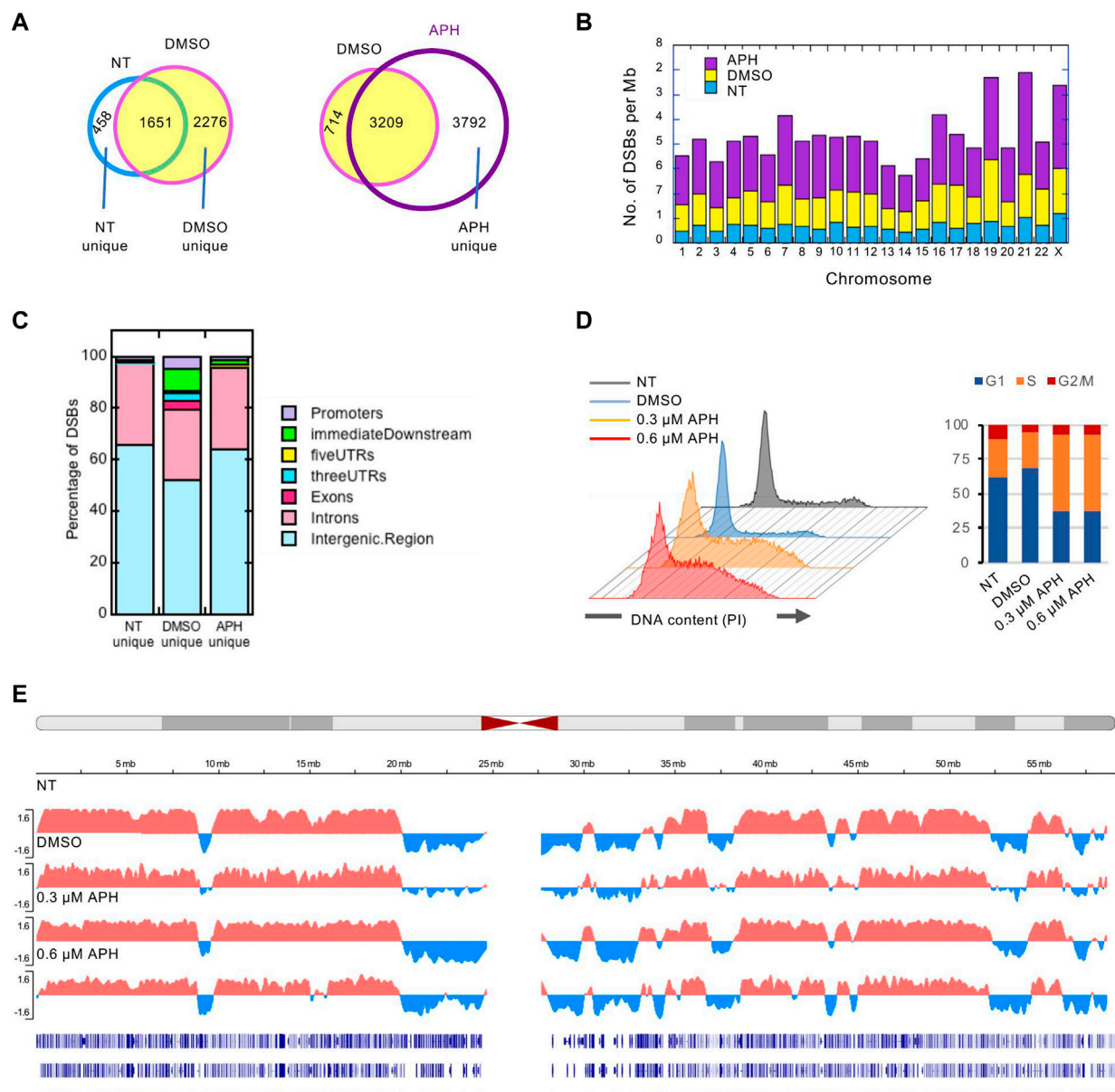
<https://www.encodeproject.org/files/ENCFF529JFV/@@download/ENCFF529JFV.bigWig>.DNaseIHSS2:bigWig

DNaseI HSS_1: bigWig

<https://www.encodeproject.org/files/ENCFF709PEX/@@download/ENCFF709PEX.bigWig>

DNaseIHSS_1: bigBed narrowPeak

<https://www.encodeproject.org/files/ENCFF735IVN/@@download/ENCFF735IVN.bigBed>

**FIGURE 1**

Break-seq mapping of DSBs in normal human lymphoblastoids. **(A)** Concordance between DSBs in pair-wise comparisons and subsetting DSBs unique to each condition. **(B)** DSB density (per Mb of DNA) across each chromosome. **(C)** Distribution of DSBs unique to each condition with respect to genes. **(D)** Cell cycle analysis of cells with the indicated treatment by flow cytometry. PI: propidium iodide. **(E)** Replication timing analysis under replication stress. Representative replication timing profile plotting the RT Log2 Ratio for chr19 is shown for each treatment. Bottom tracks represent RefSeq genes.

DNaseI HSS_2: bigBed narrowPeak

<https://www.encodeproject.org/files/ENCFF043DMN/@@download/ENCFF043DMN.bigBed>

CTCF ChIP-seq: bigWig

<https://www.encodeproject.org/files/ENCFF469OOI/@@download/ENCFF469OOI.bigWig>

CTCF ChIP-seq: bed narrowPeak

<https://www.encodeproject.org/files/ENCFF276JDQ/@@download/ENCFF276JDQ.bed.gz>

H3K4me3 ChIP-seq: bigWig

<https://www.encodeproject.org/files/ENCFF965GIX/@@download/ENCFF965GIX.bigWig>

H3K4me3 ChIP-seq: bed narrowPeak
<https://www.encodeproject.org/files/ENCFF357ALO/@@download/ENCFF357ALO.bed.gz>
 H3K27me3 ChIP-seq: bigWig
<https://www.encodeproject.org/files/ENCFF533NLA/@@download/ENCFF533NLA.bigWig>
 H3K27me3 ChIP-seq: bed narrowPeak
<https://www.encodeproject.org/files/ENCFF554UCC/@@download/ENCFF554UCC.bed.gz>
 H3K36me3 ChIP-seq: bigWig
<https://www.encodeproject.org/files/ENCFF324YFT/@@download/ENCFF324YFT.bigWig>
 H3K36me3 ChIP-seq: bed narrowPeak
<https://www.encodeproject.org/files/ENCFF372NOF/@@download/ENCFF372NOF.bed.gz>
 Bubble-seq origins of replication data
 (GSE38809_GM_combined_RD_bubbles.bedgraph) were
 downloaded from GEO accession number GSE38809.

MEME suite searches

The commands used for the following search engines are listed below.

AME: ame-verbose 1 -oc. -scoring avg-method fisher-hit-lo-fraction 0.25 -evaluate-report-threshold 10.0 -control-shuffle-kmer 2 DSBFILE.fasta db/HUMAN/HOCOMOCOv11_core_HUMAN_mono_meme_format.meme

STREME: streme-verbosity 1 -oc. -dna-totallength 4000000 -time 14400 -minw 8 -maxw 15 -thresh 0.05 -align center-p DSBFILE.fasta

Tomtom: tomtom -no-ssc -oc. -verbosity 1 -min-overlap 5 -mi 1 -dist pearson -evaluate -thresh 10.0 -time 300 query_motifs db/HUMAN/HOCOMOCOv11_core_HUMAN_mono_meme_format.meme

Random simulation test for correlation between DSBs and chromatin features

The association between a DSB and a chromatin feature was determined by whether the two regions overlap by at least 1 bp using intersectBed function in BEDTools. Random DSB sequences were generated by shuffleBed with the same number of DSBs per chromosome. The shuffled DSBs were then compared to the chromatin feature for overlaps. The random simulation was performed 1000 times in each test. The number of simulations in which the shuffled DSBs overlapped with the chromatin feature at an equal or higher frequency than the real data was divided by the number of simulations to calculate the *p* value.

GP-seq score calculation for DSBs

GP-seq scores were downloaded from the GEO database under the accession number GSE135882 for experiment 5 (GSE135882_Exp5.1 Mb). GP-seq scores in each unique DSB region were tallied, averaged, and plotted.

Random forest prediction of classifier for DSB formation

We used the R package “randomForest” to build and train the random forest models employed in this study. We left mtry on default and used a value of 700 for ntree which represents the number of decision trees used by the model. We set importance to TRUE in order to access the MeanDecreaseAccuracy and MeanDecreaseGini values as well as the variable importance plots. The ‘pROC’ package was installed to plot the ROC (receiver operating characteristic curve) and PR (Precision Recall) curves. We used the BigWigAverageOverBed function to calculate the feature signals over DSB regions and used the mean values as input variables. Random sampling of DSBs from the genome was accomplished using the shuffleBed function from the BEDTools suite.

Repli-seq

At least 2×10^6 cells were used for each Repli-seq experiment. Cells were treated with DMSO, or 0.3 μ M, or 0.6 μ M APH, or nothing at all for 24 h. BrdU was then added at 100 μ M and cells were incubated for 2 h before washing with PBS and harvesting, followed by ethanol fixation. Fixed cells were then sorted by flow cytometry into early and late-replicating fractions. BrdU-labeled DNA from each fraction was immunoprecipitated, followed by preparation for sequencing libraries as previously described (Rivera-Mulia et al., 2022). Two biological replicates were produced for each sample, with similar results. Replication timing profiles from one replicate are shown.

Results

APH-induced DSBs are a composite of those induced by replication stress through APH and transcriptional upregulation by DMSO

We recently generated high resolution mapping of genome-wide DSB sites in human lymphoblastoid (GM06990) cells (Chakraborty et al., 2020). We used conditions known to induce CFS formation, i.e., mild level of replication stress by APH, with equal volume of the vehicle, dimethyl sulfoxide (DMSO), or no treatment (NT) at all as controls. This study produced 2111 NT or spontaneous DSBs, and 3927 and 7002 DSBs in DMSO- and APH-treated cells, respectively, demonstrating drug-specific induction of DSBs (Figure 1A). Spontaneous DSBs did not show apparent chromosomal bias; in contrast, DMSO-treated cells showed an enrichment of DSBs on chr19 whereas APH-treated cells had the highest density of DSBs on chr21, followed by chr19 (Figure 1B). We compared the DSBs to derive those shared and those uniquely induced by each condition. The vast majority (>78%) of the spontaneous DSBs were also present in the DMSO- and APH-treated cells, suggesting that an integral component of the CFSs are those regions of the genome that are intrinsically susceptible to DSBs (Figure 1A).

However, DMSO apparently elicited a strong induction of 2276 DSBs not present in the NT sample (“DMSO-unique”). We have shown that the genic association of DSBs increased from 33 to 41% from untreated cells to DMSO-treated cells (Chakraborty et al., 2020). Here we further showed that ~50% of the “DMSO-unique” DSBs occur in genic regions (Figure 1C), suggesting that transcription increase in DMSO cells caused DSBs. In contrast, the genic association level dropped to ~38% for “APH-unique” DSBs, though still a slightly higher level compared to NT-unique DSBs (Figure 1C), suggesting that transcription repression by APH-induced replication stress. Thus, we concluded that DSBs in APH samples are a composite effect of transcription induction by DMSO and repression by APH.

Previously, we have shown that DMSO- and APH-induced DSBs, but not spontaneous DSBs, are significantly enriched in late-replicating domains, using published Repli-Seq data for the GM06990 cell line (Hansen et al., 2010) to establish autosomal early- and late-replicating domains (Chakraborty et al., 2020). These observations are consistent with the known characteristics of the CFSs. To test if the same DSB-associated late-replicating regions remain late-replicating in DMSO or APH treatment, we performed Repli-seq experiments (Rivera-Mulia et al., 2022) with cells treated with DMSO, 0.3 μ M or 0.6 μ M APH, or nothing at all. Cells were transiently (2 h) labeled by BrdU, followed by sorting into early and late-replicating fractions by flow cytometry. BrdU-labeled nascent DNA was then immunoprecipitated from both fractions and subjected to next-generation sequencing, producing replication timing profiles represented as the Log₂ ratios of sequence reads of early vs. late fraction. The result showed that 0.3 μ M APH, the dosage at which the majority of our Break-seq experiments were conducted, caused S-phase arrest compared to the NT and DMSO controls (Figure 1D). Treatment with 0.6 μ M APH also induced S-phase arrest (Figure 1D). Nevertheless, genome-wide replication timing profiles demonstrated that most of the normally late-replicating regions remain late-replicating in drug-treated samples (Figure 1E), with few exceptions in the 0.6 μ M APH treatment that produced some local advanced replication timing among a large late-replicating region (more later). These results are consistent with our previous findings of global preservation of replication timing after APH treatment in other cell types (Sarni et al., 2020). We next asked how these DSBs are distributed in the CFSs.

APH-induced DSBs demarcate CFS core sequences and are not enriched in large genes

It was estimated that approximately one-third of CFSs are associated with large genes (Smith et al., 2006). A recent study systematically defined these long (>300 kb) transcribed genes which experience significantly delayed replication in APH and showed that they correspond to CFS core regions reported in the

literature (Brison et al., 2019). These CFS core regions were defined by higher resolution FISH (fluorescence *in situ* hybridization) probes in fourteen molecularly characterized CFSs (Savelyeva and Brueckner 2014). We first compared the DSBs in our Break-Seq experiments to these large transcribed genes with delayed replication (henceforth “large genes”), and found no significant correlation ($p < 0.001$) (Chakraborty et al., 2020). Moreover, the APH-induced DSBs mapped by another sequencing-based method called BLESS (Crosetto et al., 2013) did not show correlation with large genes either, while they showed significant correlation with DSBs mapped in our study ($p < 0.001$) (Chakraborty et al., 2020). We then systematically examined the fourteen molecularly characterized CFSs and we observed only 1 to 2 discrete spots of APH-induced DSBs in all of these CFS regions, even in large genes (e.g., *FHIT* and *WWOX*, sized 1502 and 1113 kb, respectively) associated with FRA3B and FRA16D, respectively (Figure 2A, “Break-seq” tracks). This relatively low density of DSBs in CFSs was also observed for the BLESS data set (Figure 2A, “BLESS_APH” track). Importantly, transcriptomic analysis showed that these CFS-associated genes were expressed at similar levels in all conditions (Figure 2B), with the exception of *CAV2* at FRA7G, where we did not detect expression, nor any DSB, in APH (Figures 2A,B). *LRP2* at FRA2G showed only moderate expression in APH and not in NT or DMSO samples, and the DSBs at the FRA2G locus were detected elsewhere with expressed genes (Figures 2A,B). Thus, despite a genome-wide transcription repression by APH, CFS-associated genes remain active.

Interestingly, both the Break-seq and BLESS-derived APH-induced DSBs tend to demarcate the boundaries of the mapped CFSs or CFS cores (Figure 2A). The relatively sparse nature of DSBs at these CFSs might be due to the relatively low APH concentrations (0.03 and 0.3 μ M) employed by Break-seq experiments. Notably, we observed that when cells were treated with 0.6 μ M APH it resulted in local late-to-early replication timing changes at 9 of the 14 CFS core regions, the exceptions being FRA2Ctel, 2G, 7H, 7G, and 8I (Figure 3). FRA2Ctel actually showed further delay in replication timing within the CFS core. Among the CFS cores with advanced timing, six (FRA1E, 3B, 7K, 13A, 16D and XB) clearly demonstrated initiation events within the core (Figure 3). These advanced timing changes suggested that the CFS core regions might contain dormant origins that are activated upon replication stress. They further suggested that altered replication dynamics within the CFS contribute to the DSB formation at higher APH dosages (see discussion). We then proceeded to further investigate the relationship between DSBs and genes in different size groups defined as 1–100, 100–300, 300–800, and >800 kb. Results showed that the number of DSBs as well as DSB density (number of DSBs per Mb of DNA) decreased with increasing gene size (Figures 4A,C). This trend was observed in both spontaneous DSBs and drug-induced ones. Notably, DSB density in all gene groups increased with DMSO/APH treatment compared to “NT” control (Figure 4B). However, it was the gene group of 1–100 kb, the smallest of all, that saw the biggest increase

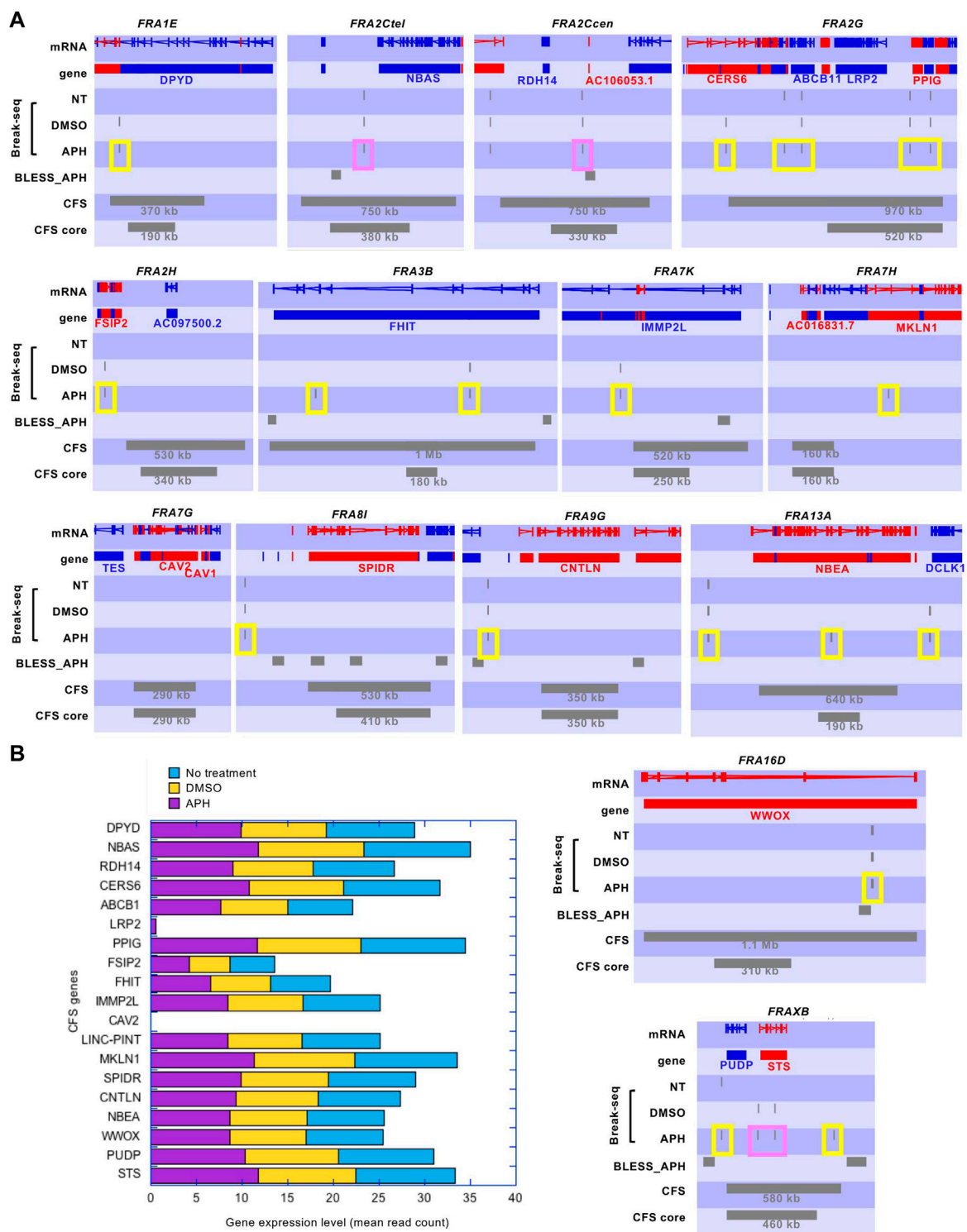


FIGURE 2 Break-seq signals tend to demarcate the boundaries of molecularly characterized CFSs. **(A)** SeqMonk profiles for DSB tracks. Genes encoded on the + strand and –strand are represented by red and blue blocks, respectively. DSB signals and CFS or CFS core regions are represented by grey blocks. Break-seq samples of NT (not treated)-, DMSO-, or APH-treated normal human lymphoblastoids (GM06990) are shown. The published APH-induced DSBs by BLESS are also shown. Note that Break-seq is more sensitive and have higher resolution than BLESS. The size of CFS and CFS core regions are shown. Those Break-seq signals flanking the CFS or CFS core regions are labeled by a yellow box, and those located within the CFS or CFS core regions are labeled by a pink box. **(B)** Expression levels of genes at CFSs measured by RNA-seq in cells treated with 0.3 μ M APH, or DMSO, (Continued)

FIGURE 2 (Continued)

or nothing at all, for 24 h. Detailed RNA-seq data analysis and raw data are described elsewhere (Chakraborty et al., 2020, in press) and accessible from the GEO accession number GSE124403.

in DMSO-treated cells (from 24 per Mb of DNA in untreated cells to 33 per Mb of DNA), followed by a decrease with APH treatment (down to 27 per Mb of DNA). This result led us to hypothesize that the smaller gene groups, particularly the 1–100 kb group, were physically clustered and the gene cluster underwent DMSO-induced transcription and APH-induced replication fork impediments, which collectively led to the increase in DSB density. Thus, our data strongly suggest an underlying role of genome organization in drug-induced DSB formation.

CTCF binding site density mirrors that of DSB density across gene groups

A recent study has shown that APH-induced CFSs correspond to TAD boundaries that are significantly delayed in replication timing (Sarni et al., 2020). TAD boundaries are delineated by CCCTC binding factor (CTCF) to form chromatin loops containing sequences with similar transcriptional regulation. We reasoned that large genes would require relatively fewer CTCF binding events for organization. In contrast, a similarly sized gene cluster housing many small genes might be organized into multiple smaller chromatin loops with distinct environment, which would require proportionally more CTCF binding sites. In other words, the density of chromatin loops should correlate with gene density. Indeed, we showed that the number and the density of CTCF binding sites decreased as gene size increased, a pattern that is almost identical to that of DSBs, particularly for those induced by APH (Figures 4C,D). This result suggested that DSB formation is determined by chromatin organization through CTCF binding and related events. Therefore, we sought to further test our hypothesis by investigating the relationship between DSBs and known chromatin accessibility markers.

Drug-induced DSBs locations are strongly correlated with TSS and the histone marker H3K36me3

We took advantage of the ENCODE project which cataloged a large set of genome-scale experiments of mapping chromatin structures. We specifically focused on those data generated from GM06990, the same cell line used in our Break-seq mapping experiments. We systematically analyzed the distribution of DSBs over each of the chromatin features including TSS,

CTCF binding sites, DNaseI HSSs, H3K4me3, H3K27me3, H3K36me3, and origins of replication (Supplementary Figure S1A), whose genomic distributions were summarized in Supplementary Figure S1B. We began by analyzing the distribution of DSBs over a 20-kb window centered on each feature in order to assess if DSBs were likely to associate with any of the features. Note that all these features were mapped in untreated cells, therefore these analyses allowed us to specifically test if and how epigenomic features in otherwise normal cells impact the DSBs seen in DMSO- or APH-treated cells. Spontaneous DSBs showed moderate association with only CTCF binding sites and origins, and no apparent association with the other features (Figure 5). Interestingly, DSBs in DMSO- and APH-treated cells showed an even stronger association with CTCF binding sites and origins (Figure 5, note the different y-axis scale on each plot). It is important to note that the association between CTCF binding sites and DSBs was not statistically significant when compared to randomized DSBs, suggesting that the comparison of DSBs to CTCF binding sites alone might not be sufficient to discern the relationship between DSBs and epigenomic features.

We next observed that drug-induced DSBs showed strong correlation with TSS and H3K36me3, moderately with H3K4me3 and DNaseI HSS, but not with H3K27me3, suggesting that DSBs are dependent on active transcription and not merely high chromatin accessibility (Figure 5). Moreover, we demonstrated that the drug-induced DSBs are correlated with TSS and H3K36me3 genome-wide by analyzing the distribution of DSBs overall features in the genome and not just the nearest ones (Supplementary Figure S2A). Specifically, 543 (13.8%) and 970 (13.9%) of the DMSO- and APH-induced DSBs, respectively, overlapped with H3K36me3 sites. These overlaps were significantly higher compared to randomized sequences in a random simulation test with >1000 iterations ($p < 0.001$). Because H3K36me3 is associated with active transcription and is enriched in the gene bodies downstream from TSS (Supplementary Figure S2B), the observed correlation of DSBs with both TSS and H3K36me3 suggested a strong dependence on transcription.

Furthermore, CTCF binding sites are strongly associated with TSSs and a small fraction of CTCF binding sites are also associated with H3K36me3 (Supplementary Figures S2C,D). Therefore, it appears that a subset of CTCF binding sites and a subset of H3K36me3 are associated with each other at the TSS, whereas the remainder of H3K36me3 are distributed downstream of TSS in gene bodies. These results led us to

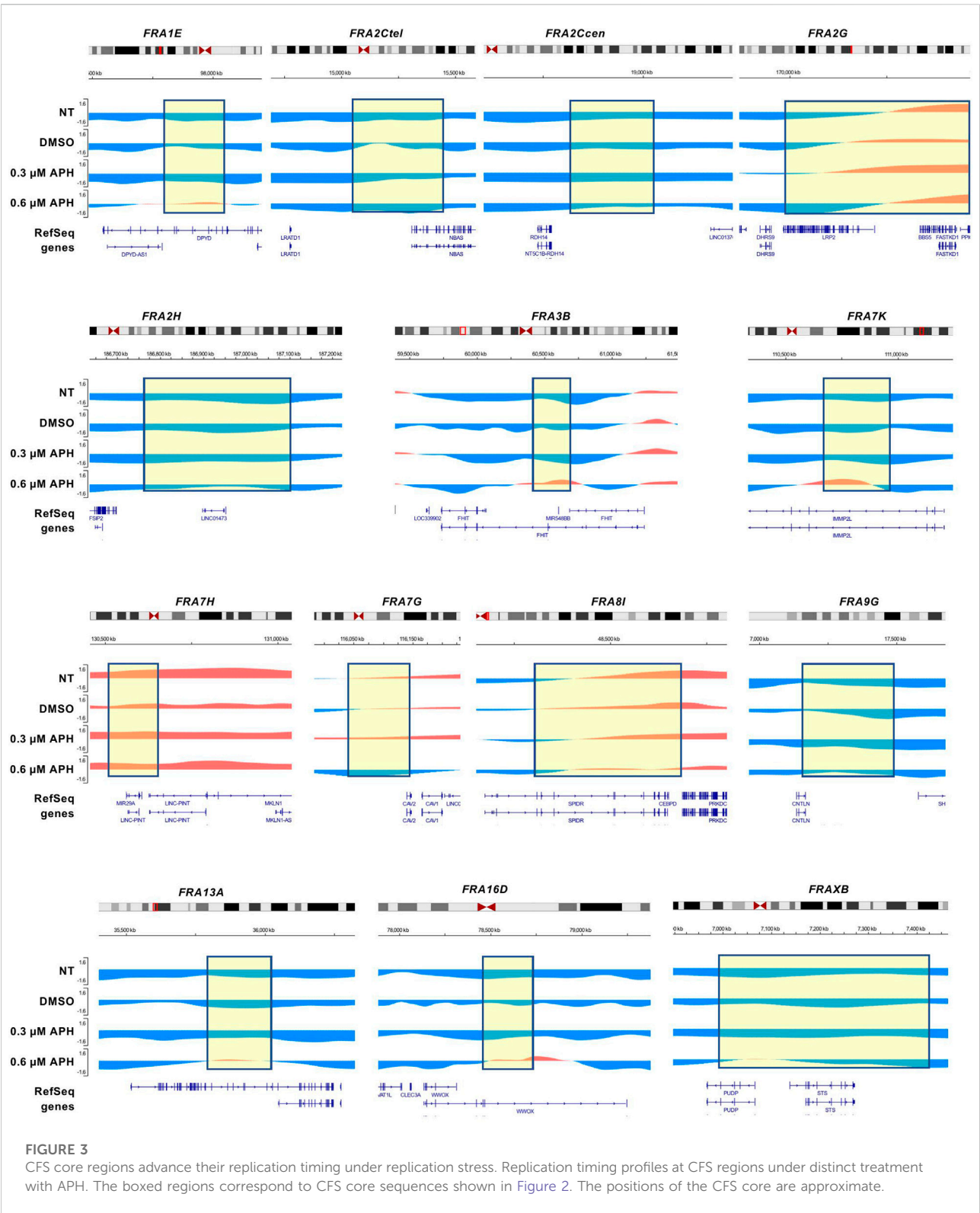
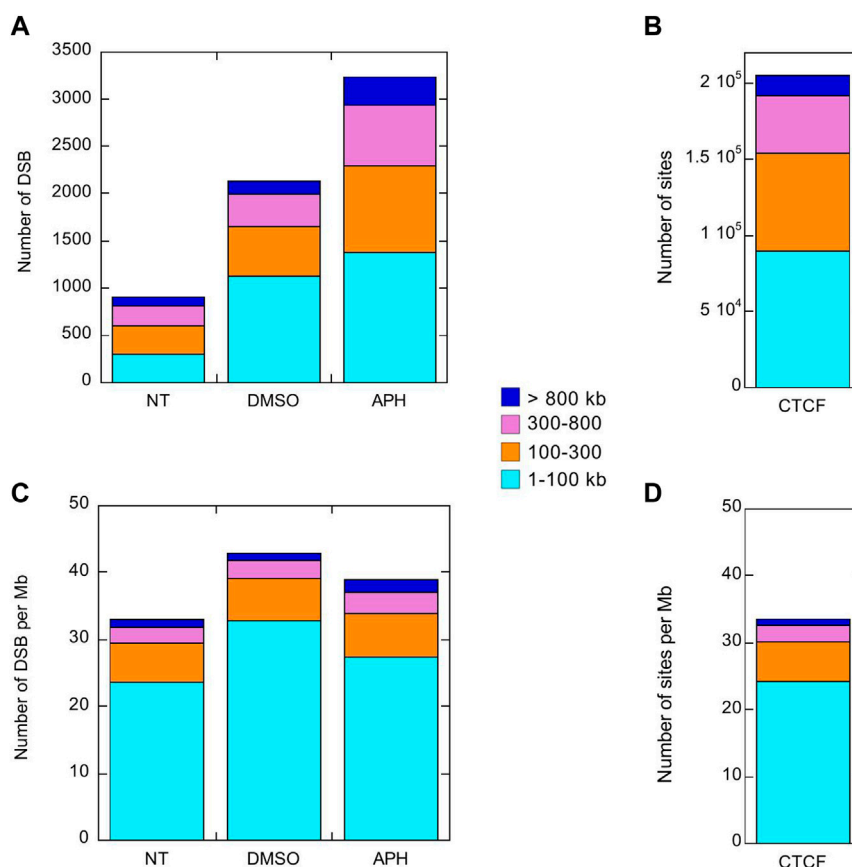


FIGURE 3
CFS core regions advance their replication timing under replication stress. Replication timing profiles at CFS regions under distinct treatment with APH. The boxed regions correspond to CFS core sequences shown in Figure 2. The positions of the CFS core are approximate.

conclude that DSBs are associated with both H3K36me3 sites at the TSS as well as those downstream from TSS, suggesting a transcription-dependent mechanism of DSB formation.

Moreover, APH-induced DSBs were better correlated with origins compared to spontaneous or DMSO-induced DSBs (Figure 5), consistent with induced replication stress

**FIGURE 4**

DSBs induced by DMSO or APH are present but not enriched in large genes >300 kb. Stacked column plots of the number of DSBs (A) and CTCF binding sites (B) in the indicated gene size groups. Stacked column plots of the density of DSBs (C) and CTCF binding sites (D), *i.e.*, number per Mb of DNA, in the indicated gene size groups.

impacting forks emanating from the origins. Finally, we showed that origins are broadly distributed around H3K36me3, CTCF binding sites, and TSS, in descending order of proximity (Supplementary Figures S3A–C). Therefore, it suggests that DSB formation at both TSS and gene bodies was also regulated by origin activities and replication fork movement.

Chromosomes enriched for DMSO- or APH-induced DSBs tend to be located in the radial center of the nucleus

To further discover defining features for these stress-induced DSBs we derived a list of DSBs overlapping with all features including H3K36me3, CTCF binding site, and origin (Data File S1). We found 24, 187, and 138 such DSBs in the untreated, DMSO-, and APH-treated samples, respectively. We first asked if there was any chromosomal bias of the DSBs. Remarkably, the chromosome with the highest enrichment of DSBs in all

conditions was Chr19. While spontaneous DSBs were only enriched on Chr19, drug-induced DSBs also showed enrichment on Chr5, 17, and 22 for DMSO and Chr15, 17, and 22 for APH treatment. These chromosomes all tend to be located near the radial center of the nucleus based on an elegant study using Genome Positioning (GP)-seq to analyze the 3D chromosome positioning (Girelli et al., 2020). Among them Chr19 has the highest GP-seq score, *i.e.*, the shortest radial distance from the center of the nucleus (Girelli et al., 2020). Indeed, compared to genomic average distribution of GP-seq score, the DSBs in our study showed skewed distribution towards higher GP-seq scores, hence nearer to the radial center of the nucleus (Figure 6). Girelli et al. further demonstrated that DSB level, using γ H2A.X as a proxy, was the highest in the center of the nucleus (Girelli et al., 2020). It has been shown previously that the deterministic parameter for radial positioning of chromatin in the nucleus is regional gene density (Kupper et al., 2007). Therefore, these results support our hypothesis that DMSO- and APH-induced DSBs are enriched in gene-dense

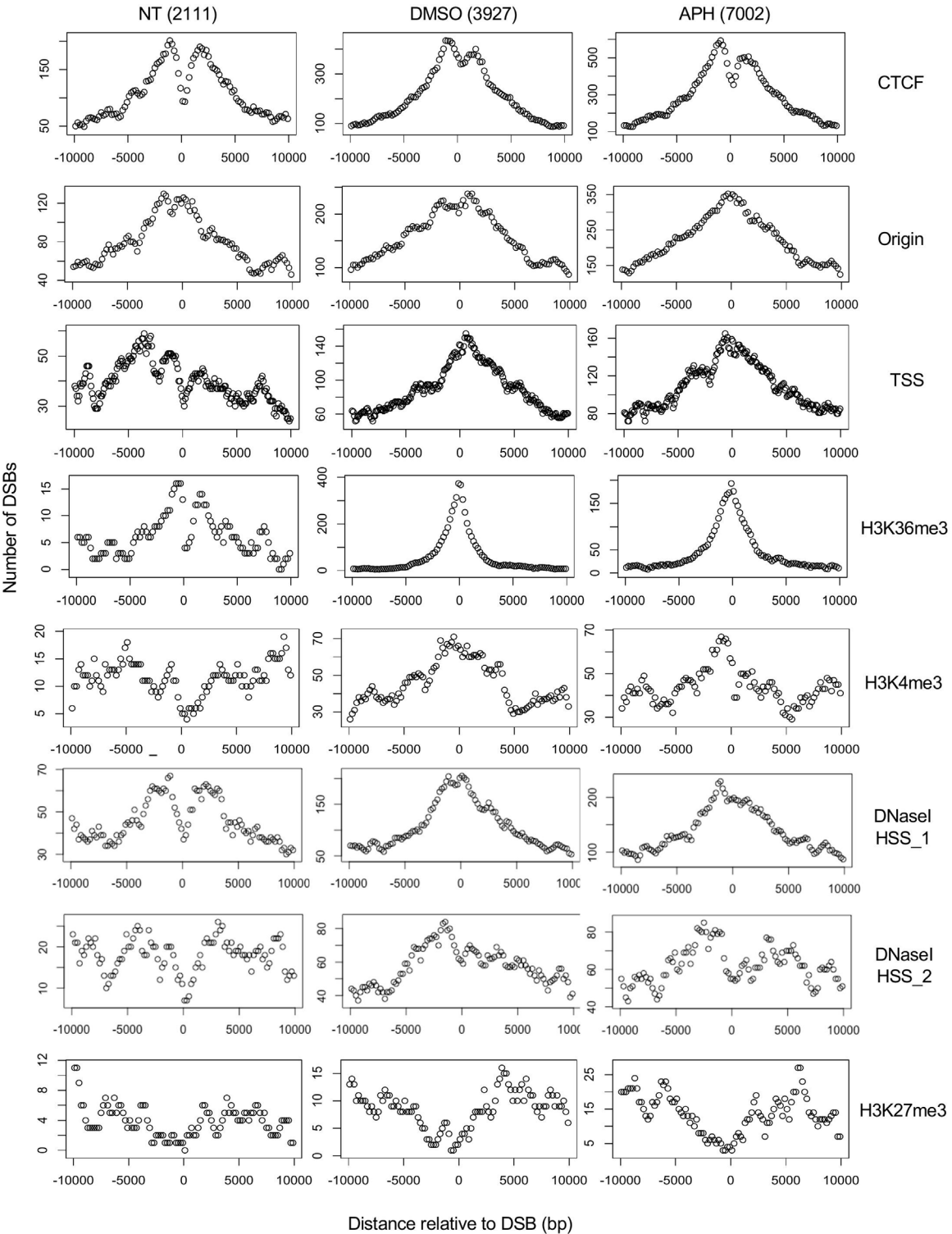
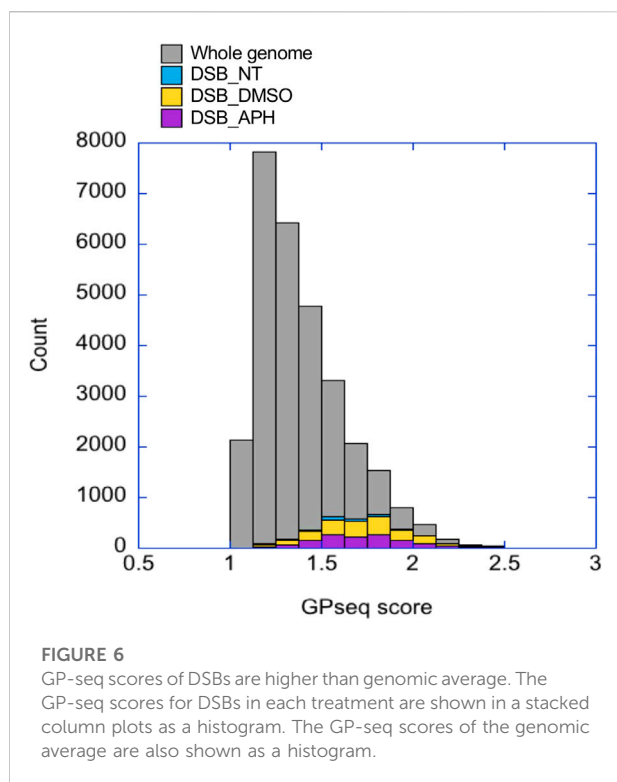


FIGURE 5
Aggregated plots of DSBs around the nearest chromatin marker. The number of DSBs in each of the 50 bins across a 20,000 bp window centered on the given chromatin marker are scored and plotted on the Y-axis against the relative distance to the center of the chromatin marker (X-axis). “TSS” and “H3K36me3” are the only two markers that were found associated with DSBs in DMSO- and APH-treated samples at a genomic scale (see [Supplementary Figure S2](#)).



regions with high level of transcription. We next investigated conservation of sequence motifs within the DSB regions in an effort to identify additional evidence for transcription-dependent mechanism of DSB formation.

Sequence motif discovery in replication stress-induced DSBs revealed binding sites for transcription factors implicated in maintaining 3D chromatin architecture














To specifically delineate the effect of DMSO-mediated transcription induction and APH-mediated replication stress on DSB formation, we analyzed the DMSO- and APH-specific DSBs, respectively. Motif search using AME (Analysis of Motif Enrichment, (McLeay and Bailey 2010)) identified a dearth of transcription factor (TF) binding sites in both groups. The top ten motifs enriched in DMSO only DSBs are binding sites for *ZNF582*, *SMARCA5*, *ZNF770*, *STAT5B*, *ZNF121*, *PAX5*, *STAT5A*, *PRDM6*, *TAF1*, and *ZNF418* ($p < 5.59 \times 10^{-42}$). The top ten motifs enriched in APH only DSBs are binding sites for *ETS2*, *EGR2*, *EGR1*, *NFATC1*, *ETV5*, *LEF1*, *TBX21*, *ZNF341*, *BCL11A*, and *ZNF121* ($p < 5.16 \times 10^{-124}$). Importantly, CTCF binding motif was also found as enriched in APH only DSBs ($p = 5.74 \times 10^{-5}$). Many of the proteins above (e.g., PRDM histone methyltransferase, the SMARCA subgroup of genes belonging to the SWI1/SNF1 chromatin remodelers, etc.) have been implicated in chromatin modifications and remodeling.

Others have been implicated in the maintenance of 3D chromatin architecture. For instance, *ZNF770* and *ZNF121* are significantly enriched at “insulator loops” mediated by CTCF to protect genes from emanating potentially harmful signals (Trieu et al., 2020). *PAX5*, a transcription factor essential for B-cell identity and function, changes 3D chromatin architecture (van Schoonhoven et al., 2020). Yet other TFs themselves are subjected to regulation for expression by chromatin architecture. For instance, CTCF has been shown to regulate and promote expression of *EGR1* and *EGR2* by establishing chromatin interaction loops between enhancer and promoter regulatory elements (Sekiya et al., 2019; Wang et al., 2020). We also used STREME (Bailey 2021) to discover new motifs in these groups, followed by identification of similar known motifs by Tomtom (Gupta et al., 2007) (Table 1). The results confirmed that DSB regions in both groups were enriched for sites for TF binding, thus supporting the notion that DSBs are transcription-dependent.

Identifying classifier(s) for DMSO- and APH-induced DSBs using machine learning

As a final endeavor, we asked if we could predict DSBs using these genomic and epigenomic features (henceforth “features” for simplicity). Motivated by Mourad et al.’s use of a machine learning method, Random Forest, to distinguish spontaneous DSBs from randomly sampled genomic regions, we developed a similar pipeline to validate features for replication stress-induced DSBs (Breiman 2001; Mourad et al., 2018). To best delineate the classification, we focused on the DSBs unique to each treatment rather than the entire cohort of DSBs in each treatment (Figure 1A). We then randomly sampled the hg19 genome for non-DSB regions with the same chromosome and length distributions as those DSBs in each treatment group for comparisons, i.e., non-NT/NT, non-DMSO/DMSO, and non-APH/APH. We assessed the variable importance of distance to DSBs by each feature. Note, we also analyzed the variable importance of feature signals over DSBs, computed as the sum or mean of signals across the DSBs from downloaded ChIP-seq bigWig data of GM06990 from the ENCODE project (Supplementary Figure S4). We then trained Random Forest to recognize the “real” DSBs, with reasonable accuracy (AUROC > 0.8, class errors < 0.3, i.e., > 70% accuracy) using feature signals over DSBs (Figure 7), and with less accuracy using distance to features (Figure 7). By far the best predictor for NT, or spontaneous DSBs was DNaseI HSS (experiment 1) (Figure 7A). This result agrees with the findings by Mourad et al. and suggests that spontaneous DSBs occur at regions with high chromatin accessibility. The best predictor for both DMSO- and APH-specific DSBs was H3K27me3, followed by H3K36me3 (Figure 7B), indicating that both transcription induction and repression are correlated with DSBs. However, it is unclear whether the repressive histone marker is a consequence rather

TABLE 1 Select discovered motifs in DMSO unique DSBs and APH unique DSBs.

Motif ($P < 1e-08$)	Logo	p -Value	Sites	Similar Motifs ($p < 0.01$)
DMSO-specific DSBs				
GCCTCAGCCTCCCA		2.3e-10	318 (14%)	ZN770, IKZF1, ZN281, SALL4, CRX, KLF4, SMAD3, TBX21
AATCTGCAAGTGGAT		3.7e-09	336 (14.8%)	PO2F2, ELF1, ELF3, ELF2, EHF, PRDM1, IRF2, CLOCK, IRF1, ETS1, SPI1, ELF5, ETV5, ERG
CACTGCACTCCAGCC		3.7e-09	278 (12.2%)	ZSC31, TEAD1, SMCA5, NR2C1, NKX21, NKX25, ZN502
GGTTCAACTCTGTGA		7.4e-09	315 (13.9%)	ZN768, ZF331, VDR, ZKSC1
AACTGCTCWATCAA		7.4e-09	314 (13.8%)	NF2L2, MAF, MAFG, MAFF, MAFB, CEBPD, PDX1, MAFK
AAACTTCTTTGTGAT		7.4e-09	312 (13.7%)	NF2L2, ZN680, SMCA1
APH-specific DSBs				
CCTCAGCCTCCCA		4.4e-16	499 (14.6%)	ZN770, IKZF1, ZN281, CRX, SALL4, ETS2, TBX21, WT1, SMAD3, SRBP2
CCAGCCTGGGCRACA		4.5e-14	539 (15.8%)	PAX5, SUH, ZN121, RFX2
GCTGGGATTACAGGC		2.3e-13	483 (14.1%)	ZN264, GFI1, GFI1B
CAGTGAGCYGAG		1.3e-12	469 (13.7%)	RARG, SRBP1, NR1H3, NFIC, ZN331, SRBP2, BRAC, RXRB
GAATYGCTTGAAC		1.2e-10	369 (10.8%)	ZN140, PBX2, ZN490, ZN329
TCTACWAAAA		1.4e-10	386 (11.3%)	TBP, MEF2C, MEF2A, ANDR, ZN490
GARACCCCRCTC		4.0e-09	327 (9.6%)	RUNX1, ZBT7A

than a cause for the DSB. Moreover, high feature density might bias the results using feature signal over DSBs. Therefore, we then analyzed the results based on feature distance from DSB. In comparison, the class errors for feature distance were moderately higher than feature signal (Figures 7A,B). The best predictors for DSBs in NT, DMSO, and APH samples are

H3K4me3, H3K36me3, and CTCF, respectively (Figure 7B). Among these, the role of active transcription in driving DMSO-specific DSBs was most apparent. These results corroborated our conclusions insofar as active transcription, as well as TAD domain boundaries, played an important part in induction of DSB formation. In summary, the inclusion of the Random Forest

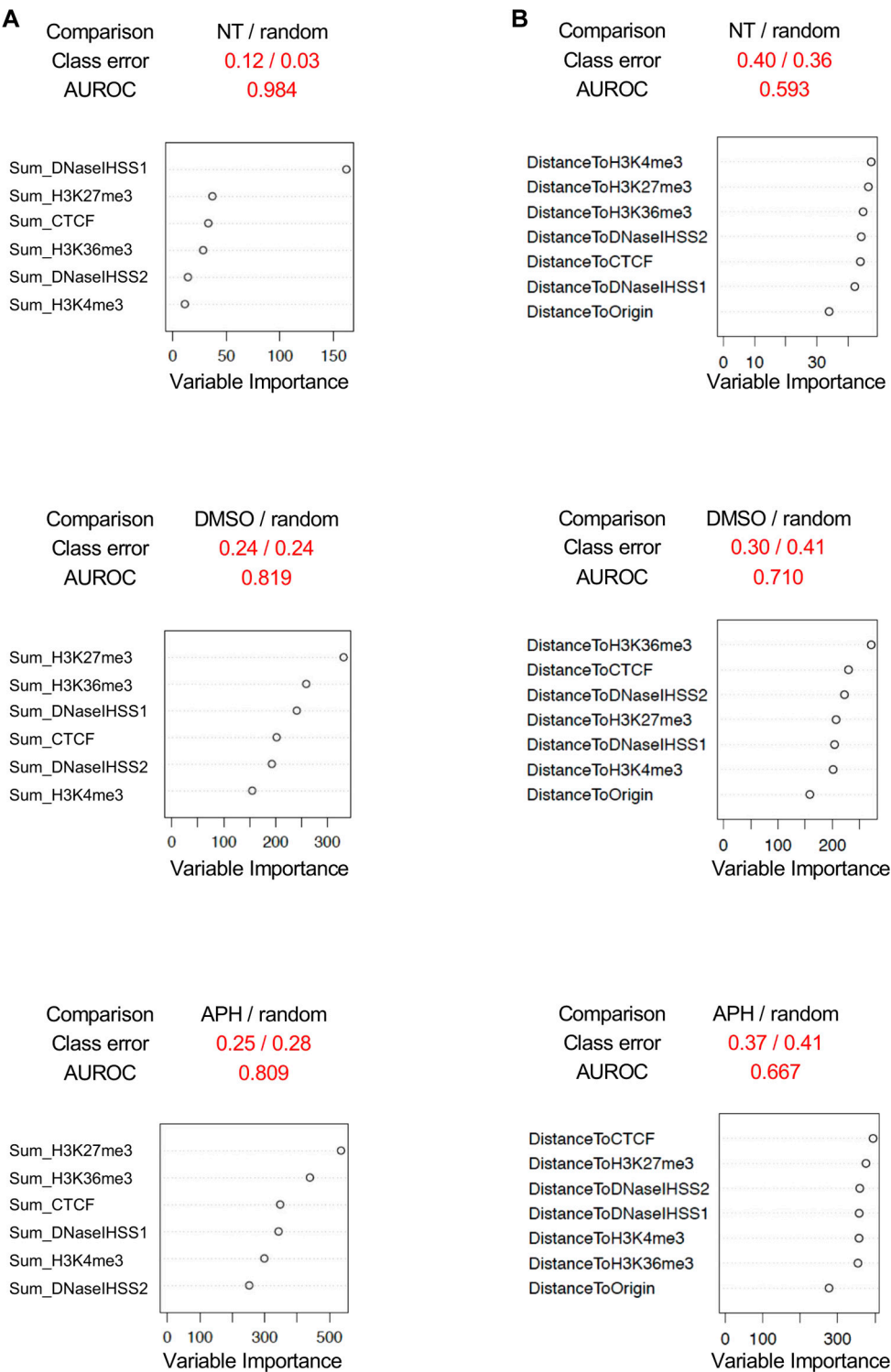
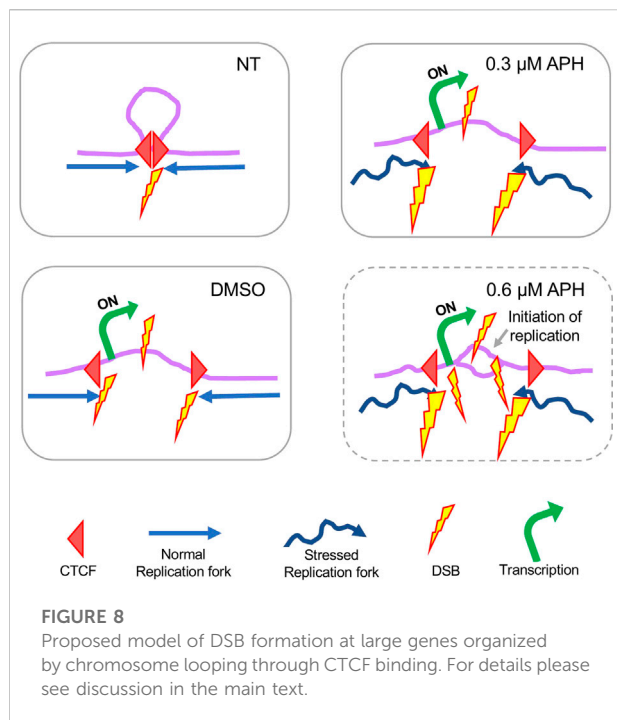


FIGURE 7
Random forest analysis. **(A)** Comparisons between DSBs specific to each treatment, NT, DMSO, and APH, and randomly sampled DSBs from the genome using feature signals over DSBs. **(B)** Comparisons between DSBs specific to each treatment, NT, DMSO, and APH, and randomly sampled DSBs from the genome using feature distance to DSBs. AUROC, area under the receiver operating characteristic curve.



model in this study proved to be a promising tool in determining specific correlations between genomic and epigenomic features within the spontaneous and replication stress-induced DSBs.

Discussion

In this report, we detail the computational analysis of spontaneous and replication stress-induced DSBs in GM06990 lymphoblastoids, which were recently mapped by Break-seq (Chakraborty et al., 2020). We focused on understanding the relationship between DSB formation and epigenomic and genomic signatures. Importantly, by parsing the DSBs to those specifically induced by DMSO, or by APH, it allowed us to dissect the effect of the replication inhibitor from its solvent independently.

The first key finding was that DSBs were closely associated with active transcription histone marker H3K36me3 and TAD boundaries. Specifically, spontaneous DSBs are associated with CTCF binding sites. In contrast, we observed a correlation between DMSO-induced DSBs and markers associated with active transcription, including TSSs, H3K36me3 and CTCF binding sites. Upon APH treatment, there was a dampening of the transcriptional response as evidenced by a decrease in the levels of all markers above, as well as a decrease of DSBs in genic regions compared to DMSO treatment. Yet, the APH-induced DSBs showed strong association with H3K36me3, suggesting

that despite an overall dampening of transcription by APH DSBs nevertheless took place at actively transcribing regions of the genome.

The second key finding related to the first one was that APH-induced DSBs in our study as well as the BLESS study (Crosetto et al., 2013; Chakraborty et al., 2020) were not enriched in large genes (>300 kb in size) and did not correlate with the CFS core sequences. Instead, we observed a tendency of the APH-induced DSBs to demarcate the boundaries of CFSs or the CFS core sequences. We suggest that DSBs occur at TAD boundaries marked by CTCF. Because the density of CTCF binding at large genes is relatively lower than that at smaller genes, DSB frequency is accordingly lower at large genes. However, we note that our Break-seq data were generated with relatively low level of APH (0.03 and 0.3 μ M), while the BLESS study used 0.4 μ M APH. Therefore, it is possible that the DSBs within the CFS cores might be more readily detected at higher level of APH.

Overall, these findings led us to hypothesize that DSBs were the result of CTCF-mediated chromosome remodeling due to transcription. We propose the following model to explain spontaneous and drug-induced DSB formation, specifically at the large genes where CFSs tend to reside (Figure 8). In untreated cells, DSBs occur at those CTCF binding sites involved in chromosome looping as replication forks originated from either side of the loop progress towards them. Thus, these DSBs form a bifurcated distribution over CTCF binding sites, as observed in Figure 5. Upon DMSO treatment, transcription induction causes the chromosome loop to disassemble, exposing the CTCF binding sites to replication forks on both sides and reducing the appearance of a bifurcated distribution pattern of DSBs over these sites. Meanwhile, active transcription within the chromosome loop also causes DSBs, either independently or when encountering approaching replication forks. In 0.3 μ M APH, transcription remains active at large genes despite a global dampening of transcription. Additionally, unstable replication forks increase the probability of DSBs at CTCF binding sites (denoted by larger DSB icons) and/or active transcription sites. Finally, replication timing study indicated that in cells treated with 0.6 μ M APH there were localized initiation events within the CFS core regions, suggesting that DSBs were produced by unscheduled replication termination with forks initiated from outside the CFS region. This observation provided an extended explanation for chromosome breakage at the CFS core regions. This model is consistent with a recent study presenting strong evidence that initiation of origin activation involves transcription-induced reorganization of the TAD demarcated by CTCF binding, which presents select origins within the TAD to move to the periphery for efficient activation (Li et al., 2021).

In summary, our study provided a comprehensive overview of genomic and epigenomic features associated with replication stress-induced DSBs. It also laid out a framework for future studies to expand DSB mapping to more well-chosen cell lines and with simultaneous queries for active origins of replication and epigenomic features. Such an experimental design will promise to deliver important insights into the mechanisms of replication stress-induced DSB formation.

Data Availability Statement

The original contributions presented in the study are included in the article/Supplementary Materials, further inquiries can be directed to the corresponding author.

Author contributions

WF conceptualized the study. SK, SM-N, JR-M, AC, and WF performed data collection and analyses. WF and SK wrote the manuscript. All authors critically read and edited the manuscript.

Funding

This work was supported by the National Institute of Health GM118799-01A1 grant to WF and R35GM137950 to JR-M.

References

- Alt, F. W., and Schwer, B. (2018). DNA double-strand breaks as drivers of neural genomic change, function, and disease. *DNA Repair (Amst)* 71, 158–163. doi:10.1016/j.dnarep.2018.08.019
- Alt, F. W., Wei, P. C., and Schwer, B. (2017). Recurrently breaking genes in neural progenitors: Potential roles of DNA breaks in neuronal function, degeneration and cancer. in *Genome Editing in Neurosciences*, Editors R. Jaenisch, F. Zhang, and F. Gage, (Cham (CH): Springer), 63–72.
- Arlt, M. F., Casper, A. M., and Glover, T. W. (2003). Common fragile sites. *Cytogenet. Genome Res.* 100 (1–4), 92–100. doi:10.1159/000072843
- Arlt, M. F., Durkin, S. G., Ragland, R. L., and Glover, T. W. (2006). Common fragile sites as targets for chromosome rearrangements. *DNA Repair (Amst)* 5 (9–10), 1126–1135. doi:10.1016/j.dnarep.2006.05.010
- Bailey, T. L. (2021). Streme: Accurate and versatile sequence motif discovery. *Bioinformatics* 37, 2834–2840. doi:10.1093/bioinformatics/btab203
- Barlow, J. H., Faryabi, R. B., Callen, E., Wong, N., Malhowski, A., Chen, H. T., et al. (2013). Identification of early replicating fragile sites that contribute to genome instability. *Cell* 152 (3), 620–632. doi:10.1016/j.cell.2013.01.006
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45 (1), 5–32. doi:10.1023/a:1010933404324
- Brison, O., El-Hilali, S., Azar, D., Koundrioukoff, S., Schmidt, M., Nahse, V., et al. (2019). Transcription-mediated organization of the replication initiation program across large genes sets common fragile sites genome-wide. *Nat. Commun.* 10 (1), 5693. doi:10.1038/s41467-019-13674-5
- Canela, A., Maman, Y., Jung, S., Wong, N., Callen, E., Day, A., et al. (2017). Genome organization drives chromosome fragility. *Cell* 170 (3), 507–521. doi:10.1016/j.cell.2017.06.034
- Chakraborty, A., Jenjaroenpun, P., Li, J., El Hilali, S., McCulley, A., Haarer, B., et al. (2020). Replication stress induces global chromosome breakage in the fragile X genome. *Cell Rep.* 32 (12), 108179. doi:10.1016/j.celrep.2020.108179
- Crosetto, N., Mitra, A., Silva, M. J., Bienko, M., Dojer, N., Wang, Q., et al. (2013). Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat. Methods* 10 (4), 361–365. doi:10.1038/nmeth.2408
- El Achkar, E., Gerbault-Seureau, M., Muleris, M., Dutrillaux, B., and Debatisse, M. (2005). Premature condensation induces breaks at the interface of early and late replicating chromosome bands bearing common fragile sites. *Proc. Natl. Acad. Sci. U. S. A.* 102 (50), 18069–18074. doi:10.1073/pnas.0506497102
- Feng, W., and Chakraborty, A. (2017). Fragility extraordinaire: Unsolved mysteries of chromosome fragile sites. *Adv. Exp. Med. Biol.* 1042, 489–526. doi:10.1007/978-981-10-6955-0_21
- Girelli, G., Custodio, J., Kallas, T., Agostini, F., Wernersson, E., Spanjaard, B., et al. (2020). GPSeq reveals the radial organization of chromatin in the cell nucleus. *Nat. Biotechnol.* 38 (10), 1184–1193. doi:10.1038/s41587-020-0519-y
- Glover, T. W., Arlt, M. F., Casper, A. M., and Durkin, S. G. (2005). Mechanisms of common fragile site instability. *Hum. Mol. Genet.* 14, R197–R205. doi:10.1093/hmg/ddi265
- Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L., and Noble, W. S. (2007). Quantifying similarity between motifs. *Genome Biol.* 8 (2), R24. doi:10.1186/gb-2007-8-2-r24
- Handt, M., Epplen, A., Hoffjan, S., Mese, K., Epplen, J. T., and Dekomien, G. (2014). Point mutation frequency in the FMR1 gene as revealed by fragile X syndrome screening. *Mol. Cell. Probes* 28 (5–6), 279–283. doi:10.1016/j.mcp.2014.08.003
- Hansen, R. S., Thomas, S., Sandstrom, R., Canfield, T. K., Thurman, R. E., Weaver, M., et al. (2010). Sequencing newly replicated DNA reveals widespread

Acknowledgments

We thank the Feng lab for general support and discussions.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.907547/full#supplementary-material>

- plasticity in human replication timing. *Proc. Natl. Acad. Sci. U. S. A.* 107 (1), 139–144. doi:10.1073/pnas.0912402107
- Hellman, A., Rahat, A., Scherer, S. W., Darvasi, A., Tsui, L. C., and Kerem, B. (2000). Replication delay along FRA7H, a common fragile site on human chromosome 7, leads to chromosomal instability. *Mol. Cell. Biol.* 20 (12), 4420–4427. doi:10.1128/mcb.20.12.4420-4427.2000
- Helmrich, A., Ballarino, M., and Tora, L. (2011). Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol. Cell* 44 (6), 966–977. doi:10.1016/j.molcel.2011.10.013
- Kimura, H. (2013). Histone modifications for human epigenome analysis. *J. Hum. Genet.* 58 (7), 439–445. doi:10.1038/jhg.2013.66
- Kupper, K., Kolbl, A., Biener, D., Dittrich, S., von Hase, J., Thormeyer, T., et al. (2007). Radial chromatin positioning is shaped by local gene density, not by gene expression. *Chromosoma* 116 (3), 285–306. doi:10.1007/s00412-007-0098-4
- Le Beau, M. M., Rassool, F. V., Neilly, M. E., Espinosa, R., 3rd, Glover, T. W., Smith, D. I., et al. (1998). Replication of a common fragile site, FRA3B, occurs late in S phase and is delayed further upon induction: Implications for the mechanism of fragile site induction. *Hum. Mol. Genet.* 7 (4), 755–761. doi:10.1093/hmg/7.4.755
- Le Tallec, B., Koundrioukoff, S., Wilhelm, T., Letessier, A., Brison, O., and Debatisse, M. (2014). Updating the mechanisms of common fragile site instability: How to reconcile the different views? *Cell. Mol. Life Sci.* 71 (23), 4489–4494. doi:10.1007/s00018-014-1720-2
- Le Tallec, B., Millot, G. A., Blin, M. E., Brison, O., Dutrillaux, B., and Debatisse, M. (2013). Common fragile site profiling in epithelial and erythroid cells reveals that most recurrent cancer deletions lie in fragile sites hosting large genes. *Cell Rep.* 4 (3), 420–428. doi:10.1016/j.celrep.2013.07.003
- Li, Y., Xue, B., Zhang, M., Zhang, L., Hou, Y., Qin, Y., et al. (2021). Transcription-coupled structural dynamics of topologically associating domains regulate replication origin efficiency. *Genome Biol.* 22 (1), 206. doi:10.1186/s13059-021-02424-w
- McLeay, R. C., and Bailey, T. L. (2010). Motif enrichment analysis: A unified framework and an evaluation on ChIP data. *BMC Bioinforma.* 11, 165. doi:10.1186/1471-2105-11-165
- Mesner, L. D., Valsakumar, V., Cieslik, M., Pickin, R., Hamlin, J. L., and Bekiranov, S. (2013). Bubble-seq analysis of the human genome reveals distinct chromatin-mediated mechanisms for regulating early- and late-firing origins. *Genome Res.* 23 (11), 1774–1788. doi:10.1101/gr.155218.113
- Mourad, R., Ginalski, K., Legube, G., and Cuvier, O. (2018). Predicting double-strand DNA breaks using epigenome marks or DNA at kilobase resolution. *Genome Biol.* 19 (1), 34. doi:10.1186/s13059-018-1411-7
- Ozeri-Galai, E., Bester, A. C., and Kerem, B. (2012). The complex basis underlying common fragile site instability in cancer. *Trends Genet.* 28 (6), 295–302. doi:10.1016/j.tig.2012.02.006
- Ozeri-Galai, E., Tur-Sinai, M., Bester, A. C., and Kerem, B. (2014). Interplay between genetic and epigenetic factors governs common fragile site instability in cancer. *Cell. Mol. Life Sci.* 71 (23), 4495–4506. doi:10.1007/s00018-014-1719-8
- Palakodeti, A., Han, Y., Jiang, Y., and Le Beau, M. M. (2004). The role of late/slow replication of the FRA16D in common fragile site induction. *Genes Chromosom. Cancer* 39 (1), 71–76. doi:10.1002/gcc.10290
- Palumbo, E., and Russo, A. (2019). Common fragile site instability in normal cells: Lessons and perspectives. *Genes Chromosom. Cancer* 58 (5), 260–269. doi:10.1002/gcc.22705
- Pelliccia, F., Bosco, N., Curatolo, A., and Rocchi, A. (2008). Replication timing of two human common fragile sites: FRA1H and FRA2G. *Cytogenet. Genome Res.* 121 (3–4), 196–200. doi:10.1159/000138885
- Rivera-Mulia, J. C., Trevilla-Garcia, C., and Martinez-Cifuentes, S. (2022). Optimized repli-seq: Improved DNA replication timing analysis by next-generation sequencing. *Chromosome Res.* 2022. doi:10.1007/s10577-022-09703-7
- Sarni, D., and Kerem, B. (2016). The complex nature of fragile site plasticity and its importance in cancer. *Curr. Opin. Cell Biol.* 40, 131–136. doi:10.1016/j.ccb.2016.03.017
- Sarni, D., Sasaki, T., Irony Tur-Sinai, M., Miron, K., Rivera-Mulia, J. C., Magnuson, B., et al. (2020). 3D genome organization contributes to genome instability at fragile sites. *Nat. Commun.* 11 (1), 3613. doi:10.1038/s41467-020-17448-2
- Savelyeva, L., and Brueckner, L. M. (2014). Molecular characterization of common fragile sites as a strategy to discover cancer susceptibility genes. *Cell. Mol. Life Sci.* 71 (23), 4561–4575. doi:10.1007/s00018-014-1723-z
- Sekiya, T., Kato, K., Kawaguchi, A., and Nagata, K. (2019). Involvement of CTCF in transcription regulation of EGR1 at early G1 phase as an architecture factor. *Sci. Rep.* 9 (1), 329. doi:10.1038/s41598-018-36753-x
- Smith, D. I., Zhu, Y., McAvoy, S., and Kuhn, R. (2006). Common fragile sites, extremely large genes, neural development and cancer. *Cancer Lett.* 232 (1), 48–57. doi:10.1016/j.canlet.2005.06.049
- Trieu, T., Martinez-Fundichely, A., and Khurana, E. (2020). DeepMILO: A deep learning approach to predict the impact of non-coding sequence variants on 3D chromatin structure. *Genome Biol.* 21 (1), 79. doi:10.1186/s13059-020-01987-4
- van Schoonhoven, A., Huylebroeck, D., Hendriks, R. W., and Stadhouders, R. (2020). 3D genome organization during lymphocyte development and activation. *Brief. Funct. Genomics* 19 (2), 71–82. doi:10.1093/bfpg/elz030
- Wang, J., Wang, J., Yang, L., Zhao, C., Wu, L. N., Xu, L., et al. (2020). CTCF-mediated chromatin looping in EGR2 regulation and SUZ12 recruitment critical for peripheral myelination and repair. *Nat. Commun.* 11 (1), 4133. doi:10.1038/s41467-020-17955-2
- Wang, L., Darling, J., Zhang, J. S., Huang, H., Liu, W., and Smith, D. I. (1999). Allele-specific late replication and fragility of the most active common fragile site, FRA3B. *Hum. Mol. Genet.* 8 (3), 431–437. doi:10.1093/hmg/8.3.431
- Zhang, T., Cooper, S., and Brockdorff, N. (2015). The interplay of histone modifications - writers that read. *EMBO Rep.* 16 (11), 1467–1481. doi:10.15252/embr.201540945

Frontiers in Genetics

Highlights genetic and genomic inquiry relating to all domains of life

The most cited genetics and heredity journal, which advances our understanding of genes from humans to plants and other model organisms. It highlights developments in the function and variability of the genome, and the use of genomic tools.

Discover the latest Research Topics

[See more →](#)

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact

