

***HsMCM8 and HsMCM9: Essential for Double-Strand Break Repair  
and Normal Ovarian Function***

by

Elizabeth Paladin Jeffries

Bachelor of Science, Indiana University of Pennsylvania, 2009

Submitted to the Graduate Faculty of

The Kenneth P. Dietrich School of Arts & Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

The Kenneth P. Dietrich School of Arts & Sciences

This dissertation was presented

by

Elizabeth P. Jeffries

It was defended on

May 4, 2015

and approved by

Xinyu Liu, Assistant Professor, Department of Chemistry

Aleksandar Rajkovic, Professor and Chair, Department of Obstetrics, Gynecology and  
Reproductive Sciences

Dissertation Co-Advisor: Seth Horne, Associate Professor, Department of Chemistry

Dissertation Co-Advisor: Michael Trakselis, Adjunct Associate Professor, Department of  
Chemistry, University of Pittsburgh

***HsMCM8 and HsMCM9: Essential for Double-Strand Break Repair***

**and Normal Ovarian Function**

Elizabeth Paladin Jeffries, PhD

University of Pittsburgh, 2015

Copyright © by Elizabeth P. Jeffries

2015

*Hs*MCM8 AND *Hs*MCM9: ESSENTIAL FOR DNA DOUBLE-STRAND BREAK REPAIR  
AND NORMAL OVARIAN FUNCTION

Elizabeth Jeffries, PhD

University of Pittsburgh, 2015

The minichromosome maintenance (MCM) family of proteins is conserved from archaea to humans, and its members have roles in initiating DNA replication. MCM8 and MCM9 are minimally characterized members of the eukaryotic MCM family that associate with one another and both contain conserved ATP binding and hydrolysis motifs. The MCM8-9 complex participates in repair of DNA double-strand breaks by homologous recombination, and MCM8 is implicated in meiotic recombination. We identified a novel alternatively spliced isoform of *Hs*MCM9 that results in a medium length protein product (MCM9<sup>M</sup>) that eliminates a C-terminal extension of the fully spliced product (MCM9<sup>L</sup>). Quantitative real-time reverse transcriptase PCR (qRT-PCR) of the relative mRNA isoform abundances across cell lines reveals MCM9<sup>L</sup> transcript is more abundant than MCM9<sup>M</sup>. The expression of both isoforms is cell cycle regulated, as they are most abundant in S-phase. Wild-type MCM9<sup>L</sup> forms DNA damage-dependent nuclear foci, while MCM9<sup>M</sup> is cytoplasmic and MCM9<sup>Cterm</sup> is diffuse throughout the nucleus. We have identified and verified a putative nuclear localization signal (NLS), and Rad51-interacting motif (BRCv) in the C-terminus of MCM9. GFP-tagged MCM9 NLS<sup>-</sup> is solely cytoplasmic, GFP-tagged MCM9 BRCv<sup>-</sup> is diffuse throughout the nucleus. A combination of SNP arrays, comparative genomic hybridization arrays, and whole-exome

sequencing analyses identified homozygous pathogenic variants in MCM8 (MCM8 c.446C>G; p.P149R) and MCM9 (MCM9 c.1732.2T>C and MCM9 c.394C>T; p.R132\*), all located within regions of homozygosity in women afflicted by premature ovarian failure (POF). The MCM9 c.1732.2T>C variant alters a splice donor site, resulting in abnormal alternative splicing and truncated forms of MCM9 that are unable to be recruited to sites of DNA damage. In the second family, MCM9 c.394C>T (p.R132\*) results in a predicted loss of functional MCM9. Compared with fibroblasts from unaffected family members, chromosomal break repair was deficient in fibroblasts from all affected individuals, likely due to inhibited recruitment of mutated MCM8 or MCM9 to sites of DNA damage. Our cumulative results suggest that MCM8-9 have evolved to act late in both mitotic recombination pathways to aid in crossover migration and/or strand resolution.

# TABLE OF CONTENTS

Preface.....	xv
1.0 Introduction .....	1
1.1 Eukaryotic DNA Replication.....	1
1.1.1 Eukaryotic DNA Replication and MCM Proteins.....	1
1.1.1.1 Replication Fork and Replication Origins .....	1
(a) Pre-Replication Complex .....	2
(i) Origin Licensing .....	4
1.1.1.2 Minichromosome Maintenance (MCM) Proteins.....	5
(a) MCM2-7 Replicative Helicase .....	7
(b) MCM8 and MCM9.....	9
1.1.2 Repair of DNA Double-Strand Breaks.....	13
1.1.2.1 DNA-damage Checkpoints.....	14
1.1.3 Non-Homologous End Joining for Repair of DSBs.....	14
1.1.4 Exogenous DNA Damaging Agents.....	17
1.1.5 Homologous Recombination for Repair of DSBs .....	17
1.1.6 Poly(ADP-Ribose) Polymerases Detect DNA Strand Breaks.....	19
1.1.7 Meiotic Recombination .....	20
1.2 Gene Transcription and RNA Splicing.....	23
1.3 Premature Ovarian Failure.....	24
1.3.1 Clinical Diagnosis and Known Causes .....	24
2.0 Identification, Quantification, and Evolutionary Analysis of a Novel Isoform of MCM9.....	25

2.1	Identification of a Novel Isoform of MCM9 .....	25
2.1.1	Research Design .....	25
2.1.2	Results and Discussion .....	26
2.1.2.1	Identification of a Novel Human MCM9 Isoform .....	26
2.1.2.2	Separation and Quantification of Expression Levels of MCM9 <sup>M</sup> and MCM9 <sup>L</sup> .....	27
2.1.2.3	MCM9 Isoforms are most abundant in S-phase .....	33
2.1.2.4	MCM9 Isoform Expression Increases During Crosslink Damage .....	37
2.1.2.5	Analysis of Intron-Exon Boundaries of M.....	39
2.1.2.6	Evolutionary Analysis of MCM9 .....	41
2.1.3	Conclusion.....	44
2.1.4	Experimental Procedures.....	45
2.1.4.1	Materials .....	45
2.1.4.2	Human Cell Culture and RNA Isolation.....	46
2.1.4.3	Semi-quantitative RT-PCR Detection of MCM9 Splice Variants .....	46
2.1.4.4	Quantitative Real-Time Reverse Transcriptase PCR (qRT-PCR) .....	47
2.1.4.5	Fluorescence Activated Cell Sorting (FACS) Analysis.....	48
3.0	Pathogenic Mutations of MCM8 and MCM9 .....	50
3.1	MCM8 and MCM9 Mutations Underly Ovarian Failure and Chromosomal Instability.....	50
3.1.1	Premature Ovarian Failure in Families with MCM8 or MCM9 Mutations .....	50
3.1.2	Research Design .....	59
3.1.2.1	Identification of MCM8 and MCM9 Mutations .....	59
3.1.2.2	Impairment of DNA Break Repair in Families A and C .....	67
3.1.2.3	Impairment of DNA Damage-Dependent Foci in Families A and C and Impairment of DNA Binding in Family C .....	75
3.1.2.4	Evolutionary Comparison of the Selected MCM8 Region .....	78
3.1.2.5	Homology Model of Human MCM8 and Crystallization Trials .....	78
3.1.2.6	DNA Binding Domain Mutation in MCM8 Impairs DNA Damage- Dependent Foci .....	81

3.1.3	Conclusion .....	83
3.1.4	Experimental.....	86
3.1.4.1	DNA Mapping and Whole Exome Sequencing.....	86
3.1.4.2	SNP Arrays and Whole Exome Sequencing to Identify MCM8 c.446C>G .....	89
3.1.4.3	Verification of <i>MCM9</i> c.1732+2T>C Splice Site Mutation.....	89
3.1.4.4	Identification of mRNA Isoforms Resulting from MCM9 c. 1732+2T>C Mutation.....	90
3.1.4.5	Foci Formation Assays .....	90
3.1.4.6	Functional Analysis of Chromosomal Instability .....	91
3.1.4.7	DNA Binding Assay.....	91
3.1.4.8	Homology Model of Human MCM8.....	93
3.1.4.9	Crystallization of MCM8 N-Terminal Domain .....	93
4.0	Towards the Activity and Function of MCM9 Motifs .....	94
4.1	MCM9 C-terminus is Essential for Response to DNA Damage .....	94
4.1.1	Research Design.....	94
4.1.2	Results.....	95
4.1.2.1	MCM9M Is Cytoplasmic and MCM9Cterm Imports to Nucleus But Does not Form Damage-Dependent Nuclear Foci .....	95
4.1.2.2	MCM9 NLS <sup>-</sup> Disrupts Nuclear Localization .....	97
4.1.2.3	MCM9 BRCV <sup>-</sup> Disrupts Damage Induced Foci Formation.....	98
4.1.3	Conclusion.....	101
4.1.4	Experimental.....	102
4.1.4.1	Transfections and Foci Formation Assays.....	102
5.0	MCM8 c.446C>G/p.P149R Mutation Causes Sensitivity to MMC But Not to PARP Inhibitors.....	103
5.1.1	Research Design.....	103
5.1.2	Results and Discussion.....	105

5.1.2.1	MCM8 c.446C>G/p.P149R Cells Are Acutely Sensitive to MMC Damage.....	105
5.1.2.2	MCM8 c.446C>G/p.P149R Does Not Sensitize Lymphocytes to BMN673 or ABT-888 .....	108
5.1.2.3	PARP inhibitors and MMC Have No Synthetic Lethal Effect in MCM c.446C>G/p.P149R.....	109
5.1.3	Conclusion.....	111
5.1.4	Experimental.....	112
5.1.4.1	Fibroblast MMC Treatment and Surviving Fraction Calculations.....	112
5.1.4.2	EBV Lymphocyte MMC and BMN673 Treatment and Percent Viability Calculations.....	113
6.0	Future Perspectives.....	114
6.1.1	Define MCM8 and MCM9 protein interaction sites.....	114
6.1.2	Examine the corecruitment of MCM8 p.P149R and MCM9 c.1732+2T>C or p.R132* patient cell lines .....	115
6.1.3	Test MCM8 and MCM9 biochemical activities.....	116
6.1.4	Continue the structural characterization of MCM8 and MCM9.....	116
6.1.5	Examine trans-acting MCM9 splicing regulation and cytoplasmic role of MCM9 <sup>M</sup> .....	117
6.1.6	Create genetically modified embryos for women afflicted by POF.....	117
6.1.7	Investigate further coincidence of POF and recombination defects.....	118
	Bibliography.....	119

## LIST OF TABLES

Table 2.1: qRT-PCR $C_T$ values normalized to GAPDH $C_T$ values, reported as $\Delta C_T$ .....	30
Table 2.2: qRT-PCR $\Delta C_T$ values for synchronized HeLa cells normalized to GAPDH $C_T$ values. .....	36
Table 2.3: HeLa qRT-PCR $\Delta C_T$ Values Following MMC or HU Treatment.....	38
Table 2.4: Genomic sequence overlapping splice sites in <i>HsMCM9</i> .....	40
Table 2.5: Oligonucleotides Used to Amplify MCM9 Isoforms. ....	46
Table 3.1: Clinical Laboratory Investigations of Affected Family C Daughters (IV-1, IV-6, and IV-9).....	57
Table 3.2 Clinical Laboratory Investigations of Affected Individuals Prior to Hormone Replacement Therapy, Families A and B .....	58
Table 3.3 Whole Exome Sequencing Variant Calls in Families A and B .....	61
Table 3.4 Regions of Homozygosity Based on SNP Array, Family C .....	65
Table 3.5 Chromosomal Breakage as Induced by DEB treatment in Patient Lymphocytes (Raw Data).....	67
Table 3.6 WES Variants in genes known to be involved in chromosomal instability or primary amenorrhea, Family C.....	69
Table 3.7 WES Variants in Genes Known to Be Involved in Chromosomal Instability or Primary Amenorrhea, Families A and B. ....	74
Table 3.8 Quality Metrics for Sequencing by Sample, Families A and B.....	87
Table 3.9 Whole Exome Sequencing Alignment Statistics, Families A and B .....	88

Table 3.10 Whole Exome Sequencing Variant Filtration, Families A and B..... 88

Table 3.11 Average number of Chromosomal Breaks Per Cell After MMC Treatment, Families A  
and B ..... 91

## LIST OF FIGURES

Figure 1.1 Pre-replication complex formation.....	3
Figure 1.2 Minichromosome maintenance (MCM) protein family. ....	5
Figure 1.3 Alignment of domains and motifs of MCM8 and MCM9 proteins.....	6
Figure 1.4 Dendromer of <i>HsMCM</i> proteins. ....	12
Figure 1.5 DNA Double-strand break repair by non-homologous end joining (NHEJ) or homologous recombination (HR) pathways (9).....	16
Figure 1.6 Function of PAR in DNA DSB detection and signaling. ....	19
Figure 1.7 Comparison of the purposes, causes and outcomes of meiotic recombination and HR for double-strand break repair.....	21
Figure 2.1 Identification of MCM9 <sup>M</sup> (A) PCR amplification products of (B) short overlapping segments of the first 10 exons from HeLa cDNA, all producing a single product band. ....	26
Figure 2.2 Alternative splicing of MCM9. ....	27
Figure 2.3 Separate quantification of MCM9 <sup>L</sup> , MCM9 <sup>M</sup> and MCM9 <sup>SML</sup> . ....	29
Figure 2.4: MCM9 <sup>L</sup> or MCM9 <sup>M</sup> abundances in cancer cell lines.....	31
Figure 2.5 Relative cancer cell line abundances of MCM9 <sup>L</sup> and MCM9 <sup>M</sup> . ....	32
Figure 2.6 FACS cell cycle profiles of double thymidine synchronized HeLa cells.....	33
Figure 2.7 FACS cell cycle profiles of nocodazole synchronized HeLa cells.....	34
Figure 2.8 Comparison of the abundances of MCM9 <sup>L</sup> and MCM9 <sup>M</sup> , assessed by qRT-PCR.....	35
Figure 2.9: MCM9 <sup>L</sup> and MCM9 <sup>M</sup> abundances following MMC or HU treatment. ....	38
Figure 2.10 Conservation of exon-intron boundaries for <i>HsMCM9</i> . ....	41

Figure 2.11 MCM9 <sup>M</sup> C-terminus sequence alignment. ....	43
Figure 3.1 Pedigree of a family with three daughters afflicted by premature ovarian failure and homozygous for the <i>MCM8</i> c.446C>G variant. ....	51
Figure 3.2 Pathogenic homozygous recessive variants in <i>MCM9</i> in two consanguineous families of Turkish descent. ....	54
Figure 3.3 Complex <i>de novo</i> chromosomal abnormality in AII7. ....	56
Figure 3.4 Region of homozygosity on chromosome 6 in Family A. ....	62
Figure 3.5 <i>MCM9</i> c.1732+2T>C results in abnormal alternative splicing. ....	63
Figure 3.6 Sanger sequencing identified abnormal splice variants of <i>MCM9</i> in family members carrying the c.1732+2T>C mutation. ....	64
Figure 3.7 Evolutionary comparison of the selected MCM8 region across species. ....	66
Figure 3.8 MCM8 mutation impairs DNA break repair. ....	70
Figure 3.9 Mutations in human <i>MCM9</i> impair repair of MMC-induced chromosomal breaks. ..	71
Figure 3.10 MCM8 and MCM9 Mutations Disrupt Foci Formation and DNA Binding. ....	77
Figure 3.11 <i>MCM8</i> c.446C>G (p.P149R) inhibits DNA binding by EMSA. ....	78
Figure 3.12 Homology model of human MCM8. ....	80
Figure 3.13 MCM8(1-368) microcrystals. ....	81
Figure 3.14 MCM8 p.E341K disrupts damage-dependent foci formation. ....	82
Figure 4.1 Formation of MCM9 foci at sites of DNA double-strand breaks. ....	96
Figure 4.2 MCM9 NLS <sup>-</sup> disrupts nuclear import. ....	98
Figure 4.3 MCM9 BRCv and NLS motifs. ....	100
Figure 4.4 MCM9 BRCv <sup>-</sup> disrupts damage-dependent foci formation. ....	101
Figure 5.1 Chemical structures of cytotoxic agents. ....	104

Figure 5.2 MMC cytotoxicity in cells homozygous for MCM8 p.P149R.....	109
Figure 5.3 BMN673 cytotoxicity in EBV-immortalized lymphocytes homozygous for MCM8 p.P149R.....	109
Figure 5.4 Synergistic cytotoxicity of MMC and BMN673 in EBV-immortalized lymphocytes homozygous for MCM8 p.P149R.....	110
Figure 5.5 Cytotoxicity of BMN673 in PE01 cells.....	111

## PREFACE

I am grateful for the growth that has come about through my graduate education. I have received far more than technical training, as I have learned to creatively design research projects and to appreciate challenges. This growth would not have been possible without the relationships I have been fortunate to develop. I am forever grateful to my husband Mark, for his boundless optimism and kindness, his unending support and for partnering with me in all endeavors.

I would like to thank Dr. Trakselis for the opportunity to be a part of this research group, for encouraging me to persistently address challenges and for his positive outlook and creative approach to overcoming obstacles. His mentorship has allowed me to grow toward independence as a scientist throughout my graduate training. The Trakselis lab is an environment that fosters creative exploration, and I am grateful to all current and former Trakselis group members for their input, guidance and friendship.

I would like to thank Aleksandar Rajkovic and Michelle Wood-Trageser for pursuing our collaboration and approaching research with persistence, creativity and cooperation. I am grateful to all Rajkovic group members for including me in group meeting discussions and welcoming me as a collaborative group member.

I would like to thank my dissertation committee members, Seth Horne, Xinyu Liu and Aleksandar Rajkovic for their mentorship, constructive criticism and guidance.

This work is the result of significant collaboration. I would like to thank Andrea Berman and Roni Lahr for collaborating to crystallize MCM8, Simon Watkins for collaborating on fluorescence imaging, Rob Sobol for collaborating to design lentiviral infection systems and

Saleem Khan, JoAnne Flynn and Amy Fraser for graciously allowing me to use laboratory instrumentation.

## **1.0 INTRODUCTION**

### **1.1 EUKARYOTIC DNA REPLICATION**

#### **1.1.1 Eukaryotic DNA Replication and MCM Proteins**

DNA replication is an essential and delicately complex process in all living cells. It is vital to cellular survival and genomic stability that faithful genome duplication occurs exactly once per cell cycle in S phase (1-4). Aberrant DNA replication has been linked to several cancer types, and certain replication factors, when overabundant, have been identified as potential clinical biomarkers in cancerous tissues (2, 5-8). Through a variety of mechanisms, eukaryotic cells limit DNA replication to S phase (2). Both timing and location of DNA replication initiation are used as mechanisms of cell cycle control. Minichromosome Maintenance (MCM) proteins are known to play important roles in normal replication as well as homologous recombination for repair of double-strand DNA breaks.

##### **1.1.1.1 Replication Fork and Replication Origins**

Eukaryotic genomes are replicated from multiple origins of replication, located throughout the genome (9-11). The controlled recruitment and function of a diverse group of catalytically active proteins to multiple origins of replication is necessary to initiate faithful genome replication during S-phase (12). The assembly of these proteins, termed the pre-replication complex (pre-RC), is

complete upon recruitment of the hexameric MCM2-7 replicative helicase (12, 13) to chromatin; a process defined as the licensing of replication origins (14, 15). Activation of MCM2-7 as well as replication fork progression occur through the concerted action of Cdc45 and the protein complex GINS (16, 17). GINS is a complex named for the numbers 5-1-2-3 in Japanese (*go-ichi-ni-san*), which refer to its four components, Sld5, Psf1, Psf2, and Psf3 (18, 19). The components of GINS were identified in a series of genetic screens in *Saccharomyces cerevisiae*, originally aiming to understand the function of Dpb11, a regulator of DNA polymerase  $\epsilon$  (20). This screen led to isolation of several Sld (synthetic lethal with Dpb1-11) genes, including Sld5 (20). Subsequent genetic studies identified “Partners of Sld5”, or Psf proteins (21). The GINS complex has been shown to mediate interactions between MCM2-7 and other key factors (19). Following recruitment of the remaining members of the replisome, replication forks progress along double-stranded DNA until they meet head-on, terminating replication (22).

### **(a) Pre-Replication Complex**

The first member of the pre-RC to load onto chromatin is the origin recognition complex (ORC) (**Fig. 1.1**). Chromatin binding of ORC is followed by loading of replication factors Cdc6, Cdt1 and MCM2-7 (23) and is responsible for selecting genomic sites at which to begin replication (24). It is important that DNA replication initiation is coordinated with cell cycle progression and that the pre-RC is formed sequentially. Each step of the pre-RC formation occurs only after completion of the previous step, and this ordered nature of pre-RC formation aids in the control of origin licensing (25). ORC is a six subunit protein complex that mediates the loading of MCM2-7 onto chromatin and the subsequent initiation of DNA replication (26, 27). The six subunits of ORC are AAA+ (ATPases Associated with a variety of Cellular Activities) ATPases and require ATP to

bind the genomic origin of replication (28). ORC facilitates recruitment of Cdt1-bound MCM2-7 to chromatin by hydrolyzing ATP only when complexed to Cdc6.

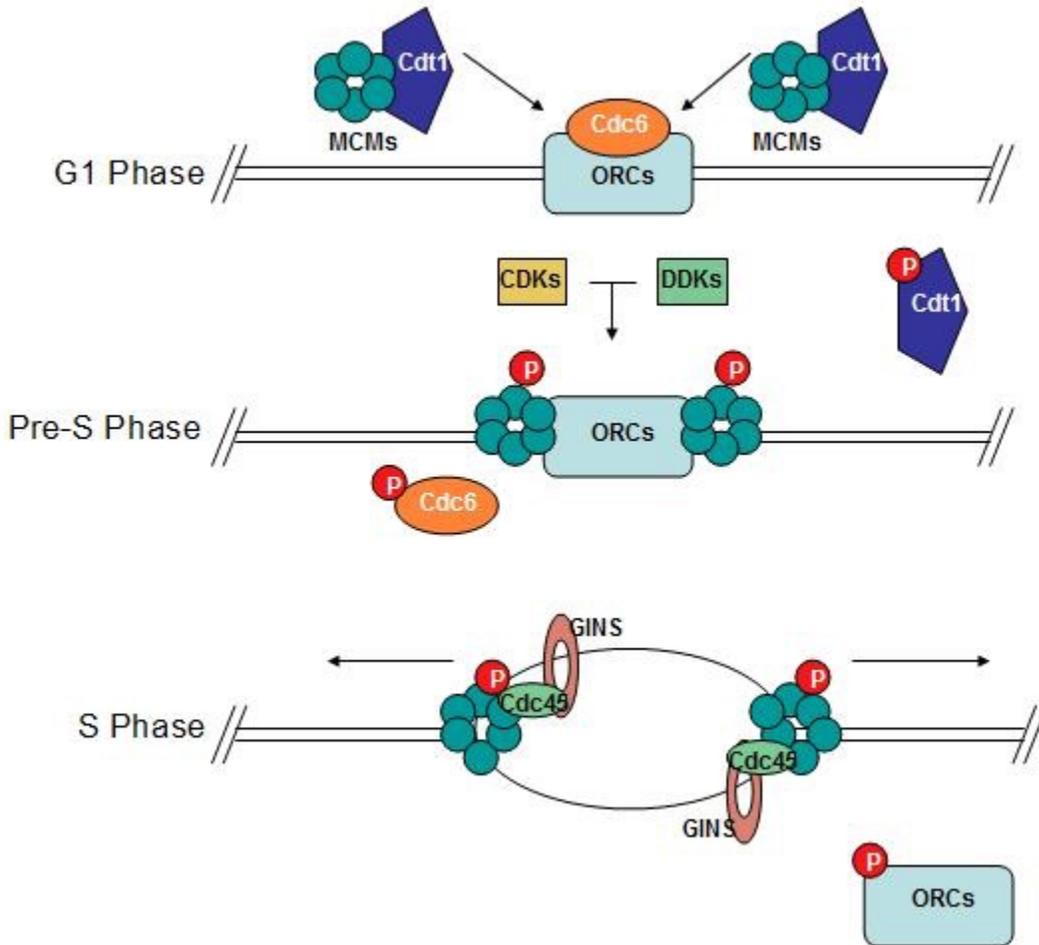


Figure 1.1 Pre-replication complex formation.

Initial stages of pre-replication complex formation occur in G1 phase, when ORCs and Cdc6 bind to chromatin, recruiting Cdt1-bound MCM complexes. CDK and DDK regulate phosphorylation of MCM complexes, Cdc6 and Cdt1, which promotes dissociation of Cdt1 and Cdc6 from chromatin. The replication fork is established in S phase upon recruitment of Cdc45 and GINS, to complete the CMG complex.

Cdc6 is a member of the pre-RC which also contains AAA+ ATPase motifs essential for proper DNA replication (**Fig. 1.1**) (29). The localization of the protein is cell cycle regulated in mammalian cells, being exported from the nucleus as a cell enters S phase (30). This regulation is likely to aid in prevention of unlicensed pre-RC formation. Cdc6 only associates with chromatin after ORC is bound and is required for MCM2-7 chromatin association (31, 32).

Cdt1 was first identified in *Schizosaccharomyces pombe* as a gene necessary for replication. Cdt1 expression is regulated in a cell cycle dependent manner, peaking in G1 phase and decreasing in abundance as cells pass through S phase (33). Cdt1 has since been shown to be a necessary member of the pre-RC, associating with replication factor Cdc6 to cooperatively promote MCM2-7 association with chromatin (34). Cdt1 is conserved throughout eukaryotes and homologs have been identified in *Xenopus laevis*, humans and *Drosophila melanogaster* (4, 34, 35).

### **(i) Origin Licensing**

Cell-cycle coordinated control of pre-RC protein recruitment controls origin licensing. Pre-RC formation is regulated by two protein kinases, cyclin-dependent kinase (CDK) and Dbf4-dependent kinase (DDK) (36). CDK may inhibit MCM loading onto chromatin by phosphorylating ORC, MCM and Cdc6 (37). CDK is also essential for helicase activation, phosphorylating the GINS components Sld2 and Sld3 to promote MCM loading onto chromatin. Both functions of CDK prevent recruitment of MCM2-7 to chromatin (38). DDK promotes replication by phosphorylating MCM4 and MCM6. The activity of these kinases is cell cycle-regulated in order to temporally control the firing of replication origins (36).

Geminin has a suggested role in controlling origin licensing by binding and directly inhibiting the function of Cdt1. Geminin abundance is regulated in a cell-cycle dependent manner, targeted for degradation during M phase by the anaphase-promoting complex (APC) (39). Geminin binds and stabilizes Cdt1 in G2/M phases of the cell cycle, blocking it from reinitiating origins of replication (40, 41). It has been suggested that the stoichiometry of the geminin-Cdt1 complex regulates DNA licensing (42, 43). Geminin accumulates on chromatin prior to genome duplication, following Cdt1 release from chromatin (44). The full-length isoform of the MCM9 protein has

been shown to form a stable complex with Cdt1 in *Xenopus* egg extract assays and this complexation is necessary for recruitment of the MCM2-7 complex to chromatin (41, 45).

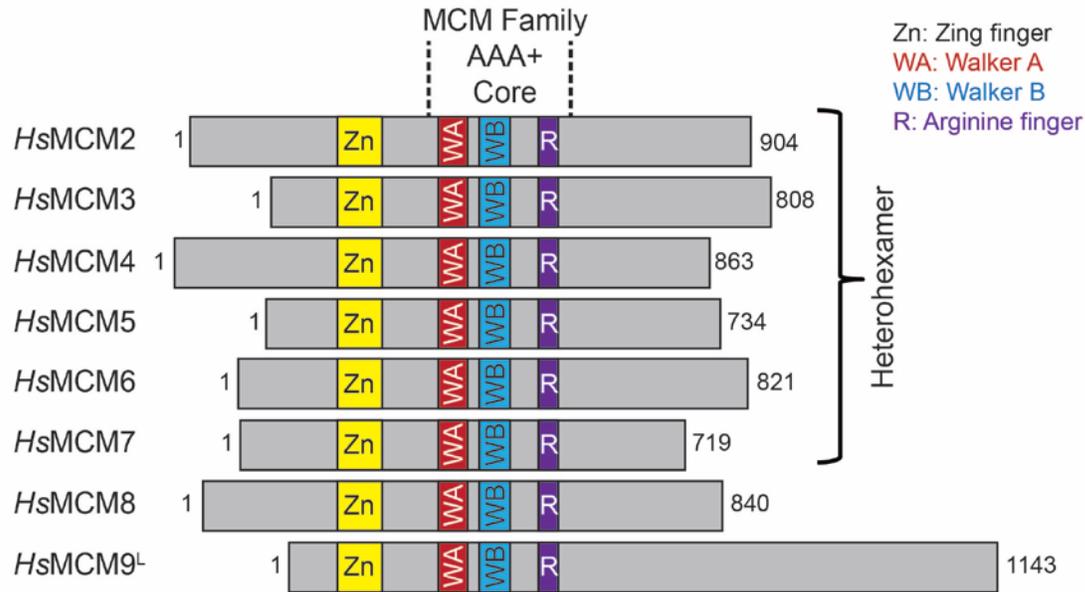


Figure 1.2 Minichromosome maintenance (MCM) protein family. Schematic shows alignment of conserved genetic motifs, including the zinc finger domain (Zn) and AAA<sup>+</sup> core, containing the Walker A (WA), Walker B (WB) and Arginine finger (R) motifs.

### 1.1.1.2 Minichromosome Maintenance (MCM) Proteins

Minichromosome maintenance (MCM) proteins (**Fig. 1.2**) were discovered by randomly mutating a pool of samples of *Saccharomyces cerevisiae* and identifying mutants defective in the maintenance of plasmids (46). Mutations within the MCM genes result in a cell's inability to replicate foreign plasmids. MCM2-9 are classified within the “pre-sensor 1  $\beta$ -hairpin” (PS1 $\beta$ HP) subset of the AAA+ ATPase class of proteins. These proteins are characterized by a conserved core of amino acid motifs (29, 47). The basis of ATPase biochemical activity is an ability to hydrolyze the  $\beta$ - $\gamma$  phosphate bond of a bound adenosine triphosphate molecule, releasing free energy (48, 49). AAA+ ATPases contain a conserved phosphate binding segment (“P-loop”) that interacts with the  $\gamma$  phosphate group of a bound ATP molecule (47, 50). Contained within the P-

loop is the Walker A motif (GxxGxGKS), abbreviated “WA” in **Fig. 1.2** (51). The Walker B motif (DExx, abbreviated “WB” in **Fig. 1.2**), also conserved among AAA+ ATPases is responsible for binding a Mg<sup>2+</sup> cation, which facilitates ATP hydrolysis (48, 51). Sensor-1 (S1, **Fig. 1.3**) contains a conserved polar amino acid residue that senses ATP molecules, and mediates conformational changes in the enzyme in response to ATP (47, 50). In the active MCM2-7 hexameric helicase the conserved polar residue in Sensor-1 is positioned to allow for hydrogen bonding with the  $\gamma$ -phosphate group of a bound ATP molecule (50). Sensor-2 (S2, **Fig. 1.3**), a conserved arginine residue also interacts specifically with the  $\gamma$  phosphate of a bound ATP molecule (47). An arginine finger (RF, **Fig. 1.3**) is contained within the AAA+ conserved sequence, and in hexamerized MCM2-7, protrudes from one member into the active site of another, inducing conformational changes necessary for ATP hydrolysis (47, 52). A zinc finger domain (ZF, **Fig. 1.3**) is contained in the N-terminus of all MCM family members, which serves a role in DNA binding. Also labeled in **Fig. 1.3** are the winged helix (WH) domain and N-terminus to C-terminus linker domains, defined for their predicted structural roles based on their homology to motifs in *Sso*MCM (discussed in Chapter 3).

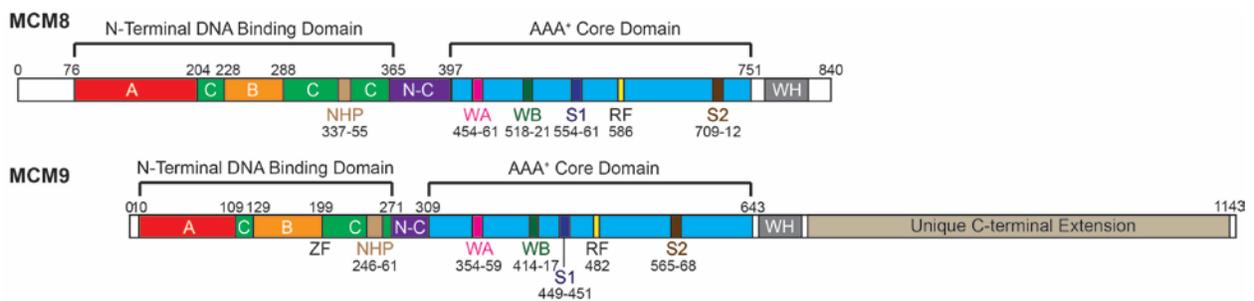


Figure 1.3 Alignment of domains and motifs of MCM8 and MCM9 proteins. ZF: zinc finger domain; NHP: N-terminal Hairpin; N-C: N-terminal to C-terminal linker; WA: Walker A; WB: Walker B; S1: Sensor-1; RF: Arginine Finger; S2: Sensor-2 WH: winged-helix.

### (a) MCM2-7 Replicative Helicase

The MCM proteins that have been most extensively characterized are MCM2-7. The MCM2-7 subunits interact with one another, forming a stable heterohexamer in solution. The MCM2-7 hexamer functions as a replicative helicase, melting double-stranded DNA and allowing for subsequent elongation by a DNA polymerase (25, 53). Helicase activity of the heterohexamer was initially predicted on the basis of genetic motifs, and has been demonstrated in biochemical assays multiple times (13, 54-57). Biochemical assays have shown that an excess of chromatin-bound MCMs is necessary to allow a cell to withstand stress and further, not every MCM2-7 hexamer loaded on chromatin leads to a licensed replication origin (58). This finding raises a question regarding the function of these additional MCM2-7 hexamers, termed the “MCM paradox.” The current prevailing interpretation is that these excess complexes license dormant origins to be used as a backup in the presence of replicative stress (58).

The archaeal MCM proteins serve as simplified models for understanding the function and structure of eukaryotic MCM2-7. The MCM homohexameric helicases found in *Methanothermobacter thermoautotrophicus* (*Mth*) and *Sulfolobus solfataricus* (*Sso*) have been studied to elucidate mechanisms of ATP hydrolysis, DNA binding and unwinding for this class of proteins (59). The AAA+ domain, located in the C-terminus of archaeal MCM proteins, contains conserved motifs essential for ATP hydrolysis, as a lysine residue mutation within this core abolishes unwinding activity (52). Mutational analyses of *Mth*MCM have revealed essential motifs for DNA binding, including the N-terminal hairpin, presensor 1 hairpin and helix 2 insert (59).

MCM hexamers possess a common architecture consisting of a central channel through which the DNA strand passes, an N-terminal domain which stabilizes the DNA and a C-terminal AAA+ ATPase domain (60-62). There is 3'-5' polarity for MCM helicase unwinding, and residues

at the tips of the  $\beta$ -hairpins are critical for DNA binding and ATP hydrolysis in order to direct processivity (52, 63). The MCM hexamer contains two channels through which DNA may pass, the central channel and side channel. Three main models have been proposed for the mechanism of DNA unwinding by MCM hexamers. The proposed models are steric exclusion and wrapping (SEW) (64), steric exclusion and side channel extrusion (60). In the steric exclusion model, the double-stranded DNA is separated before entering the MCM central channel (65). The side channel extrusion model proposes that the MCM hexamer encircles both strands of DNA and strand separation occurs in the core of the MCM hexamer, then the displaced DNA strand is extruded through the side channel (61). Neither model proposes a role for the excluded or extruded strand. The SEW model of DNA unwinding proposes that the displaced DNA strand interacts with key residues on the exterior of the MCM hexamer, contributing to the activity of the helicase. In the development of the model, it was shown that the length of the 5' tail impacts helicase processivity (64).

While the archaeal MCM homohexamer serves as a simplified model by which to understand human MCM2-7, it is clear that the individual subunits of MCM2-7 are different but all essential for eukaryotic helicase activity (59). The three dimer interfaces in the heterohexamer form ATPase active sites, and mutational analysis reveals that these sites contribute unequally to the activity of the helicase, emphasizing the importance of all members of the heterohexamer (66). While a specific subset of MCM subunits (MCM467) is sufficient for *in vitro* helicase activity, this subset differs from MCM2-7 in its ability to bind circular DNA. Biochemical mutational analyses show that the MCM2/5 interface is necessary for MCM2-7 loading and activation (13). This interface forms a “gate” which opens to allow the hexamer to load onto circular ssDNA *in vitro* (13).

### **(b) MCM8 and MCM9**

MCM8 and MCM9 are present only in higher eukaryotes, and their functions remain poorly understood (67). These proteins form a stable nuclear complex, demonstrated by chromatin immunoprecipitation experiments (68). Despite its lack of a nuclear localization signal, MCM8 is localized to the nucleus throughout the cell cycle and bound to chromatin in S phase (69). Further, MCM8 mRNA accumulates in G1/S phase of the cell cycle, suggesting a role for the protein in DNA replication (69). Human MCM8 is implicated in pre-RC formation, as it interacts with the ATPase Cdc6, important for the licensing of origins of replication (70). Studies in *Xenopus* identified a role for MCM8 in the elongation step of DNA replication, but found the protein to be unnecessary for origin licensing (71). In contrast to these studies, experiments in *Drosophila* identified a role for MCM8 in meiosis but showed that the protein is not essential for DNA replication (72).

The initial discovery of MCM9 from homology to expressed sequence tags (ESTs) identified a short isoform (MCM9<sup>S</sup>) of only 391 amino acids that lacks complete conservation of the AAA<sup>+</sup> MCM core motifs (73). Full-length MCM9 protein was subsequently identified in *Xenopus*, and a homology search revealed the fully spliced human isoform (MCM9<sup>L</sup>), 1143 amino acids in length (67). It was shown that MCM9 shares highest total-amino acid identity with MCM8 out of all MCM protein family members. Both MCM8 and MCM9 were initially implicated in DNA replication. Immunodepletion of *X*MCM9 from egg extracts causes interruption of pre-RC formation resulting in inhibition of DNA replication (45). *X*MCM9 binds chromatin in a manner dependent on ORC recruitment and forms a stable complex with Cdt1 presumably within the pre-RC (45). Depletion of MCM8 in HeLa cells inhibits S-phase entry during normal replication,

suggesting a role for MCM8 in formation of the pre-RC (69, 70). MCM8 binds chromatin during S phase, following the formation of the replication fork (69).

The necessity of MCM9 for pre-RC formation cannot be generalized to all vertebrates. Knockouts of MCM8 and MCM9 in mice have shown that neither MCM8 nor MCM9 are essential for DNA replication or organism survival but are needed for efficient homologous recombination (HR) as well as germ-line stem-cell maintenance and/or meiotic recombination (74, 75). Removal of MCM8 or MCM9 results in genomic instabilities, sex-specific tumor development, and sterility. MCM8 and MCM9 have been shown to interact directly and this complex is thought to participate in either preventing or repairing double strand breaks (74, 76). Both MCM8 and MCM9 are localized in cells as foci at DNA damage sites and are essential for the recruitment of other HR genes, including Rad51 (68). It is proposed that the MCM8-9 complex acts in some capacity as an ATP-dependent helicase to resolve complex DNA structures during HR. The contradiction regarding the necessity of MCM9 in eukaryotes could be explained by differences in *Xenopus* cell-free extract replication systems versus assays with whole cells or organisms. Alternatively, specific differences in alternative isoforms of MCM9 may provide differential or synergistic roles.

Although it was initially thought that MCM9 was a vertebrate-specific protein due to its absence in *Drosophila*, nematodes, and yeast (67), it has now been identified in the genome of the invertebrate sea squirt (77) as well as many plants, parasites, and even fungi (78, 79) suggesting wider eukaryotic evolution (80). The collective findings suggest that MCM8 and MCM9 are a more recent evolutionary addition to the MCM family that is required for preventing or repairing DNA damage during replication that may lead to chromosomal instabilities and tumor development.

Comparative genomics and phylogenetic analysis been used to re-construct the evolutionary history of MCM2-9 (78). This analysis revealed that MCM2-7 are likely to have evolved from a single common ancestor, an archaeal-like MCM. MCM8 and MCM9 are widespread in eukaryotes, as they are present in five of the six eukaryotic supergroups (78). It is therefore likely that these proteins arose early in the eukaryotic evolution. MCM8 and MCM9 are co-present or co-absent in all organisms except *Drosophila*, suggesting co-evolution of the two (67). The co-absence or co-presence of MCM8 and MCM9 suggests related functions for the two proteins. MCM8 and MCM9 bear more sequence homology to one another than to any other MCM proteins, as shown by the dendrogram in **Fig. 1.4**, constructed using ClustalW, by comparing the *Homo sapiens* MCM amino acid sequences to one another.

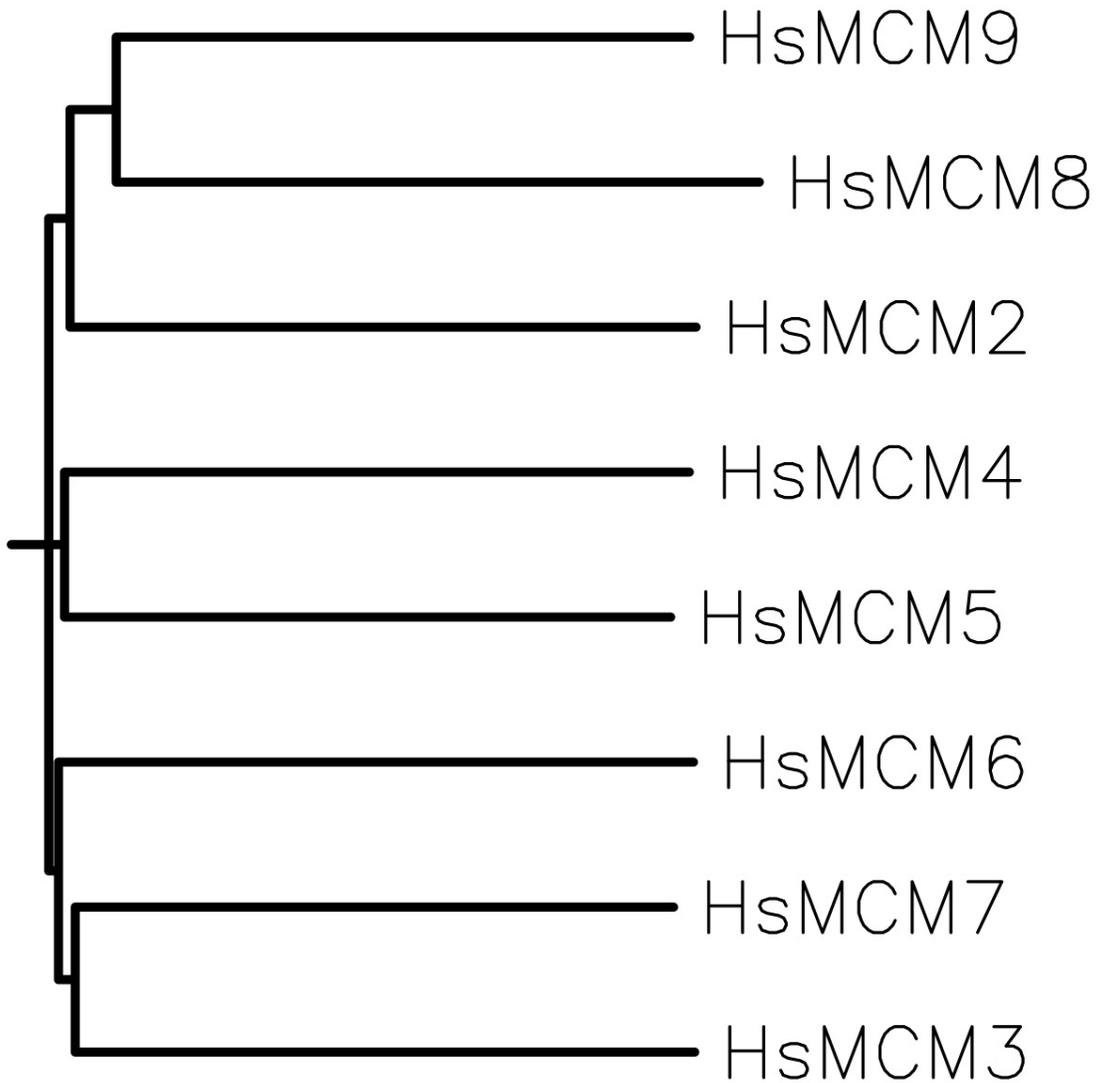


Figure 1.4 Dendrogram of *HsMCM* proteins.  
Dendrogram was constructed using ClustalW, showing the high sequence homology between *HsMCM8* and *HsMCM9*.

In addition to the role established for MCM8-9 in repair of DNA DSBs, the protein complex is likely required for the resolution of double stranded DNA breaks that occur during

homologous recombination between homologous chromosomes in pachytene of meiosis I (discussed in 1.1.6) (72).

Genome-wide association studies (GWAS) directed at identifying factors that determine reproductive lifespan have revealed single nucleotide polymorphisms (SNPs) in several genes, including *MCM8*, which are significantly involved in age of natural menopause (81). In addition to the sterility of *MCM8* knockout mice, it has been established that homozygous mutations in *MCM8* lead to primary amenorrhea in females and small testes and azoospermia in males (82). A nonsynonymous SNP in exon 9 of *MCM8* (rs16991615), p.E341K, is highly associated with the age of natural menopause, and meta-analysis of menopause GWAS studies strongly implicates DNA repair pathways (81, 83-87). The functional relevance of p.E341K in reproductive senescence has not been elucidated.

### **1.1.2 Repair of DNA Double-Strand Breaks**

DNA damage is an inevitable and potentially lethal threat to the genome of a cell. Lesions in the backbone of the cellular DNA may be caused by replication failure or by exogenous agents, such as ionizing radiation, ultraviolet light or chemical agents. Normal cellular functions also cause DNA lesions, specifically by production of reactive oxygen species as metabolic by-products. Double-strand breaks (DSBs), in which both strands of DNA are severed, are particularly detrimental to a cell. A single unrepaired DSB can result in cell death, (88) and misrepair of DSBs may lead to chromosomal rearrangements, causing genomic instability and cancer (89). Because of the significant threat posed by DSBs, cells have developed pathways for detecting and repairing DSBs in a timely, accurate manner. Two main pathways are available for the repair of DSBs: non-homologous end joining (NHEJ) and homologous recombination (HR). HR requires large regions

of homology and is thus the prevalent repair pathway in S and G2 phases of the cell cycle, due to the availability of homologous donor sister chromatids during these phases.

### **1.1.2.1 DNA-damage Checkpoints**

Regardless of the repair pathway used by a cell, all DSBs must first be sensed to allow for subsequent repair. Checkpoints are delays in cell cycle progress to allow for repair to take place. These checkpoints can signal delay or apoptosis and consist of phosphorylation cascades in which DNA-damage sensors detect unrepaired DSBs and recruit transducer kinases. In mammalian cells, the transducer kinases ATM (ataxia telangiectasia-mutated) and ATR (ataxia telangiectasia and Rad3-related) together with DNA-PKcs (DNA-dependent protein kinase catalytic subunit) play a central role in triggering activation of effector proteins, which repair the DNA damage (9). Several mediator proteins, including BRCA1 modulate the activity of the transducer kinases by interactions with members of the DNA damage response pathway (9, 90, 91). These checkpoint proteins direct DSB repair into one of two pathways (NHEJ or HR) dependent on the availability of a sister chromatid.

### **1.1.3 Non-Homologous End Joining for Repair of DSBs**

In the absence of homology to a sister chromatid, NHEJ repairs DSBs by re-ligating the two broken ends together. As NHEJ does not require regions of homology, it is a more versatile and efficient DSB repair pathway than HR. However, the NHEJ pathway is rather error –prone as significant processing of the DSB ends is required before ligation. If DNA synthesis is necessary for gap filling, the accuracy of this pathway decreases dramatically (9). NHEJ is the primary repair mechanism in vertebrates due to the inefficiency of the homology search given the size and

repetitive nature of the genome (9). This is attributed to the inefficiency of the homology search given the size and repetitive nature of the genome. If the ends of the break contain overhangs, NHEJ may stabilize the ends of the DSB with base pairing. In order to facilitate this, terminal base degradation may be necessary. This degradation must not be extensive because NHEJ end joining is only efficient when overhangs consist of less than four bases (9, 92). It may be necessary for gaps to be filled in by DNA synthesis prior to ligation, although this step often results in insertions or deletions (93). The NHEJ pathway is inherently error prone and can lead to a variety of further genomic instabilities including chromosomal rearrangements.

NHEJ is accomplished by three main protein complexes that function in structural stabilization as well as DNA degradation, polymerization and ligation. The three protein complexes are Ku (including DNA-dependent protein kinase, abbreviated DNA-PK), DNA ligase complexes and the MRN complex, composed of Mre11, Rad50 and Nbs1 (**Fig. 1.5B-D**). The MRN complex tethers the ends together, forming a bridge to structurally stabilize the DNA ends (**Fig. 1.5A**). Ku and MRN are responsible for preventing degradation of the DSB ends by binding them shortly after formation of the DSB (**Fig. 1.5 B**). Ku and MRN also are involved in recruiting, stabilizing and stimulating the ligase complexes at the DSB (**Fig. 1.5C**) (9). If DNA end-processing is necessary to facilitate ligation, Ku and MRN recruit DNA modifying factors to allow for this. The NHEJ process is dynamic, and until ligation occurs successfully, end-processing factors may continue to be recruited intermittently (9).

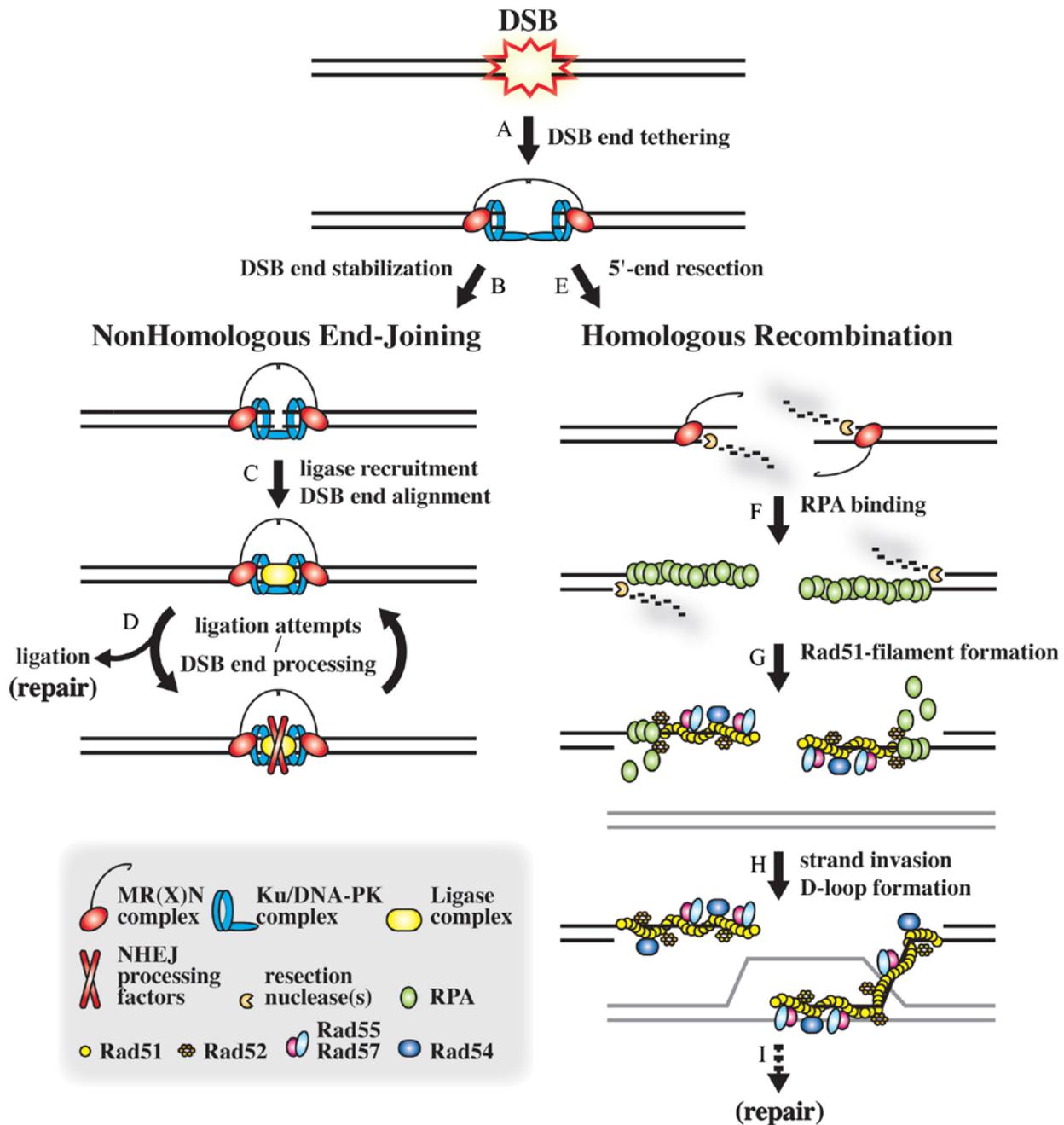


Figure 1.5 DNA Double-strand break repair by non-homologous end joining (NHEJ) or homologous recombination (HR) pathways (9).

The initial step in both pathways is to tether the ends of the DSB by the MRN complex and Ku (specific to the NHEJ pathway) as well as DNA-PK complexes (A). (B) In the NHEJ pathway, the DSB ends are further stabilized by MRN and Ku/DNA-PK. (C) MRN and Ku/DNA-PK recruit DNA ligase and ends of the DSB are aligned. (D) DSB ends are ligated or are processed prior to ligation. (E) In the HR pathway, 5' DSB ends are resected by MRN and other nucleases. (F) RPA binds to ssDNA ends remaining following resection. (G) RPA-coated ssDNA is a substrate for Rad51-filament formation. (H) Rad51-filament homology search and strand invasion produces the D-loop. (I) From the D-loop, different specific HR pathways can result in DSB repair.

#### **1.1.4 Exogenous DNA Damaging Agents**

Cellular exposure to several exogenous agents is known to cause DNA double-strand breaks by a variety of mechanisms. Mitomycin C (MMC) is a potent DNA damaging agent that efficiently crosslinks DNA with specificity for CpG sites (94). While MMC was initially discovered as an antibacterial agent, its detrimental effects extend to mammalian cells as well. MMC causes mutagenesis and chromosome breakage in mammalian cells and has been used clinically as an antitumor agent (95). MMC operates by a mechanism in which it crosslinks the complementary strands of the DNA double helix, and is highly lethal to the cell (94). If left unrepaired, interstrand crosslinks are converted to DNA double-strand breaks during DNA replication (96). When the replication fork encounters these single-strand interruptions, the fork collapses, leading to double-strand DNA ends.

Exposure to ionizing radiation directly causes breaks in double-stranded DNA by breaking chemical bonds in the nucleotide structures (97). Often, free radicals are generated as by-products of the chemical bond breakage, and participate in further reactions to damage the molecular DNA structure. Exposure of a cell to camptothecin causes cytotoxicity by inhibiting the function of DNA topoisomerase I, preventing the enzyme from properly relaxing and reannealing newly replicated DNA strands (98).

#### **1.1.5 Homologous Recombination for Repair of DSBs**

HR is defined by three stages: presynapsis, synapsis and postsynapsis (99). The presynapsis stage of HR is identical to the beginning of NHEJ, and DSBs are detected by an interaction of the DNA ends with the MRN complex (100). Once the DSB is detected, a short 3' ssDNA tail is created.

The BLM helicase acts to further resect the end of the DNA break, lengthening the 3' overhang (**Fig. 1.5E**) (101). Following DNA resection, RPA, a single-stranded DNA binding protein is recruited to the overhang to stabilize and protect it (**Fig. 1.5F**). Mediator proteins, including Rad51B-Rad51C, Rad51D-XRCC, Rad52, and BRCA2, promote formation of a Rad51 filament (**Fig. 1.5G**). Rad51B-Rad51C, along with Rad51D-XRCC form a tetrameric complex that binds ssDNA, preserving it during the homology search (91). Rad51C dimerizes with XRCC3, and this heteroprotein complex functions to bind ssDNA and mediate recombination (91). Rad52 interacts with both Rad51 and RPA, and has ssDNA binding and annealing activity. BRCA2 interacts with Rad51, RPA and numerous other factors involved in the Fanconi anemia pathway, and also possesses ssDNA binding activity. Rad51 interacts with BRCA2 through 8 BRC repeat motifs, which are significantly conserved throughout mammalian BRCA2 (102, 103). Homologous motifs, classified as Rad51-interacting peptides (RIP) have been identified in RecQ15 (104). In addition to Rad51 association, there is evidence for these motifs playing a role in presynaptic filament stabilization (105, 106). RAD51 facilitates the search for a homologous segment on the sister chromatid, allowing for strand invasion, formation of the D-loop and homologous recombination (**Fig. 1.5H**) (91). Rad51 is not involved in the NHEJ pathway, but is unique to HR. BRCA1, a breast cancer susceptibility gene, associates and colocalizes with Rad51 (107). If BRCA1 or BRCA2 are depleted or mutated, a cell experiences genomic instability and cancer susceptibility (107). The synapsis stage consists of strand invasion, which forms a D-loop, and displacement of Rad51 to allow for DNA synthesis. The postsynapsis stage consists of branch migration and completion of repair by a crossover or noncrossover pathway (99). It has been demonstrated that MCM8 and MCM9 are important for recruitment of Rad51 to sites of DNA damage (68).

### 1.1.6 Poly(ADP-Ribose) Polymerases Detect DNA Strand Breaks

Poly(ADP-ribose) polymerases (PARPs) are nuclear enzymes that detect and signal DNA strand breaks. While there are a total of 18 members of the PARP family of proteins, PARP-1 has been studied most extensively. PARP-1 efficiently binds DNA breaks through its N-terminal zinc finger and recruits DNA repair proteins to the site of damage (108-110) (**Fig. 1.6**). Upon binding, a post-translational modification occurs on histones H1 and H2B, causing chromatin to relax and DNA to become more accessible (111). PARP-1 undergoes a conformational change and uses NAD<sup>+</sup> as a substrate for synthesis of poly(ADP-ribose) (PAR) oligomers. PAR is responsible for modifying chromatin architecture and recruiting proteins for further chromatin modification as well as recruiting SSB repair enzymes. PARP-1 catalyzes the synthesis of PAR using NAD<sup>+</sup> as a substrate. PAR serves two main roles in response to DNA strand breaks: relaxing chromatin to increase accessibility of DNA and recruiting DNA repair proteins (111, 112). Addition of PAR oligomers usually leads to successful and efficient repair of DNA damage, but in cases of extreme damage may induce apoptosis (113).

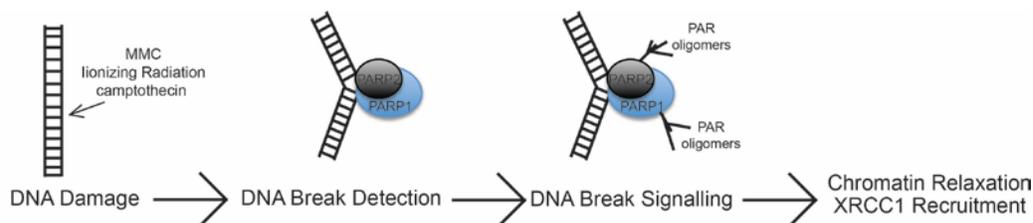


Figure 1.6 Function of PAR in DNA DSB detection and signaling.

After introduction of a DSB by agents such as MMC, camptothecin or ionizing radiation, PARP-1 and PARP-2 bind to the site of damage and poly (ADP-ribose) oligomers are produced on the surface of PARP-1 and PARP-2, which function both in relaxing chromatin to spatially accommodate repair factors and to recruit XRCC1 to the site of damage.

When PARP-1 is inhibited, single-strand breaks collapse replication forks and homologous recombination is triggered to repair the DNA break (114, 115). Because of the initiation of HR in

the absence of functional PARP-1, cells deficient in HR proteins are acutely sensitive to PARP inhibitors. This heightened susceptibility has been demonstrated in BRCA2-deficient tumor cells (115), PTEN-deficient tumor cells (116) and PTEN-mutated cells (117). Because of the high dependence of PARP-inhibited cells on HR, many forms of cancer are more dependent on PARP than are healthy cells. Inhibitors of PARP have been shown to be effective in tumor suppression as they enhance the effects of ionizing radiation and DNA-damaging chemotherapy agents. Currently, several PARP inhibitors are being developed and implemented for clinical treatment of cancer (118, 119).

### **1.1.7 Meiotic Recombination**

Meiosis is a type of cell division by which an organism halves its chromosome number in order to maintain its genomic size in each generation. The process consists of a single round of DNA replication followed by two rounds of chromosome segregation. Most cells contain two copies of each chromosome, but for sexual reproduction these two copies must separate from one another to produce sperm or eggs with one copy of each chromosome. Homologous recombination plays a role in this segregation of chromosomes as it allows for physical connections between chromosome pairs to create genetic diversity. Only physically connected chromosomes will orient properly on the meiotic spindle (120). A role for MCM8-9 in meiotic recombination has been proposed (72), which supports the role of the protein complex in homologous recombination.

Studies in *Saccharomyces cerevisiae* reveal fundamental similarities between double-strand break repair and meiotic recombination (121). While the purposes of these two pathways differ, the underlying mechanisms share many protein factors in common, including Rad51 and BRCA2 (122). The commonality of Rad51 in particular suggests a role for MCM8-9 in meiosis,

given the established role of the complex in HR for double-strand break repair, even though MCM8-9 are absent in *Saccharomyces cerevisiae*. A comparison of the purpose, timing and outcomes of meiotic recombination and HR for repair of double-strand breaks is shown in

**Fig. 1.7.**

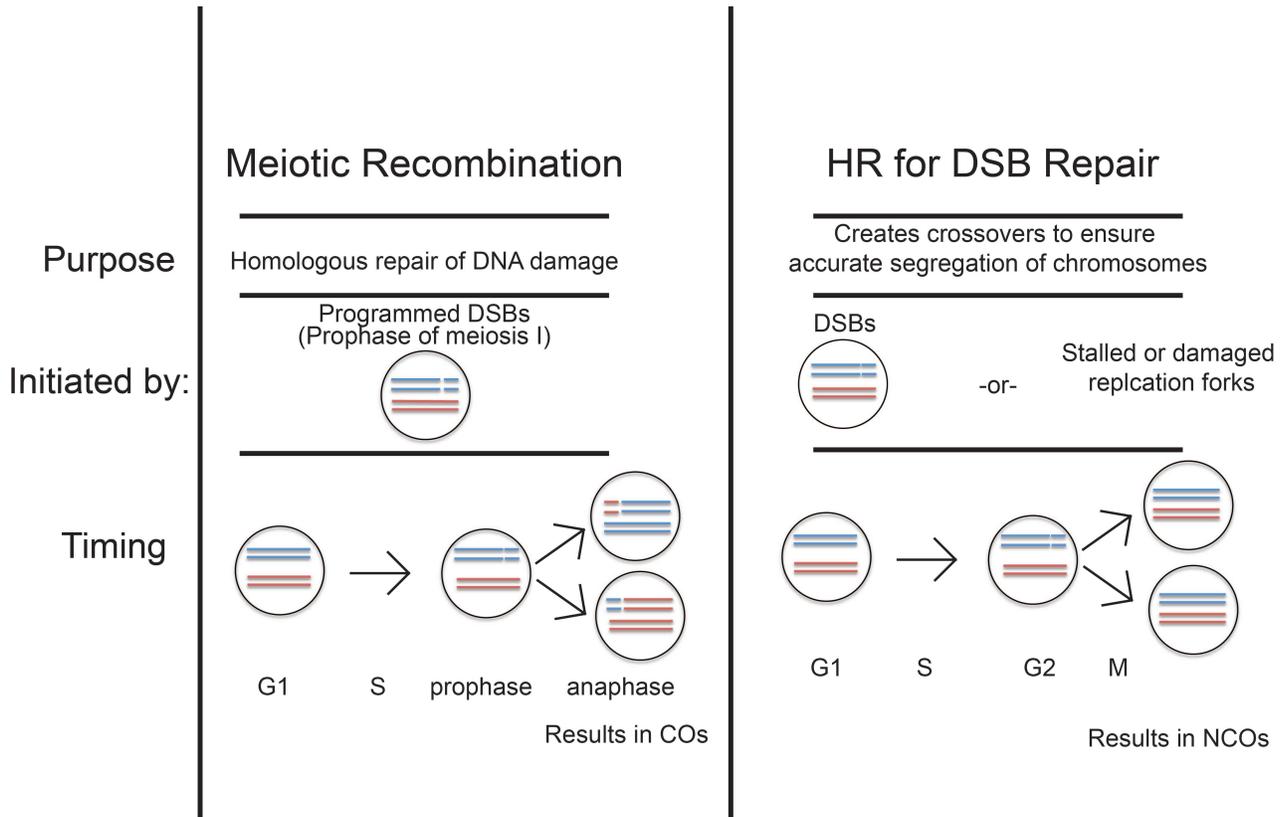


Figure 1.7 Comparison of the purposes, causes and outcomes of meiotic recombination and HR for double-strand break repair.

These processes respectively result in crossover (CO) or non-crossover (NCO) products.

Rad51 and BRCA2 are common to HR for DSB repair and meiotic recombination. Rad51 along with Dmc1, a protein factor unique to meiotic recombination, carry out the homology search in meiotic recombination (122). BRCA2 is important for single strand invasion in *Arabidopsis thaliana*, in meiotic recombination. As discussed above, the MCM8-9 complex colocalizes with Rad51 to sites of DNA damage and is essential for Rad51 recruitment to damage sites. Given the importance of MCM8-9 in repairing DSBs by HR, the MCM8-9 complex is likely involve in the

resolution of double stranded DNA breaks that occur between homologous chromosomes in pachytene of meiosis I. While MCM9 is absent from *Drosophila*, MCM8 has been shown to be essential for 95% of *Drosophila melanogaster* meiotic crossovers (72). In light of this role of MCM8, the effect on meiosis of mutations in *Arabidopsis thaliana* MCM8 was investigated. Interestingly, the introduction of these mutations in MCM8 lead to chromosome fragmentation after the formation of meiotic DSBs, suggesting a role for MCM8 in the repair of DSBs (123). In the absence of Dmc1, mutated MCM8 produces even more chromosome fragmentation. Chromosomal crossover remains uninhibited in the absence of functional MCM8, however, but does not occur in the absence of Dmc1. These findings established a role for MCM8 in resolution of *Drosophila* meiotic DSBs, but not chromosomal crossover. It is suggested that MCM8 and Rad51 act in parallel with Dmc1, resolving DSBs, but are unable to produce crossovers when Dmc1 is absent (123).

Studies in *MCM8*- and *MCM9*-null mice directly reveal a role for MCM8 and MCM9 in meiosis (74). While mouse knockout models of *MCM8* and *MCM9* are viable, all except *MCM9*<sup>-/-</sup> males are sterile (74). In *MCM8*<sup>-/-</sup> males, spermatocytes have persistent DNA damage and are defective in HR, exemplified by their unsynapsed chromosomes. *MCM8*<sup>-/-</sup> ovaries exhibit a block in follicle development as well as tumor development. *MCM9*<sup>-/-</sup> ovaries lack oocytes and testes exhibit a proliferation defect of germ cells. Further, *MCM8*- and *MCM9*-null mouse embryonic fibroblasts exhibit growth defects, chromosomal damage and heightened sensitivity to replication fork stress. Knockout MEFs are also defective in HR-mediated repair of DNA breaks, including an inability to recruit HR factors Mre11, RPA, and Rad51 (74).

## 1.2 GENE TRANSCRIPTION AND RNA SPLICING

Eukaryotic genes are discontinuous, consisting of coding exons separated by noncoding intronic regions (124). When eukaryotic genes are transcribed into precursor mRNA (pre-mRNA), particular exons may be included or excluded, allowing for production of multiple alternatively spliced pre-mRNA isoforms from a single gene. Alternative splicing is a regulated process during transcription that results in a single gene coding for multiple proteins. This is a means of significantly increasing the diversity of proteins synthesized from limited genomic information (125). The control of splicing is an essential aspect of eukaryotic gene regulation, as it is only through alternative splicing that an organism is able to produce the diversity of proteins necessary for the accomplishment of the various functions it requires (126). Alternative splicing is widespread throughout eukaryotes, with 100% of human genes producing more than one alternative mRNA isoform and an average of three alternatively spliced transcripts produced by each human gene (125).

There are multiple modes of alternative splicing, and exon skipping, in which a single exon is alternatively excluded or included in a transcript, is the most common (127). In addition to exon skipping, exons may be mutually excluded to produce alternative pre-mRNA isoforms. In this mode, mRNAs include one of several exons, but no more (127). Alternative donor (5' junction) sites and acceptor (3' junction) sites may also be used to produce multiple mRNA transcripts from a single gene (128, 129). Alternatively spliced transcripts may arise from inclusion of intronic regions, which code for amino acids in these cases(129). All modes of alternative splicing can cause a frame shift in the resulting protein, which can lead to a premature stop codon and loss of protein function.

## **1.3    PREMATURE OVARIAN FAILURE**

### **1.3.1   Clinical Diagnosis and Known Causes**

Premature ovarian failure (POF) affects 1-4% of women, and is defined as a cessation of menses prior to age 40, with elevated follicle stimulating hormone (FSH) and low serum estradiol levels (130). Women with POF present with amenorrhea (primary or secondary), and hypogonadotropic symptoms. POF is genetically heterogeneous (131), with few genes identified, and can be idiopathic and non-syndromic or part of a genetic syndrome. While hereditary factors have been identified in POF, genetic causes are still being investigated.

In 90% of POF cases, the cause is unknown. There are a few causative genetic factors that have been identified. Single genes, such as bone morphogenetic protein 15 [BMP15], diaphanous homolog 2 [DIAPH2], and inhibin alpha subunit [INHA] have been associated with POF and structural abnormalities in the X chromosome may also cause POF (130). Many of the genes associated with POF are involved in repair of DNA breaks by homologous recombination, suggesting that mutations in HR genes increase the likelihood of developing POF. Efforts are being made to define genetic biomarkers of ovarian reserve which may serve as a prediction of ovarian function. Genome-wide association studies have identified approximately 20 genomic loci which correlate with POF, but no single cause has been identified (131).

## 2.0 IDENTIFICATION, QUANTIFICATION, AND EVOLUTIONARY ANALYSIS OF A NOVEL ISOFORM OF MCM9<sup>1</sup>

### 2.1 IDENTIFICATION OF A NOVEL ISOFORM OF MCM9

#### 2.1.1 Research Design

We have identified a novel mRNA isoform of *HsMCM9* resulting from an alternative splicing event in precursor mRNA (pre-mRNA) producing a medium isoform (MCM9<sup>M</sup>) more consistent with the amino acid size of the rest of the MCM family members. Full length MCM9<sup>L</sup> contains 12 contiguous exons, while in MCM9<sup>M</sup>, exon 11 is excluded. This exclusion results in a frame shift from exon 10 to 12, causing a premature stop codon and truncation of the unconserved C-terminal extension. For the first time, we have separately quantified the individual mRNA expression levels of the two functional AAA<sup>+</sup> MCM9 isoforms (MCM9<sup>L</sup> and MCM9<sup>M</sup>) in a panel of human cell lines. The expression profiles of both MCM9 isoforms are cell cycle regulated, peaking during S-phase and returning to basal levels as the cell cycle progresses through G2/M and into G1. The cell cycle profile of MCM9 differs slightly between the two isoforms examined and significant differences are noted in early S-phase and G2/M. Induction of DNA crosslinks using mitomycin C leading to double strand breaks causes MCM9<sup>L</sup> specifically to be transcriptionally activated, whereas stalled replication forks from hydroxyurea treatment had no effect on MCM9

---

<sup>1</sup> This section is taken from “Jeffries EP, Denq WI, Bartko JC & Trakselis MA (2013). Identification, quantification, and evolutionary analysis of a novel isoform of MCM9. *Gene* **519**, 41-49.” Permission to reprint was obtained from Elsevier Journals.

transcription. These results identify, characterize, and validate a novel MCM9<sup>M</sup> isoform that lacks the long C-terminal extension placing it more consistent in length with the rest of the MCM family as well as medium length MCM9 variants in a variety of other eukaryotes. The cumulative expression data suggests that MCM9<sup>M</sup> and MCM9<sup>L</sup> may have different regulated roles for DNA maintenance.

## 2.1.2 Results and Discussion

### 2.1.2.1 Identification of a Novel Human MCM9 Isoform

Because MCM9 has previously been shown to exist in two isoforms (MCM9<sup>S</sup> and MCM9<sup>L</sup>), we investigated the possibility that other isoforms may also exist. By amplifying short

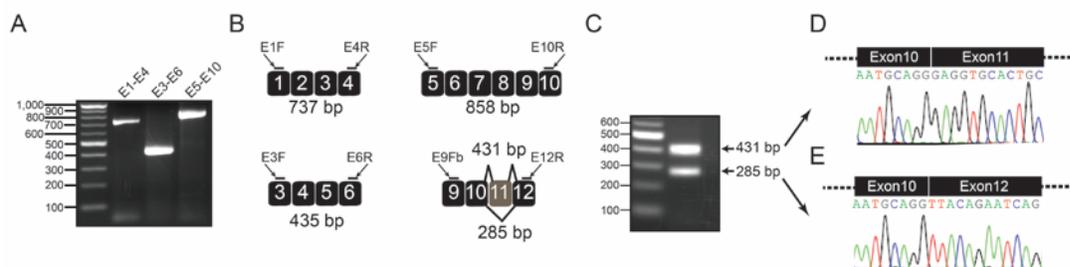


Figure 2.1 Identification of MCM9<sup>M</sup> (A) PCR amplification products of (B) short overlapping segments of the first 10 exons from HeLa cDNA, all producing a single product band. Forward and reverse primers were designed to amplify a 737-bp region spanning exons 1-4, a 435-bp region spanning exons 3-6 and an 858-bp region spanning exons 5-10. (C) Detection of MCM9<sup>M</sup> and MCM9<sup>L</sup> splice variants using E9Fb forward and E12R reverse primers giving 285 and 481 bp products, respectively. DNA sequencing results of (D) MCM9<sup>L</sup> shows the junction of exons 10 and 11 and (E) MCM9<sup>M</sup> shows the junction of exons 10 and 12, skipping exon 11.

regions of MCM9 cDNA spanning multiple exons, we identified a unique mRNA isoform of *HsMCM9* that results from an alternative splicing event (**Fig. 2.1**). Amplification of the overlapping regions spanning exons 1-4, 3-6, and 5-10 all resulted in a single predicted product band, indicating that contiguous splicing occurs within exons 1-10 of *HsMCM9* (**Fig. 2.1A-B**). However, amplification of the region spanning exons 9-12 produced two distinctly sized DNA

products from a single primer pair (E9Fb and E12R). DNA sequencing revealed that the longest fragment (431 bp) contains a fully spliced product, while the shorter fragment (285 bp) lacked all 146 base pairs of exon 11 (**Fig. 2.1 C-E**).

In addition to the two previously identified isoforms of MCM9 ( $MCM9^S$  and  $MCM9^L$ ) (73, 132), we have detected a third isoform ( $MCM9^M$ ) that is the result of an alternative splicing event that effectively removes exon 11. Exons 10 and 12 are directly spliced, shifting the reading frame and resulting in a premature nonsense stop codon within exon 12 after translation of 43 unique amino acids at the C-terminal end. This is predicted to produce an isoform that eliminates the uncharacteristic C-terminal extension of  $MCM9^L$  and is more consistent in size with the rest of the MCM family members (**Fig. 2.2**). Splicing of  $MCM9^M$  is fundamentally different than  $MCM9^S$ , where the intron following exon 6 is retained and becomes part of the mRNA reading frame.

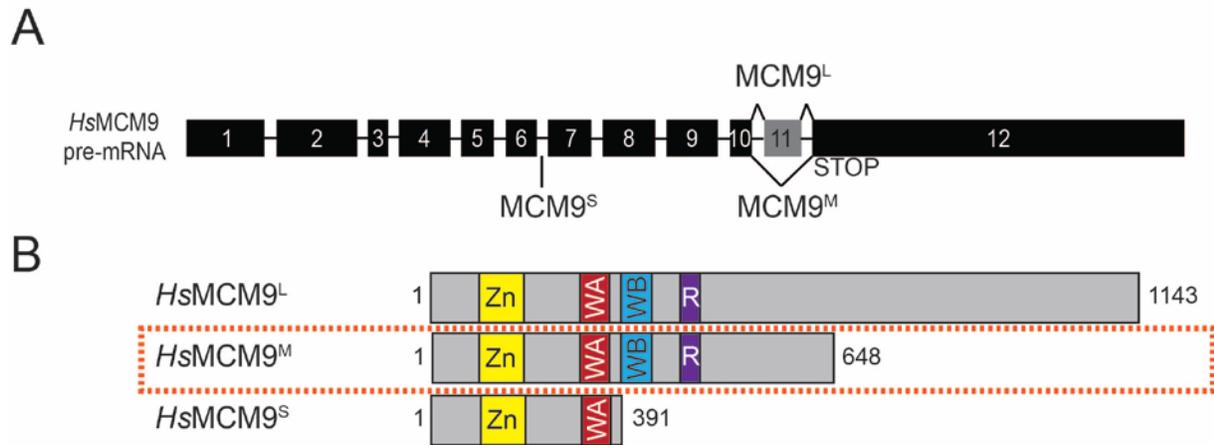


Figure 2.2 Alternative splicing of MCM9.

(A) Alternatively spliced MCM9 pre-mRNAs give rise to (B) multiple predicted protein products. *HsMCM9<sup>M</sup>*, outlined by the dashed red line, is the newly discovered isoform of MCM9.

### 2.1.2.2 Separation and Quantification of Expression Levels of $MCM9^M$ and $MCM9^L$

We optimized the DNA primers to allow for specific amplification of short but similarly sized segments of  $MCM9^{SML}$  (270 bp),  $MCM9^L$  (275 bp) and  $MCM9^M$  (261 bp) (**Fig. 2.3A, C-D**).

Primers E3F and E5R anneal specifically to the region universally conserved among all MCM9

isoforms, producing a single PCR product (**Fig. 2.3B**). For specific amplification of MCM9<sup>M</sup> and MCM9<sup>L</sup>, a common forward primer (E9Fa) is paired with unique regions in the two isoforms: exon 11 in MCM9<sup>L</sup> (E11R) or the region overlapping exons 10 and 12 in MCM9<sup>M</sup> (E11SkipR). The two reverse primers were designed to have similar melting temperatures, and E11SkipR was optimized with regard to the number of bases annealed to either exon 10 or 12 (from 8-16 bases in exon 10 and 6-11 bases in exon 12). To show specificity, each cDNA isoform (MCM9<sup>M</sup> or MCM9<sup>L</sup>) was cloned into pGEMT, and PCR was performed with each possible forward and reverse primer pair giving only the appropriately sized products without any significant background (**Fig. 2.3E**). Total RNA was isolated from cultures of five cancer cell lines: HeLa, MDA-MB-231, MCF-7, PCI-13 and A549 as well as normal skin keratinocytes (HaCaT). Both MCM9<sup>M</sup> and MCM9<sup>L</sup> were specifically detected from cDNA from each of these cell lines in semi-quantitative RT-PCR reactions (**Fig. 2.3F**).

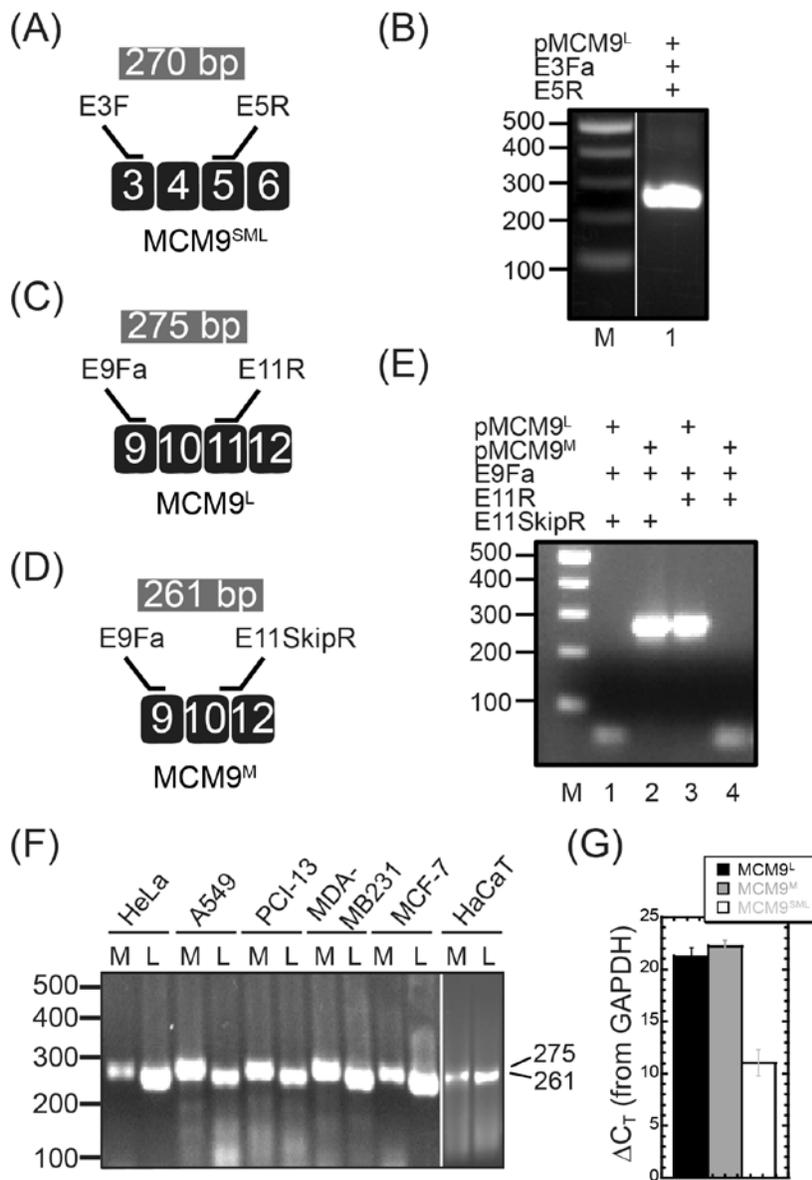


Figure 2.3 Separate quantification of MCM9<sup>L</sup>, MCM9<sup>M</sup> and MCM9<sup>SML</sup>.

Sites of specific RT-PCR primer annealing for (A) MCM9<sup>SML</sup>, (C) MCM9<sup>L</sup> and (D) MCM9<sup>M</sup> resulting in 270 bp, 275 bp and 261 bp products, respectively. (B) Specific PCR amplification of MCM9<sup>SML</sup> from a control plasmid containing MCM9<sup>L</sup>. (E) Specificity of PCR amplification of MCM9 splice variants from control plasmids containing each MCM9 isoform. Lane 1: pGEMT-MCM9<sup>L</sup> with E9Fa and E11SkipR. Lane 2: pGEMT-MCM9<sup>M</sup> amplified with E9Fa and E11SkipR. Lane 3: pGEMT-MCM9<sup>L</sup> amplified with E9Fa and E11R. Lane 4: pGEMT-MCM9<sup>M</sup> with E9Fa and E11R. (F) Two primer sets (E9Fa/E11SkipR for MCM9<sup>M</sup> or E9Fa/E10R for MCM9<sup>L</sup>) were used to specifically amplify each isoform of MCM9 from cDNA isolated from human cancer cells (HeLa, A549, PCI-13, MDA-MB231, and MCF-7) as well as normal skin keratinocytes (HaCaT) in semiquantitative RT-PCR experiments. (G) The three primer sets shown in (A), (C) and (D) were used to amplify MCM9<sup>SML</sup>, MCM9<sup>L</sup> and MCM9<sup>M</sup> by qRT-PCR from HeLa cDNA, respectively.

To more accurately quantify the relative amounts of each mature MCM9 mRNA isoform (MCM9<sup>M</sup> or MCM9<sup>L</sup>) compared with the sum of all isoforms (MCM9<sup>SML</sup>), we used quantitative real-time reverse transcriptase PCR (qRT-PCR). To provide a standard data set and for quantification of relative expression levels, DNA primers were also designed to amplify a 188 bp region of the housekeeping gene GAPDH (glyceraldehyde 3-phosphate dehydrogenase) (133). GAPDH cycle threshold ( $C_T$ ) values were subtracted from *Hs*MCM9  $C_T$  values obtained from a single HeLa RNA template and the resulting  $\Delta C_T$  values are reported in **Fig. 2.3G** and **Table 2.1**. For each averaged  $\Delta C_T$  data point, the error bars are the standard deviation of the mean of at least three independent experiments each performed in triplicate. The level of expression of MCM9<sup>SML</sup> in HeLa cells is consistent with the expression profile data compiled by the European Bioinformatics Institute which quantifies total MCM9 (sum of all isoforms) at least 9 logs lower than GAPDH and 6 logs lower than MCM2 in a variety of cell and tissue types (134).

Cell Line	$\Delta C_T^1$		
	MCM9 <sup>L</sup>	MCM9 <sup>M</sup>	MCM9 <sup>SML</sup>
<b>N. Cervix</b> <sup>2</sup>	12.0 ± 1.1	15.0 ± 1.1	ND <sup>3</sup>
<b>A549</b>	17.6 ± 0.6	23.9 ± 0.8	ND <sup>3</sup>
<b>HeLa</b>	21.2 ± 0.9	22.2 ± 0.6	11.0 ± 1.3
<b>MDA-MB-231</b>	23.0 ± 0.5	24.3 ± 0.5	ND <sup>3</sup>
<b>PCI-13</b>	16.2 ± 0.7	23.1 ± 1.4	ND <sup>3</sup>
<b>MCF-7</b>	21.7 ± 1.2	26.3 ± 0.8	ND <sup>3</sup>

<sup>1</sup>Averages taken from at least three independent experiments performed in triplicate. <sup>2</sup>Normal cervix RNA was purchased. RNA from all other cells lines was isolated from cells grown in culture. <sup>3</sup>ND- not determined.

Table 2.1: qRT-PCR  $C_T$  values normalized to GAPDH  $C_T$  values, reported as  $\Delta C_T$ .

Importantly, the  $\Delta C_T$  values for MCM9<sup>SML</sup> (targeting common exons 3-5) are roughly 10 logs less than for either MCM9<sup>M</sup> or MCM9<sup>L</sup> alone highlighting that multiple isoforms exist and can be quantified either simultaneously or independently with drastically different results.

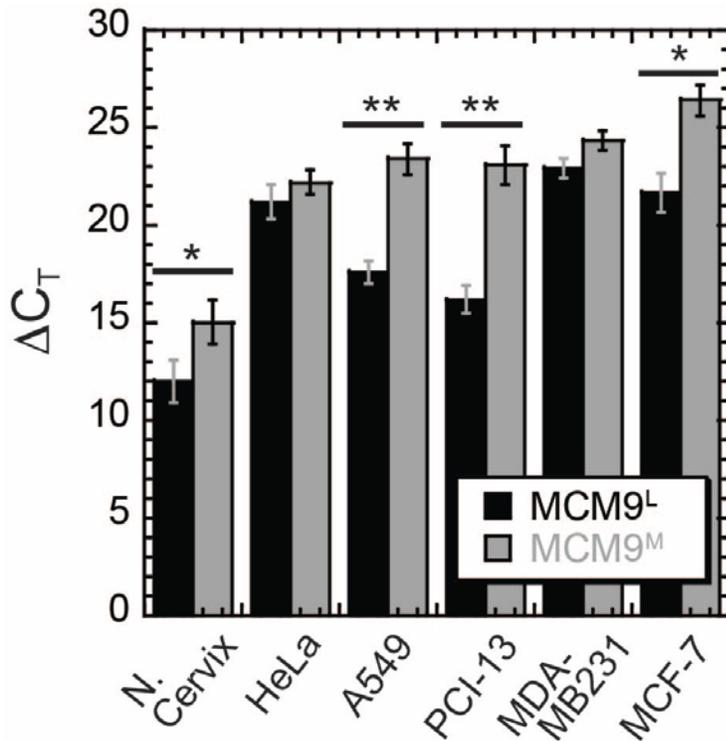


Figure 2.4: MCM9<sup>L</sup> or MCM9<sup>M</sup> abundances in cancer cell lines. qRT-PCR experimental results from total RNA extracted from indicated human cell lines. GAPDH  $C_T$  values were subtracted from MCM9  $C_T$  values giving reported  $\Delta C_T$ . Statistical significance ( $p$ -values) was evaluated for each pair of MCM9 isoforms, MCM9<sup>L</sup> (black) or MCM9<sup>M</sup> (grey), in each cell line (\* =  $p < 0.05$ ; \*\* =  $p < 0.01$ ). At least three individual experiments performed in triplicate were used to calculate the standard deviation.

Further, we individually quantified MCM9<sup>M</sup> and MCM9<sup>L</sup> mRNA levels in each of the above cell lines as well as a commercial normal cervical RNA sample using qRT-PCR (Fig. 2.4). For each RNA template, a  $p$ -value was calculated to determine the statistical significance of the difference between the  $\Delta C_T$  values of the two MCM9 isoforms (\* =  $p < 0.05$ ; \*\* =  $p < 0.01$ ). Both isoforms are in low abundance compared with GAPDH reflected in the positive  $\Delta C_T$  values, but interestingly, the abundance of MCM9<sup>L</sup> is greater than MCM9<sup>M</sup> in every cell line tested.

Statistically significant differences between MCM9<sup>L</sup> and MCM9<sup>M</sup> expression levels were detected at the  $p < 0.05$  (normal cervix and MCF-7) and  $p < 0.01$  (A549 and PCI-13) levels and can be seen when directly compared using a  $2^{-\Delta\Delta C_T}$  plot (**Fig. 2.5**). All relative abundances represented in **Fig. 2.5** represent experiments performed in triplicate. In those RNA templates with significant differences, we found MCM9<sup>L</sup> to be more abundant by 2-7 logs over MCM9<sup>M</sup>. The two breast cancer cell lines (MCF-7 and MDA-MB231) exhibit the lowest abundances of MCM9<sup>L</sup> and MCM9<sup>M</sup>. Expression levels of MCM9 isoforms from commercial normal cervix RNA are significantly greater than any of the other cell lines, but these absolute values should be interpreted with caution as the cells were not grown under the same conditions. qRT-PCR was also attempted on HaCaT RNA without consistently reliable results above threshold values, suggesting even lower expression levels in these cells.

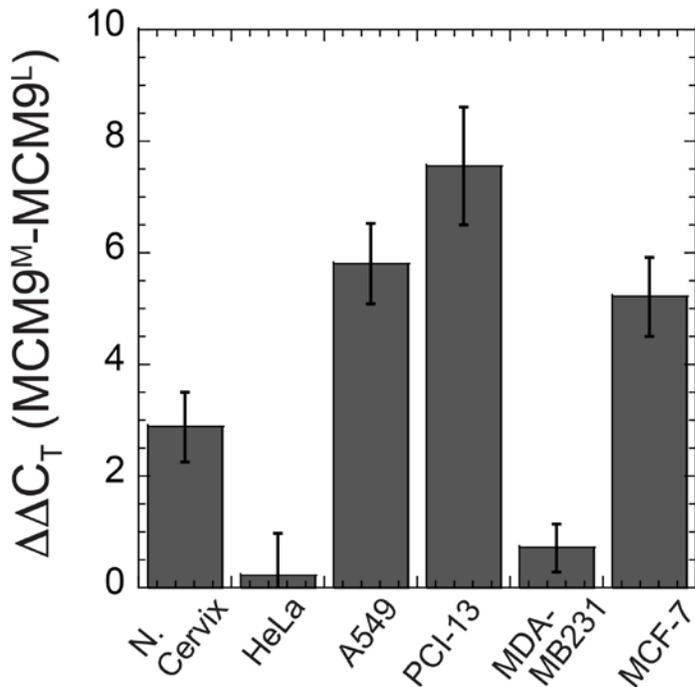


Figure 2.5 Relative cancer cell line abundances of MCM9<sup>L</sup> and MCM9<sup>M</sup>.

MCM9<sup>M</sup>  $\Delta C_T$  values were subtracted from MCM9<sup>L</sup>  $\Delta C_T$  values, as shown in Figure 4 and Table 1, to yield  $\Delta\Delta C_T$  values for each cell line. At least three individual experiments performed in triplicate were used to calculate the standard deviations.

Quantification of the mRNA transcripts shows that MCM9<sup>M</sup> is ubiquitously present at a concentration equal to or significantly lower than MCM9<sup>L</sup> in a variety of cell types.

### 2.1.2.3 MCM9 Isoforms are most abundant in S-phase

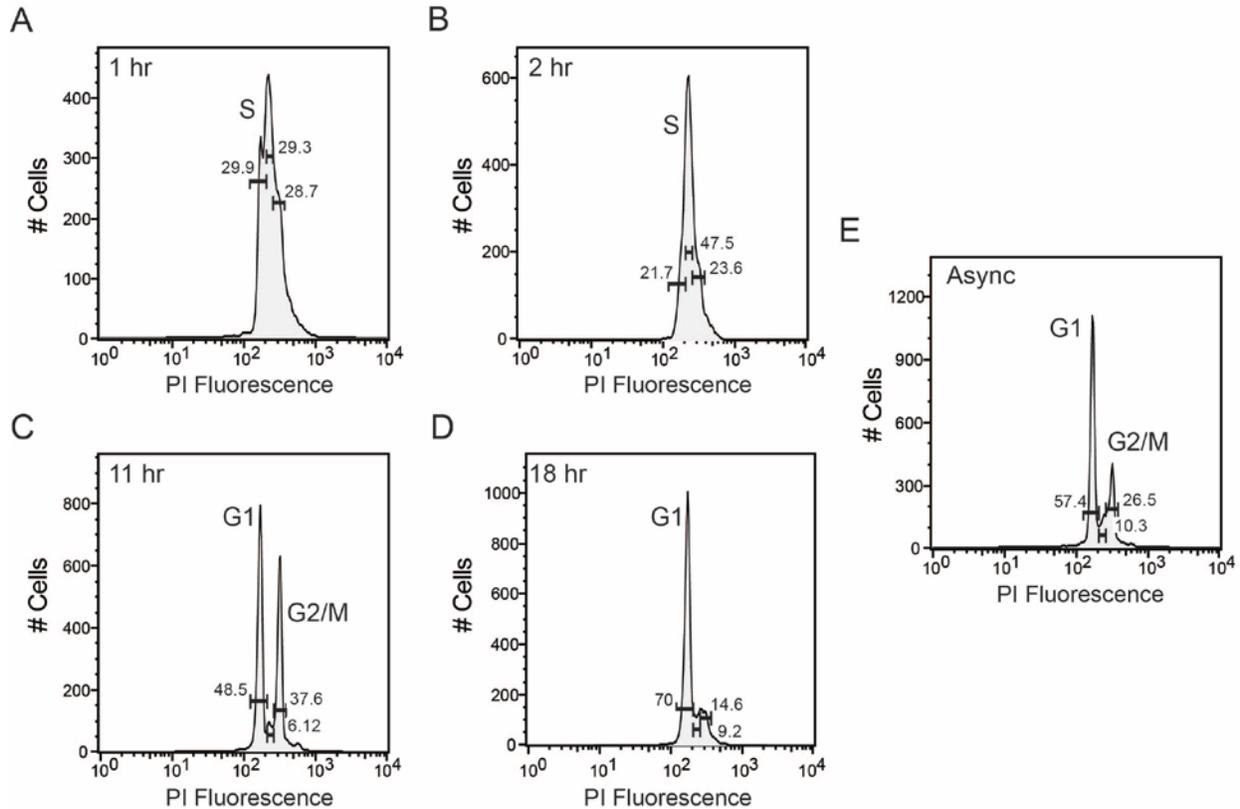


Figure 2.6 FACS cell cycle profiles of double thymidine synchronized HeLa cells. Profiles are of synchronized HeLa cells collected at individual time points after release from double thymidine, stained with propidium iodide (PI) at the following time points: (A) 1 Hr, (B) 2 Hr, (C) 11 Hr, (D) 18 Hr, and (E) asynchronous. The relative populations of cells in each cell cycle phase (S, G2/M, or G1) gated from the asynchronous population are quantified. A total of 10,000 cells were analyzed for each timepoint.

To investigate the cell cycle expression profile of either MCM9 isoform, we synchronized HeLa cells at the beginning of S-phase using a double thymidine block and at the G2-M boundary using nocodazole. RNA was isolated at various time points following release covering the entire cell cycle. The HeLa cell line was chosen for synchronization because of the similarity between MCM9<sup>M</sup> and MCM9<sup>L</sup> expression levels in asynchronous cells. Fluorescence activated cell sorting

(FACS) was used to verify the cell cycle phases of the points after release from synchronization (Fig. 2.6 and 2.7). Release from nocodazole and progress through G1 is slightly longer than normal presumably due to incomplete or slow removal of nocodazole but progressed into S-phase normally and synchronously.

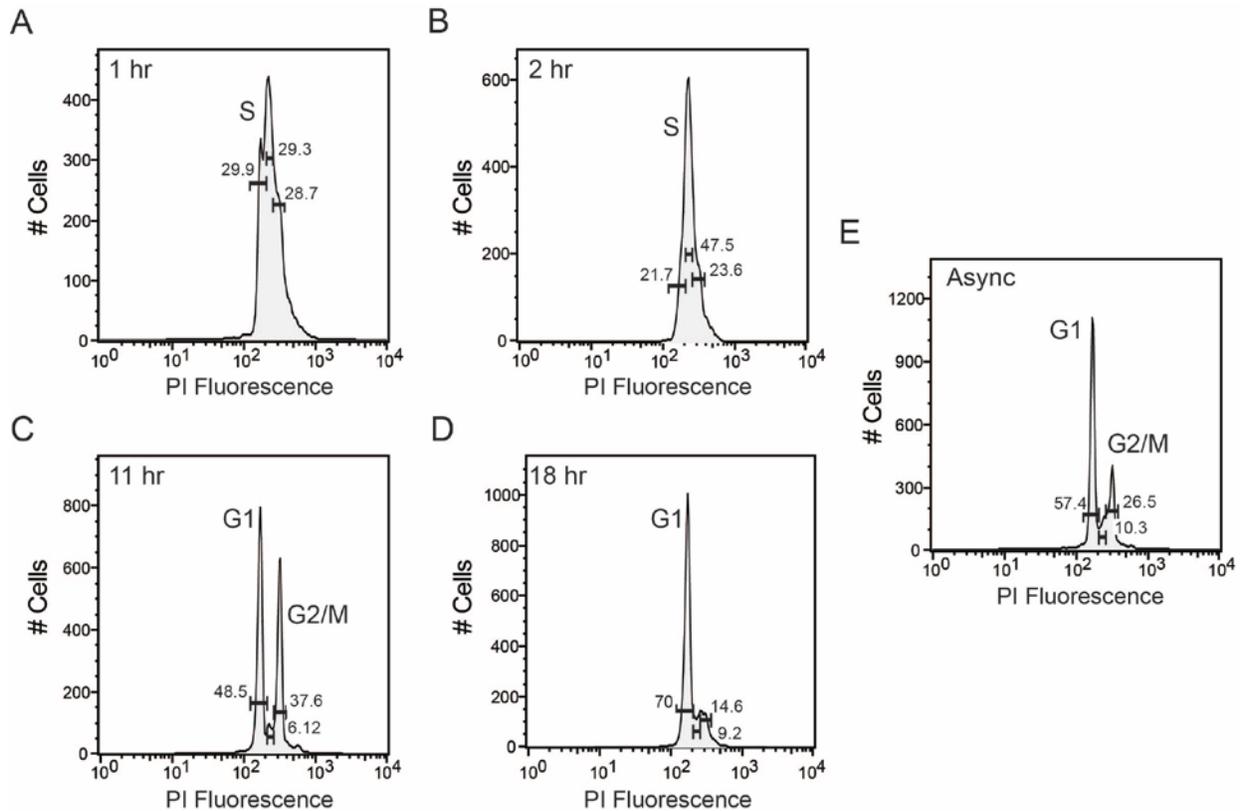


Figure 2.7 FACS cell cycle profiles of nocodazole synchronized HeLa cells.

Profiles are of individual time points after release from nocodazole synchronized HeLa cells, stained with propidium iodide (PI) at the following time points: (A) 4 Hr, (B) 8 Hr, (C) 12 Hr, (D) 16 Hr, and (E) asynchronous. The relative populations of cells in each cell cycle phase (S, G2/M, or G1) gated from the asynchronous population are quantified. A total of 20,000 cells were analyzed for each timepoint.

qRT-PCR experiments were performed as described above for each MCM9 isoform on the isolated RNA from the synchronized time points corresponding to each phase of the cell cycle. Fig. 2.8 and Table 2.2 show the  $\Delta C_T$  values for the two MCM9 splice variants relative to GAPDH across the cell cycle. Statistical significance of transcript abundances are indicated by p-values (\* =  $p < 0.05$ ). Consistent with the relative isoform abundance observed in asynchronous cells,

MCM9<sup>L</sup> is more abundant than MCM9<sup>M</sup> at all points in the cell cycle. Strikingly, both MCM9<sup>M</sup> and MCM9<sup>L</sup> are induced significantly during S-phase before decreasing to basal levels during G2 and into M-phase. However, the trends in expression levels of the two isoforms are not exactly identical. The magnitude of the expression level change is greater for MCM9<sup>L</sup> than MCM9<sup>M</sup> at the start of S-phase. The decreases in expression levels of MCM9<sup>L</sup> and MCM9<sup>M</sup> after S-phase are significantly different compared points in G1 highlighting true changes in transcription after DNA replication has begun. This, along with the lower abundance of MCM9<sup>M</sup> in all RNA samples, suggests that there are significant differences in the transcriptional regulation of MCM9<sup>M</sup> and MCM9<sup>L</sup>.

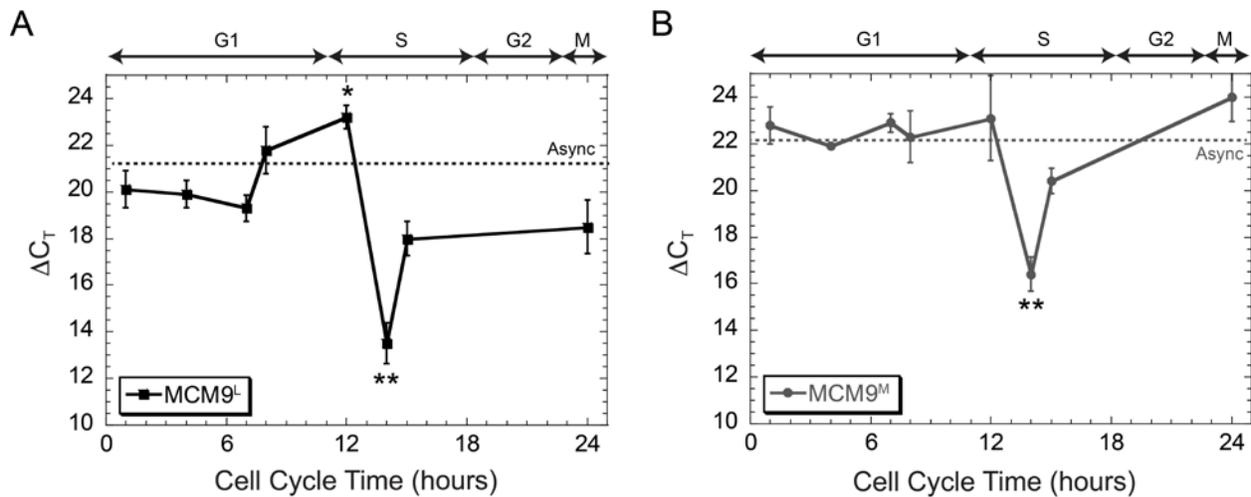


Figure 2.8 Comparison of the abundances of MCM9<sup>L</sup> and MCM9<sup>M</sup>, assessed by qRT-PCR.

Assay was performed using total RNA extracted from time points after release of cell cycle synchronized HeLa cells. The FACS profiles for each time point is shown in Supplemental Figure S1. GAPDH C<sub>T</sub> values were subtracted from those for MCM9 isoforms at each time point and reported as ΔC<sub>T</sub> (also in Table 2). At least three individual experiments performed in triplicate were used to calculate the standard deviations. Time = 0 on the x-axis corresponds to final release from nocodazole (\* =  $p < 0.05$ ).

<b>SSync</b>	<b>Cell Cycle Phase</b>	<b><math>\Delta C_T</math><sup>2</sup></b>	
		<b>MCM9<sup>L</sup></b>	<b>MCM9<sup>M</sup></b>
<b>DDT<sup>1</sup></b>	Mid S	13.6 +/- 0.9	16.4 +/- 0.7
	Late S	18.0 +/- 0.7	20.4 +/- 0.5
	M	17.1 +/- 1.2	24.0 +/- 1.0
	Late G <sub>1</sub>	19.3 +/- 0.6	23.0 +/- 0.4
<b>NN<sup>3</sup></b>	Early G <sub>1</sub>	20.0 +/- 0.8	22.8 +/- 2.2
	Mid G <sub>1</sub>	19.9 +/- 0.6	21.9 +/- 0.1
	Late G <sub>1</sub>	21.8 +/- 1.0	22.3 +/- 1.1
	Start S	23.2 +/- 0.6	23.1 +/- 1.8
<b>Asynchronous</b>		21.2 +/- 0.9	22.2 +/- 0.6

<sup>1</sup>DT – double thymidine synchronization. <sup>2</sup>Averages taken from at least three independent experiments performed in triplicate. <sup>3</sup>N – nocodazole synchronization

Table 2.2: qRT-PCR  $\Delta C_T$  values for synchronized HeLa cells normalized to GAPDH  $C_T$  values.

Expression of MCM9 is induced both in S-phase and after exposure to DNA damaging conditions, validating its importance in DNA maintenance activities. The exact role or roles of the MCM9 isoforms in the cell are currently unknown, but significant differences in expression levels between MCM9<sup>M</sup> and MCM9<sup>L</sup> in some cell types and at certain stages of the cell cycle may indicate that they are individually regulated or utilized. Interestingly, both breast cancer cell lines, MDA-MB231 and MCF-7, have the lowest abundance of both MCM9 isoforms compared to all other cell lines tested. The variations in expression levels between MCM9<sup>M</sup> and MCM9<sup>L</sup> across cell types do not correlate with cytogenetic data for chromosome 6. Therefore, transcriptional levels are regulated directly by specific elements for each cell type.

Our cell cycle results are also consistent with a previous expression profile specifically targeting MCM9<sup>S</sup> from mouse fibroblasts using semi-quantitative RT-PCR that detected transcriptional induction in S-phase after release from G<sub>0</sub> (73). The similarity between our results

for MCM9<sup>M</sup> and MCM9<sup>L</sup> and those for MCM9<sup>S</sup> would suggest that both transcriptional activation and pre-mRNA transcript processing are important for function during S-phase. It is possible that the large induction of MCM9 expression relative to GAPDH in S-phase is a consequence of halting the cell cycle with thymidine for synchronization (135), but this protocol has been used routinely to synchronize cells and assess changes in expression for a variety of transcripts and proteins. Normalization to GAPDH as  $\Delta C_T$  values also minimizes any global expression differences due to synchronization. There are also significant differences in the cell cycle expression profile for MCM9<sup>M</sup> compared with MCM9<sup>L</sup>, but the basic trends are parallel, suggesting that splicing and not transcription provides for these differences.

#### **2.1.2.4 MCM9 Isoform Expression Increases During Crosslink Damage**

After two recent reports that implicated MCM9 in the homologous recombination or crosslink repair pathways (74, 76), we assessed the expression profiles of MCM9<sup>M</sup> and MCM9<sup>L</sup> after exposure to mitomycin C (MMC) or hydroxyurea (HU) treatment. We exposed HeLa cells to 2  $\mu\text{g}/\text{mL}$  MMC for 6 hours or 150  $\mu\text{M}$  HU for 48 hours and individually quantified the expression levels of MCM9<sup>M</sup> and MCM9<sup>L</sup>. MMC induces DNA crosslinking, causing replication to stall, forks to collapse, and the occurrence of subsequent double strand DNA breaks. Treatment with HU at these concentrations (136) acts to deplete dNTP pools stalling replication forks without significant double strand breaks (137). Both MCM9 isoforms show increases in expression after MMC exposure, although only changes in MCM9<sup>L</sup> are statistically significant (**Table 2.3, Fig. 2.9**), as indicated by p-values (\* =  $p < 0.05$ ). The expression of MCM9<sup>M</sup> has a 1.5-log increase while MCM<sup>L</sup> has a 5-log increase. On the other hand, treatment of cells with HU has no significant effect on the expression of either MCM9 isoform (**Fig. 2.9**).

Treatment	$\Delta C_T^1$	
	MCM9 <sup>L</sup>	MCM9 <sup>M</sup>
Untreated	21.2 +/- 0.9	22.2 +/- 0.6
MMC <sup>2</sup>	16.3 +/- 0.3	21.2 +/- 0.9
HU <sup>3</sup>	21.7 +/- 0.07	21.2 +/- 0.2

<sup>1</sup>Averages were calculated from experiments performed in triplicate.  
<sup>2</sup>MMC – mitomycin C, <sup>3</sup>HU – hydroxyurea.

Table 2.3: HeLa qRT-PCR  $\Delta C_T$  Values Following MMC or HU Treatment

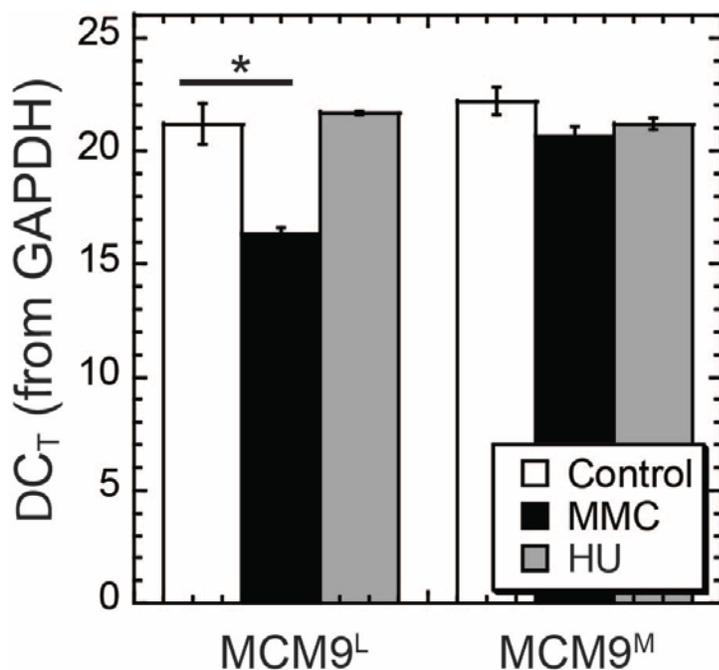


Figure 2.9: MCM9<sup>L</sup> and MCM9<sup>M</sup> abundances following MMC or HU treatment. qRT-PCR experimental results from total RNA extracted from HeLa cells treated with 2  $\mu$ g/ml of mitomycin C (MMC) for 6 hours, 150  $\mu$ M hydroxyurea (HU) for 48 hours, or no treatment. GAPDH  $C_T$  values were subtracted from MCM9  $C_T$  values giving reported  $\Delta C_T$ . Experiments performed in triplicate were used to calculate the standard deviations. Statistical significance ( $p$ -values) of the difference in MCM9 expression between MMC treated cells and untreated cells was evaluated for each MCM9 isoform (\* =  $p < 0.05$ ).

In direct contrast to the work in *Xenopus*, multiple reports show that *Mus musculus* MCM9 (*MmMCM9*) is not required for DNA replication or cell survival although MCM9-deficient mice

had elevated genomic instabilities such as an increased occurrence of sex-specific cancers including carcinomas (74, 75). Qualitative RT-PCR analysis showed expression of MCM9<sup>S</sup> and MCM9<sup>L</sup> over a variety of tissue types and through early development, but did not quantify MCM9<sup>M</sup> expression (75). Genetic ablation mutants designed to target MCM9<sup>SL</sup> or MCM9<sup>L</sup> isoforms also unknowingly disrupted MCM9<sup>M</sup> and resulted in viable mice with germ cell losses and predispositions to cancer (74, 75). Similarly, MCM9 depleted chicken DT40 cells were viable but sensitive to DNA crosslinks. The effects of both mitomycin C and hydroxyurea on MCM9<sup>L</sup> expression quantified in this manuscript are consistent with the increased protein levels and foci found in DT40 cells and validate a transcriptional response of MCM9<sup>L</sup> to crosslink induced double strand breaks (76). Importantly, the combined findings have directly implicated MCM9<sup>L</sup> as a member of the homologous recombination machinery acting to repair DNA breaks or resolve complex DNA structures during replication, but it is not yet clear whether MCM9<sup>M</sup> also participates in this pathway. Future experiments will be aimed at detecting and separating MCM9 protein isoforms and identifying their individual influences on DNA repair once appropriate antibodies are available. It will be important to examine whether the unique C-terminal domain of MCM9<sup>L</sup> is responsible for directing activities through specific interactions and whether any potential role exists for MCM9<sup>M</sup> in cells.

#### **2.1.2.5 Analysis of Intron-Exon Boundaries of MCM9**

Bioinformatics has indicated that up to 90% of genes with multiple exons are involved in alternative splicing in the human genome (128), yet very few alternatively spliced products have been implicated directly in regulating eukaryotic DNA replication or repair (138). Alternative splicing can occur due to many molecular reasons, but both sequence based mechanisms and cis/trans acting elements that either positively or negatively regulate splicing are the predominant

theories (127, 139, 140). Conservation at both the 5' and 3' splice site junctions (AG | GU) is required for efficient sequence based splicing mechanisms (141, 142). The intron-exon junctions are essentially conserved in *HsMCM9* (**Fig. 2.10** and **Table 2.4**). Sequence logos shown in **Fig. 2.10** represent conservation at the intron-exon boundaries for *HsMCM9* for the 3' (**Fig. 2.10A**) and 5' (**Fig. 2.10B**) splice site junctions shown in **Table 2.4**. Logos were made using Weblogo (<http://weblogo.berkeley.edu>).

<b>Exon No.</b>	<b>Intron – exon boundaries (3' splice site – 5' splice site)</b>		
<b>1</b>		AUGAAU – UAUCAG	guaaau
<b>2</b>	caucag	GUUUGC – GAACAG	guaaau
<b>3</b>	ugcuag	GUUCA – AAUCUG	guaaga
<b>4</b>	uccuag	GUGAUG – UUGCAG	guuugu
<b>5</b>	uccuag	GAAGGA – UCAGAG	guuagu
<b>6</b>	uccuag	GAGAAU – GUGCAG	guauug
<b>7</b>	ccucag	GUCUGA – GGCUGG	guaaga
<b>8</b>	uggcag	CCUCGU – AUAAAG	guacgu
<b>9</b>	auacag	GUUACC – CAGAAG	gucuau
<b>10</b>	ucuuag	CUCAUG – AUGCAG	guaagg
<b>11</b>	cuucag	GGAGGU – UGAAAG	gugagg
<b>12</b>	cuacag	GUUACA – UCAUAG	

Table 2.4: Genomic sequence overlapping splice sites in *HsMCM9*

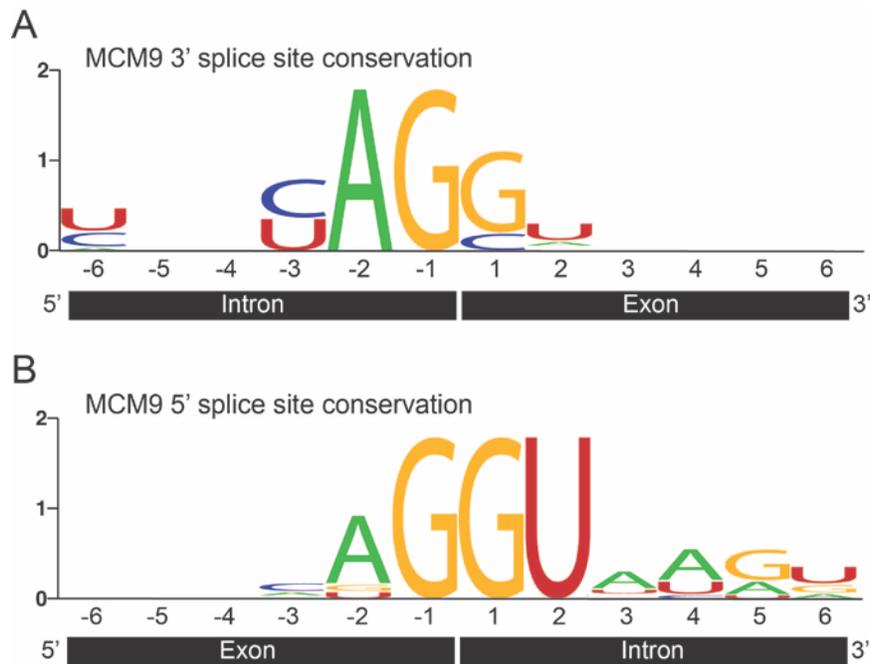


Figure 2.10 Conservation of exon-intron boundaries for *HsMCM9*. Sequence logos showing the conservation at all the intron-exon boundaries for *HsMCM9* for the (A) 3' and (B) 5' splice site junctions shown in Supplemental Table 2. Logos were made using Weblogo (<http://weblogo.berkeley.edu>).

The only slight anomaly occurs at the 3' splice site of Exon 11 which has a G at the +2 position instead of a U, although the +2 residue of the exon is not universally conserved nor required for efficient splicing (142). Instead of a sequence-based mechanism, alternative splicing of *MCM9<sup>M</sup>* is controlled by yet unidentified cis or trans acting elements that repress efficient splicing of exon 11.

#### 2.1.2.6 Evolutionary Analysis of MCM9

Alternative splicing and exclusion of exon 11 will cause a frame shift when translating from exon 10 directly to exon 12, resulting in a premature stop codon at 1944 bp at the beginning of exon 12 (**Fig. 2.2A**). This results in a deletion of the region encoding the unconserved C-terminal extension of *MCM9<sup>L</sup>*. The protein product translated from this alternatively spliced isoform (*MCM9<sup>M</sup>*) will be 648 amino acids in length, compared to the 1143 amino acid full length protein (*MCM9<sup>L</sup>*) with

a long C-terminal extension (**Fig. 2.2B**). MCM9<sup>M</sup> retains all of the conserved DNA and ATP binding motifs and is more consistent in size with the rest of the MCM family. Although we have attempted to detect protein levels of MCM9 using both custom and commercial antibodies targeting the N-terminal domains, we have been unable to reproducibly or specifically detect a band corresponding to 72 kDa (MCM9<sup>M</sup>) or 127 kDa (MCM9<sup>L</sup>) from a variety of cell types. Previous successful detection of MCM9 utilized custom antibodies specifically targeting the C-terminal domain present only in MCM9<sup>L</sup> (74, 76).

The C-terminal extension of MCM9<sup>L</sup> has no known homology or motif that would predict function but may be important for binding other factors (32, 67). After searching the database and aligning sequences using CLUSTAL W2 (<http://www.ebi.ac.uk/Tools/clustalw2>), we found an expressed sequence tag (EST XP\_003586510) from *Bos taurus* that is extremely similar to HsMCM9<sup>M</sup> (**Fig. 2.11A**). The predicted protein sequence of fully spliced *Bt*MCM9 with 13 exons contains 1139 amino acids but does not include the C-terminal amino acid sequence from the EST. After examination of the genomic sequence (143), it is clear that this EST is the result of alternative splicing similar to HsMCM9<sup>M</sup>. In this case, exon 12 was skipped and exons 11 and 13 were spliced producing a frame shift and resulting in the predicted protein sequence from the EST. The result is an extremely similar alternatively spliced isoform of MCM9<sup>M</sup> between humans and cattle.



from these other eukaryotes not only have a similar size, but they also have some conservation of the predicted amino acid sequence at the C-terminus to *HsMCM9<sup>M</sup>* and *DreMCM9<sup>L</sup>*. Due to the translational frame shift that occurs as a result of alternative splicing, this region of *HsMCM9<sup>M</sup>* is unique and not present in *HsMCM9<sup>L</sup>*. More complete amino acid homology at the C-terminus would be predicted if an additional isoform of *HsMCM9* were to be detected that lacks splicing to the terminal exon 12. Therefore, evolution of MCM9 seems to have most recently occurred at the 3' end of the pre-mRNA transcript either through addition of the terminal exon, alternative splicing, or both.

**Fig. 2.11** shows grouped representatives from animalia (blue), fungi (red), and plantae (green). Boxes are color coded for absolutely conserved residues (orange) and highly conserved residues (yellow). Alternative splicing of specific exons is shown in **Fig. 2.11A** for *Hsa*, *Homo sapiens* and *Bta*, *Bos taurus* MCM9<sup>M</sup>. **Fig. 2.11B** shows incomplete splicing of *Dre*, *Danio rerio* MCM9 isoforms that either excludes or includes coding from exon 12. **Fig. 2.11C** shows examples of fully spliced but similarly sized MCM9<sup>M</sup> isoforms from Fungi: *Bde*, *Batrachochytrium dendrobatidis* or Plants: *Rco*, *Ricinus communis*; *Ptr*, *Populus trichocarpa*; *Ath*, *Arabidopsis thaliana*; *Vvi*, *Vitis vinifera*; *Bdi*, *Brachypodium distachyon*; *Osa*, *Oryza sativa*.

### 2.1.3 Conclusion

Although both *HsMCM9<sup>M</sup>* and *HsMCM9<sup>L</sup>* are expressed in a variety of cell and tissue types at a similar but low relative abundance, their cellular importance has only been revealed recently through a variety of knockout studies (74-76). Although we and others (67) suspect that the long C-terminal extension of MCM9<sup>L</sup> will be important for directing other protein interactions, both isoforms contain the essential motifs seen in other ATPases and DNA helicases. Further studies of

the enzymatic roles and regulation of both MCM9<sup>M</sup> and MCM9<sup>L</sup> protein isoforms in controlling DNA replication initiation or repair are needed to identify the mechanisms of action, and future experiments should be designed to differentiate between the multiple MCM9 isoforms to describe their individual roles.

## **2.1.4 Experimental Procedures**

### **2.1.4.1 Materials**

Fetal bovine serum and Dulbecco's Modified Eagle Medium (DMEM) supplemented with D-glucose, L-glutamine and sodium pyruvate were from Atlanta Biologicals (Atlanta, GA). Penicillin-streptomycin and trypsin were from Invitrogen (Carlsbad, CA). Adult human cervix RNA was from Agilent Technologies (Santa Clara, CA). Nocodazole and hydroxyurea were from Fisher Scientific (Waltham, MA) and Mitomycin C was from Roche (Indianapolis, IN). Taq DNA polymerase was from Biorline (Tauton, MA). Propidium iodide was from Sigma-Aldrich (St. Louis, MO). Ultraspec RNA isolation reagent was from Biotecx (Houston, TX). iScript One-Step RT-PCR Kit with SYBR Green was from Bio-Rad (Hercules, CA, USA). M-MuLV reverse transcriptase was from New England Biolabs. All other reagents were analytical grade or better.

DNA Primers	Sequence (5'-3')
E9Fa	5' -GATGAAAACCTATTTCTGCCTCAT-3'
E11R	5' -GCATTCACACCTCCTAGCAGTGCA-3'
E11SkipR	5' -CTCTGATTCTGTAACCTGCATT-3'
E9Fb	5' -CTCATAAGGAATCTGCAGCCCACACT-3'
E12R	5' -AGTAGTCTCTACCTCCAATACCCGTGGTT-3'
GAPDHF	5' -CAGCCTCAAGATCATCAGCA-3'
GAPDHR	5' -TGTGGTCATGAGTCCTTCCA-3'
E1F	5' -GATCTCGAGCATGAATAGCGATCAAGTTACACTGG3'
E3F	5' -GGTTCAAAGGCTATCTGTTGGAAGTAT-3'
E4R	5' -TGCATTACAATCCCGTAAATAGTGAGG-3'
E5F	5' -GGAATGTAATATTGGCTAGCTTGTGCC-3'
E5R	5' -GCTCTTATAGTATTCCCAAAAATCTTCG-3'
E6R	5' -CCCAACCAATAAAAGATGAGATTCTCCT-3'
E10R	5' -GTTACAGTATCACGAAACATCAGGCG-3'

Table 2.5: Oligonucleotides Used to Amplify MCM9 Isoforms.

#### 2.1.4.2 Human Cell Culture and RNA Isolation

Cervical cancer (HeLa), lung cancer (A549), head and neck cancer (PCI-13), breast cancer (MDA-MB-231 and MCF-7) and normal skin keratinocyte (HaCaT) cell lines were maintained in a humidified 5% CO<sub>2</sub> incubator at 37°C. Cells were cultured in Dulbecco's Modified Eagle Medium supplemented with 4.5 g/L D-glucose, 4.5 g/L L-glutamine and 4.5 g/L sodium pyruvate, 10% fetal bovine serum and 5% penicillin-streptomycin. Mitomycin C at 300 nM was added and cells were incubated for six hours, before washing and harvesting. Hydroxyurea at 150 µM was added to the media for 48 hours before washing and harvesting. Cells were grown, harvested by trypsinization, and total RNA was extracted using the Ultraspec RNA isolation system (Biotex, Houston, TX) according to the manufacturer's instructions. Total RNA concentrations were calculated from the absorbance at 260 nm using a DU730 UV-visible spectrophotometer (Beckman Coulter Inc., Brea, CA).

#### 2.1.4.3 Semi-quantitative RT-PCR Detection of MCM9 Splice Variants

cDNA was generated from total RNA using reverse transcriptase according to the manufacturer's instructions. Four regions of MCM9 (E1-E4, E3-E6, E4-E10, and E9-E12) were specifically

amplified from cDNA using Taq. The forward and reverse primers (sequences in Table 5) were designed to overlap exons in the amplified regions. The A-tailed products of MCM9<sup>M</sup> and MCM9<sup>L</sup> were cloned into pGEM-T (Promega, Fitchburg, WI) and sequenced using the DNA sequencing facility at the University of Pittsburgh. For HeLa, A549, PCI-13, MDA-MB-231, MCF-7, and HaCaT cDNA, two sets of primers were used to specifically amplify each isoform. A single forward primer complementary to exon 9 (E9Fa) was used for amplification of both isoforms. A reverse primer specific to exon 11 (E11R) was used to amplify MCM9<sup>L</sup> and a reverse primer complementary to the region bridging exons 10 and 12 (E11SkipR) was used to amplify MCM9<sup>M</sup>.

#### **2.1.4.4 Quantitative Real-Time Reverse Transcriptase PCR (qRT-PCR)**

Human cells were grown and harvested by trypsinization. DNaseI-treated total RNA (8 µg) was subjected to qRT-PCR analysis using iScript One-Step RT-PCR Kit with SYBR Green (Bio-Rad, Hercules, CA, USA) and the real-time thermocycler iQ5 (Bio-Rad, Hercules, CA, USA). GAPDH was used as a reference gene. For qRT-PCR analysis of MCM9<sup>L</sup> and MCM9<sup>M</sup>, oligos E11R and E11SkipR were used, respectively with E9Fa (**Table 2.5**). For amplification of reference housekeeping gene GAPDH, GAPDHF and GAPDHR were used and for simultaneous amplification of MCM9<sup>S</sup>, MCM9<sup>M</sup> and MCM9<sup>L</sup>, oligos E3F and E5R were used (**Table 2.5**). The cycle threshold or C<sub>T</sub> value was determined for each amplified RNA sample and is the number of PCR cycles completed when the amplification curve intersects a specified threshold value. qRT-PCR was performed in triplicate on each plate and relative expression of mRNAs was evaluated by the difference in cycle threshold values from GAPDH ( $\Delta C_T$ ) or between isoforms ( $\Delta\Delta C_T$ ) from at least three independent experiments (100). For all sets of data, *p*-values were calculated comparing the  $\Delta C_T$  value for MCM9<sup>M</sup> with that for MCM9<sup>L</sup> or between different cell samples for

the same isoform using a Student's *t*-test in Excel. A *p*-value of <0.05 was considered statistically significant. The abundance of each target mRNA was determined by  $\Delta C_T = (C_{T(\text{TARGET})} - C_{T(\text{GAPDH})})$  or  $\Delta\Delta C_T = \Delta C_{T(\text{TARGET1})} - \Delta C_{T(\text{TARGET2})}$ .

**HeLa S-Phase and G2/M Synchronization** HeLa cells were synchronized at the beginning of S-phase using a double thymidine block. Adherent cells were grown to 40% confluency in 150 cm<sup>2</sup> flasks with 30 mL growth medium. A final concentration of 2 mM thymidine (Sigma) was added and cells were incubated 20 hours. Cells were washed four times with 10 mL PBS, fresh growth medium was added and cells were incubated nine hours. 2 mM thymidine was added a second time, cells were incubated 15 hours, washed four times with PBS, and harvested by trypsinization at various time points following final release from thymidine with fresh DMEM. To produce cells synchronized at the G2 to M-phase boundary, cells were first synchronized using the double thymidine block described above, and following release from the second thymidine treatment, were washed four times with PBS and incubated in fresh growth medium 6 hours. A final concentration of 0.2 µg/ml nocodazole was then added and cells were incubated 6 hours before being washed four times with PBS and harvested by trypsinization at various time points following release from nocodazole with fresh DMEM. An asynchronous flask of cells was grown simultaneously and harvested at ~90% confluency by trypsinization. Total RNA was extracted and the concentration was determined as described above.

#### **2.1.4.5 Fluorescence Activated Cell Sorting (FACS) Analysis**

Samples of synchronized HeLa cells were permeabilized in 70% ethanol, resuspended in 75 µM propidium iodide, 10 mM Tris pH 7.5, 5 mM MgCl<sub>2</sub> and 10 µg/ml RNase A, and incubated at 37°C for 30 min. Data from 10,000 or 20,000 events were collected with a FACSCalibur (Becton Dickinson Immunocytometry Systems, San Jose, CA) flow cytometer. Forward scatter (FSc) and

side scatter (SSc) were used to gate the cell populations, selecting for unaggregated live cells. Quantification of the cell populations occupying each cell cycle phase from the propidium iodide signal were determined using gating from the asynchronous sample and quantified using FlowJo software.

### 3.0 PATHOGENIC MUTATIONS OF MCM8 AND MCM9<sup>2</sup>

#### 3.1 MCM8 AND MCM9 MUTATIONS UNDERLY OVARIAN FAILURE AND CHROMOSOMAL INSTABILITY

##### 3.1.1 Premature Ovarian Failure in Families with MCM8 or MCM9 Mutations

The role of MCM2-7 in DNA replication elongation and prereplication complex formation has been established. (145) MCM8 and MCM9 are interacting partners that have not yet been characterized so extensively, although they have also been implicated in DNA replication. (69) A pathogenic mutation of MCM8 (c.446C>G; p. P149R) was identified in three sisters presenting with premature ovarian failure (POF) (**Fig. 3.1**), and two pathogenic mutations in MCM9 (c.1732+2T>C and c.394C>T) (**Fig. 3.2**) in two separate families, whose daughters also present with POF. These families will be referred to as Family A (MCM9 c.1732+2T>C), Family B (MCM9 c.394C>T) and Family C (MCM8 c.446C>G; p. P149R).

---

<sup>2</sup> This section is taken from “AlAsiri S, Basit S, Wood-Trageser MA, Yatsenko SA, Jeffries EP, et al. (2015). Exome sequencing reveals MCM8 mutation underlies ovarian failure and chromosomal instability. *J. Clin. Invest.* **125**, 258-262.” and “Wood-Trageser Michelle A, Gurbuz F, Yatsenko Svetlana A, Jeffries Elizabeth P, Kotan LD, et al. (2014). MCM9 mutations are associated with ovarian failure, short stature, and chromosomal instability. *Am. J. Hum. Genet.* **95**, 754-762.” Permission to reprint was obtained from Elsevier Journals. Jeffries was responsible for cloning the mutated forms of MCM8 and MCM9, and performing transfection and foci formation experiments, as well as assessing DNA binding activity of MCM8(1-368).

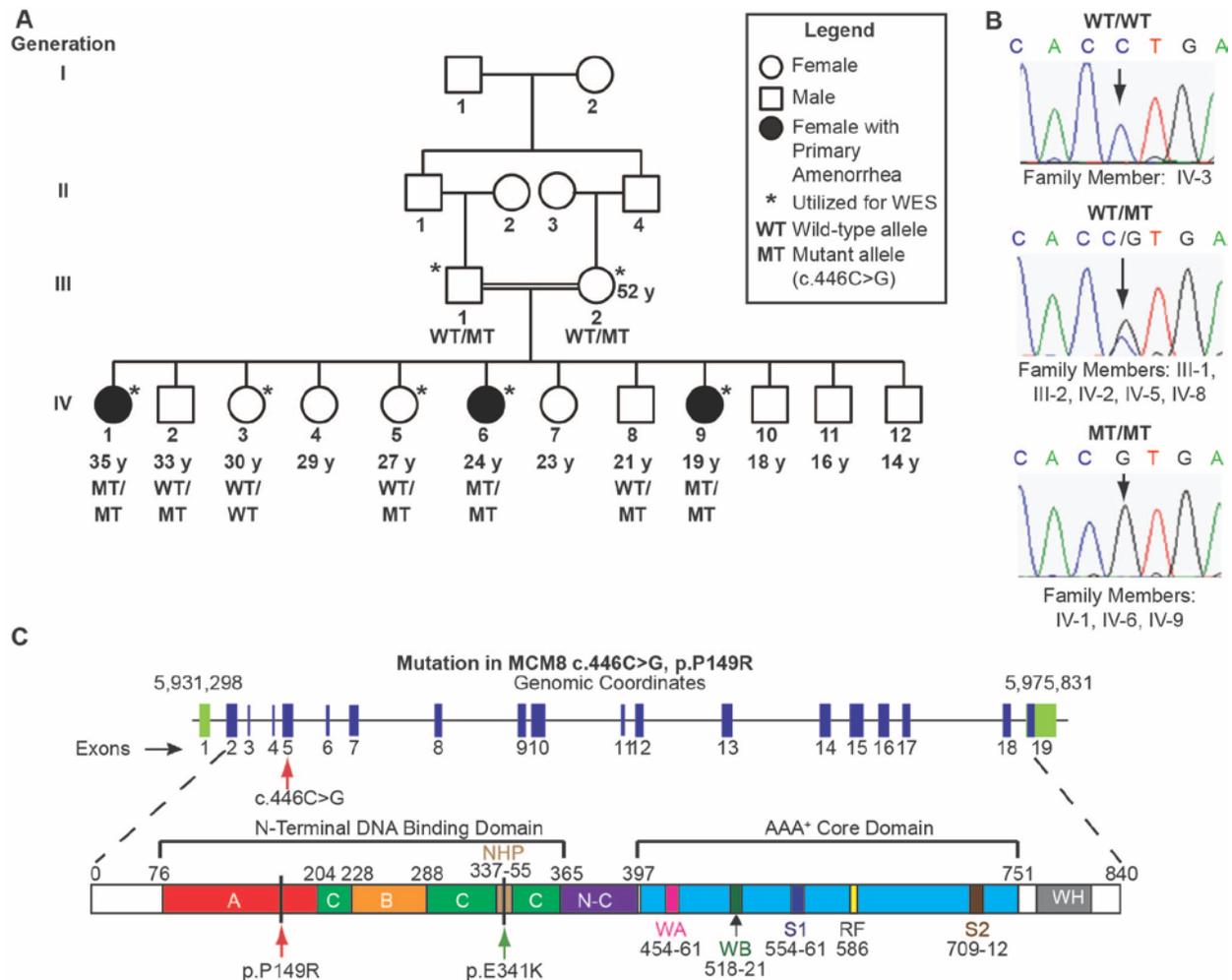


Figure 3.1 Pedigree of a family with three daughters afflicted by premature ovarian failure and homozygous for the *MCM8* c.446C>G variant.

(A) Family members are designated by Arabic numerals. Horizontal lines between individuals represent marriage. Double horizontal lines indicate consanguinity in a marriage. Vertical lines represent lineage. Below each individual, current age (if known) and *MCM8* genotype are provided. (B) Sanger sequencing was used to validate genotypes and representative chromatograms are shown. Individuals who are heterozygous for the c.446C>G *MCM8* variant show overlapping C and G peaks (middle). Individuals homozygous for the c.446C>G *MCM8* variant have a single G peak (bottom). (C) *MCM8* is encoded on chr20: 5,931,293-5,975,831 (NCBI37/hg19). The *MCM8* c.446C>G variant in exon 5 (red arrow). Full boxes represent exons (blue = coding sequences, green = noncoding sequences). Introns are indicated as lines. *MCM8* consists of an N-terminal DNA binding domain and AAA<sup>+</sup> core domain. The c.446C>G substitution causes a change in the amino acid sequence, p.P149R, within the predicted DNA binding domain (red arrow). *MCM8* p.E341K (green arrow) is located within the N-terminal hairpin of the DNA binding domain. All domains are color coded with the homology model (Figure 3.12). NHP: N-terminal hairpin; WA: Walker A; WB: Walker B; S1: Sensor-1; RF: Arginine Finger; S2: Sensor-2.

Premature ovarian failure (POF) affects 1-4% of women, and is defined as a cessation of menses prior to age 40, with elevated follicle stimulating hormone (FSH) and low serum estradiol levels (130). Women with POF present with amenorrhea (primary or secondary), and

hypoestrogenic symptoms (i.e. hot flashes, vaginal dryness, premature osteoporosis). POF is genetically heterogeneous (131), with few genes identified, and can be idiopathic and non-syndromic or part of a genetic syndrome.

Three independent consanguineous families (Family A, Family B and Family C) presented separately for evaluation of their daughters. Affected daughters in these families experience amenorrhea as well as various other symptoms of POF. In the investigation of the cause of symptoms in Families A and B, peripheral blood samples were obtained from affected (AII4, AII6, and BII1) and unaffected family members (AI1, AI2, AII1, AII2, AII3, AII5, AII7, BI1, BI2, BII2, BII3, and BII4). The study was approved by the Ethics Committee of Cukurova University, Faculty of Medicine and informed written consent was obtained from all participating subjects.

Two sisters in Family A and one sister in Family B presented with amenorrhea, as well as short stature and a 46,XX karyotype, all symptoms of POF. A combination of single nucleotide polymorphism arrays, comparative genomic hybridization arrays, and whole exome sequencing analyses identified homozygous pathogenic variants in *MCM9* in affected sisters of both families. In Family A, the *MCM9* c.1732+2T>C variant converts a splice donor site from GT to GC. The splice site mutation causes abnormal alternative splicing that results in transcripts predicted to encode deleted and truncated forms of MCM9 protein. While wild-type MCM9 readily forms foci at sites of DNA damage, the protein resulting from the mutation in which exon 9 is skipped is unable to form damage-induced foci. In Family B, the *MCM9* c.394C>T variant inserts a stop codon in exon 2 (p.R132X) with a predicted loss of functional MCM9 protein. Chromosome break repair was impaired in lymphocytes from affected as compared to unaffected females in both families, consistent with *MCM9* gene function in homologous recombination. Autosomal recessive variants in *MCM9* cause a new genomic instability syndrome associated with hypergonadotropic

hypogonadism and short stature. Preferential sensitivity of the germline meiosis to MCM9 functional deficiency, as well as compromised DNA repair in the somatic component, likely account for the ovarian failure and short stature.

Three sisters (IV-1, IV-6, and IV-9) from Family C are also affected by POF (**Fig. 3.1**). All three sisters have a normal 46,XX karyotype, elevated FSH levels, infantile uteri, and small ovaries (**Table 3.1**). Secondary sexual characteristics were delayed. All three probands are currently treated with estrogen and progesterone replacement therapy and experiencing regular menstrual cycles. All three patients were also diagnosed with hypothyroidism and responded to thyroxine. The mother (III-2) entered menarche at age 14, reported normal pubertal development, regular menstrual periods (26-28 day menstrual cycles) until her mid-40s, and her last recorded menstrual period was at the age of 49. There is no known family history of anemia, blood dyscrasias, photosensitivity, immunodeficiency, or malignancies. We ruled out autoimmune polyendocrinopathy-candidiasis-ectodermal-dystrophy syndrome (APECED [MIM 240300]) by the absence of mucocutaneous candidiasis, hypocalcemia, hypoglycemia, hypotension, vitiligo, alopecia, anemia, or hepatitis in affected daughters. We did not identify pathogenic variants in the gene that causes APECED, *AIRE* (MIM 607358). Anti-thyroid and anti-adrenal gland antibodies were not detected in affected daughters. In summary, the three daughters had idiopathic hypergonadotropic primary amenorrhea with hypothyroidism, atrophic ovaries, and normal female karyotype.

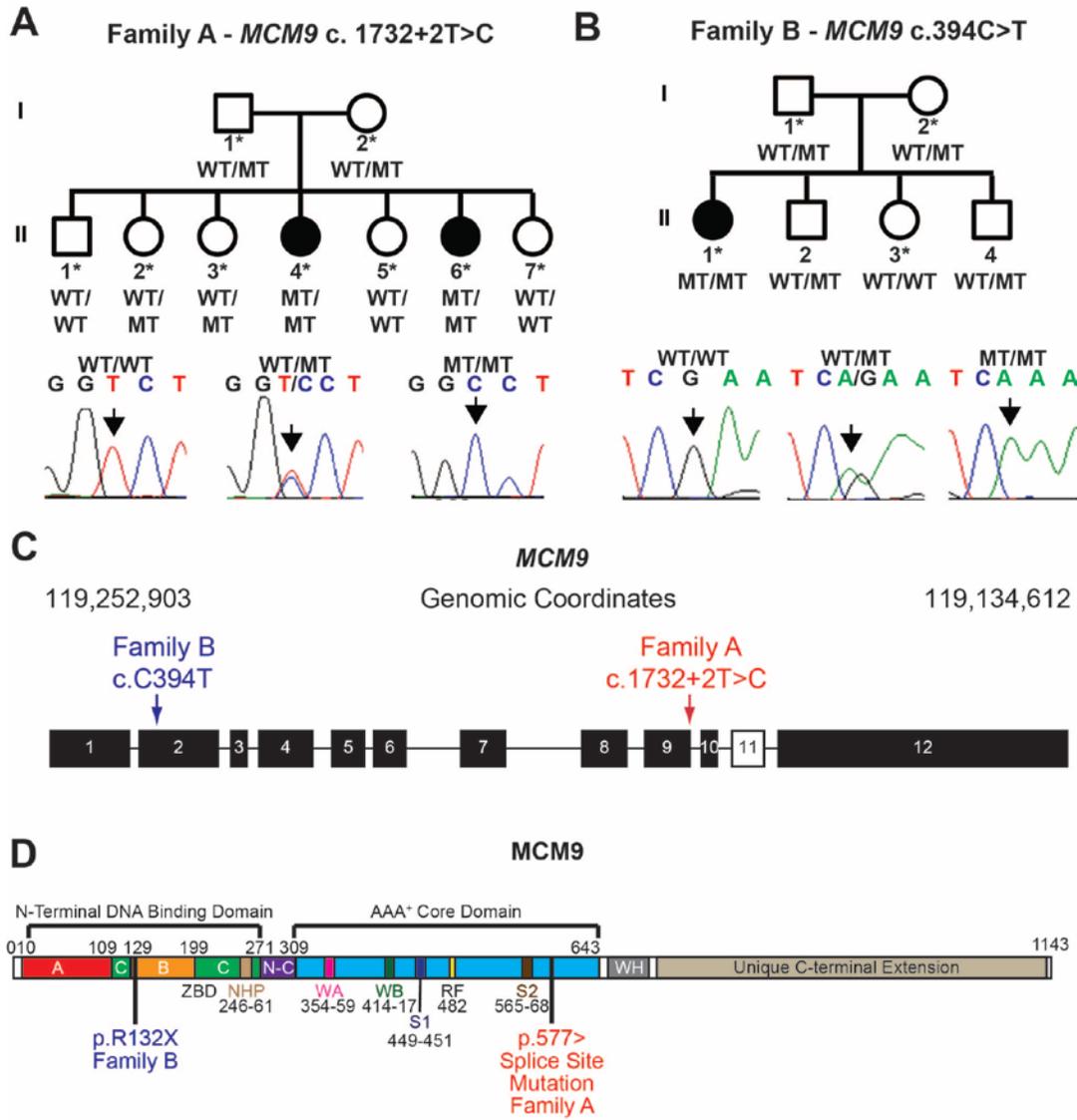


Figure 3.2 Pathogenic homozygous recessive variants in *MCM9* in two consanguineous families of Turkish descent. (A-B) Each generation is designated by a Roman numeral and each family member by an Arabic numeral. Circles represent female family members while squares represent males. Filled circles represent family members affected by primary amenorrhea. Horizontal lines between individuals represent marriage. Double horizontal lines indicate consanguinity in a married couple. Vertical lines represent lineage. Below each individual, their *MCM9* genotype is provided. WT denotes wild type allele. MT indicates either (A) the c.1732+2T>C variant or (B) the c.394C>T variant of the *MCM9* gene. Samples utilized for WES are indicated by an asterisk. Sanger sequencing was used to validate genotypes and representative chromatograms are shown. (A-B) Genotype is indicated as WT/WT, WT/MT or MT/MT for each individual. (B) Chromatograms shown are from sequencing in the reverse orientation of the gene. Arrows point to the nucleotide peak of interest. (C) The *MCM9* c.1732+2T>C splice variant lies in the splice donor site of exon 9 (red arrow). The c.394C>T mutation lies near the 5' end of exon 2 (blue arrow). Exons are indicated as full boxes (black). Alternative splice isoforms of *MCM9* either include or exclude Exon 11 (white). Introns are indicated as lines. (D) *MCM9* consists of an N-terminal DNA binding domain and AAA<sup>+</sup> core domain. The c.1732+2T>C substitution is predicted to disrupt normal splicing after amino acid 577. The c.394C>T mutation causes insertion of a stop codon in the place of an arginine at p.132. This likely results in a truncated protein, if one is produced. Yellow box – Zn Finger Domain, Red box - Walker A, blue box - Walker B, purple box - Arginine finger. NHP: N-terminal hairpin; ZF: Zinc Finger; WA: Walker A; WB: Walker B; RF: Arginine Finger; S1: Sensor-1; S2: Sensor-2.

In family A, parents are first degree cousins. Mother (AI2) entered menarche at age 15, reported normal pubertal development and continues to have regular menstrual periods (26-28 day menstrual cycles) at 49 years of age. She had no difficulties becoming pregnant and is currently 150 cm tall and weighs 68 kg. She gave birth to six girls and one boy. Two of her daughters (AII4 and AII6) presented initially for clinical evaluation due to a Turner-like phenotype due to primary amenorrhea and short stature (**Fig. 3.2A**). The proband (AII4, **Fig. 3.2A**) is now a 23 year old woman who presented with hypergonadotropic primary amenorrhea and short stature at the age of 16, and had normal 46,XX karyotype (**Table 3.2**). Her estradiol levels were low at 7.4 pg/ml with elevated luteinizing hormone (LH; 33.9 mIU/ml) and post-menopausal levels of FSH (73 mIU/ml). Her prolactin level was 9.7 ng/ml (reference range: 3.3-26.7). She had short stature (136.1 cm, -4.3 SD) and low weight (42.3 kg, -2.4 SD). Ovaries were not identified on pelvic ultrasound and the uterus was infantile. The patient is currently regularly menstruating on combined estrogen-progestin replacement therapy and at 23 years of age has achieved a bone age of 18 years, pubic hair stage V, and breast development Tanner stage V. The second affected individual in Family A is a 19 year-old woman (AII6, **Fig. 3.2A**), who was diagnosed at 16 years of age with hypergonadotropic primary amenorrhea, short stature and absent breast development. She presented with short stature (height = 134.4 cm, -4.8 SD), low weight (29.4 kg, -5.7 SD), and a bone age of 12 years (**Table 3.2**). Her basal estradiol level was low (5 pg/ml) with elevated LH and FSH levels (47.3 and 122.5 mIU/ml, respectively). Prolactin levels were within the normal range at 7.1 ng/ml. Ovaries were not detected on pelvic ultrasound. Cytogenetic studies showed 46,XX karyotype. She had stage III pubic hair development, stage II axillary hair, and absent breast development. She is currently treated with estrogen and progesterone replacement therapy, and experiencing regular menstrual cycles, and breast development. There is no known family history

of anemia, blood dyscrasias, photosensitivity, immunodeficiency, or malignancies. In summary, the two sisters had hypergonadotropic primary amenorrhea, short stature, normal uterine anatomy with atrophic ovaries, and normal female karyotype. Three unaffected sisters (AII2, AII3, and AII5), age 22-26, all entered menarche between the ages of 12 and 13 and are currently experiencing regular menstrual cycles. The youngest sister, AII7 presented with developmental delay and was excluded from further genetic analyses due to abnormal *de novo* chromosomal microarray findings (**Figure 3.3**).

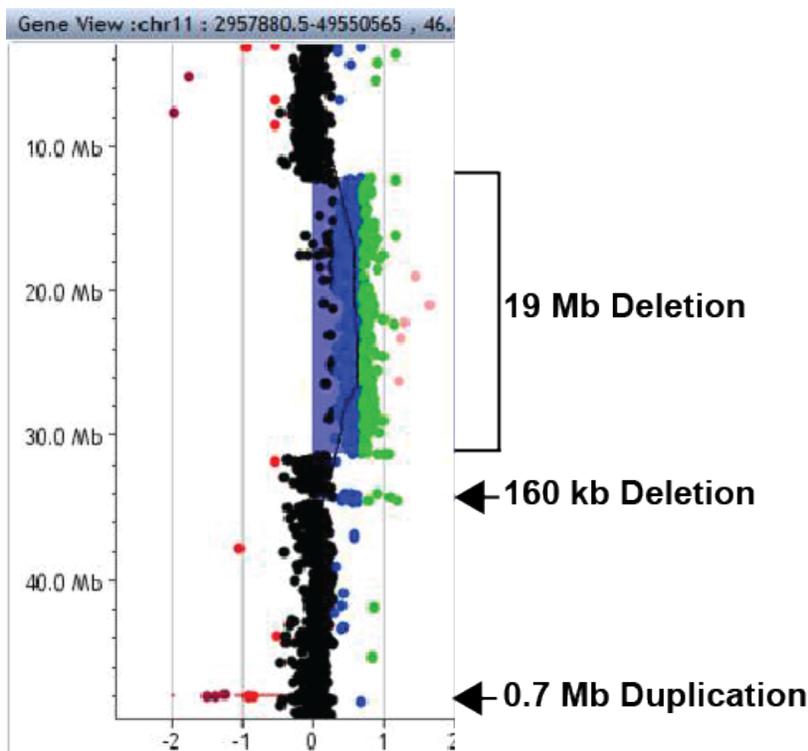


Figure 3.3 Complex *de novo* chromosomal abnormality in AII7.

AII7 presented with developmental delay (Wechsler intelligence scale for children, score = 62) and short stature. She has cubitus valgus, shield chest, widely-spaced nipples, nevi on the face, long eyelashes, and anteverted nares, with normal female external genitalia. Use of CGH array on this patient identified a large 19 Mb duplication (12,258,762-31,345,582), a smaller, 0.7 Mb duplication (33,952,093-34,660,948) and a small deletion (160 kb; 47,912,081-48,073,069) all on chromosome 11 p. The G-banding karyotype showed a 46,XX, dup11p15. This was a *de novo* complex rearrangement, as neither parent carried this complex rearrangement. Black dots represent normal copy number. Blue, green and pink dots represent loss of normal copy number (deletion). Red dots indicate more than normal copy number. Genomic coordinates are based on hg19.

	<b>Normal Range<sup>a</sup></b>	<b>IV-1</b>	<b>IV-6</b>	<b>IV-9</b>
FSH (IU/ml)	1.8-22.5 <sup>b</sup>	73.25	95.09	42.13
AMH (ng/ml)	<6.9 <sup>c</sup>	<0.6	<0.6	<0.6
LH (IU/ml)	1.2-100 <sup>b</sup>	22.10	28.93	12.05
Estradiol (pg/ml)	30 to 300 <sup>b</sup>	1.11	10.37	10.82
TSH (mIU/ml)	0.3-5.0 <sup>b</sup>	7.68	6.51	6.24
Prolactin (ng/ml)	3-27	2.5	11.93	14.67
Uterine Volume (cm <sup>3</sup> ) <sup>d</sup>	20-100(146)	18.5	14.6	3.7
Right Ovarian Volume (cm <sup>3</sup> )	6.6±0.19(147)	Not visualized	Not visualized	0.7
Left Ovarian Volume (cm <sup>3</sup> )	6.6±0.19(147)	4.3	2.7	0.4
Anti-Thyroid Antibodies	Negative	Negative	Negative	Negative
Karyotype	46,XX	46,XX	46,XX	46,XX

<sup>a</sup>All hormone measures provided were prior to hormone replacement therapy.

<sup>b</sup>Mayo Clinic, Mayo Medical Laboratories, Rochester, MN.

<sup>c</sup>Reference Range for Adult Females, Esoterix Lab, Calabasas, CA.

<sup>d</sup>Uterine volume calculated by using the formula for prolate ellipsoid: longitudinal diameter x anterioposterior diameter x transverse diameter x 0.5233. Range based on Tanner stage 4-5 or adult uterine volumes.

Table 3.1: Clinical Laboratory Investigations of Affected Family C Daughters (IV-1, IV-6, and IV-9).

	Normal Range	AI2	AII4	AII6	BI2	BII1
Current Age (years)	-	49	23	19	39	18
Age at Menarche (years)	-	15	-	-	13	-
Age at Diagnosis (years)	-	-	16	16	-	16
Karyotype		-	46,XX	46,XX	-	46,XX
Pre-hormone Replacement Therapy						
FSH (mIU/ml)	1.8-22.5*	-	73	122.5	-	140.1
LH (mIU/ml)	1.2-100*	-	33.9	47.3	-	32.42
Estradiol (pg/ml)	30 to 300*	-	7.4	5	-	6
Prolactin (ng/ml)	3.3-26.7	-	9.7	7.08	-	12.58
Height (cm)	158-168**	150	136.1	134.4		141.5
Weight (kg)	47-62**	68	42.3	29.4	61	35.5
IGF-1 (ng/ml)	130-595***	-	N/A	408	-	200
IGFBP-3 (□g/ml)	3.4-9.5***	-	N/A	4.81	-	3.76
GH (ng/ml)	0.01-3.61**	-	N/A	9.5	-	3.03
Uterine Length (cm)	7.3±0.8(148)	-	Not visible	Not visible	-	2.9
Ovarian Volume (cm <sup>3</sup> )	6.6 <sup>f</sup>	-	Not visible	Not visible	-	Not visible
Tanner Stage (Pubic Hair)	-	-	2	2-3	-	2-3
Tanner Stage (Breast)	-	-	1	1	-	1
Bone Age (years)	-	-	11	12	-	11
Post-hormone Replacement Therapy						
FSH (mIU/ml)	1.8-22.5*	-	30.13	159.2	-	34.1
LH (mIU/ml)	1.2-100*	-	30.15	50.1	-	4.3
Estradiol (pg/ml)	30 to 300*	-	73.1	19	-	<20
Height (cm)	158-168**	-	144.3	143.1	-	150.4
Weight (kg)	47-62**	-	46	42.3	-	41.7
Tanner Stage (Pubic Hair)	-	-	5	5	-	5
Tanner Stage (Breast)	-	-	5	5	-	5
Bone Age (years)	-	-	18	17	-	16
*Mayo Clinic, Mayo Medical Laboratories, Rochester, MN. **Ranges based on CDC Clinical Growth Charts. ***Reference Range for 16 year old female, Mayo Clinic, Mayo Medical Laboratories.						

Table 3.2 Clinical Laboratory Investigations of Affected Individuals Prior to Hormone Replacement Therapy, Families A and B

An unrelated Family B presented with a 16-year old girl (**Fig. 3.2B**) who had primary amenorrhea, short stature, and lacked breast development. Her past medical history was otherwise unremarkable. On physical examination she was found to have minimal breast development, stage III (P3) pubic hair, and stage II axillary hair development. She had short stature with a height of 141.5 cm (-3.27 SD), weighed 35.5 kg (-3.65 SD), and her bone age was 11 years (**Table 3.2**).

Pelvic ultrasonography revealed an infantile uterus (about 2.9 cm in length, expected length=7.3 ± 0.8 cm) and no visible ovaries. Her basal estradiol level was low (6 pg/ml) with elevated LH and FSH levels (respectively 32.42 and 140.07 mIU/ml). She had a normal 46,XX female karyotype. Prolactin levels measured in the normal range at 12.58 ng/ml and 8.11 ng/ml (reference range: 3.34-26.72). She has regular menstrual bleeding with combined estrogen-progestin replacement therapy. Hormone replacement therapy also induced growth to a height of 150.4 cm (-1.9 SD), a weight of 41.7 kg (-2.4 SD), and breast development to Tanner stage V. The proband's mother (BI2) is currently 39 years of age. She exhibited signs of thelarche at 10 years of age and had normal first menarche at 13 years of age. She continues to cycle normally. She is 157.5 cm tall and weighs 61 kg. The 13-year-old unaffected sister (subject BII3) had normal first menarche at 12 years of age and has had regular menstrual cycles. Her height was 150.1 cm (-1.0 SD), her weight was 37.4 kg (-1.1 SD). Additionally, the proband has two healthy brothers (BII2 and BII4) that have no known medical problems.

A nonsynonymous SNP in exon 9 of *MCM8* (rs16991615), p.E341K, has been highly associated with the age of natural menopause through genome-wide association studies is strongly implicates DNA repair pathways (81, 83-87). The mutation is contained within the N-terminal hairpin of the DNA binding domain of MCM8 (**Fig. 3.14**), predicted to be essential for the DNA binding activity of the protein.

### **3.1.2 Research Design**

#### **3.1.2.1 Identification of MCM8 and MCM9 Mutations**

To identify the genetic cause of ovarian dysgenesis and primary amenorrhea in Family A, we performed DNA copy number analysis and homozygosity mapping in two affected individuals

(AII4 and AII6) as well as four unaffected (AII2, AII3, AII5, AII7) female siblings. Two contiguous regions of homozygosity were identified on chromosomes 1 and 6: 1p34.3 (chr1:35,187,876-36,929,988; hg19) and 6q21-q22.33 (chr6:114,543,815-127,262,938; NCBI37/hg19), that were shared by the two affected daughters but not shared by unaffected individuals (**Fig. 3.4**). These two regions of homozygosity did not contain genes known to be involved in human ovarian failure and short stature. We therefore undertook an unbiased whole exome sequencing (WES) approach to identify pathogenic variants.

Gene	Location	Coding Mutation	Protein Mutation	dbSNP	Conservation (PhyloP)	Protein Prediction Change	
						SIFT	Poly-phen2
<b>Family A</b>							
<i>CEP85L</i>	chr6: 118887215	c.A497T	p.D166V	rs9489444	N	D	P
<i>GEMIN4</i>	chr17: 648664	c.C2619A	p.H873Q	rs19177812 7	N/A	N/A	B
<i>MCM9</i>	chr6: 119149088	c.1732+2 T>C	Splicing mutation	-	N/A	N/A	N/A
<b>Family B</b>							
<i>AZ11</i>	chr17: 79169670	c.G1981A	p.E661K	rs61741549	C	T	B
<i>CCDC137</i>	chr17: 79639666	c.C802T	p.R268W	rs11269964 8	N	D	D
<i>DNAH14</i>	chr1: 225237950	c.A1951G	p.I651V	rs78320839	N/A	N/A	N/A
<i>DNAH17</i>	chr17: 76481652	c.C7463T	p.T2488M	rs37438791 3	N/A	N/A	N/A
<i>ENGASE</i>	chr17: 77080696	c.G1519A	p.V507I	rs20179280 0	C	T	B
<i>FER1L6</i>	chr8: 124992866	c.C1225T	p.R409W	rs20125412 9	C	N/A	N/A
<i>MCM9</i>	chr6: 119245203	c.C394T	p.R132X	-	C	N/A	N/A
<i>MUC5B</i>	chr11: 1271321	c.C13211G	p.A4404G	rs2943517	N/A	N/A	N/A
<i>OXLD1</i>	chr17: 79632592	c.C83A	p.P28Q	rs11195939 3	N	T	B
<i>ST6GALN AC1</i>	chr17: 74622747	c.G1297A	p.V433M	rs14030002 5	C	D	P
<i>TMEM63 A</i>	chr1: 226055624	c.G478A	p.V160I	rs11560089 6	N	T	B
Abbreviations for PhyloP, SIFT, and Polyphen-2: C = Conserved; N = Not Conserved; D = Damaging; P=Probably Damaging; B = Benign; T = Tolerated.							

Table 3.3 Whole Exome Sequencing Variant Calls in Families A and B

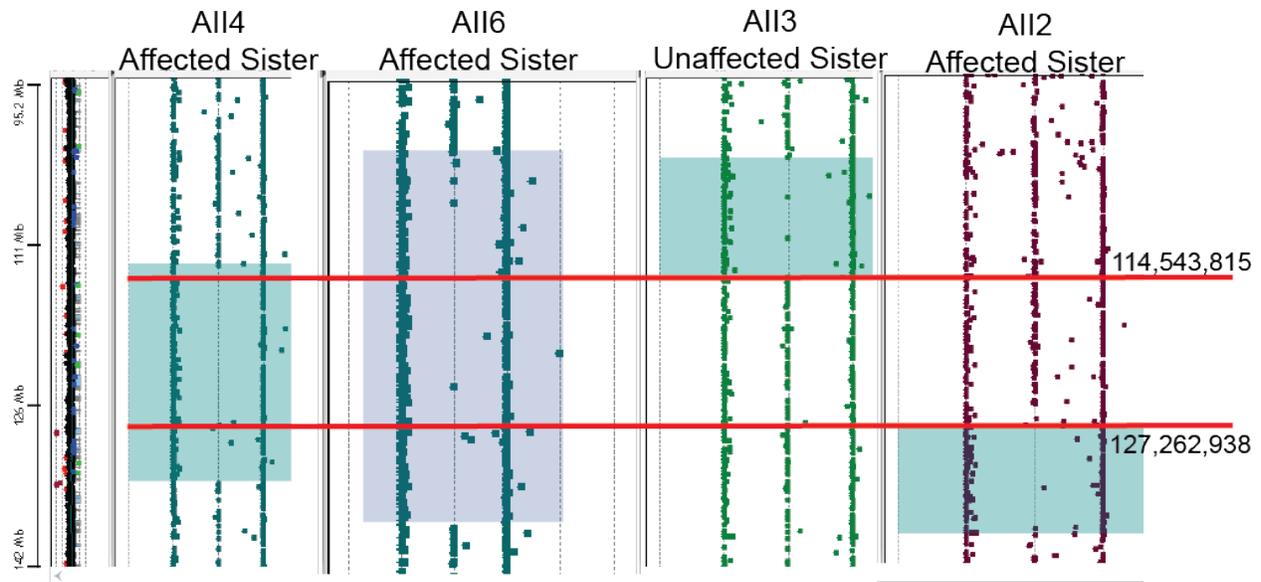


Figure 3.4 Region of homozygosity on chromosome 6 in Family A.

Use of a SurePrint G3 ISCA design CGH+SNP, 4x180K oligonucleotide microarray (Agilent, Santa Clara, CA) allowed the identification of a region of homozygosity between chr6: 114,543,815-127,262,938 (hg19, red lines) that was found both in AII4 and AII6 (affected sisters) that was not present in AII3 and AII2 (unaffected sisters). AII4 and AII6 show an absence of dots down the middle of the plot which are present in AII3 and AII2.

WES was conducted on three unaffected daughters (AII2, AII3, AII5), both parents (AII1 and AII2) and two affected siblings (AII4 and AII6). Three variants in *CEP85L* (6q22.31), *GEMIN4* (17p13.3), and *MCM9* (6q22.31), fit the homozygous autosomal recessive model (**Table 3.3**). Of these three, *CEP85L* and *MCM9* mapped to the region of homozygosity shared by affected sisters on chromosome 6 (**Fig. 3.4**). Homozygosity in AII4 and AII6 is indicated by the absence of dots down the center of the plot for these two patients. The *CEP85L* variant in affected sisters (c.A497T, p.D166V) was previously reported as a SNP (rs9489444), with 3% allele frequency, and was not conserved across species. We therefore eliminated *CEP85L* as a likely cause of observed phenotype.

The c.1732+2T>C variant in *MCM9* remained as the only pathogenic variant in Family A. The *MCM9* variant c.1732+2T>C (chr6:119,149,088, NCBI37/hg19) changed the highly

conserved splice donor site GT to GC located right after exon 9 (**Fig. 3.2C, D**). This variant has not been previously reported in either the Exome Variant Server or 1000 Genome databases.

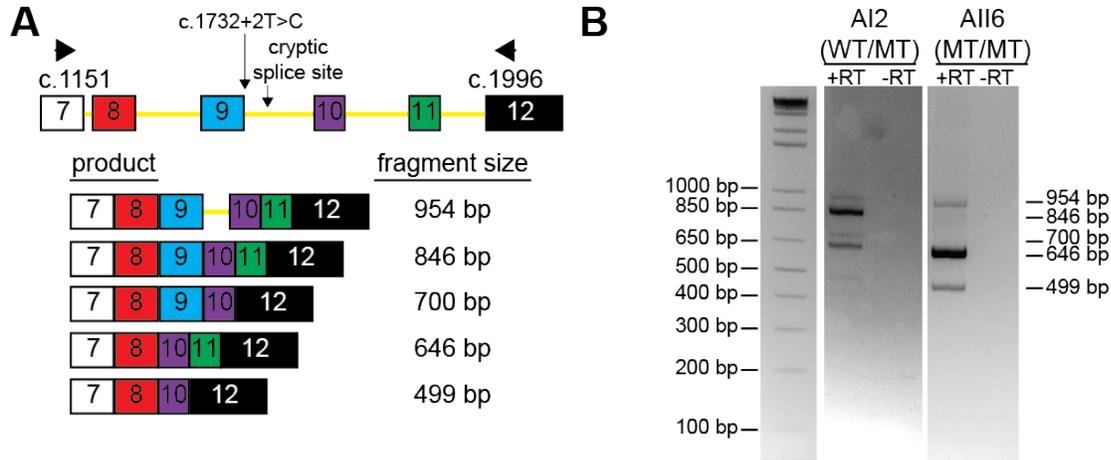


Figure 3.5 *MCM9* c.1732+2T>C results in abnormal alternative splicing.

(A) Schematic of primer design to detect splicing aberrations and band sizes of each product analyzed. Wild-type product is predicted to produce a 846 bp fragment and known alternative splicing of exon 11 results in a 700 bp fragment. These fragments were amplified in the heterozygote AI2, but not in homozygous AII6 individual. A 954 bp fragment is an abnormal splicing product that contains 108 bp of intron 9 and is predicted to introduce a premature stop codon. A 646 bp fragment is an abnormal splicing product in which exon 9 is deleted and a smaller, 499 bp fragment lacks both exons 9 and 11. (B) Representative gel image of DNA products fractionated on agarose gel following PCR amplification of cDNA derived from RNA of lymphoblastoid cell lines from heterozygous mutation carrier, AI2, and from affected homozygous mutation carrier, AII6, using primers (shown in A, 1151 Forward: 5' – GAT CTA CTA GTG CAG GTC TGA CG – 3'; 1996 Reverse: 5' – GCA CAC TCT GAT TCT GTA ACC TTT – 3').

We examined the effect of *MCM9* c.1732+2T>C mutation on *MCM9* RNA alternative splicing products. We predicted that loss of the exon 9 splice donor site would alter splicing downstream of exon 9 (corresponding to amino acid 577; **Fig. 3.2C, D**). Using oligonucleotide primers that bracket c.1151-1996 (exon 7 to exon 12) in *MCM9* cDNA are expected to amplify a wild type 846 bp product as well as the 700 bp band resulting from exon 11 skipping, a known alternatively spliced mRNA variant from the wild type *MCM9* locus (149) (**Fig. 3.5A and 3.6B, C**). The heterozygous mother contained the wild type (846 bp), alternative wild type (700 bp) product (faint), and three additional splicing products: 954 bp (faint), 646 bp, and 499 bp (faint) (**Fig. 3.5A,B**). The wild type 846 bp and 700 bp bands were absent from the homozygous daughter's DNA but the three mutant products were detected. Sanger sequencing of the predicted

846 bp wild-type band contained exon 8, 9, and 10 in sequence, and the 700 bp alternative wild type product was lacking exon 11, as expected (**Fig. 3.6A, B**). Sequencing of the 954 bp band revealed an abnormal splicing product comprising exon 9 plus 108 bp of downstream intronic sequence spliced onto exon 10 (**Fig. 3.6A**). This intronic region ends at a cryptic splice site and is predicted to encode an additional 29 amino acids followed by a premature stop codon. This stop codon likely results in a truncated protein lacking the winged helix domain and unique C-terminal extension of MCM9. The 646 bp fragment was a result of abnormal splicing that skipped exon 9 (**Fig. 3.6C**). Exon 9 encodes a portion of MCM9 that is homologous to the Sensor 2 domain of Archaeal MCM ATPase domain, which is predicted to be involved in ATP binding (60). The loss of exon 9 is predicted to delete 67 amino acids from the wild type MCM9 protein. The 499 bp fragment was a result of alternative splicing that skipped both exons 9 and 11 (**Fig. 3.5A,B and 3.6B,C**). Our results show that *MCM9* c.1732+2T>C genomic mutation causes abnormal splicing products predicted to disrupt MCM9 protein function.

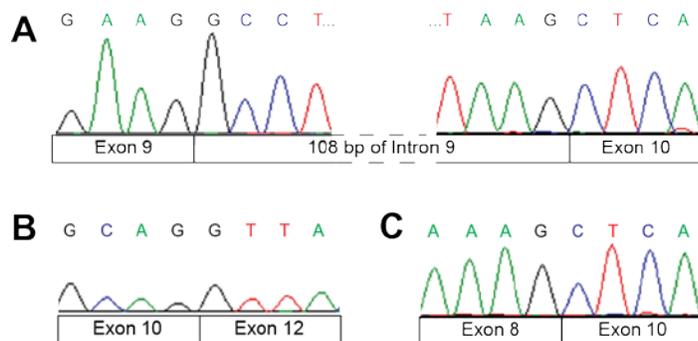


Figure 3.6 Sanger sequencing identified abnormal splice variants of *MCM9* in family members carrying the c.1732+2T>C mutation.

Primers that bracket the c.1732+2T>C mutation were used to amplify cDNA from lymphoblastoid cells of family members AI2, heterozygous mother, and AII6, homozygous daughter. PCR products were fractionated on an agarose gel (as shown in Figure 3.5). The nucleotide sequence of each product was evaluated by Sanger sequencing. Representative chromatograms are shown for the DNA fragments that were identified. The 846 bp DNA fragment revealed the wild-type sequence containing exons 7, 8, 9, 10, 11, and 12 in order. (A) The largest DNA fragment, 954 bp, is the result of abnormal splicing which includes exon 9, 108 bp of intron 9, and exon 10, in sequence. Intron 9 is 1062 bp in size and contains a cryptic splice site at nucleotide 108. (B) The 700 bp fragment revealed a loss of exon 11 from the wild-type sequence, as had been previously described. (C) The 646 bp fragment was the result of abnormal splicing and revealed a loss of exon 9 from the wild-type sequence. The smallest fragment, 499 bp, revealed a loss of both exons 11 (B) and 9 (C).

In Family B, we conducted WES on the unaffected daughter (BII3), both parents (BI1 and BI2), and the affected daughter (BIII1). After variant filtration, eleven variants remained that fit the homozygous autosomal recessive model (Table 3.3). Of these eleven, only the *MCM9* (c.394C>T, chr6:119,245,203, NCBI37/hg19) variant was novel, while the rest were known SNPs. The *MCM9* c.394C>T variant has not been previously reported in either the Exome Variant Server or 1000 Genome databases, it is highly conserved, and it results in insertion of a premature stop codon in exon 2 of *MCM9*, p.R132X (Fig. 3.2C, D). The *MCM9* c.394C>T variant was verified by Sanger Sequencing (Fig. 3.2B, bottom) and was absent in 200 fertile women.

Chromosome	Coordinates (hg19)	Siblings in Which Region Identified	LOD Score
Chr 1	2010994-4953528	Unaffected and Affected	1.53
Chr 12	81229423-82251957	Unaffected and Affected	-0.65
Chr 19	21157697-28750407	Unaffected and Affected	0.66
Chr 20	65288-1136217	Affected Only	1.58
Chr 20	2163414-4313037	Affected Only	1.18
Chr 20	4861939-6976715	Affected Only	2.77

Table 3.4 Regions of Homozygosity Based on SNP Array, Family C

In Family C, SNP analyses identified a 3.3 Mb region of homozygosity on chromosome 20p13-p12.3 flanked by rs1547618 and rs1012891 present only in affected subjects (**Table 3.4**). No known POF genes lay in this interval and we undertook whole exome sequencing (WES) to identify pathogenic variants. WES revealed two non-synonymous variants, *FERMT1* (MIM 607900) and *MCM8* (MIM 608187), which met autosomal recessive inheritance filter criteria and mapped to the region of homozygosity with the highest LOD score on chromosome 20. The *FERMT1* variant (NM\_017671:c.293G>A, p.R98H) was previously reported (rs137862671) in public databases and is likely a benign variant. Pathogenic variants in *FERMT1* cause Kindler



Cell Number	Patient		
	Control	IV-1	IV-6
1	0	0	0
2	0	1	0
3	0	0	0
4	1	0	0
5	0	0	0
6	0	0	0
7	0	0	0
8	0	0	0
9	0	0	0
10	0	0	0
Total Number of Breaks	1	1	0
Ave. No. Breaks/ Cell	0.1	0.1	0

Table 3.5 Chromosomal Breakage as Induced by DEB treatment in Patient Lymphocytes (Raw Data).

### 3.1.2.2 Impairment of DNA Break Repair in Families A and C

Fibroblasts from *Mcm8* deficient mice show hypersensitivity to agents that induce dsDNA breaks and DNA crosslinks resulting in a higher number of broken chromosomes, a hallmark of genomic instability (74). We assayed DNA repair capabilities of cultured fibroblasts derived from affected and unaffected family members (**Fig. 3.8A-C**) (150). Fibroblasts from unaffected sister IV-3 (WT genotype) showed few chromosome breaks at 150 nM and 300 nM concentrations of MMC (**Fig. 3.8A, D**). Cells from III-2 (unaffected mother, heterozygous WT/MT) showed a significantly increased number of chromosomal breaks/cell when compared to IV-3 (unaffected sister, WT genotype) at all concentrations of MMC (**Fig. 3.8B, D**): 50 nM ( $0.9 \pm 0.3$  vs.  $0.1 \pm 0.1$ ,  $p = 0.02$ ), 150 nM ( $5.3 \pm 0.4$  vs.  $2.6 \pm 0.3$ ,  $p < 0.001$ ), and 300 nM ( $8.1 \pm 0.5$  vs.  $5.0 \pm 0.4$ ,  $p < 0.001$ ). In both affected sisters (IV-1 and IV-6; homozygous MT/MT), the number of chromosomal breaks/cell was at least 8-10 fold higher than in heterozygote carriers (**Fig. 3.8C, D**,  $p < 0.01$ ). At 150 nM MMC, an average of  $22.4 \pm 1.5$  and  $15.4 \pm 0.8$  chromosome breaks/cell were observed for affected sisters IV-1 and IV-6, respectively ( $p < 0.001$ ). More than 50 chromosomal breaks/cell

were observed at 300 nM of MMC exposure in cells from affected sisters IV-1 and IV-6 ( $p < 0.001$ ). In the majority of cells homozygous for the mutation, multiple complex chromosomal aberrations were observed (**Fig. 3.8C, D**). Therefore, *MCM8* c.446C>G homozygous mutation impedes repair of MMC induced chromosomal breaks. Exposure of peripheral lymphocytes to diepoxybutane (0.1  $\mu\text{g/ml}$ ) did not cause chromosomal breakage in cells from the same family members (IV-1 and IV-6) homozygous for the *MCM8* c.446C>G mutation. Ten metaphase cells/patient were evaluated but the number of breaks detected was less than 0.2 breaks/cell (**Table 3.5**). Diepoxybutane results suggest that *MCM8* is not part of the Fanconi anemia DNA repair pathway. WES data analysis from affected sisters did not reveal additional variants that would affect protein function in genes known to be associated with chromosomal instability and ovarian failure (**Table 3.6**).

<b>Gene</b>	<b>Associated Syndrome</b>	<b>Variant</b>
<i>ATM</i>	Ataxia-telangiectasia	Benign SNP
<i>BRCA1</i>	Breast Cancer	Benign SNP
<i>BRCA2</i>	Fanconi Anemia, Wilms tumor, and cancers	Benign SNP
<i>BRIP1</i>	Fanconi anemia, Breast Cancer	Benign SNP
<i>ERCC4</i>	Fanconi anemia	None
<i>FANCA</i>	Fanconi anemia	None
<i>FANCB</i>	Fanconi anemia	None
<i>FANCC</i>	Fanconi anemia	None
<i>FANCD2</i>	Fanconi anemia	None
<i>FANCE</i>	Fanconi anemia	None
<i>FANCF</i>	Fanconi anemia	None
<i>FANCM</i>	Fanconi anemia	None
<i>FMRI</i>	Fanconi anemia	None
<i>HGPS</i>	Hutchinson-Gilford progeria	None
<i>MLH1</i>	Mismatch repair cancer syndrome	None
<i>MLH3</i>	Colorectal Cancer, Endometrial Cancer	Benign SNP
<i>MSH2</i>	Mismatch repair cancer syndrome	None
<i>MSH3</i>	Endometrial carcinoma	Benign SNP
<i>MSH6</i>	Mismatch repair cancer syndrome	Benign SNP
<i>NBN</i>	Nijmegen breakage syndrome	Benign SNP
<i>NBS1</i>	Nijmegen Breakage Syndrome	None
<i>PALB2</i>	Fanconi anemia	None
<i>PHF9</i>	Fanconi anemia	None
<i>RAD50</i>	Nijmegen breakage syndrome-like disorder	None
<i>RAD51C</i>	Fanconi anemia	None
<i>SLX4</i>	Fanconi anemia	None
<i>TP53</i>	Multiple Cancers	Benign SNP
<i>WRN</i>	Werner syndrome	Benign SNP
<i>XPC</i>	Xeroderma pigmentosum	Benign SNP
<i>XRCC9</i>	Fanconi anemia	None

Table 3.6 WES Variants in genes known to be involved in chromosomal instability or primary amenorrhea, Family C.

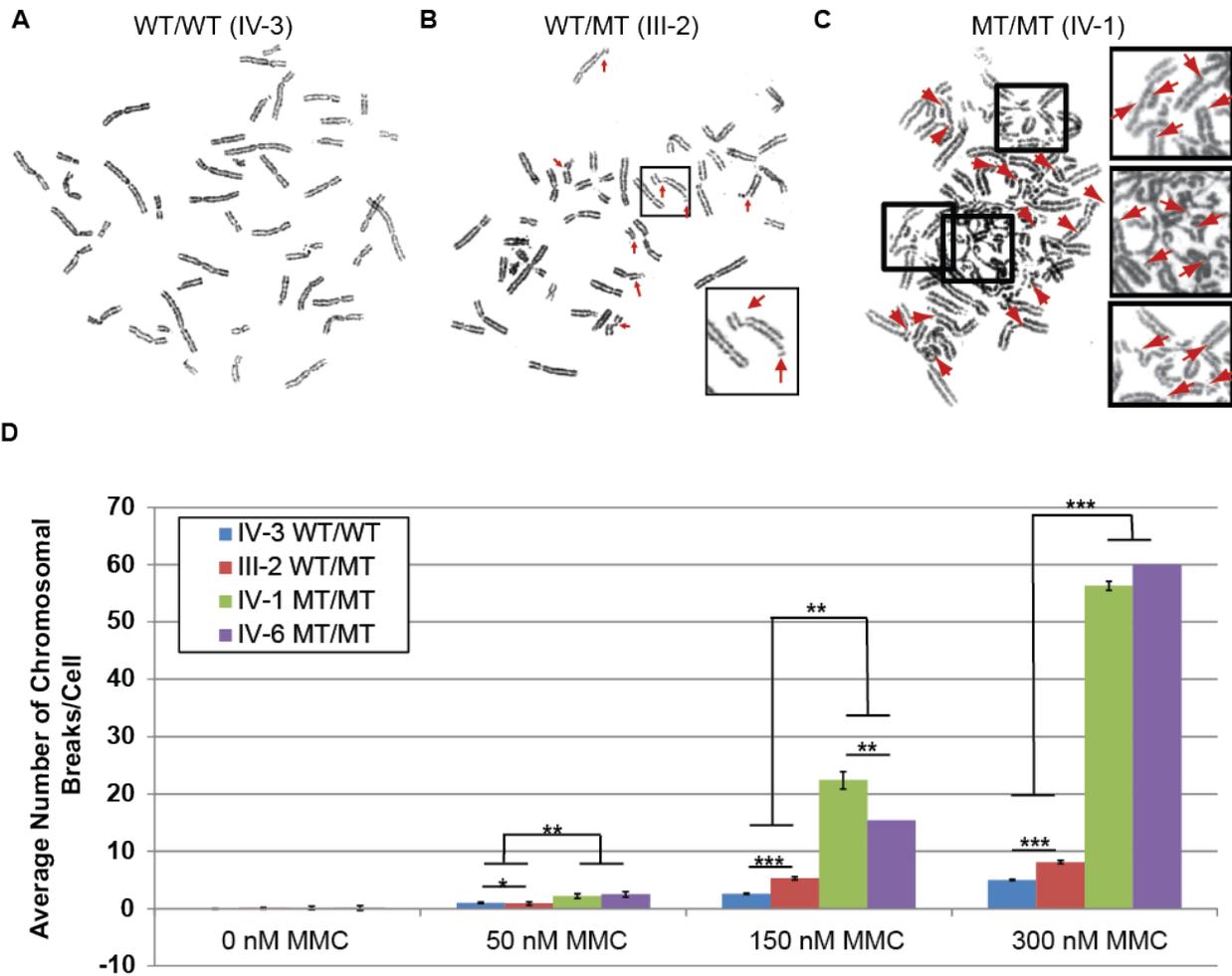


Figure 3.8 MCM8 mutation impairs DNA break repair.

Cells from homozygous *MCM8* c.446C>G individuals have impaired ability to repair dsDNA breaks induced by MMC. A total of four experimental groups were treated with MMC for each sample. Representative metaphase cells treated with 300 nM MMC are shown from (A) healthy and fertile individual with *MCM8* WT/WT genotype (family member IV-3), (B) unaffected WT/MT genotype (family member III-2), and (C) affected females homozygous for the *MCM8* c.446C>G pathogenic variant (MT/MT genotype; family member IV-1). The number of chromosomal breaks counted was limited to 60 per cell and those with more were indicated as 60+ breaks/ cell. Error bars represented the standard error of the mean. \* = p<0.01, \*\* = p<0.001, \*\*\* = p<0.0001.

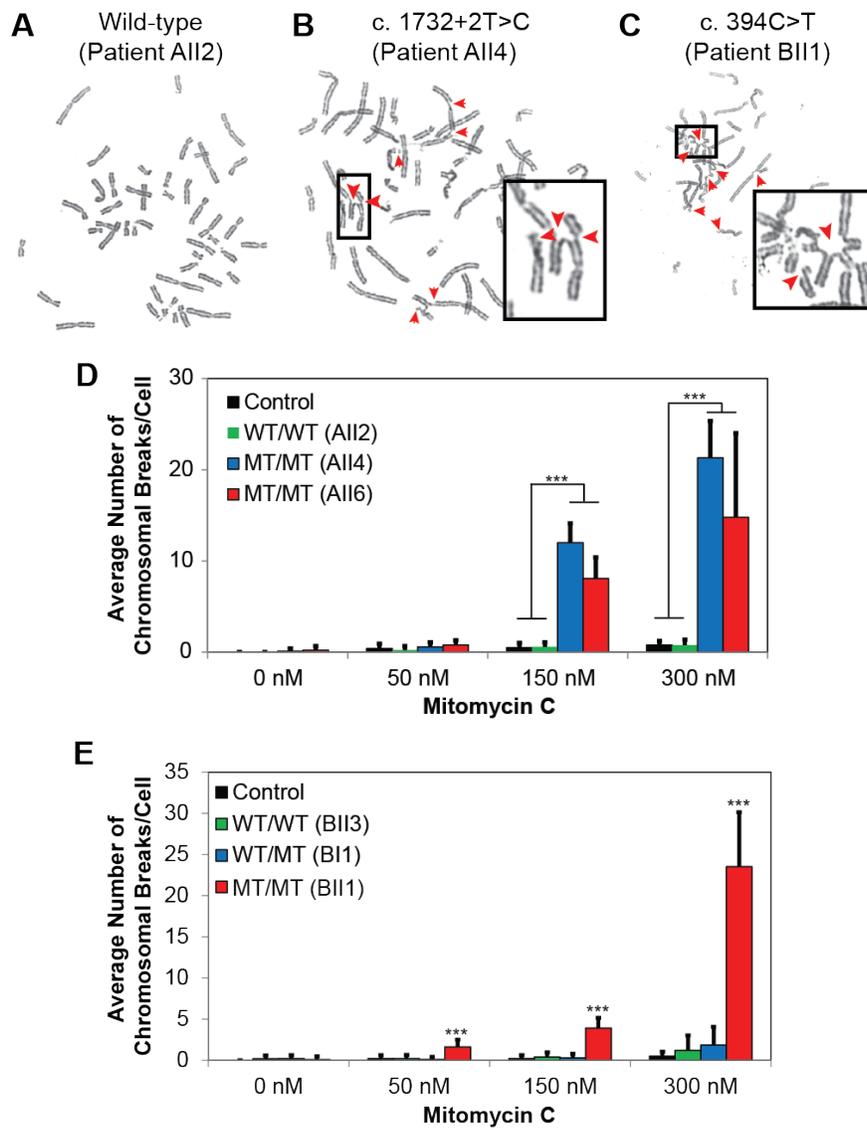


Figure 3.9 Mutations in human *MCM9* impair repair of MMC-induced chromosomal breaks.

Homozygous *MCM9* c.1732+2T>C and c.394C>T mutations associate with increased rate of chromosomal breakage. Peripheral blood lymphocytes from unaffected (wild-type or heterozygous carriers) and homozygous mutant individuals were stimulated by phytohaemagglutinin (PHA), cultured for 72 hours after treatment with 0 nM, 50 nM, 150 nM, or 300 nM Mitomycin C (MMC), and metaphase chromosome breaks were quantified. (A-C) Representative figures of metaphase spreads from wild-type (A), homozygous mutant c.1732+2T>C (B) and homozygous mutant c.394C>T (C) are shown after treatment with 300 nM MMC. Homozygous mutant cells show extensive chromosome damage with numerous breaks, non-homologous exchanges, chromatid deletions, complex aberrations, and acentric fragments (arrows) as compared to the wild-type. The mean numbers of chromosomal breaks per cell were quantified for cells from Family A, c.1732+2T>C mutation (D) and Family B, c.394C>T mutation (E). Bar graphs show the mean number of breaks per cell ( $\pm$ SD) observed for each MMC concentration. At least ten metaphase cells were studied for affected (AII4, AII6, and BII1), unaffected (AII2, BII3, and BI2) family members, and a healthy fertile unrelated female (control). ANOVA for Single Factor variation was utilized to determine the effect of drug concentration on each cell line. Two-tailed T-tests assuming unequal variance were used to compare two cell lines at a single drug concentration. \*\*\*,  $p < 0.001$ .

Similar to MCM8, it has been established that during gametogenesis, MCM9 protein is critical for homologous recombination (HR) in the repair of DNA double-stranded breaks (DSBs) and stabilization of broken or stalled DNA replication forks (68-70, 74, 76, 151). In addition to gonadal failure, fibroblasts from *MCM9* deficient mice show hypersensitivity to agents that induce DSBs and DNA crosslinks (i.e. ionizing radiation and MMC) resulting in a higher number of broken chromosomes, a hallmark of genomic instability. We assayed DNA repair capabilities of lymphocytes derived from members of both Family A (**Fig. 3.9A, B, D**) and Family B (**Fig. 3.9C, E**).

Treatment with 50 nM MMC induced significant numbers of chromosomal breaks only in cells homozygous for the p.R132X mutation from BII1 when compared to unrelated control ( $p = 6.93 \times 10^{-5}$ ) or heterozygous parent BI2 ( $p = 2.69 \times 10^{-5}$ ; **Fig. 3.9E**). Treatment with 150 nM MMC revealed significantly more chromosomal breaks in cells from patients homozygous for the splicing mutation c.1732+2T>C and stop codon mutation p.R132X (AII4, AII6 or BII1) when compared with cells from an unrelated control ( $p = 2.6 \times 10^{-7}$ ,  $p = 1.6 \times 10^{-14}$ , and  $p = 3.58 \times 10^{-8}$ , respectively) or heterozygous carriers (AII4 vs. AII2,  $p = 2.6 \times 10^{-7}$ ; AII6 vs. AII2,  $p = 5.4 \times 10^{-14}$ ; BII1 vs. BI2,  $p = 3.24 \times 10^{-8}$ ; **Fig. 3.9D,E**). Treatment with 300 nM MMC revealed a similar trend with significantly more chromosomal breaks in cells from affected AII4 or AII6 compared with control ( $p = 7.5 \times 10^{-8}$  or  $p = 1.6 \times 10^{-7}$ , respectively) or heterozygous AII2 ( $p = 1.6 \times 10^{-7}$ ; **Fig. 3.8D**). Similarly, cells from affected BII1 had significantly more chromosomal breaks than cells from heterozygote BI2 ( $p = 5.49 \times 10^{-7}$ ) or BII3 ( $p = 1.24 \times 10^{-6}$ ; **Fig. 3.9E**). These studies indicate that both the c.1732+2T>C and the c.394C>T homozygous mutations in *MCM9* impair repair of MMC induced chromosomal breakages. When comparing cells of the affected daughters from the two families to each other, BII1 (c.394C>T) appears to be more sensitive to chromosomal breakage at

50 nM (average breaks/cell =  $1.6 \pm 0.9$ ) when compared to AII4 (c.1732+2T>C; average breaks/cell =  $0.6 \pm 0.5$ ;  $p = 0.001$ ) or AII6 (c.1732+2T>C; average breaks/cell =  $0.8 \pm 0.5$ ;  $p = 0.00039$ ). AII4 and AII6 had significantly more breaks than BII1 at 150 nM ( $10.5 \pm 2.1$  vs.  $8.1 \pm 2.3$  vs.  $4.0 \pm 1.1$ , respectively; AII4 vs. BII1,  $p = 7.5 \times 10^{-6}$ ; AII6 vs. BII1,  $p = 4.96 \times 10^{-8}$ ). At the highest concentration of MMC, BII1 (average breaks/cell =  $23.5 \pm 6.6$ ) had significantly more breaks/cell on average than AII6 (average breaks/cell =  $14.79 \pm 9.23$ ;  $p = 0.0051$ ) but not AII4 (average breaks/cell =  $24.1 \pm 4.0$ ;  $p = 0.81$ ). Overall, the stop codon mutation (c.394C>T) appears to have a statistically greater effect on DNA damage repair in cells treated with MMC at 50 nM MMC than those cells with the splice site variant (c.1732+2T>C). However, at higher MMC concentrations cells of homozygous carriers fail to repair DNA damage as compared to heterozygous or wild type carriers. WES data analysis from affected sisters in either family did not reveal pathogenic variants in genes known to be associated with chromosomal instability and/or ovarian failure (**Table 3.7**).

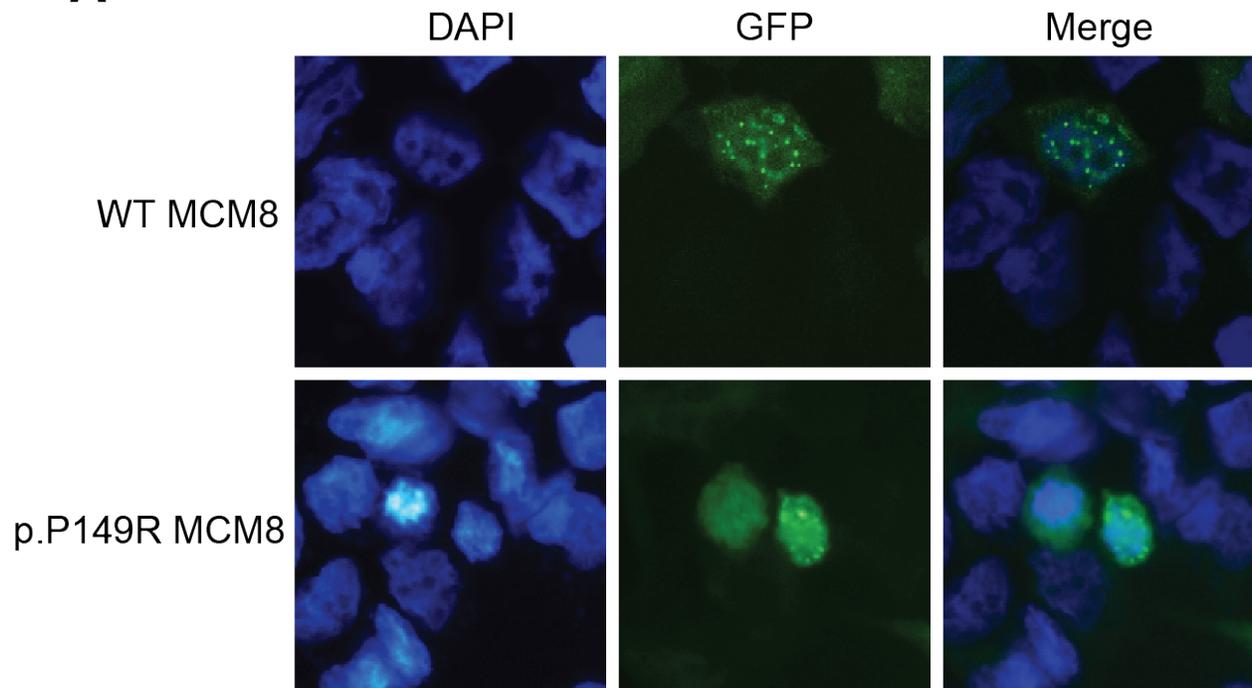
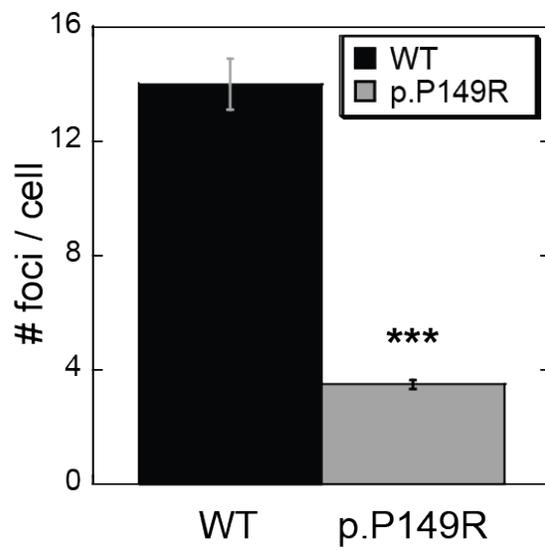
<b>Gene</b>	<b>Associated Syndrome</b>	<b>AII4</b>	<b>AII6</b>	<b>BIII</b>
<i>ATM</i>	Ataxia-telangiectasia	None	None	None
<i>BRCA1</i>	Breast Cancer	Benign SNP	Benign SNP	Likely Pathogenic SNP
<i>BRCA2</i>	Fanconi Anemia, Wilms tumor, and cancers	None	None	None
<i>BRIP1</i>	Fanconi anemia, Breast Cancer	None	None	None
<i>ERCC4</i>	Fanconi anemia	None	None	Benign SNP
<i>FANCA</i>	Fanconi anemia	None	Benign SNP	Benign SNP
<i>FANCB</i>	Fanconi anemia	Benign SNP	None	None
<i>FANCC</i>	Fanconi anemia	None	None	Benign SNP
<i>FANCD2</i>	Fanconi anemia	None	None	None
<i>FANCE</i>	Fanconi anemia	None	None	None
<i>FANCF</i>	Fanconi anemia	None	None	None
<i>FANCM</i>	Fanconi anemia	None	None	None
<i>FMR1</i>	Fanconi anemia	None	None	None
<i>HGPS</i>	Hutchinson-Gilford progeria	None	None	None
<i>MCM8</i>	Primary Amenorrhea, Endocrine Dysfunction	None	None	None
<i>MLH1</i>	Mismatch repair cancer syndrome	None	None	None
<i>MLH3</i>	Colorectal Cancer, Endometrial Cancer	None	None	None
<i>MSH2</i>	Mismatch repair cancer syndrome	None	None	None
<i>MSH3</i>	Endometrial carcinoma	None	Benign SNP	None
<i>MSH6</i>	Mismatch repair cancer syndrome	None	None	None
<i>NBN</i>	Nijmegen breakage syndrome	None	None	None
<i>NBS1</i>	Nijmegen Breakage Syndrome	None	None	None
<i>PALB2</i>	Fanconi anemia	None	None	None
<i>PHF9</i>	Fanconi anemia	None	None	None
<i>RAD50</i>	Nijmegen breakage syndrome-like disorder	None	None	Benign SNP
<i>RAD51C</i>	Fanconi anemia	None	None	None
<i>SLX4</i>	Fanconi anemia	None	None	None
<i>TP53</i>	Multiple Cancers	None	None	None
<i>WRN</i>	Werner syndrome	None	None	None
<i>XPC</i>	Xeroderma pigmentosum	None	None	None
<i>XRCC9</i>	Fanconi anemia	None	None	None

Table 3.7 WES Variants in Genes Known to Be Involved in Chromosomal Instability or Primary Amenorrhea, Families A and B.

### 3.1.2.3 Impairment of DNA Damage-Dependent Foci in Families A and C and Impairment of DNA Binding in Family C

MCM8 and MCM9 containing complexes are known to form foci at sites of DNA damage (76). To determine if the c.1732+2T>C mutation affects the ability of MCM9 to form foci at sites of DNA damage, we generated an MCM9-GFP cDNA construct lacking exon 9 (c.1732+2T>C). After MMC exposure of cells expressing wild-type MCM9-GFP, GFP expression was observed primarily in the nucleus with an average of  $20 \pm 3$  (SD) nuclear GFP foci/cell formed at sites of DNA damage (**Fig. 3.10C**, top). Interestingly, in mitomycin exposed cells expressing mutant MCM9-GFP, GFP signal was localized to the nucleus but foci could not be found at sites of DNA damage (**Fig. 3.10C**, bottom). These studies indicate that the *MCM9* c.1732+2T>C pathogenic variant, lacking exon 9, inhibits MCM9 from being recruited to sites of DNA damage.

In similar fashion, we generated a MCM8-GFP construct containing the c.446C>G/p.P149R mutation to determine if foci formation was affected. Cells were transfected with the MCM8-GFP constructs and treated with 300 nM MMC to induce DNA damage. GFP expression was primarily nuclear, with an average of  $14 \pm 0.9$  nuclear GFP foci/cell formed at sites of DNA damage in cells expressing wild-type MCM8-GFP (**Fig. 3.10A, B**). Significantly fewer nuclear GFP foci form, averaging  $3.5 \pm 0.2$  foci/cell ( $p < 0.001$ ; **Fig. 3.10A, B**) in cells expressing mutant MCM8-GFP. The GFP fluorescence was more diffuse throughout the nucleus and the cytoplasm (**Fig. 3.10A**). These experiments show that the c.446C>G/p.P149R mutation inhibits MCM8 recruitment to sites of DNA damage.

**A****B**

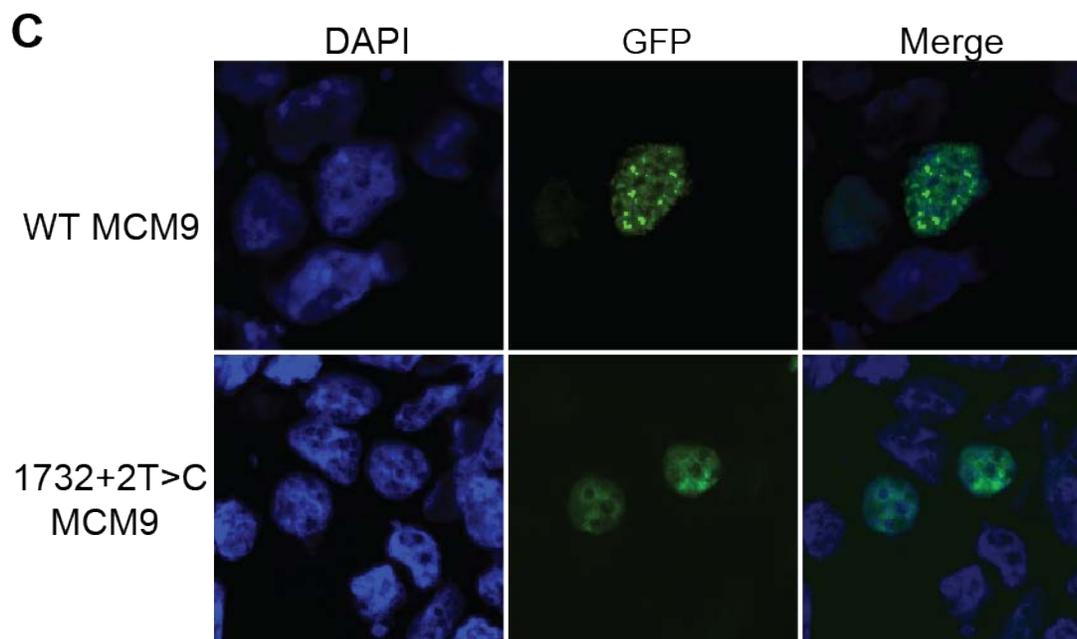


Figure 3.10 MCM8 and MCM9 Mutations Disrupt Foci Formation and DNA Binding.

(A) *MCM8* c.446C>G (p.P149R) mutation inhibits MCM8 foci formation at sites of DNA damage. HEK293T cells were transfected with WT MCM8-GFP (WT) or mutant MCM8-GFP (p.P149R) and treated with 300 nM MMC for 6 hours. Nuclei were counterstained with DAPI (blue). Nuclear foci formed in 293T cells expressing WT MCM8, but not in cells expressing mutant MCM8. Four independent experiments for transfection of MCM8-GFP (WT vs. p.P149R) coupled with DNA damage were performed. Representative confocal images are shown. (B) Twenty representative cells per condition were quantified for the number of damage induced nuclear foci. A two-tailed T-test (assuming unequal variance) revealed a statistically significant difference (\*\*\*,  $p < 0.001$ ) between the WT ( $14 \pm 0.9$  foci/cell) and c.446C>G ( $3.5 \pm 0.2$  foci/cell) expressing cells. Error bars represent standard error. (C) MCM9 1732+2T>C Disrupts Damage-Dependent Foci Formation. HEK293T cells were transfected with cDNA encoding wild-type MCM9-GFP (WT) or mutant MCM9-GFP which lacks exon 9 (1732+2T>C). Cells were treated with 300 nM MMC for 6 hours. Nuclei were counterstained with DAPI (blue). Nuclear foci formed in 293T cells expressing wild-type MCM9, but not in cells expressing mutant MCM9. Representative confocal images taken with a 40x objective are shown.

We also compared DNA binding of wild-type and mutant MCM8 protein. Wild-type and mutant *MCM8* cDNAs were generated that correspond to the N-terminus of the protein that is predicted to bind DNA (nucleotides 1-1104; aa 1-368). We chose a random 46 nt single stranded DNA (ssDNA) since RAD51 is recruited to ssDNA ends at sites of damage and MCM8 is likely to bind at these sites (68). Wild-type MCM8 showed an increase in binding to ssDNA with concentration ( $K_d = 6.7 \pm 0.8 \mu\text{M}$ ; **Fig. 3.11**). Unlike wild type, mutant MCM8 does not saturate binding over the course of the titration, making calculation of a  $K_d$  difficult ( $K_d = \text{not}$

determined). These results show that the c.446C>G/p.P149R mutation impairs DNA binding ability at the N-terminus of MCM8.

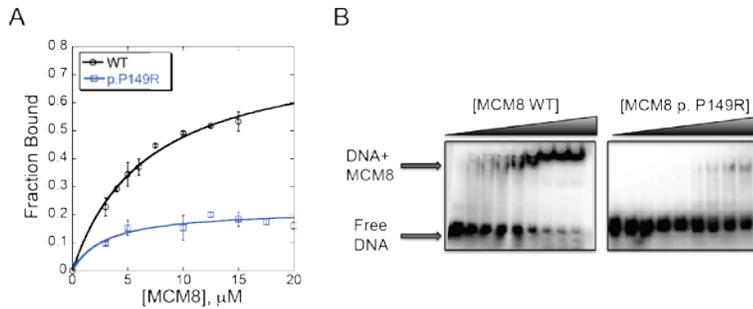


Figure 3.11 *MCM8* c.446C>G (p.P149R) inhibits DNA binding by EMSA.

(A) WT MCM8 or mutant MCM8 (p.P149R) protein was bound to a random 46 nt ssDNA oligonucleotide. Gels were imaged and quantification of the fraction of band shift was performed. Sample gels are shown in (B). Mutant MCM8 (p.P149R, blue squares) showed a significant reduction in DNA binding affinity for ssDNA at each concentration when compared to WT MCM8 (WT, black circles). Each point is the average of three replicates. Error bars represent standard deviation.

### 3.1.2.4 Evolutionary Comparison of the Selected MCM8 Region

CLUSTAL Omega web site (<http://www.ebi.ac.uk/Tools/msa/>) was used to align the sequences from MCM8 proteins in human (*Homo*, NP\_001268449. 1), mouse (*Mus*, NP\_079952.2), frog (*Xenopus*, NP\_001089437.1) and the predicted zebrafish sequence (*Danio*, XP\_002665161.2). p.P149R is found at the evolutionary conserved Proline (highlighted in yellow, **Fig. 3.7**).

### 3.1.2.5 Homology Model of Human MCM8 and Crystallization Trials

As a means of elucidating a model of the three-dimensional structure of human MCM8, archaeal *SsoMCM* (PDBID: 3F9V) was used as a template for threading the sequence of HsMCM8 using SWISS-PDB. The p.P149R mutation is located within a loop connecting two alpha-helices in the

A-domain which pack on top of the C-terminal domain (**Fig. 3.12**). Mutation of p.P149R will most likely disrupt a turn within the A-domain.

The crystal structure of *Methanothermobacter thermautotrophicus* MCM (*mtMCM*) has been solved and studied as a model for eukaryotic MCMs (62). Specifically, the *mtMCM* crystal structure provides insights into the *bob-1* mutation in yeast (P62L) MCM5, which is homologous to MCM8 p.P149R. The MCM5 *bob-1* mutation as a bypass suppressor to Cdc7/Dbf4 deletions (152). Cdc7/Dbf4 normally regulates entry into S-phase through phosphorylation of the MCM2-7 complex, but the *bob-1* mutation bypasses this requirement. Structural analysis of the homologous mutation in *mtMCM* was shown to weaken contacts in the N-terminal domain, making it more flexible, but it does not significantly affect DNA binding or ATPase activity *in vitro* (62, 153).

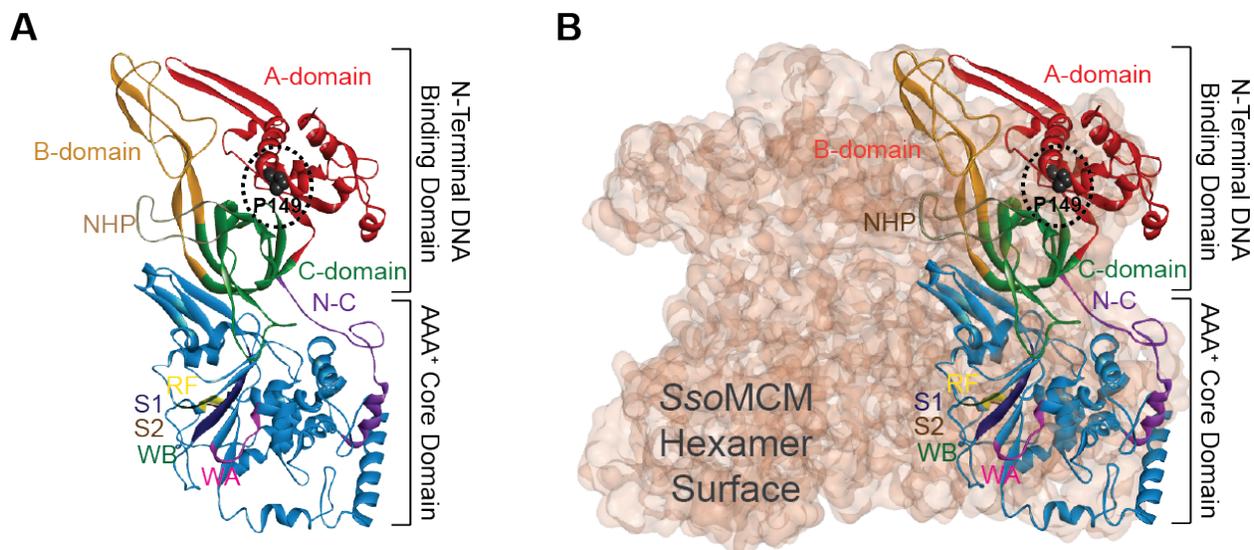


Figure 3.12 Homology model of human MCM8.

Local and global sequence alignments were performed using ClustalW2 analysis (<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi>). The homology model of human MCM8 was created by threading the adjusted alignment (WT or P149R) onto the structure of *Sulfolobus solfataricus* MCM (PDBID: 3F9V) (60) using SWISS-MODEL (154). P149R is located within a loop connecting two alpha-helices in the A-domain which pack on top of the C-domain. Proline 149 is circled by a black dashed line. NHP - N-terminal hairpin, WA - Walker A, WB - Walker B, S1 - Sensor 1, RF - Arginine finger, S2 - Sensor 2, WH - Winged helix. Colored domains correspond with linear protein diagram in Figure 3.1C.

To obtain a more accurate representation of the MCM8 structure, we have performed crystallization trials with the purified N-terminal domain of MCM8 in collaboration with Andrea Berman, PhD and Roni Lahr. Initial crystallization trials as well as optimization were performed with either MCM8(1-368) and MCM8(73-368). Initial MCM8(1-368) crystals became visible after 1 hour (**Fig. 3.13**), and were screened using FR-E rotating anode and Saturn 944 CCD image plate detector. These MCM8(1-368) crystals contained protein but diffracted very poorly due to high solvent content or poor crystal packing. We addressed this issue by performing crystallization trials of MCM8(73-368). We designed the MCM8(73-368) truncation because of the lack of secondary structure predicted in the first 73 amino acids, predicted using the PSIPRED Protein Analysis Workbench (<http://bioinf.cs.ucl.ac.uk/psipred/>). Unfortunately, only micro crystals grew in these trials. Crystal macro and micro seeding using was also attempted, but did not result in enough

crystal growth to allow for diffraction. Crystallization of alternative protein truncations or variation in buffers and methods may yield the crystal structure of MCM8 in the future.

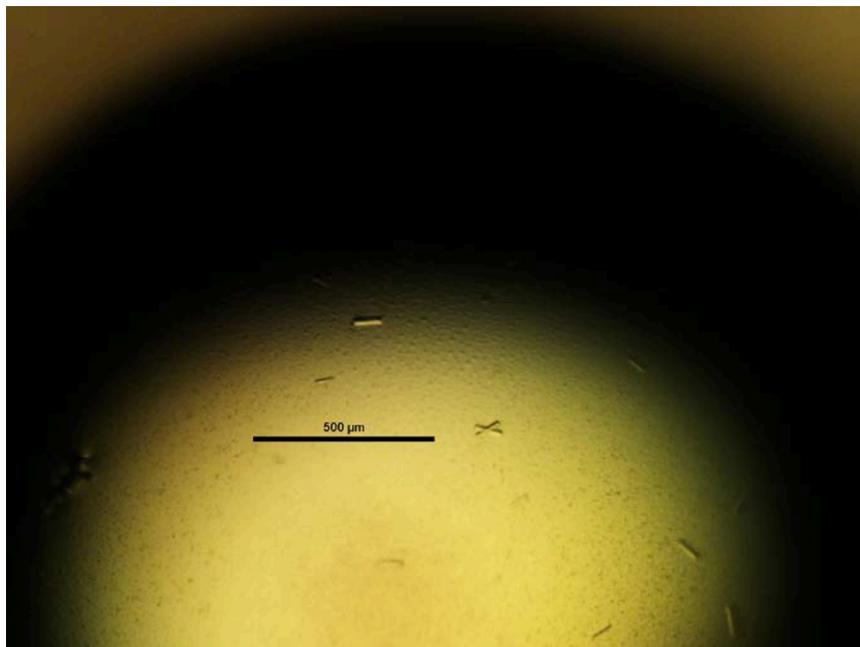
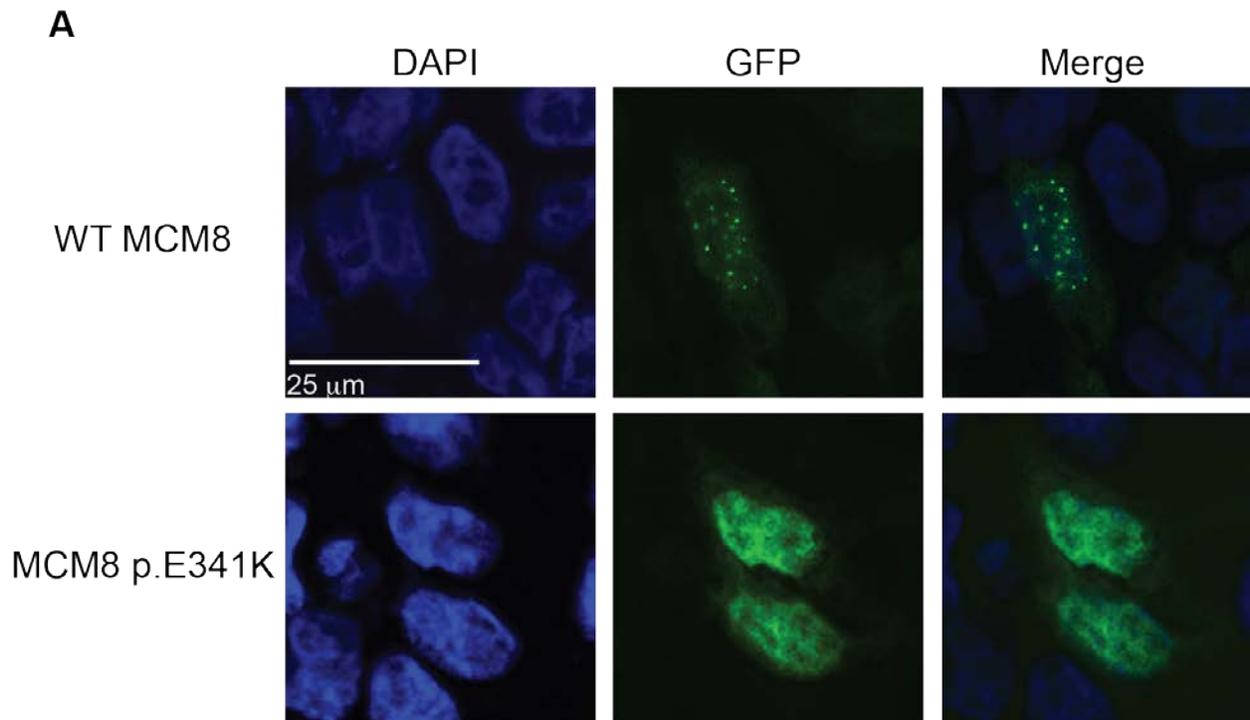


Figure 3.13 MCM8(1-368) microcrystals. Crystals were grown in 25% PEG 8000, 100mM HEPES pH 7. Crystals were screened using FR-E rotating anode and Saturn 944 CCD image plate detector.

### 3.1.2.6 DNA Binding Domain Mutation in MCM8 Impairs DNA Damage-Dependent Foci

To extend our investigation of the ability of MCM8 to form damage-dependent nuclear foci, we investigated MCM8 p.E341K, the nonsynonymous SNP previously associated with age of natural menopause (81, 83-87). The p.E341K mutation is contained within the N-terminal hairpin of the DNA binding domain (**Fig. 3.1**) of MCM8 suggesting that this mutant may be defective in responding to DNA damage and binding DNA. We generated a MCM8-GFP construct containing the p.E341K mutation to determine if foci formation was affected. Cells were transfected with the MCM8-GFP constructs and treated with 300 nM MMC to induce DNA damage. As described previously, GFP expression was primarily nuclear, with an average of  $14 \pm 0.9$  nuclear GFP foci/cell formed at sites of DNA damage in cells expressing wild-type MCM8-GFP (**Fig. 3.14A**,

**B).** Significantly fewer nuclear GFP foci form, averaging  $4.7 \pm 0.2$  foci/cell ( $p < 0.001$ ; **Fig. 3.14A, B**) in cells expressing mutant MCM8-GFP. The GFP



**B**

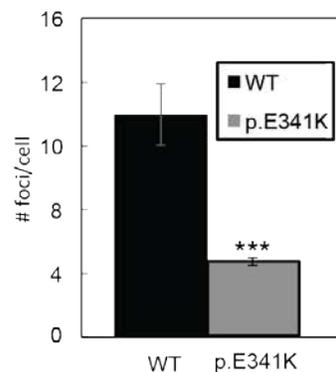


Figure 3.14 MCM8 p.E341K disrupts damage-dependent foci formation.

(A) MCM8 p.E341K inhibits MCM8 foci formation at sites of DNA damage. HEK293T cells were transfected with WT MCM8-GFP (WT) or mutant MCM8-GFP (p.E341K) and treated with 300 nM MMC for 6 hours. Nuclei were counterstained with DAPI (blue). Nuclear foci formed in 293T cells expressing WT MCM8, but not in cells expressing mutant MCM8. Four independent experiments for transfection of MCM8-GFP (WT vs. p.E341K) coupled with DNA damage were performed. Representative confocal images are shown. (B) Twenty representative cells per condition were quantified for the number of damage induced nuclear foci. A two-tailed T-test (assuming unequal variance) revealed a statistically significant difference (\*\*\*,  $p < 0.001$ ) between the WT ( $14 \pm 0.9$  foci/cell) and p.E341K ( $4.7 \pm 0.2$  foci/cell) expressing cells. Error bars represent standard error. Representative confocal images taken with a 40x objective are shown.

fluorescence was more diffuse throughout the nucleus (**Fig. 3.14A**). These experiments show that the p.E341K mutation inhibits MCM8 recruitment to sites of DNA damage in a manner similar to the pathogenic MCM8 mutant p.P149R.

### 3.1.3 Conclusion

We used homozygosity mapping, SNP Arrays and WES in a consanguineous family to identify a homozygous recessive pathogenic variants in *MCM8* (c.446C>G) and *MCM9* (c.1732+2T>C and c.394C>T). These three variants were identified as causes of a novel syndrome in humans characterized by ovarian insufficiency, hypothyroidism, and genomic instability in somatic cells. The MCM8 p.P149R substitution occurs in a highly conserved region of MCM8 predicted to bind DNA. Individuals homozygous for this variant show high numbers of chromosomal breaks when exposed to MMC and the p.P149R mutation inhibits MCM8 foci formation at sites of DNA damage. MCM8 p.P149R mutant protein shows impaired binding to DNA and likely prevents repair at these sites. Heterozygote sisters and their mother have unremarkable medical histories to date. This is consistent with the unaffected carrier state. There is no known family history of cancers. Future follow-up and phenotyping of additional individuals with *MCM8* variants will be of great importance to understand their predisposition to cancer. Given the consanguineous nature of the family, we cannot rule out a possibility that other genes contribute to the observed clinical phenotype.

The c.1732+2T>C pathogenic variant converted a conserved splice donor site GT to a GC, that resulted in aberrant splicing. The c.394C>T pathogenic variant introduces a premature stop codon in exon 2 of MCM9, at amino acid 132, in a protein that contains 1143 amino acids, and is

predicted to generate a truncated protein that lacks half of the DNA binding domain and all of the ATPase domain. We also showed that cells containing the c.1732+2T>C and c.394C>T variants are impaired in their ability to repair DNA damage likely due to the inability of MCM9 to be recruited to sites of DNA damage. Affected women in both families showed higher number of chromosomal breaks as compared to heterozygotes or wild type controls. A similar effect is observed in cells homozygous for MCM8 p.P149R (155).

Minichromosome maintenance proteins, MCM2-7 have been implicated in DNA replication elongation and pre-replication complex formation (145). MCM8 and MCM9 are the most recently discovered members of the minichromosome maintenance protein family and by association with other MCMs, were initially implicated in DNA replication (69). Unexpectedly, *MCM8* and *MCM9* deficient mice showed a phenotype primarily restricted to the germline with infertility and small gonads (74). In addition, somatic cells exhibited growth defects and chromosomal instability (74). MCM8 and MCM9 are therefore novel regulators of germ cell survival, rapidly induced and recruited to DNA damage sites, co-regulate each other's stability, and promote RAD51 recruitment to the chromatin. The mechanisms of their action, induction, and regulation in homologous recombination remain poorly understood. The phenotype of the three affected individuals from two independent families is very similar to what has been described in mouse knockouts.

The mechanism of MCM8 and MCM9 action in the maintenance of the germline remains to be elucidated. MCM8/MCM9 complex is likely required for the resolution of double stranded DNA breaks that occur during homologous recombination between homologous chromosomes in pachytene of meiosis I (72). Failure to resolve breaks will lead to oocyte death and thus very small or absent ovaries in women homozygous for these mutations. A number of mouse mutants lacking

genes involved in meiosis and DNA repair, such as *DMC1* and *Msh5*, lose oocytes rapidly in part due to their inability to process double stranded DNA breaks (156, 157). Indeed, mutations in *DMC1* and *Msh5* were identified in women with POF (158). Double stranded DNA break repair genes have been implicated in the onset of menopause by genome wide association studies, including a SNP in *MCM8* (rs16991615) (81, 83-87). Although no association has been identified between menopause and *MCM9*, the close physical interdependence of these two proteins shows the importance of this novel pathway in a spectrum of ovarian functions from gonadal development to menopause.

Genomic instability syndromes, such as Fanconi anemia (*FANCA* [MIM 227650]) (159), can associate with short stature, hypogonadism, and multiple endocrine dysfunctions (160). The mechanism of how genomic instability contributes to endocrinopathies has not been elucidated and is likely linked to complex regulatory networks in the hypothalamic-pituitary-endocrine axes of the body (160). Heterozygote sisters and mothers had unremarkable medical histories to date. The mothers of both families entered menarche between the ages of 13 and 15 years and did not develop POF. There is no known family history of cancers, however affected individuals may have an increased risk, particularly if exposed to DNA damage agents. Affected patients presented with hypergonadotropic hypogonadism, primary amenhorrea, and short stature. Future follow-up and phenotyping of additional individuals with *MCM9* variants will be of great importance to understand their predisposition to other disorders including endocrine dysfunction and cancer predisposition.

A role for *MCM8* in gonadal development and maintenance has been confirmed since the publication of these findings. Two novel homozygous mutations in *MCM8*: a splice site mutation (c.1954-1G>A) and a frameshift mutation (c.1469-1470insTA) were identified as the genetic

causes of primary amenorrhea in females and small testes and azoospermia in males (82). The two mutations are predicted to produce truncated, non-functional proteins. Quantitative analysis of MCM8 transcript revealed low abundance of the aberrant transcripts in affected cells, compared with the abundance of wild-type MCM8 transcript in unaffected cells. Similar to the findings in our families, it was found that both of these MCM8 mutations lead to difficulty in repairing chromosomal breaks induced by MMC treatment (82).

MCM8 p.E341K is highly associated with the age of natural menopause, and meta-analysis of menopause GWAS studies strongly implicates DNA repair pathways (81, 83-87). While our findings show the importance of this residue for the response of MCM8 to DNA damage, the functional relevance of MCM8 p.E341K in reproductive senescence has not yet been elucidated. Our study shows that *MCM8*, when inherited in Mendelian fashion, is essential for normal gonadal development. Future studies and follow-up on these and additional individuals will be necessary to define the spectrum of human phenotypes associated with *MCM8* variants. The role of the novel MCM8/MCM9 pathway in women with idiopathic POF needs to be explored further.

### **3.1.4 Experimental**

#### **3.1.4.1 DNA Mapping and Whole Exome Sequencing**

DNA copy number analysis and homozygosity mapping was performed using a 180K comparative genomic hybridization (CGH) + single nucleotide polymorphism (SNP) oligonucleotide array (SurePrint G3 ISCA design CGH+SNP, Agilent, Santa Clara, CA).

Whole exome sequencing was performed. Exons and splice sites were captured using the Agilent SureSelectXT Human Exon v4+UTRs Kit (Agilent, Santa Clara, CA) and 2x100 bp paired-end WES was performed on an Illumina HiSeq 2500 (Illumina Inc., San Diego, CA, USA).

Quality metrics for sequencing are shown in **Table 3.8**. Reads were prepared for analysis using Cutadapt version 1.2.1 to remove the adapters and the Fastx Toolkit 0.0.13.2 to trim the first 5 bp at 5' end of reads. Data was aligned to NCBI37/hg19 using BWA version 0.7.3a MEM (Maximum Exactly Match) (161, 162). Alignment statistics are presented in **Tables 3.8 and 3.9**. Local realignment around insertions and deletions, reads base quality recalibration, and variant calling were conducted using GATK Tool Kits version 2.6-5. GATK HaplotypeCaller (HC) was used to call variants. Filtering of variants is described in **Table 3.10**. The number of family members and availability of unaffected siblings enhanced the ability to filter gene variants that did not match an autosomal recessive inheritance model.

Sample ID	Samples in Flow Cell Lane	Yield (Mbases)	Number of Reads	% of raw clusters per lane	% of Perfect Index Reads	% of One Mismatch Reads (Index)	% of $\geq$ Q30 Bases (PF)	Mean Quality Score (PF)
AI1	1	40,014	400,142,260	99.31	97.87	2.13	89.32	34.95
AI2	1	40,103	401,028,406	96.61	95.91	4.09	88.38	34.62
AII1	3	10,564	105,639,628	25.43	99.29	0.71	88.14	34.58
AII2	1	38,942	389,418,144	99.25	98.06	1.94	89.94	35.10
AII3	2	21,150	211,502,536	53.01	98.52	1.48	89.73	35.06
AII4	1	37,010	370,102,596	99.20	99.11	0.89	90.08	35.18
AII5	1	37,489	374,888,092	99.37	99.52	0.48	90.93	35.41
AII6	1	39,231	392,306,596	99.19	98.50	1.50	89.47	34.99
AII7	2	18,405	184,053,748	46.13	99.03	0.97	89.80	35.09
BI1	2	16,244	162,437,184	41.25	79.91	20.09	86.17	33.95
BI2	2	10,532	105,319,862	26.74	79.86	20.14	86.33	33.98
BII1	2	16,802	168,018,024	42.56	77.08	22.92	84.58	33.44
BII3	2	18,121	181,214,362	45.91	77.23	22.77	84.07	33.29

Table 3.8 Quality Metrics for Sequencing by Sample, Families A and B

Sample ID	Reads Aligned to the Exome	Average Coverage Across Exome (reads)	Average Coverage of Called Variants (reads)
AI1	351,637,889	324	130
AI2	356,961,485	329	150
AII1	90,981,936	84	101
AII2	344,391,857	317	128
AII3	181,142,719	167	117
AII4	329,790,954	304	131
AII5	330,827,415	305	154
AII6	352,234,023	324	158
AII7	166,247,481	153	131
BI1	132,290,201	236	121
BI2	84,124,803	153	105
BII1	126,805,441	240	125
BII3	145,614,552	264	124

Table 3.9 Whole Exome Sequencing Alignment Statistics, Families A and B

MCM9 early stop codon mutation was identified using the same methodology described above. Primers to bracket c.394C>T variant (chr6:119,245,033-119,245,256, NCBI37/hg19) were 5' – GCC TGA GAG GCA AGT GAA TTT AG – 3' and 5' – TAC CTA AAA CCA AGG ATG TGG GA – 3'. The premature stop codon is assumed to result in loss of functional protein and therefore inhibit the function of MCM9.

	Family A	Family B
Total Variants Called in the Family	36,908	38,060
Application of GATK Quality Filters	36,056	37,430
Variants in Exons or Splice Sites	25,512	26,369
Nonsynonymous Variants	12,048	12,560
Remove Known Polymorphic Genes from NIH Recommendations	10,835	11,141
Variants with a Minor Allele Frequency <5%	1,765	1,743
Coverage >8	1,640	1,676
Homozygous Recessive	3	19

Table 3.10 Whole Exome Sequencing Variant Filtration, Families A and B

### 3.1.4.2 SNP Arrays and Whole Exome Sequencing to Identify *MCM8* c.446C>G

Regions of homozygosity within the family were mapped using Affymetrix GeneChip Human Mapping 250K Nsp array (Affymetrix). Data was deposited in the Gene Expression Omnibus (Accession Number GSE56043). Exons and splice sites were captured using the Agilent Haloplex All Exon Kit and WES was performed on an Illumina HiSeq 2500. Raw data were deposited in the Sequencing Reads Archive (ID number SRP046742). Sanger sequencing was used to confirm WES discovered variants and to evaluate 200 fertile controls for putative damaging variants.

### 3.1.4.3 Verification of *MCM9* c.1732+2T>C Splice Site Mutation

The *MCM9* c.1732+2T>C variant was confirmed by Sanger Sequencing (**Fig. 3.6A**). Verification of the *MCM9* c.1732+2T>C variant was completed via PCR amplification of the region surrounding the variant using KAPA HiFi Taq Polymerase according to manufacturer's instructions. Primers to bracket the c.1732+2T>C variant (chr6:119,148,847-119,149,165, NCBI37/hg19) were 5' – TCT AGG AGG TCC CGA GAT GG -3' and 5' – CAA AGG CAG AGT GAT TGC CG – 3'. PCR products were then sent to Beckman Coulter Genomics (Danvers, MA) for sequencing. Results were analyzed using Sequencher (Gene Codes Corporation, Ann Arbor, MI USA). Sample chromatograms are shown in **Fig. 3.6A** (bottom). A control population of Caucasian women, who had at least one live birth, was recruited at Magee-Womens Hospital. Informed written consent and samples were obtained and subsequently de-identified at the time of recruitment. The study was approved by the Institutional Review Board of the University of Pittsburgh. The *MCM9* c.1732+2T>C mutation was not present in the 200 fertile women from this cohort. The exon 9 splice donor site is conserved at the nucleotide level across species including mouse, xenopus, and zebrafish.

#### **3.1.4.4 Identification of mRNA Isoforms Resulting from *MCM9* c. 1732+2T>C Mutation**

We isolated mRNA from Epstein-Barr virus transformed lymphoblastoid cell lines using Trizol (Invitrogen) from unaffected mother AI2 (heterozygous), and affected daughter AII6 (homozygous for the mutation), and converted it to cDNA using Super Script III First Strand Synthesis kit (Invitrogen). To amplify mRNA isoforms, we used KAPA HiFi Hotstart PCR Kit with dNTPs (KAPA Biosystems).

#### **3.1.4.5 Foci Formation Assays**

Human wild-type *MCM9* was cloned into pEGFPc2 (Clontech) in frame with the N-terminal GFP tag using *XhoI* and *Sall*. pEGFPc2-*MCM9* c.1732+2T>C was generated by Quikchange introducing an *SphI* site for screening. Human wild-type *MCM8* was cloned into pEGFPc2 (Clontech) in frame with the N-terminal GFP tag using *BglII* and *BamHI*. pEGFPc2-*MCM8* p.P149R was generated by Quikchange and KAPA HiFi polymerase introducing an *NruI* site for screening. pEGFPc2-*MCM8* p.E341K was generated by Quikchange and KAPA HiFi polymerase introducing a *BamHI* site for screening. Human embryonic skin keratinocytes (293T cells) were seeded onto 12-cm glass coverslips coated in poly-L-lysine and cultured under standard conditions. Transient transfection of either wild-type *MCM9*-GFP or mutant *MCM9*-GFP (c.1732+2T>C) into 293T cells was performed using Lipofectamine 2000 (Invitrogen) according to the manufacturer's directions and cells were exposed to 300 nM mitomycin C (MMC) for 6 hours the following day to induce DNA damage. Coverslips were rinsed twice in PBS, fixed with 3% paraformaldehyde for 15 minutes, and incubated in PBS with 20% sodium dodecyl sulfate and 10% Triton-X-100 for 15 minutes. Coverslips were mounted using Fluoroshield + DAPI (Sigma) and imaged using an

Olympus Fluoview 500 or Olympus Provis AX70 confocal microscope. GFP foci were counted in 15 cells per experiment.

### 3.1.4.6 Functional Analysis of Chromosomal Instability

Peripheral lymphocytes from unaffected (IV-3, III-2; **Fig. 3.8A-B** and AII2, BI2, and BII3; **Fig. 3.9A**) and affected (IV-1; **Fig. 3.8C** and AII4, AII6, and BII1; **Fig. 3.8B,C**) family members were stimulated with phytohaemagglutinin (PHA) and cultured in the presence of MMC as described previously (150). The control for these studies was an unrelated female individual whose cells were used as a baseline for comparison in both families. Briefly, cells were cultured in the presence of MMC at concentrations of 0 nM, 50 nM, 150 nM, or 300 nM. Cells were harvested after 72 hours of incubation at 37°C. At least ten metaphase spreads per sample were evaluated for chromosome aberrations and breaks (**Fig. 3.8** and **Table 3.11**).

Sample	<i>MCM9</i> Genotype	Mitomycin C Concentration			
		0 nM	50 nM	150 nM	300 nM
Control		0±0	0.4±0.5	0.5±0.5	0.8±0.4
<b>Family A – c. 1732+2T&gt;C</b>					
AII2	WT/MT	0±0	0.2±0.4	0.6±0.5	0.8±0.6
AII4	MT/MT	0.1±0.3	0.6±0.5	10.5±2.1	24.1±4.0
AII6	MT/MT	0.2±0.4	0.8±0.5	8.1±2.3	14.8±9.2
<b>Family B – c.C394T</b>					
BI2	WT/MT	0.2±0.4	0.1±0.3	0.3±0.5	1.9±2.2
BII1	MT/MT	0.1±0.4	1.6±0.9	3.9±1.2	23.5±6.6
BII3	WT/WT	0.2±0.4	0.2±0.4	0.4±0.6	1.2±1.8

Table 3.11 Average number of Chromosomal Breaks Per Cell After MMC Treatment, Families A and B

### 3.1.4.7 DNA Binding Assay

We generated wild-type and mutant *MCM8* cDNAs corresponding to the N-terminus of the protein that is predicted to bind DNA, and compared their ability to bind DNA. The wild type *MCM8* cDNA fragment that encodes N-terminus (nucleotides 1-1104; amino acids 1-368) was amplified from HeLa cell cDNA with Accuzyme polymerase (Bioline; F: 5'-CAC CGG ATT CAT GAA

TGG AGA GTA TAG AGG CAG; R: 5' - ATT ATG CAT CTA CTG TCC TTT GCT ATT ACT AAT AGA ATT TG), and subcloned into a pET30a expression vector (EMD Chemicals) using *BamHI* and *NotI* to allow for expression of an N-terminal 6XHis tag used in purification. The c.446C>G mutation was generated using a standard Quikchange protocol with KAPA HiFi DNA polymerase. MCM8 wild-type and mutant proteins were induced in BL21(DE3) Rosetta 2 cells with IPTG (0.1 mM; 16 hours at 15°C). Cells were lysed and sonicated (buffer: 20 mM sodium phosphate buffer [pH 7.5], 300 mM NaCl, 10 mM imidazole, 10 mM β-mercaptoethanol, and lysozyme). Soluble protein was purified using a nickel column (Thermo Fisher), washed with 100 mM NaCl, eluted with 250 mM imidazole, and further purified using an AKTA Prime FPLC system with a HiTrap heparin column (GE Healthcare). Final cleanup and sizing was performed with a Superdex 200 26/60 column (GE Healthcare). The extinction coefficient of MCM8 1-368 was determined to be 40,380 M<sup>-1</sup> cm<sup>-1</sup>. Electrophoretic mobility shift assays (EMSAs) were performed. As the affinity of MCM8 to DNA remains unknown, we chose a random 46 nucleotide long single stranded DNA (ssDNA; 5'-CGA TGA GAG CGA GTC GCA TGG TAT CCC GTA AAT TGG GAT GCT TAG GCT TA - 3'), as RAD51 is recruited to ssDNA ends at sites of damage and MCM8 is likely to bind at these sites (68). EMSAs were performed in a 15 μl reaction volume containing 20 mM sodium phosphate buffer (pH 7.5), 100 mM NaCl, 20 mM β-mercaptoethanol and 5 nM DNA probe labeled at the 5'-end using a standard polynucleotide kinase reaction and <sup>32</sup>P-γ-ATP, and the indicated amount of MCM8. Binding reactions were allowed to equilibrate for 5 min followed by directly loading onto a 6% polyacrylamide/TBE gel. Gels were run for 20 minutes at 13 volts cm<sup>-1</sup> followed by imaging using a Storm phosphorimager (GE Healthscience). Quantification of the fraction of band shift was performed using the ImageQuant

software (v5.0). Data were fit using Kaleidagraph (Synergy) to a single site binding model defined by  $\Delta F[MCM8]/K_d + [MCM8]$  where  $F$  is the fraction bound and  $K_d$  is the dissociation constant.

#### **3.1.4.8 Homology Model of Human MCM8**

Modeling of MCM8 and the identification of the p.P149R substitution was created using Swiss-PDB after threading the sequence alignment onto the structure of *Sulfolobus solfataricus* MCM (PDBID:3F9V).

#### **3.1.4.9 Crystallization of MCM8 N-Terminal Domain**

Initial crystallization trials were performed by hanging drop vapor diffusion with the 12 most common crystallization buffer conditions at 2-6 mg/ml and 1:1, 1:2, 2:1 protein to mother liquor ratio. Crystals began to become visible after 1 hr, both at 4° C and room temperature. After growth, these initial crystals were screened using vapor drop vapor diffusion with a 1:1 ratio of protein at 3 mg/ml in a random sparse matrix screen (Wizard II Emerald Bio Systems). Crystals grew in two conditions: 25% PEG 3350, 100mM Tris pH 8.5, 200 mM LiSO<sub>4</sub> and 25% PEG 8000, 100 mM HEPES pH 7. Crystals were optimized at 17% PEG 3350, 100 mM Tris pH 8.5, 200 mM LiSO<sub>4</sub> and grew to 0.5 to 1 um final at 4 degrees. Crystals equilibrated to 30% PEG3350, 100 mM Tris 8.5, 200 mM LiSO<sub>4</sub> to cryoprotect and flash frozen in liquid nitrogen. Crystals were screened using FR-E rotating anode and Saturn 944 CCD image plate detector.

For crystallization of MCM8(73-368), 1:1 ratio of protein at 5, 20, 25 and 30 mg/mL was screened in random sparse matrix screen using hanging drop diffusion (Wizard I& II Emerald Bio systems, JCSG plus core suite)

Crystal macro and micro seeding was performed using these crystals. Moved single crystal into 15% PEG 3350, 100mM Tris pH 8.5, 200 mM LiSO<sub>4</sub> for 2, 5, or 10 minutes and then moved into stabilization conditions 16-25% PEG 3350, 100 mM Tris pH 8.5, 200 mM LiSO<sub>4</sub>.

## 4.0 TOWARDS THE ACTIVITY AND FUNCTION OF MCM9 MOTIFS

### 4.1 MCM9 C-TERMINUS IS ESSENTIAL FOR RESPONSE TO DNA DAMAGE

#### 4.1.1 Research Design

MCM8 and MCM9 bear significant sequence homology and form a stable nuclear complex, demonstrated in co-immunoprecipitation studies (67, 68). While both proteins were initially implicated in normal DNA replication (45, 69, 70), *MCM8* and *MCM9* knockout mice are viable but encounter difficulties in germ-line stem-cell maintenance and homologous recombination (HR) (74, 75). Depletion of MCM8 or MCM9 results in genomic instabilities, sex-specific tumor development, and sterility and the MCM8-9 complex participates in preventing or repairing DNA double strand breaks (DSBs) (74, 76). An interaction between Rad51, a key factor in HR and MCM8-9 forms the basis of this involvement in HR. The MCM8-9 complex interacts with Rad51 in cell extract, shown through co-IPs and in live cells, MCM8 and MCM9 form foci at sites of DNA damage that co-localize with Rad51 (68, 74, 76, 163). *MCM9*-null mouse embryonic fibroblasts (MEFs) are defective in HR and impaired in their ability to recruit Rad51 to sites of DNA damage (68).

*MCM9*<sup>L</sup> contains an extended C-terminus that does not possess homology to other members of the MCM family or to any other protein domain that we can identify (**Fig. 1.2**). Therefore, we set out to determine the functional significance of the extended MCM9 C-terminus. We have identified and validated specific motifs within the C-terminus that are responsible for

nuclear import as well as interactions with DNA damage foci. We predict that the role of the MCM9 C-terminus is to localize both MCM8 and MCM9 to the nucleus and interact with homologous recombination proteins during double strand break repair.

## **4.1.2 Results**

### **4.1.2.1 MCM9<sup>M</sup> Is Cytoplasmic and MCM9<sup>Cterm</sup> Imports to Nucleus But Does not Form Damage-Dependent Nuclear Foci**

To investigate the role of the unique C-terminus of MCM9<sup>L</sup> in the response to DNA damage, we reproduced previously reported MCM9<sup>L</sup> foci (76) and observed the response of MCM9<sup>M</sup> and MCM9<sup>Cterm</sup> to MMC treatment. MCM9<sup>Cterm</sup> contains only amino acids 608-1143, excluding the AAA+ core domain of the protein, which is predicted to provide the basis of the protein's enzymatic function.

We transiently expressed pEGFPc2-MCM9<sup>L</sup> in 293T cells, treated cells with the crosslinking agent mitomycin C (MMC) to form nuclear GFP-MCM9<sup>L</sup> foci that were visualized with confocal microscopy (**Fig. 4.2**). In the absence of DNA damage, MCM9<sup>L</sup> is exclusively nuclear. Interestingly, GFP-MCM9<sup>L</sup> also forms foci in the absence of damage, but the number of foci increases significantly in number (> 5-fold) upon exposure to MMC (164).

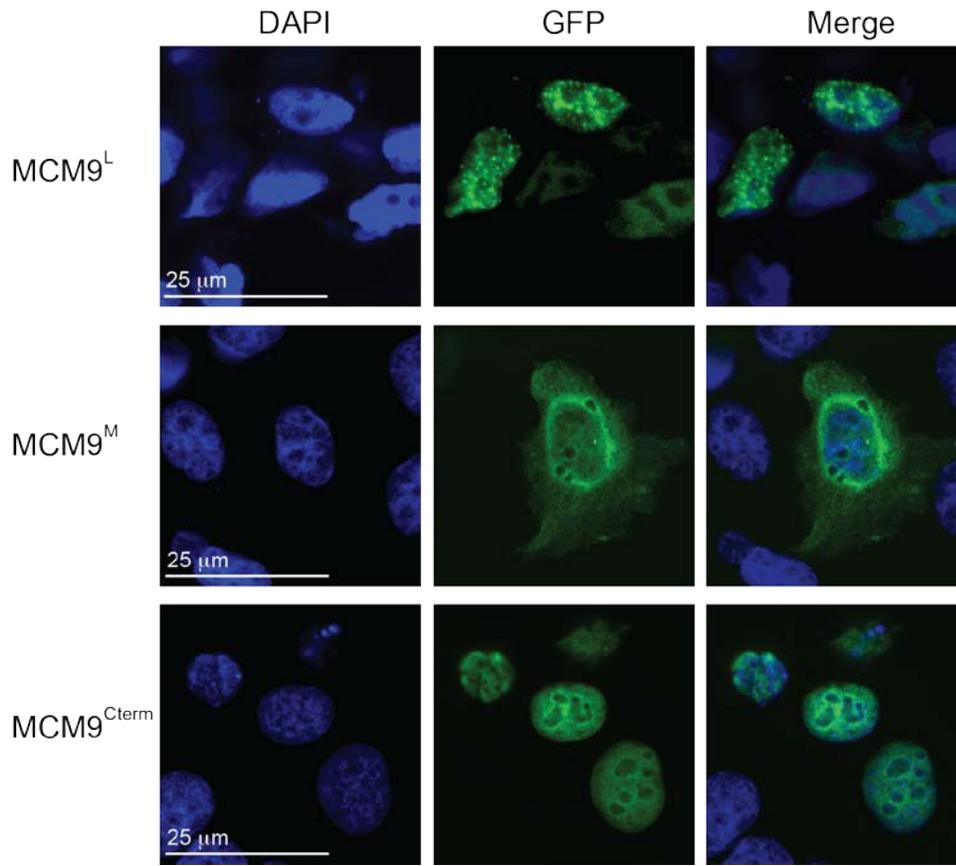


Figure 4.1 Formation of MCM9 foci at sites of DNA double-strand breaks. 293T cells were transfected with GFP-tagged MCM9 constructs and exposed to 300 nM MMC for 6 hrs to induce formation of DNA double-strand breaks. Nuclei were stained with DAPI and nuclear foci were visualized with confocal microscopy. While MCM9<sup>L</sup> readily forms nuclear damage-dependent foci, MCM9<sup>M</sup> is localized to the cytoplasm and MCM9<sup>Cterm</sup> is nuclear, but does not form damage-dependent foci.

Transiently transfected pEGFPc2-MCM9<sup>M</sup> is mainly cytoplasmic, suggesting that the unique C-terminus of the protein is important for nuclear import. MCM9<sup>M</sup> does not form damage-induced foci, and there is no nuclear localization of the isoform. This finding suggests that the C-terminus of MCM9 contains a motif necessary for nuclear localization.

A transiently expressed GFP-tagged version of the C-terminus of MCM9 (aa608-1143, **Fig. 4.3B**) localizes to the nucleus of the cell but does not form foci in response to double-strand breaks formed by exposure to MMC (**Fig. 4.1**). This finding, together with the observation that MCM9<sup>M</sup> does not enter the nucleus supports the importance of the C-terminus for nuclear import.

The inability of this construct to form damage-induced foci supports the importance of the AAA+ core domain for the response of MCM9 to DNA damage. As the C-terminal tail alone (aa 608-1143) is insufficient to form foci, we conclude that the AAA+ core domain is also necessary for damage-dependent foci formation.

#### 4.1.2.2 MCM9 NLS<sup>-</sup> Disrupts Nuclear Localization

Based on the results shown in **Fig. 4.1**, we searched for a putative nuclear localization signal in the C-terminus of MCM9 using the NLS prediction tool: [nls-mapper.iab.keio.ac.jp/](http://nls-mapper.iab.keio.ac.jp/). The nuclear localization signal (NLS) is a series of positively charged lysines or arginines on the surface of the protein that interact with the  $\alpha$ -subunit of the cytosolic receptor importin (165, 166). The  $\beta$  subunit of importin binds to the nuclear pore and allows for nuclear import (166, 167). Using the NLS prediction tool, we identified amino acids 891PKRK as a putative NLS. MCM8 is devoid of an NLS, and we expect that complexation with MCM9 allows MCM8 to enter the nucleus. We cloned MCM9 NLS<sup>-</sup>, containing the mutation in the putative NLS motif (MCM9 c.891PKRK-GKDD) into pEGFPc2. This mutation was designed to replace several amino acid residues that are positively charged at physiological pH (arginine and lysine) with negatively charged residues (aspartic acid) predicted to disrupt nuclear import (167).

A transiently expressed GFP-tagged version of MCM9 NLS<sup>-</sup> is exclusively found in the cytoplasm and does not form foci in response to DNA double-strand breaks (**Fig. 4.2**). The cytoplasmic localization of this construct supports the importance of the NLS for nuclear import. The localization of MCM9 NLS<sup>-</sup> is similar to that of MCM9<sup>M</sup>, confirming the importance of the C-terminal extension and the NLS motif in particular for nuclear import and damage-dependent foci formation.

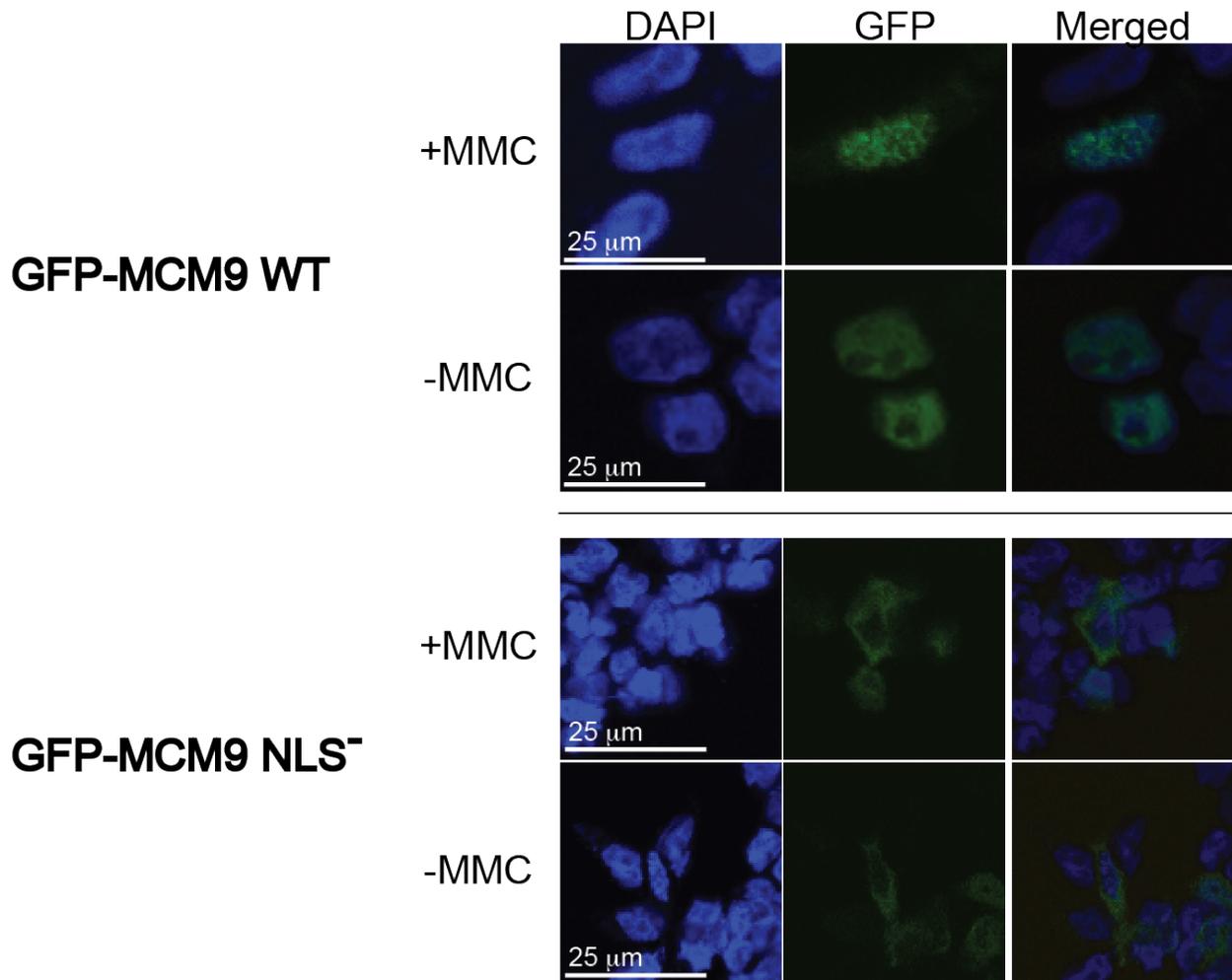


Figure 4.2 MCM9 NLS<sup>-</sup> disrupts nuclear import.

293T cells were transfected with GFP-tagged versions of mutated or wild type MCM9 and exposed to 300 nM MMC for 6 hrs to induce formation of DNA double-strand breaks. Nuclei were stained with DAPI and nuclear foci were visualized with confocal microscopy. While wild type MCM9 readily forms nuclear foci at sites of DNA damage, MCM9 NLS<sup>-</sup> does not localize to the nucleus.

#### 4.1.2.3 MCM9 BRCV<sup>-</sup> Disrupts Damage Induced Foci Formation

BRCA2 interacts with Rad51 through a series of 8 conserved BRC motifs, encoded in exon 11 of the human *BRCA2* gene (102, 103). These motifs are conserved in all mammalian BRCA2 proteins that have been sequenced (103). Yeast two-hybrid searches using various segments of the *BRCA2* gene as “bait” reveal that the interaction between Rad51 and BRCA2 arises in the BRC motifs (102). Each BRCA2 motif contains a 59-residue minimal region which forms the basis of the

Rad51 interaction. Two BRC motifs with an intervening linker region have been identified in the RecQ15 helicase and have been shown to be necessary for binding Rad51 (104). This domain, containing the two BRC motifs and linker region is considered a BRC variant (BRCv) motif. Co-IPs between RecQ15 truncations and Rad51 reveal that this region is essential for Rad51 interaction (104). D-loop formation assays including purified truncations of RecQ15 and Rad51 have provided further evidence of the *in vitro* role of the BRCv motif forming an interaction with Rad51 (104). *In vivo* assays have also supported the function of the BRCv, such as sister chromatid exchange (SCE) assays which demonstrate that the BRCv motif of RecQ15 is sufficient to rescue SCE frequency in RecQ15<sup>-/-</sup> cells (104). Further, camptothecin resistance is restored by the BRCv motif of RecQ15 in RecQ15<sup>-/-</sup> cells (104). We have identified a BRCv motif in the C-terminus of MCM9 that is homologous to the RecQ15 BRCv repeat motifs (**Fig. 4.3**) (104). We cloned MCM9 BRCv<sup>-</sup>, in which two residues homologous to key Rad51-interacting residues in RecQ15, are mutated (FR687-8AA) (104). The damage-dependent cellular localization of this construct was observed.

A) MCM9's BRC variant (BRCv) motif

BRCA2_BRC1	1005	FGGS	E	R	T	A	S	N	K	E	I	K	L	S	E	H	N	I	K	-----	K	S	K	M	F	F	K	-	D	I	E	E													
BRCA2_BRC2	1215	GFRG	F	Y	S	A	H	G	T	K	L	N	V	S	T	E	A	L	Q	-----	K	A	V	K	L	F	S	-	D	I	E	N													
BRCA2_BRC3	1424	SDTF	F	Q	T	A	S	G	K	N	I	S	V	A	K	E	S	F	N	-----	K	I	V	N	F	F	D	-	Q	K	P	E													
BRCA2_BRC4	1520	TLLG	F	H	T	A	S	G	K	V	K	I	A	K	E	S	L	D	-----	K	V	K	N	L	F	D	-	E	K	E	Q														
BRCA2_BRC5	1667	SALAF	T	S	C	S	R	K	T	S	V	S	Q	T	S	L	L	-----	E	A	K	K	W	L	R	-	E	G	I	F															
BRCA2_BRC6	1840	GPPAF	R	I	A	S	G	K	I	V	C	V	S	H	E	T	I	K	-----	K	V	K	D	I	F	T	-	D	S	F	S														
BRCA2_BRC7	1974	TCGI	F	S	T	A	S	G	K	S	V	Q	V	S	D	A	S	L	Q	-----	N	A	R	Q	V	F	S	-	E	I	E	D													
BRCA2_BRC8	2054	AFSG	F	S	T	A	S	G	K	Q	V	S	I	L	E	S	S	L	H	-----	K	V	K	G	V	L	E	-	E	F	D	L													
BRCv Motif		---	F	x	x	A	-----	LINKER	-----	h	x	-	a	-----																															
RECQL5_BRCv	662	GSCPF	F	Q	T	A	T	--	E	L	M	E	T	T	R	I	R	E	Q	A	P	Q	P	E	R	G	G	E	H	E	P	P	S	R	P	C	G	L	L	D	-	E	D	G	S
MCM9_BRCv	683	EESN	F	R	T	S	S	Q	Q	E	I	N	S	T	H	I	F	S	P	G	S	P	E	G	S	P	V	L	D	---	P	P	H	L	E	P	N	R	S	T					
consensus		---	g	f	-	t	a	s	g	k	-	v	-	v	s	-	s	l	-----	k	-	k	-	l	f	d	-	e	-																

B) MCM8 and MCM9 truncations and mutations

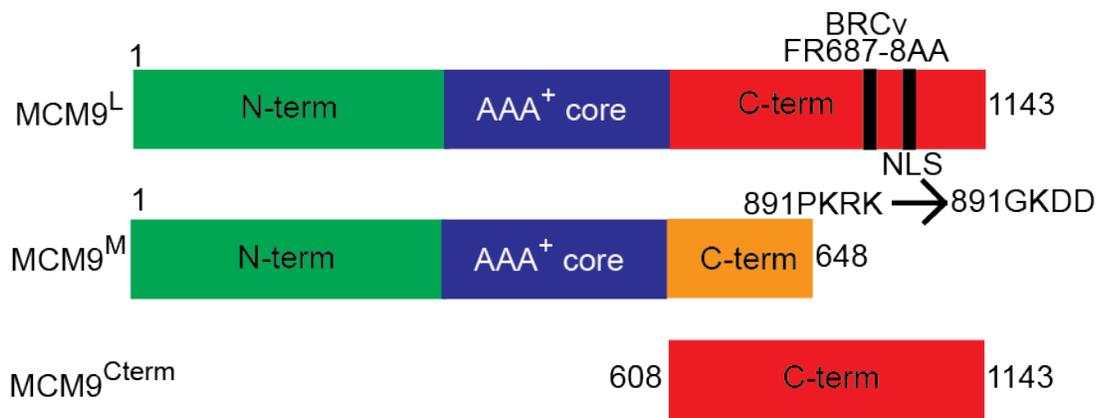


Figure 4.3 MCM9 BRCv and NLS motifs.

(A) Alignment of the MCM9 BRCv motif with the 8 BRC motifs in BRCA2 and the one BRCv motif in RECQL5. (B) Location of BRCv and NLS motifs in MCM9 linear sequence. Both motifs are present in MCM9<sup>L</sup> and MCM9<sup>Cterm</sup>, but not in MCM9<sup>M</sup>.

MCM9 BRCv<sup>-</sup> disrupts the BRCv motif which we propose to interact with Rad51 based on its homology to the RIP in BRCA2. A transiently transfected GFP-tagged version of this mutant is nuclear in damaged and undamaged cells, but does not form damage-induced foci (**Fig. 4.4**). This finding supports the importance of the BRCv motif for the response and localization of MCM9 to DNA damage.

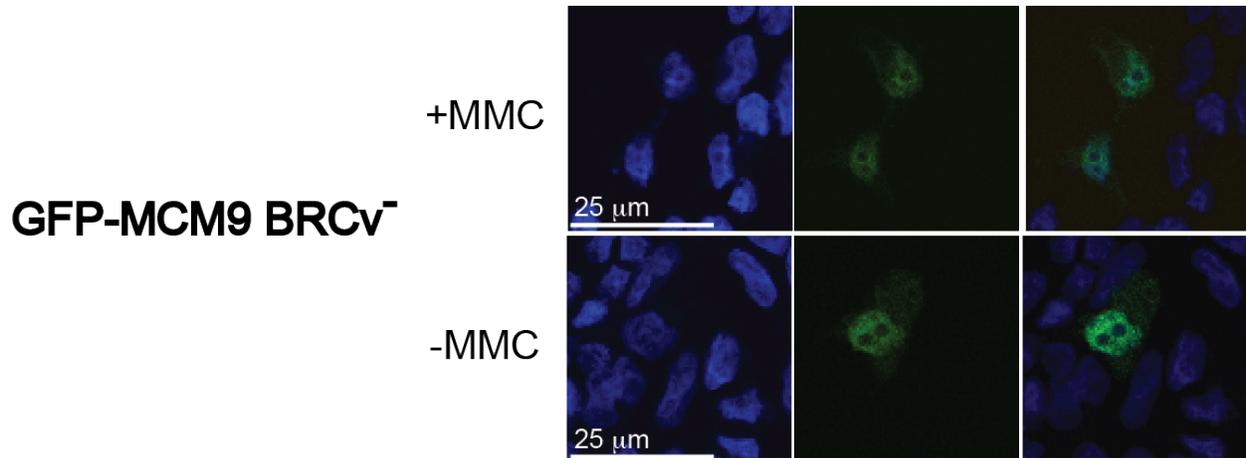


Figure 4.4 MCM9 BRCV<sup>-</sup> disrupts damage-dependent foci formation.

293T cells were transfected with GFP-tagged versions of mutant BRCV<sup>-</sup> MCM9 and exposed to 300 nM MMC for 6 hrs to induce formation of DNA double-strand breaks. Nuclei were stained with DAPI and nuclear foci were visualized with confocal microscopy. While wild type MCM9 readily forms damage-dependent nuclear foci, MCM9 BRCV<sup>-</sup> localizes to the nucleus but does not form nuclear foci.

### 4.1.3 Conclusion

We have characterized multiple isoforms of MCM9 (MCM9<sup>L</sup> and MCM9<sup>M</sup>), a truncation (MCM9<sup>Cterm</sup>) and mutated versions of the protein (MCM9 NLS<sup>-</sup> and MCM9 BRCV<sup>-</sup>). In our analysis, we have determined the essential domains and motifs of MCM9 for nuclear import and DNA damage-dependent foci formation exist in the extended C-terminus. Our observations reveal that nuclear localization requires the NLS motif (891PKRK) and DNA damage-induced foci formation depends on the AAA+ domain and the BRCV motif (FR687-688). While WT MCM9<sup>L</sup> readily forms damage-induced foci, MCM9<sup>M</sup> and MCM9 891PKRK-GKDD do not enter the nucleus, and MCM9<sup>Cterm</sup> and MCM9 FR687-688AA enter the nucleus but do not form foci. We would expect that the absence of a NLS in MCM8 would dictate that an MCM8/9 complex is imported to the nucleus through the NLS sequence in MCM9. The role of the BRCV motif in MCM9 will require further study to validate that the specific target is in fact Rad51 and that its importance is required for efficient DNA double strand break repair.

#### 4.1.4 Experimental

##### 4.1.4.1 Transfections and Foci Formation Assays

Human skin keratinocyte (293T) cells were seeded onto 12-cm glass coverslips coated in poly-L-lysine and cultured in Dulbecco's Modified Eagle Medium supplemented with 4.5 g/L D-glucose, 4.5 g/L L-glutamine, 4.5 g/L sodium pyruvate and 10% fetal bovine serum in a humidified 5% CO<sub>2</sub> incubator at 37°C. Human wild-type MCM9<sup>L</sup>, MCM9<sup>M</sup> and MCM9<sup>Cterm</sup> were each cloned into pEGFPc2 (Clontech) in frame with the N-terminal GFP tag using *Sall* and *XhoI*. pEGFPc2-MCM9 FR687-688AA was generated by Quikchange and KAPA HiFi polymerase introducing an *AatII* site for screening and pEGFPc2-MCM9 891PKRK-GKDD was generated similarly, introducing a *BspEI* site for screening. Transient transfection of plasmids in 293T cells was performed using Lipofectamine 2000 (Invitrogen) according to the manufacturer's directions. Cells were incubated with transfection mixture overnight, before adding 300 nM mitomycin C for 6 hours. Coverslips were rinsed twice in PBS, incubated in PBS supplemented with 3% paraformaldehyde for 15 minutes, and then incubated in PBS supplemented with 20% sodium dodecyl sulfate and 10% Triton-X-100 for 15 minutes. Coverslips were mounted using Fluoroshield + DAPI (Sigma) and imaged using an Olympus Fluoview 500 confocal microscope.

## 5.0 MCM8 C.446C>G/P.P149R MUTATION CAUSES SENSITIVITY TO MMC BUT NOT TO PARP INHIBITORS

### 5.1.1 Research Design

When PARP-1 is inhibited, replication forks collapse at single-strand DNA interruptions and homologous recombination (HR) is triggered to repair the DNA breaks (114, 115). Cells deficient in HR proteins are thus highly sensitive to PARP inhibitors. PARP-1 does not directly catalyze HR, but its inhibition leads to accumulation of unrepaired SSBs, which become DSBs after one round of replication (168). Because of the initiation of HR in the absence of functional PARP-1, cells deficient in HR proteins are acutely sensitive to PARP inhibitors. This heightened susceptibility has been demonstrated in BRCA2-mutated or deficient tumor cells(115), PTEN-deficient tumor cells (116) and PTEN-mutated cells (117). Knowing that MCM8 and MCM9 are important for Rad51 recruitment to sites of DNA damage, we investigated the synthetic lethal effect of MMC (**Fig. 5.1**) with either of the PARP inhibitors BMN673 or ABT-888 (**Fig. 5.1**) in cells either homozygous or heterozygous for MCM8 c.446C>G/p.P149R, in patients presenting with premature ovarian failure (discussed in further detail in Section 3.1.2.3).

We first assessed the cytotoxicity of MMC and BMN673 individually, before determining their synergistic effect. Cells homozygous for MCM8 p.P149R exhibit heightened sensitivity to MMC, compared with cells from a related individual expressing wild type MCM8. BMN673 and ABT-888, in the absence of induced DNA damage, affect cells with a MT genotype similarly as it affects WT cells.

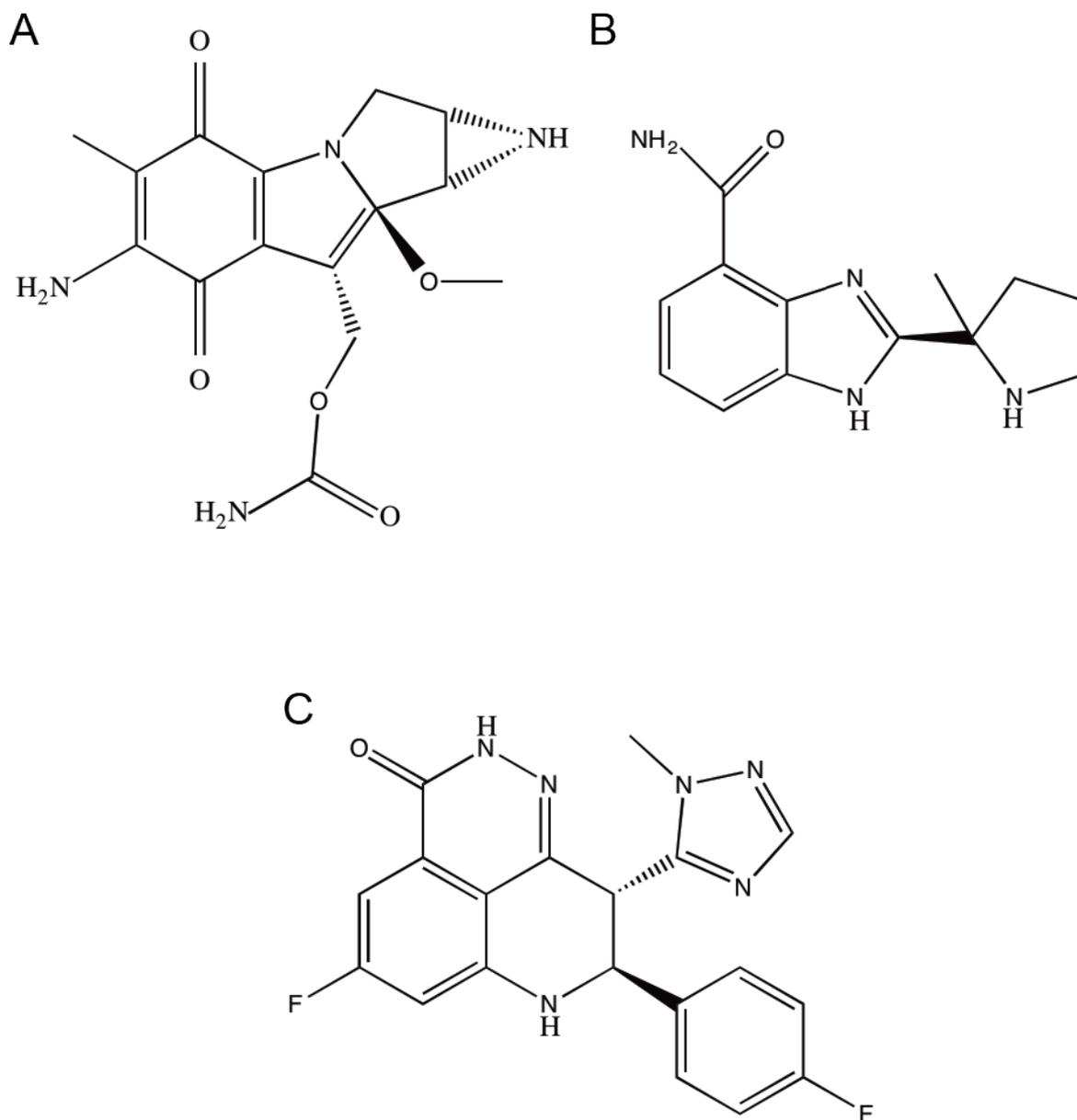


Figure 5.1 Chemical structures of cytotoxic agents. (A) The crosslinking agent mitomycin C, and the PARP inhibitors (B) ABT-888 and (C) BMN673.

We compared the synthetic cytotoxicity of MMC and PARP inhibitor in cells homozygous for MCM8 p. P149R, heterozygous for the mutation or homozygous for wild type MCM8. We tested the cytotoxicity of MMC combined with PARP inhibitor in EBV-transformed lymphocytes homozygous for either MCM8 p.P149R or wild type MCM8. We designed these experiments to reveal any involvement of MCM8 in PARP-1 DNA damage recognition. While cells from patient

IV-1 (homozygous for MCM8 p.P149R, see **Fig. 3.1**) exhibit higher sensitivity to MMC than cells heterozygous for the mutation (patient III-2, see **Fig. 3.1**), or homozygous for wild type MCM8 (patient IV-3, see **Fig. 3.1**), there is no observable PARP inhibitor sensitivity difference among cell lines.

## **5.1.2 Results and Discussion**

### **5.1.2.1 MCM8 c.446C>G/p.P149R Cells Are Acutely Sensitive to MMC Damage**

Fibroblasts homozygous for MCM8 p.P149R (IV-6, **Fig. 3.1**) and fibroblasts homozygous for wild type MCM8 (IV-3, **Fig. 3.1**) were exposed to various concentrations of MMC for 6 hours and live cells were counted 3 days after exposure (**Fig. 5.3A**). A greater dose-dependent cytotoxic effect was observed in cells homozygous for MCM8 p.P149R. After 3 days, significantly more cells homozygous for MCM8 p.P149R had died than cells homozygous for wild type MCM8 (**Fig. 5.1A**). This result shows that c.446C>G/p.P149R disrupts the role of MCM8 in HR to repair DNA damage caused by MMC, consistent with the previously observed complex chromosomal rearrangements in metaphase spreads after MMC exposure (**Fig 3.8**) (155).

The duration-dependence of MMC exposure was tested in cells homozygous or heterozygous for MCM8 p.P149R (IV-6 and III-2 respectively, **Fig. 3.1**), as well as cells homozygous for wild type MCM8 (IV-3, **Fig. 3.1**). We determined percent viability of these three EBV-transformed lymphocyte cell lines 24, 48 and 72 hours after exposure to 10 nM MMC (**Fig. 5.2B**). Lymphocytes homozygous for the mutation exhibit highest sensitivity to MMC of the three cell lines, while cells heterozygous for the mutation exhibit MMC sensitivity intermediate between that of the homozygous mutant and homozygous wild type. Sensitivity increases over the 3-day

timecourse (**Fig. 5.2B**). A MMC concentration of 10 nM was selected to produce a moderate killing effect in lymphocytes.

(A) Seed Cells  $\longrightarrow$  +MMC (6 hours)  $\longrightarrow$  Count Cells  
 (B)

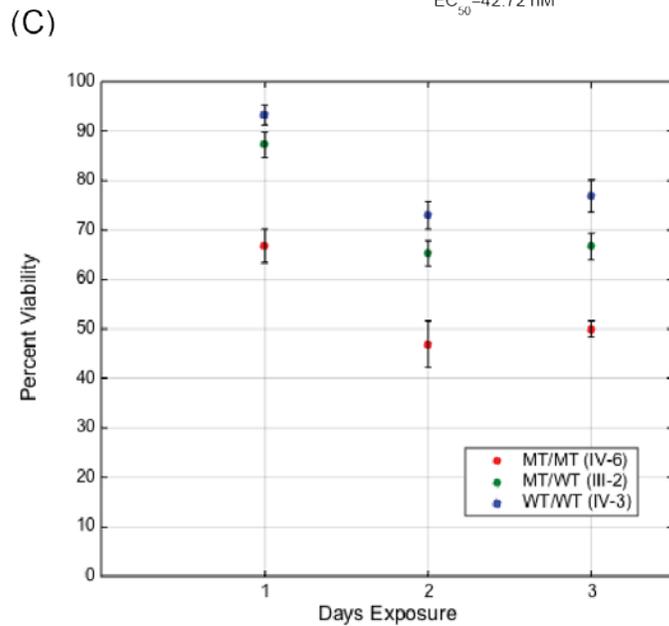
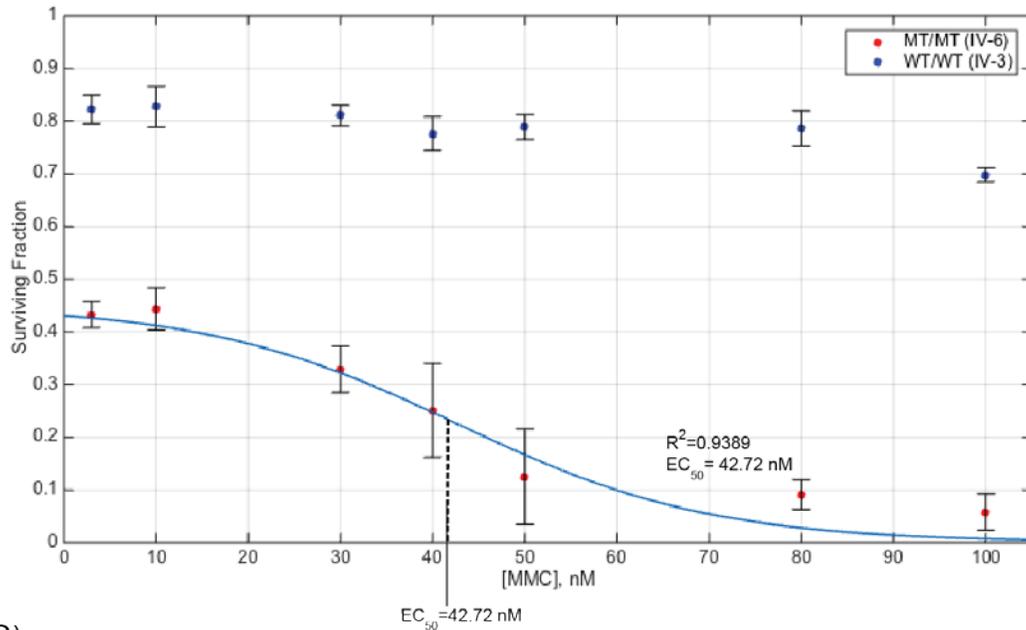


Figure 5.2 MMC cytotoxicity in cells homozygous for MCM8 p.P149R.

(A) Experimental procedure to assess cytotoxicity of MMC mutant cells, compared with cells homozygous for wild type MCM8; (B) Fibroblasts were exposed to various concentrations of MMC for 72 hours and the surviving fraction is indicated for each cell line at each concentration of MMC. (C) Cytotoxicity of 10 nM MMC was compared in EBV-immortalized lymphocytes homozygous for MCM8 p. P149R, heterozygous for the mutation of homozygous for wild type MCM8. Error bars represent standard deviation of the mean comparing three independent experiments. The  $EC_{50}$  was calculated based on a sigmoidal fit, determined to be 42.72 nM MMC. Cell line delineations are defined in **Fig. 3.1**.

### **5.1.2.2 MCM8 c.446C>G/p.P149R Does Not Sensitize Lymphocytes to BMN673 or ABT-888**

BMN673 and ABT-888 are PARP1/2 inhibitors effective for treatment of DNA repair-deficient cancers (169-171). Phase II clinical trials have been completed for ABT-888 (172) while BMN673 has not yet been as extensively characterized and is in Phase I trials (169). The sensitivity of EBV-immortalized lymphocytes containing MCM8 p.P149R to ABT-888 and BMN673 was tested. The potency of the drug at 2  $\mu$ M and 5  $\mu$ M concentrations and effectiveness of our method was demonstrated in PE01 cells, a human ovarian cancer cell line known to be sensitive to PARP inhibition (**Fig. 5.5**). Lymphocytes homozygous for MCM8 p.P149R (IV-6, **Fig. 3.1**), heterozygous for the mutation (III-2, **Fig. 3.1**) or homozygous for wild type MCM8 (IV-3, **Fig. 3.1**) were exposed to various concentrations of BMN673 in DMSO solvent, as well as DMSO alone as a control. Percent viability was determined 24, 48 and 72 hours following treatment. While PE01 cells are sensitive to BMN673 treatment, MCM8 p.P149R does not cause sensitivity to BMN673 (**Fig. 5.3**). Exposure of lymphocytes to 2  $\mu$ M ABT-888 produced a similar result, confirming that MCM8 p.P149R does not confer sensitivity to PARP inhibitors (data not shown).

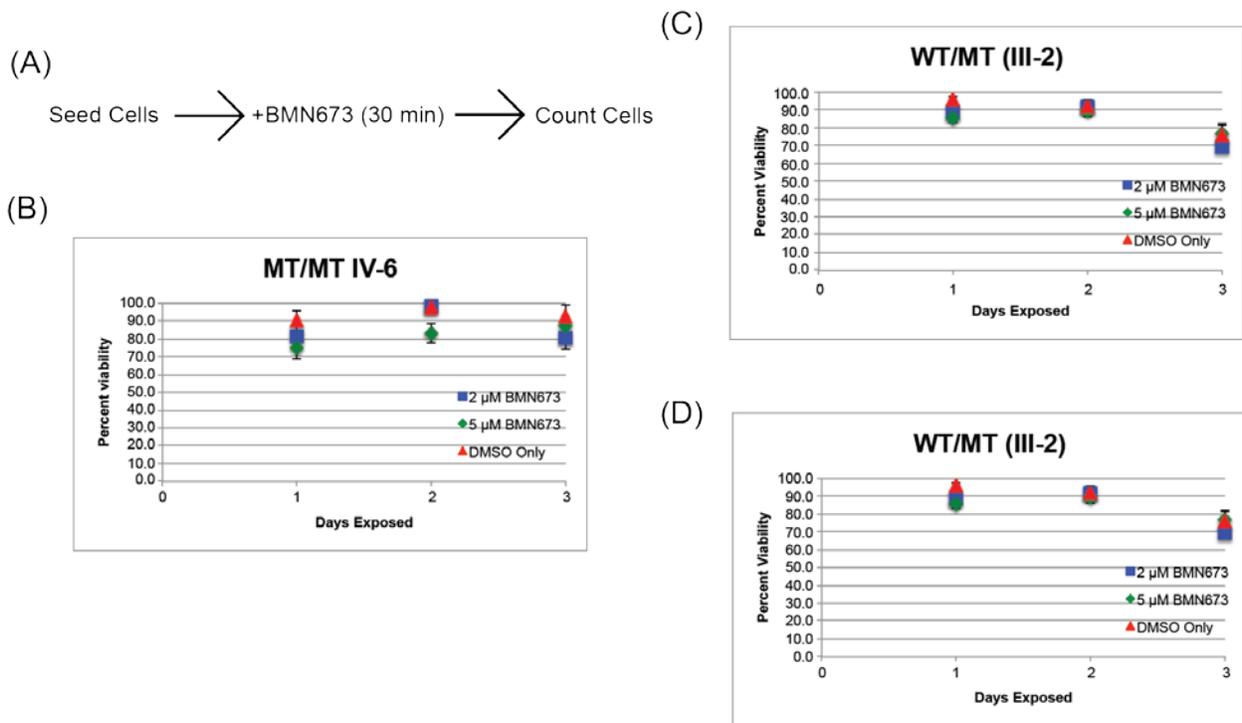


Figure 5.3 BMN673 cytotoxicity in EBV-immortalized lymphocytes homozygous for MCM8 p.P149R. (A) Experimental procedure to assess cell sensitivity to BMN673. (B) MCM8 p.P149R does not sensitize EBV-immortalized lymphocytes to the PARP inhibitor BMN673 in cells homozygous for the mutation, (C) heterozygous for the mutation (D) or homozygous for wild type MCM8. Error bars represent standard deviation of the mean of three independent experiments. Cell line delineations are defined in **Fig. 3.1**.

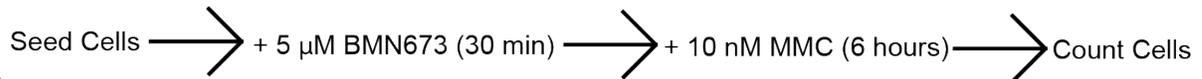
### 5.1.2.3 PARP inhibitors and MMC Have No Synthetic Lethal Effect in MCM8 c.446C>G/p.P149R

The synergistic effect of MMC and the PARP inhibitors ABT-888 or BMN673 was tested in EBV-immortalized lymphocytes. Cells were treated with MMC and PARP inhibitor simultaneously to determine whether MCM8 p.P149R is necessary for HR when the pathway is triggered by inhibition of PARP1/2. A MMC concentration of 10 nM was chosen as this was found to produce a moderate killing effect in cells heterozygous for MCM8 p.P149R (**Fig. 5.2**). This moderate concentration of MMC was combined with 2  $\mu\text{M}$  BMN673, a concentration that had an observable killing effect in PE01 cells. There was no observable synergistic effect of MMC and BMN673 (**Fig. 5.5**). A similar procedure exposing lymphocytes to MMC in combination with 2  $\mu\text{M}$  ABT-

888 produced a similar result, confirming that MCM8 p.P149R does not confer sensitivity to PARP inhibitors (data not shown).

This finding suggests that while MCM8 p.P149R heightens sensitivity to DNA DSBs, this sensitivity does not increase when PARP is inhibited. It is possible that when PARP1/2 are inhibited, another factor(s) performs the role normally performed by wild type MCM8 in HR. Alternatively, recognition of DNA damage by PARP may be entirely separate from the action of MCM8 in the HR pathway.

(A)



(B)

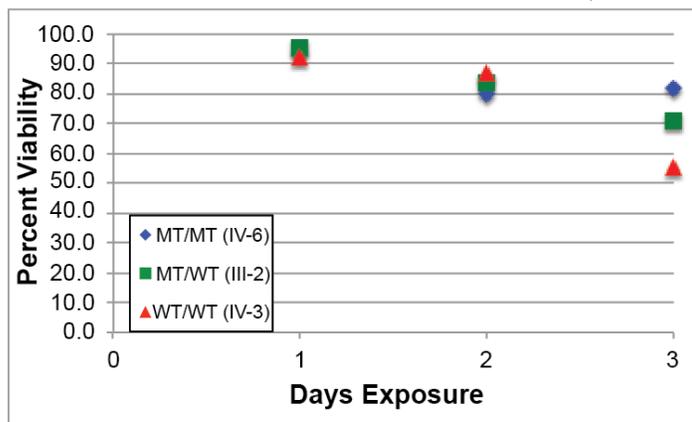
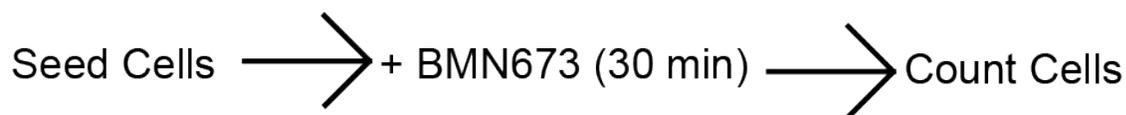


Figure 5.4 Synergistic cytotoxicity of MMC and BMN673 in EBV-immortalized lymphocytes homozygous for MCM8 p.P149R.

(A) Experimental procedure to assess cell sensitivity to BMN673 following exposure to MMC; (B) There is no significant difference in the synergistic effect of MMC and BMN673 on cells homozygous for MCM8 p.P149R, heterozygous for the mutation or homozygous for wild type MCM8. Error bars represent standard deviation of the mean of three independent experiments. Cell line delineations are defined in **Fig. 3.1**.

(A)



(B)

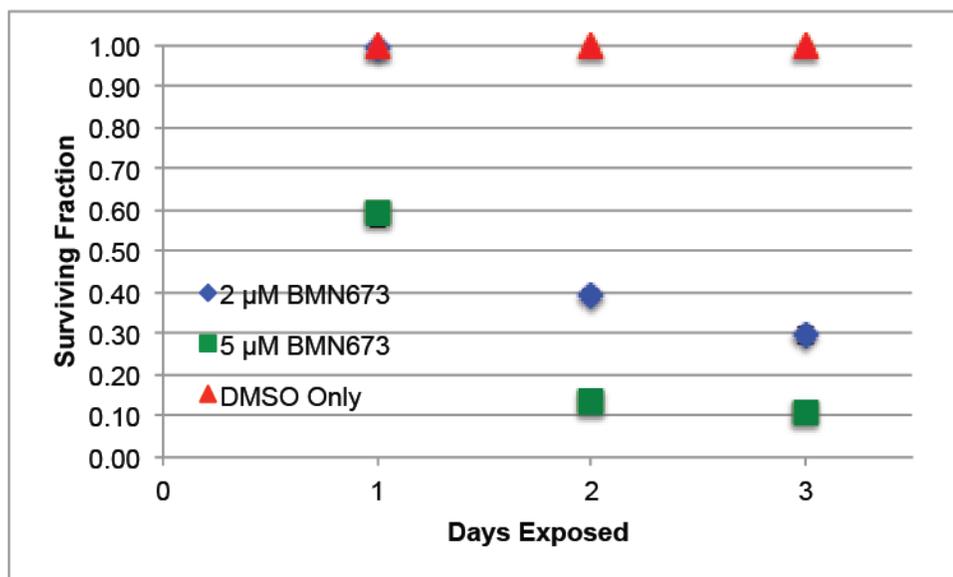


Figure 5.5 Cytotoxicity of BMN673 in PE01 cells.

(A) Experimental procedure to assess PE01 cell sensitivity to exposure to various concentrations of BMN673; (B) PE01 cells were exposed to 2  $\mu$ M or 5  $\mu$ M BMN673 and surviving fraction of cells was calculated each day after treatment, for a total of 3 days. PE01 cells are significantly sensitive to BMN673, compared with sensitivity to DMSO alone. Error bars represent standard deviation of the mean of three independent experiments.

### 5.1.3 Conclusion

Cells homozygous for MCM8 p.P149R are highly sensitive to DNA DSBs, supporting the role that has previously been proposed for MCM8 in HR repair of DSBs as well as the detrimental impact of this pathogenic mutant. The role of MCM8 in HR coupled with PARP1/2 inhibition is inconclusive, as cells homozygous for MCM8 p.P149R respond to PARP1/2 inhibition in a similar way to cells homozygous for wild type MCM8. Our fluorescence microscopy studies (see Section 3) show that MCM8 p.P149R does not form foci in response to DNA damage. While it has been

shown in *Xenopus* egg extract that MCM8-9 are important for recruitment of Rad51 to sites of DNA damage, it is possible that in a cellular context, other factors assume this recruitment role.

Alternatively, it may be that MCM8/9 act later in the HR pathway (synapsis or postsynapsis) where inhibition of PARP1/2 is not synergistic with any defects in MCM8/9 because of temporal separation in the pathway. Many HR proteins that are synergistic with PARPi, such as BRCA1-2 are proposed to act early in the pathway where their detrimental effects combined with PARP inhibition will sensitize cells towards death. As MCM8/9 are proposed to act later in the pathway, recruitment of HR factors after polyadenylation of PARP1 has already begun. At this point, we would propose that the role of PARP1 is finished and HR repair is underway. Mutations in MCM8 or MCM9 would therefore have temporally separate effects from HR proteins acting earlier in the pathway, and cell death would result solely from the inability to resolve DNA crossover intermediates that have occurred showing no synergistic effect with PARP inhibition.

#### **5.1.4 Experimental**

##### **5.1.4.1 Fibroblast MMC Treatment and Surviving Fraction Calculations**

Fibroblast cells were cultured in DMEM/10%FCS/penicillin-streptomycin and equal numbers of cells were seeded into wells of 24-well plates. Cells were exposed to MMC for 6 hours, washed and media was replaced. They were cultured in a humidified incubator with 5% CO<sub>2</sub> for 72 hours. Cells were then washed and live cells were stained with crystal violet (0.5% w/v) in gluteraldehyde (6% v/v). Plating efficiency (PE) was calculated for an untreated well, defined as  $PE = \frac{\#colonies}{\#cells\ seeded}$ . Surviving fraction (SF) following MMC treatment is defined as  $SF = \frac{\#colonies\ after\ treatment}{(\#cells\ seeded)(PE)}$ . This analysis was done as previously described (173). To determine the dose-dependent cytotoxicity of MMC, surviving fraction data

points were fitted to a sigmoidal expression,  $(0.45)/(1+10^{((-\text{EC}_{50}-x)^*a)})$  where  $\text{EC}_{50}$  is the [MMC] corresponding to a half-maximal effect.

#### **5.1.4.2 EBV Lymphocyte MMC and BMN673 Treatment and Percent Viability Calculations**

EBV lymphocyte suspension cultures were grown in single wells of a 24-well plate in Dulbecco's Modified Eagle Medium supplemented with 4.5 g/L D-glucose, 4.5 g/L l-glutamine, 4.5 g/L sodium pyruvate and 10% fetal bovine serum in a humidified 5%  $\text{CO}_2$  incubator at 37°C. MMC and BMN673 in DMSO solutions were added to individual wells, including a DMSO-only control sample for each cell line. 24, 48 and 72 hours following exposure, a 10  $\mu\text{l}$  sample of each was removed and mixed with 10  $\mu\text{l}$  0.4% trypan blue solution in PBS. Live and dead cells were counted using a hemacytometer for each condition and percent viability was calculated for each (Live Cells/Total Cells x 100).

## 6.0 FUTURE PERSPECTIVES

Our investigation of MCM8 and MCM9 reveals their importance in homologous recombination (HR), which occurs both during meiotic recombination to exchange genetic material and in mitotic recombination as a means of repairing DNA double-strand breaks. We have identified pathogenic mutants of MCM8 and MCM9 in patients with premature ovarian failure and confirmed the previously proposed importance of these two proteins in promoting successful HR. As a means of advancing toward the long-term goal of defining mechanisms of DNA repair and relating this understanding to human disease and development of therapies, we aim to investigate the roles of MCM8 and MCM9 in HR. Specifically, I propose defining regions of MCM8 and MCM9 that participate in protein-protein interactions, and characterize the biochemical activities of the two proteins, including the *in vivo* interdependence of these functions. With a goal of enabling women who express pathogenic mutants of MCM8 and MCM9 to produce healthy offspring, strategies will be developed for generation of induced pluripotent stem cells from primary cell samples.

### 6.1.1 Define MCM8 and MCM9 protein interaction sites

As a step toward understanding the functions of MCM8 and MCM9 domains and motifs, the regions of MCM8 and MCM9 that engage in protein-protein interactions need to be defined. While MCM8 and MCM9 form a stable nuclear complex (68), the specific regions of the two proteins participating in this interaction are unknown. Protein complex immunoprecipitation assays between truncations of the two proteins would allow for the interacting regions to be identified. Defining the interacting regions of the two proteins will be a significant step toward elucidating

the structure and orientation of the two proteins as they perform their functions. We have demonstrated that the C-terminal MCM9 BRCv motif is essential for DNA damage-dependent foci formation, and we predict that this motif directly interacts with Rad51. While colocalization at sites of DNA damage as well as a direct *in vitro* protein-protein interaction between wild type MCM9 and Rad51 have been previously observed (68), the region of MCM9 that interacts with Rad51 has yet to be defined. Protein complex immunoprecipitation assays between purified Rad51 and truncated, mutated, or wild type MCM9 will directly define the interacting regions of the proteins. Testing for protein-protein interactions between MCM8-9 and other HR factors including BRCA1-2 would direct further investigation of the role of MCM8-9 in the HR mechanism.

### **6.1.2 Examine the corecruitment of MCM8 p.P149R and MCM9 c.1732+2T>C or p.R132\* patient cell lines**

As a means of understanding the *in vivo* interdependence of MCM8 and MCM9 functions, I propose testing the effect of MCM8 mutations or depletion on the ability of MCM9 to form damage-dependent nuclear foci and vice versa. We predict that the C-terminal MCM9 NLS facilitates nuclear importation of both proteins, and one way of testing this is to transiently express MCM9 NLS<sup>-</sup> and GFP-tagged MCM8 in *MCM9*<sup>-/-</sup> cells and observe localization of MCM8. Using a similar approach, I would propose testing the effect of MCM8 p.P149R on the DNA damage response of wild type MCM9 or the effect of MCM9 c.1732+2T>C or p.R132\* on wild type MCM8. We have demonstrated the reduced affinity of MCM8 p.P149R for ssDNA, and the reduced ability of MCM8 and MCM9 mutants to form damage-induced nuclear foci. The major mRNA products formed from MCM9 c.1732+2T>C and p.R132\* are truncated forms that lack the NLS and BRCv motifs, which we predict will disrupt their involvement in HR. Assessing the effect

of this loss of function on the ability of the interacting partner to form damage-dependent nuclear foci will reveal the interdependence of the cellular functions of MCM8 and MCM9.

### **6.1.3 Test MCM8 and MCM9 biochemical activities**

We have demonstrated the ability of purified MCM8(1-368) to bind ssDNA, but ATP hydrolysis and DNA unwinding activities of MCM8, MCM9<sup>L</sup> and MCM9<sup>M</sup> have not been tested *in vitro*. These biochemical activities should be tested with purified proteins, and the affinity of MCM8 and MCM9 isoforms for various types of DNA substrates will also be tested. Codon-optimized versions of MCM8 and MCM9 expressed in *E. coli* could be used to facilitate purification of soluble protein. Coexpression/ purification of MCM8 and MCM9 should also be tried to stabilize the predicted complex.

### **6.1.4 Continue the structural characterization of MCM8 and MCM9**

Efforts should be continued to optimize crystallization conditions for the two purified proteins, aiming to define a crystal structure for MCM8 and MCM9. Our initial attempts to crystallize N-terminal domains of MCM8 resulted in small and unstable crystal forms. Possibly expressing longer constructs using codon optimization will result in more stable proteins amenable to crystallization. This will further our understanding of the molecular structure of the MCM8-9 protein complex and will provide a structural context for its protein and DNA interactions.

### **6.1.5 Examine trans-acting MCM9 splicing regulation and cytoplasmic role of MCM9<sup>M</sup>**

Our studies do not reveal a specific function for MCM9<sup>M</sup>, and show that this protein isoform is localized in the cytoplasm, excluding it from any direct role in HR. These observations raise a question regarding the role and necessity of MCM9<sup>M</sup>. Assays to determine the interacting partners and biochemical activities of the MCM9 protein isoforms would reveal the specific importance of MCM9<sup>M</sup> in the cell, directing our future study of its function. While no function of MCM9<sup>M</sup> has been established, these future studies may elucidate a function of MCM9<sup>M</sup> in the cytoplasm, completely distinct from the function of full length MCM9<sup>L</sup>. While MCM9<sup>M</sup> and MCM9<sup>L</sup> mRNA transcript abundances follow a similar trend throughout a variety of cell types and at particular stages during the progression of the cell cycle, DNA damage differentially controls abundance of MCM9<sup>M</sup> and MCM<sup>L</sup> transcripts. Our observation that MMC treatment of HeLa cells increases the abundance of MCM9<sup>L</sup> mRNA but has no effect on MCM9<sup>M</sup> mRNA suggests that the MCM9 splicing mechanism is controlled by the presence of DNA double-strand breaks, favoring an increase in abundance of the fully spliced isoform when DNA damage is present. Further study of the splicing mechanism under conditions of DNA damage is necessary to understand this differential splicing of the MCM9 gene.

### **6.1.6 Create genetically modified embryos for women afflicted by POF**

While hormone replacement therapy is currently used for symptom management in patients afflicted with POF and homozygous for MCM8 p. P149R or MCM9 c.1732+2T>C or p.R132\*, no therapy has been implemented to allow the affected individuals to produce viable offspring. By combining recently available techniques, there is a possibility for these women to have viable

offspring without passing on the pathogenic MCM8 or MCM9 mutations. For example, the MCM8 or MCM9 mutation could be corrected by newly developed CRISPR technologies in a cell line isolated from the patient, and used to generate induced pluripotent stem cells. This technique has already been implemented, as induced pluripotent stem cells have been generated from human fibroblasts (174). From these stem cells, an oocyte may be developed, using a similar approach to what has been successfully developed in mice (175). The generated oocyte may be fertilized *in vitro*, producing embryos whose genome will contain wild type MCM8 or MCM9 instead of the pathogenic mutant.

#### **6.1.7 Investigate further coincidence of POF and recombination defects**

The expected outcome of these efforts is a further understanding of the mechanism of HR, as well as the deleterious outcomes of defects in this mechanism. While there is a known relationship between premature ovarian failure and defects in the HR involvement of MCM8 and MCM9, we expect that a greater understanding of HR will have an impact beyond understanding this disease. Meiotic recombination is a fundamental process in all sexually reproducing organisms, and defects in this process have profound effects on the cell, including aneuploidy and mis-segregation of chromosomes. Defects in HR also lead to an inability to repair double-strand breaks, which may lead to unregulated cell division and cancerous tumor development. Due to the essential and foundational nature of HR, we expect that a further understanding of its mechanism will have a broad positive impact on the field of molecular biology.

## BIBLIOGRAPHY

1. Sivaprasad U, Machida YJ & Dutta A (2007). APC/C - the master controller of origin licensing? *Cell Div.* **2**, 8.
2. Machida YJ, Hamlin JL & Dutta A (2005). Right place, right time, and only once: Replication initiation in metazoans. *Cell* **123**, 13-24.
3. Ritzi M & Knippers R (2000). Initiation of genome replication: assembly and disassembly of replication-competent chromatin. *Gene* **245**, 13-20.
4. Nishitani H, Lygerou Z, Nishimoto T & Nurse P (2000). The Cdt1 protein is required to license DNA for replication in fission yeast. *Nature* **404**, 625-628.
5. Freeman A, Morris LS, Mills AD, Stoeber K, Laskey RA, et al. (1999). Minichromosome maintenance proteins as biological markers of dysplasia and malignancy. *Clin. Cancer Res.* **5**, 2121-2132.
6. Williams GH, Romanowski P, Morris L, Madine M, Mills AD, et al. (1998). Improved cervical smear assessment using antibodies against proteins that regulate DNA replication. *Proc. Natl. Acad. Sci. USA* **95**, 14932-14937.
7. Coleman N & Laskey RA (2009). Minichromosome maintenance proteins in cancer screening. *Eur. J. Cancer* **45**, Supplement **1**, 416-417.
8. Gonzalez MA, Tachibana KK, Chin S-F, Callagy G, Madine MA, et al. (2004). Geminin predicts adverse clinical outcome in breast cancer by reflecting cell-cycle progression. *J. Pathol.* **204**, 121-130.
9. Pardo B, Gómez-González B & Aguilera A (2009). DNA repair in mammalian cells. *Cell. Mol. Life Sci.* **66**, 1039-1056.
10. Douglas Max E & Diffley John FX (2012). Replication timing: The early bird catches the worm. *Curr. Biol.* **22**, R81-R82.
11. Diffley JFX (2011). Quality control in the initiation of eukaryotic DNA replication. *Philos. T. R. Soc. B.* **366**, 3545-3553.
12. Bielinsky A-K (2003). Replication origins: Why do we need so many? *Cell Cycle* **2**, 306-308.
13. Bochman ML & Schwacha A (2008). The MCM2-7 complex has *in vitro* helicase activity. *Mol. Cell* **31**, 287-293.
14. Remus D & Diffley JFX (2009). Eukaryotic DNA replication control: Lock and load, then fire. *Curr. Opin. Cell Biol.* **21**, 771-777.
15. Tsakraklides V & Bell SP (2010). Dynamics of pre-replicative complex assembly. *J. Biol. Chem.* **285**, 9437-9443.
16. Bruck I & Kaplan DL (2011). Origin single-stranded DNA releases Sld3 protein from the MCM2-7 complex, allowing the GINS tetramer to bind the MCM2-7 complex. *J. Biol. Chem.* **286**, 18602-18613.
17. Ilves I, Petojevic T, Pesavento JJ & Botchan MR (2010). Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. *Mol. Cell* **37**, 247-258.
18. MacNeill S (2010). Structure and function of the GINS complex, a key component of the eukaryotic replisome. *Biochem. J.* **425**, 489-500.

19. Labib K & Gambus A (2007). A key role for the GINS complex at DNA replication forks. *Trends Cell Biol.* **17**, 271-278.
20. Kamimura Y, Masumoto H, Sugino A & Araki H (1998). Sld2, which interacts with Dpb11 in *Saccharomyces cerevisiae*, is required for chromosomal DNA replication. *Mol. Cell Biol.* **18**, 6102-6109.
21. Takayama YK, Yoichiro; Okawa, Mariko; Muramatsu, Sachiko; Sugino, Akio; Araki, Hiroyuki (2003). GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast. *Genes Dev.* **17**, 1153-1165.
22. Davey MJ & O'Donnell M (2000). Mechanisms of DNA replication. *Curr. Opin. Chem. Biol.* **4**, 581-586.
23. Tsuyama TT, S.; Watanabe, S.; Seki, M.; Enomoto, T. (2005). Licensing for DNA replication requires a strict sequential assembly of Cdc6 and Cdt1 onto chromatin in *Xenopus* egg extracts. *Nucleic Acids Res.* **33**, 765-775.
24. DePamphilis ML (2003). The 'ORC cycle': a novel pathway for regulating eukaryotic DNA replication. *Gene* **310**, 1-15.
25. Bell SP & Dutta A (2002). DNA replication in eukaryotic cells. *Annu. Rev. Biochem.* **71**, 333-374.
26. Bowers JLR, J.C.W.; Chen, S.; Bell, S.P. (2004). ATP hydrolysis by ORC catalyzes reiterative MCM2-7 assembly at a defined origin of replication. *Mol. Cell* **16**, 967-978.
27. Speck CC, Z.; Li, H.; Stillman, B. (2005). ATPase-dependent, cooperative binding of ORC and Cdc6p to origin DNA. *Nat. Struct. Mol. Biol.* **12**, 965-971.
28. Lee DGB, S.P. (1997). Architecture of the yeast origin recognition complex bound to origins of DNA replication. *Mol. Cell Biol.* **17**, 7159-7168.
29. Neuwald AFA, L.; Spouge, J.L.; Koonin, E.V. (1999). AAA+: A class of chaperone-like ATPases associated with the assembly, operation and disassembly of protein complexes. *Genome Res.* **9**, 27-43.
30. Fujita MY, C.; Goto, H.; Yokoyama, N.; Kuzushima, K.; Inagaki, M. and Tsurumi, T. (1999). Cell cycle regulation of human CDC6 protein. *J. Biol. Chem.* **274**, 25927-25932.
31. Coleman TRC, P.B. (1996). The *Xenopus* Cdc6 protein is essential for the initiation of a single round of DNA replication in cell-free extracts. *Cell* **87**, 53-63.
32. Maiorano DL, Malik; Méchali, Marcel (2006). MCM proteins and DNA replication. *Curr. Opin. Cell Biol.* **18**, 130-136.
33. Hofmann JFB, D. (1994). *cdt1* is an essential target of the Cdc10/Sct1 transcription factor: requirement for DNA replication and inhibition of mitosis. *EMBO J.* **13**, 425-434.
34. Maiorano DM, J. Mechali, M. (2000). *XICDT1* is required for the assembly of pre-replicative complexes in *Xenopus laevis*. *Nature* **404**, 622-625.
35. Whittaker AJR, I.; Orr-Weaver, T.L. (2000). *Drosophila* double parked: a conserved, essential replication protein that colocalizes with the origin recognition complex and links DNA replication with mitosis and the down-regulation of S phase transcripts. *Genes Dev.* **14**, 1765-1776.
36. Yeeles JTP, Deegan TD, Janska A, Early A & Diffley JFX (2015). Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature* **519**, 431-435.
37. Chen SB, Stephen P. (2011). CDK prevents MCM2-7 helicase loading by inhibiting Cdt1 interaction with Orc6. *Genes Dev.* **25**, 363-372.
38. Cox LSE *Molecular Themes in DNA Replication*. (The Royal Society of Chemistry, Cambridge; 2009).

39. McGarry TJK, M.W. (1998). Geminin, an inhibitor of DNA replication, is degraded during mitosis. *Cell* **93**, 1043-1053.
40. Kulartz MK, R. (2004). The replicative regulator protein geminin on chromatin in the HeLa cell cycle. *J. Biol. Chem.* **279**, 41686-41694.
41. Maiorano D, Rul W & Méchali M (2004). Cell cycle regulation of the licensing activity of Cdt1 in *Xenopus laevis*. *Exp. Cell Res.* **295**, 138-149.
42. Lutzmann MM, D.; Mechali, M. (2006). A Cdt1-geminin complex licenses chromatin for DNA replication and prevents rereplication during S phase in *Xenopus*. *EMBO J.* **25**, 5764-5774.
43. De Marco V, Gillespie PJ, Li A, Karantzelis N, Christodoulou E, et al. (2009). Quaternary structure of the human Cdt1-geminin complex regulates DNA replication licensing. *P. Natl. Acad. Sci. USA* **106**, 19807-19812.
44. Rowles AT, S.; Blow, J.J. (1999). Changes in association of the *Xenopus* origin recognition complex with chromatin on licensing of replication origins. *J. Cell Sci.* **112**, 2011-2018.
45. Lutzmann M & Méchali M (2008). MCM9 binds Cdt1 and is required for the assembly of prereplication complexes. *Mol. Cell* **31**, 190-200.
46. Maine GTS, P.; Tye, B-K (1984). Mutants of *S. Cerivisiae* defective in the maintenance of minichromosomes. *Genetics* **106**, 365-385.
47. Iyer LML, D.D.; Koonin, E.V. & Aravind, L. (2004). Evolutionary relationships and structural mechanisms of AAA+ proteins. *J. Struct. Biol.* **146**, 11-31
48. Vetter IRW, A. (1999). Nucleoside triphosphate-binding proteins: different scaffolds to achieve phosphoryl transfer. *Q. Rev. Biophys.* **32**, 1-56.
49. Duderstadt KEB, J.M. (2008). AAA+ ATPases in the initiation of DNA replication. *Crit. Rev. Biochem. Mol. Biol.* **43**, 163-187.
50. Guenther BO, R.; Sali, A.; O'Donnell, M.; Kuriyan, J. (1997). Crystal structure of the  $\delta$  subunit of the clamp-loader complex of *E. coli* DNA polymerase III. *Cell* **91**, 335-345.
51. Walker J.E. S, M., Runswick, M.J., and Gay, N.J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1**, 945-951.
52. McGeoch ATT, M.A.; Laskey, R.A. and Bell, S.P. (2005). Organization of the archaeal MCM complex on DNA and implications for the helicase mechanism. *Nat. Struct. Mol. Biol.* **12**, 756-762.
53. Tye B (1999). MCM proteins in DNA replication. *Annu. Rev. Biochem.* **68**, 649-686.
54. Kearsey SL, K (1998). MCM proteins evolution, properties, and role in DNA replication. *Biochim. Biophys. Acta* **1398**, 113-136.
55. Forsburg SL (2004). Eukaryotic MCM proteins: Beyond replication initiation. *Microbiol. Mol. Biol. Rev.* **68**, 109-131.
56. Labib KD, JFX (2001). Is the MCM2-7 complex the eukaryotic DNA replication fork helicase? *Curr. Opin. Genet. Dev.* **10**, 64-70.
57. Takahashi TSW, D.B.; Walter, J.C. (2005). Pumps, paradoxes and ploughshares: mechanism of the MCM2-7 DNA helicase. *Trends Biochem. Sci.* **30**, 437-444.
58. Das M, Singh S, Pradhan S & Narayan G (2014). MCM paradox: Abundance of eukaryotic replicative helicases and genomic integrity. *Mol. Biol. Int.* **2014**, 574850.
59. Bochman ML & Schwacha A (2009). The MCM complex: Unwinding the mechanism of a replicative helicase. *Microbiol. Mol. Biol. R.* **73**, 652-683.

60. Brewster AS, Wang G, Yu X, Greenleaf WB, Carazo JM, et al. (2008). Crystal structure of a near-full-length archaeal MCM: Functional insights for an AAA+ hexameric helicase. *P. Natl. Acad. Sci. USA* **105**, 20191-20196.
61. Costa A, Pape T, van Heel M, Brick P, Patwardhan A, et al. (2006). Structural basis of the *Methanothermobacter thermoautotrophicus* MCM helicase activity. *Nucleic Acids Res.* **34**, 5829-5838.
62. Fletcher RJ, Bishop BE, Leon RP, Sclafani RA, Ogata CM, et al. (2003). The structure and function of MCM from archaeal *M. Thermoautotrophicum*. *Nat. Struct. Mol. Biol.* **10**, 160-167.
63. Barry ER, McGeoch AT, Kelman Z & Bell SD (2007). Archaeal MCM has separable processivity, substrate choice and helicase domains. *Nucleic Acids Res.* **35**, 988-998.
64. Graham BW, Schauer GD, Leuba SH & Trakselis MA (2011). Steric exclusion and wrapping of the excluded DNA strand occurs along discrete external binding paths during MCM helicase unwinding. *Nucleic Acids Res.* **39**, 6585-6595.
65. Patel SSaP, K.M. (2000). Structure and function of hexameric helicases. *Annu. Rev. Biochem.* **69**, 651-697.
66. Bochman ML & Schwacha A (2010). The *Saccharomyces cerevisiae* MCM6/2 and MCM5/3 ATPase active sites contribute to the function of the putative MCM2-7 'gate'. *Nucleic Acids Res.* **38**, 6078-6088.
67. Lutzmann M, Maiorano D & Méchali M (2005). Identification of full genes and proteins of MCM9, a novel, vertebrate-specific member of the MCM2–8 protein family. *Gene* **362**, 51-56.
68. Park JL, D.T.; Lee, K.Y.; Abbas, T.; Shibata, E.; Negishi, M.; Luo, Y.; Schimenti, J.C.; Gambus, A.; Walter, J.C.; Dutta, A. (2013). MCM8-9 complex promotes RAD51 recruitment at DNA damage sites to facilitate homologous recombination. *Mol. Cell Biol.*
69. Gozuacik DC, M.; Lagorce, D.; Faivre, J.; Murakami, Y.; Poch, O.; Biermann, E.; Knippers, R.; Brechot, C.; Paterlini-Brechot, P. (2003). Identification and functional characterization of a new member of the human MCM protein family: hMCM8. *Nucleic Acids Res.* **31**, 570-579.
70. Volkening MH, I. (2005). Involvement of human MCM8 in prereplication complex assembly by recruiting hcdc6 to chromatin. *Mol. Cell Biol.* **25**.
71. Maiorano D, Cuvier O, Danis E & Méchali M (2005). MCM8 Is an MCM2-7-related protein that functions as a DNA helicase during replication elongation and not initiation. *Cell* **120**, 315-328.
72. Blanton HL, Radford SJ, McMahan S, Kearney HM, Ibrahim JG, et al. (2005). REC, *Drosophila* MCM8, drives formation of meiotic crossovers. *PLoS Genet.* **1**, e40.
73. Yoshida K (2005). Identification of a novel cell-cycle-induced MCM family protein MCM9. *Biochem. Biophys. Res. Commun.* **331**, 669-674.
74. Lutzmann MG, C.; Traver, S.; Ganier, O.; Maya-Mendoza, A.; Ranisavljevic, N.; Bernex, F.; nishiyama, A.; Montel, N.; Gavois, E.; Forichon, L.; de Massy, B.; Mechali, M. (2012). MCM8- and MCM9-deficient mice reveal gametogenesis defects and genome instability due to impaired homologous recombination. *Mol. Cell* **47**, 1-12.
75. Hartford SAL, Y.; Southard, T.L.; Min, I.M.; Lis, J.T.; Schimenti, J.C. (2011). Minichromosome maintenance helicase paralog MCM9 is dispensible for DNA replication but functions in germ-line stem cells and tumor supression. *Proc. Natl. Acad. Sci. USA*.

76. Nishimura KI, M.; Horikawa, K.; Fukagawa, T.; Takata, M.; Takisawa, H.; Kanemaki, M. (2012). Mcm8 and Mcm9 form a complex that functions in homologous recombination repair induced by DNA interstrand crosslinks. *Mol. Cell* **47**, 1-12.
77. Dehal PS, Y.; Campbell, R.K. (2002). The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* **298**, 2157-2167.
78. Liu Y, Richards TA & Aves SJ (2009). Ancient diversification of eukaryotic MCM DNA replication proteins. *BMC Evol. Biol.* **9**, 60.
79. Iyer LMA, L. The evolutionary history of proteins involved in pre-replication complex assembly DNA Replication and Human Disease. (ed. M.L. DePamphilis) 751-757 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY; 2006).
80. Aves SJ, Liu Y & Richards TA Evolutionary diversification of eukaryotic DNA replication machinery, The Eukaryotic Replisome: a Guide to Protein Structure and Function, Vol. 62. (ed. S. MacNeill) 19-35 (Springer Netherlands, 2012).
81. Carty CL, Spencer KL, Setiawan VW, Fernandez-Rhodes L, Malinowski J, et al. (2013). Replication of genetic loci for ages at menarche and menopause in the multi-ethnic Population Architecture using Genomics and Epidemiology (PAGE) study. *Hum. Reprod.* **28**, 1695-1706.
82. Tenenbaum-Rakover Y, Weinberg-Shukron A, Renbaum P, Lobel O, Eideh H, et al. (2015). Minichromosome maintenance complex component 8 (MCM8) gene mutations result in primary gonadal failure. *J. Med. Genet.*
83. Chen CTL, Fernández-Rhodes L, Brzyski RG, Carlson CS, Chen Z, et al. (2012). Replication of loci influencing ages at menarche and menopause in Hispanic women: the Women's Health Initiative SHARe Study. *Hum. Molec. Genet.* **21**, 1419-1432.
84. He C, Kraft P, Chasman D, Buring J, Chen C, et al. (2010). A large-scale candidate gene association study of age at menarche and age at natural menopause. *Hum. Genet.* **128**, 515-527.
85. Schuh-Huerta S, Johnson N, Rosen M, Sternfeld B, Cedars M, et al. (2012). Genetic markers of ovarian follicle number and menopause in women of multiple ethnicities. *Hum. Genet.* **131**, 1709-1724.
86. Spencer KL, Malinowski J, Carty CL, Franceschini N, Fernández-Rhodes L, et al. (2013). Genetic variation and reproductive timing: African American women from the population architecture using genomics and epidemiology (PAGE) study. *PLoS One* **8**, e55258.
87. Stolk L, Zhai G, van Meurs JBJ, Verbiest MMPJ, Visser JA, et al. (2009). Loci at chromosomes 13, 19 and 20 influence age at natural menopause. *Nat. Genet.* **41**, 645-647.
88. Rich T, Allen RL & Wyllie AH (2000). Defying death after DNA damage. *Nature* **407**, 777-783.
89. Srivastava M & Raghavan Sathees C (2015). DNA double-strand break repair inhibitors as cancer therapeutics. *Chem. Biol.* **22**, 17-29.
90. Hopfner K-P, Putnam CD & Tainer JA (2002). DNA double-strand break repair from head to tail. *Curr. Opin. Struct. Biol.* **12**, 115-122.
91. San Filippo J, Sung P & Klein H (2008). Mechanism of eukaryotic homologous recombination. *Annu. Rev. Biochem.* **77**, 229-257.
92. Boulton SJ & Jackson SP (1996). *Saccharomyces cerevisiae* Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways. *EMBO J.* **15**, 5093-5103.

93. Daley JM & Wilson TE (2005). Rejoining of DNA double-strand breaks as a function of overhang length. *Mol. Cell Biol.* **25**, 896-906.
94. Tomasz M (1995). Mitomycin C: small, fast and deadly (but very selective). *Chem. Biol.* **2**, 575-579.
95. Iyer VN & Szybalski W (1963). A molecular mechanism of mitomycin action: Linking of complementary DNA strands *Proc. Natl. Acad. Sci. USA* **50**, 355-362.
96. Kuzminov A (2001). Single-strand interruptions in replicating chromosomes cause double-strand breaks. *Proc. Natl. Acad. Sci. USA* **98**, 8241-8246.
97. Udayakumar D, Horikoshi N, Mishra L, Hunt C & Pandita T (2015). Detecting ATM-dependent chromatin modification in DNA damage response. *Methods Mol. Biol.* **1288**, 317-336.
98. Liu Y-Q, Li W-Q, Morris-Natschke SL, Qian K, Yang L, et al. (2015). Perspectives on biologically active camptothecin derivatives. *Med. Res. Rev.*, n/a-n/a.
99. Heyer W-D, Ehmsen KT & Liu J (2010). Regulation of homologous recombination in eukaryotes. *Annu. Rev. Genet.* **44**, 113-139.
100. Hopfner K-P, Craig L, Moncalian G, Zinkel RA, Usui T, et al. (2002). The Rad50 zinc-hook is a structure joining Mre11 complexes in DNA recombination and repair. *Nature* **418**, 562-566.
101. Imamura O & Campbell JL (2003). The human Bloom syndrome gene suppresses the DNA replication and repair defects of yeast *dna2* mutants. *Proc. Natl. Acad. Sci. USA* **100**, 8193-8198.
102. Wong AKC, Pero R, Ormonde PA, Tavtigian SV & Bartel PL (1997). RAD51 interacts with the evolutionarily conserved BRC motifs in the human breast cancer susceptibility gene BRCA2. *J. Biol. Chem.* **272**, 31941-31944.
103. Bignell G, Micklem G, Stratton MR, Ashworth A & Wooster R (1997). The BRC repeats are conserved in mammalian BRCA2 proteins. *Hum. Mol. Genet.* **6**, 53-58.
104. Islam MN, Paquet N, Fox D, Dray E, Zheng X-F, et al. (2012). A variant of the breast cancer type 2 susceptibility protein (BRC) repeat is essential for the RECQL5 helicase to interact with RAD51 recombinase for genome stabilization. *J. Biol. Chem.* **287**, 23808-23818.
105. Galkin VE, Esashi F, Yu X, Yang S, West SC, et al. (2005). BRCA2 BRC motifs bind RAD51-DNA filaments. **102**, 8537-8542.
106. Petalcorin MIR, Galkin VE, Yu X, Egelman EH & Boulton SJ (2007). Stabilization of RAD-51-DNA filaments via an interaction domain in *Caenorhabditis elegans* BRCA2. *Proc. Natl. Acad. Sci. USA* **104**, 8299-8304.
107. Ashworth A (2008). A synthetic lethal therapeutic approach: Poly(ADP) ribose polymerase inhibitors for the treatment of cancers deficient in DNA double-strand break repair. *J. Clin. Oncol.* **26**, 3785-3790.
108. Lindahl T, Satoh MS, Poirier GG & Klungland A (1995). Post-translational modification of poly(ADP-ribose) polymerase induced by DNA strand breaks. *Trends Biochem. Sci.* **20**, 405-411.
109. D'Amours D, Desnoyers S, D'Silva I & Poirier GG (1999). Poly(ADP-ribosylation) reactions in the regulation of nuclear functions. *Biochem. J.* **342**, 249-268.
110. El-Khamisy SF, Masutani M, Suzuki H & Caldecott KW (2003). A requirement for PARP-1 for the assembly or stability of XRCC1 nuclear foci at sites of oxidative DNA damage. *Nucleic Acids Res.* **31**, 5526-5533.

111. Dantzer F, Amé JC, Schreiber V, Nakamura J, Ménissier- de Murcia J, et al. (2006). Poly(ADP- ribose) polymerase–1 activation during DNA damage and repair. *Methods Enzymol.* **409**, 493-510.
112. Peralta-Leal A, Rodríguez-Vargas JM, Aguilar-Quesada R, Rodríguez MI, Linares JL, et al. (2009). PARP inhibitors: New partners in the therapy of cancer and inflammatory diseases. *Free Radic. Biol. Med.* **47**, 13-26.
113. Tallis M, Morra R, Barkauskaite E & Ahel I (2014). Poly(ADP-ribosyl)ation in regulation of chromatin structure and the DNA damage response. *Chromosoma* **123**, 79-90.
114. Graziani G & Szabó C (2005). Clinical perspectives of PARP inhibitors. *Pharmacol. Res.* **52**, 109-118.
115. Bryant HE, Schultz N, Thomas HD, Parker KM, Flower D, et al. (2005). Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **434**, 913-917.
116. Chatterjee P, Choudhary GS, Sharma A, Singh K, Heston WD, et al. (2013). PARP inhibition sensitizes to low dose-rate radiation TMPRSS2-ERG fusion gene-expressing and PTEN-deficient prostate cancer cells. *PLoS One* **8**, e60408.
117. Mendes-Pereira AM, Martin SA, Brough R, McCarthy A, Taylor JR, et al. (2009). Synthetic lethal targeting of PTEN mutant cells with PARP inhibitors. *EMBO Mol. Med.* **1**, 315-322.
118. Liu JF, Konstantinopoulos PA & Matulonis UA (2014). PARP inhibitors in ovarian cancer: Current status and future promise. *Gynecol. Oncol.* **133**, 362-369.
119. Liu X, Shi Y, Maag DX, Palma JP, Patterson MJ, et al. (2012). Iniparib nonselectively modifies cysteine-containing proteins in tumor cells and is not a *bona fide* PARP inhibitor. *Clin. Cancer Res.* **18**, 510-523.
120. Neale MJ & Keeney S (2006). Clarifying the mechanics of DNA strand exchange in meiotic recombination. *Nature* **442**, 153-158.
121. Andersen SL & Sekelsky J (2010). Meiotic versus mitotic recombination: Two different routes for double-strand break repair: The different functions of meiotic versus mitotic DSB repair are reflected in different pathway usage and different outcomes. *BioEssays* **32**, 1058-1066.
122. Seeliger K, Dukowic-Schulze S, Wurz-Wildersinn R, Pacher M & Puchta H (2012). BRCA2 is a mediator of RAD51- and DMC1-facilitated homologous recombination in *Arabidopsis thaliana*. *New Phytol.* **193**, 364-375.
123. Crismani W, Portemer V, Froger N, Chelysheva L, Horlow C, et al. (2013). MCM8 is required for a pathway of meiotic double-strand break repair independent of DMC1 in *Arabidopsis thaliana*. *PLoS Genet.* **9**, e1003165.
124. Sharp PA (1994). Split genes and RNA splicing. *Cell* **77**, 805-815.
125. Lee Y & Rio DC (2015). Mechanisms and regulation of alternative pre-mRNA splicing. *Annu. Rev. Biochem.* **84**, 31.31-31.33.
126. Nilsen TWG, B.R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463**, 457-463.
127. Black DL (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **72**, 291-336.
128. Pan Q, Shai O, Lee LJ, Frey BJ & Blencowe BJ (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**, 1413-1415.

129. Matlin AJ, Clark F & Smith CWJ (2005). Understanding alternative splicing: Towards a cellular code. *Nat. Rev. Mol. Cell Biol.* **6**, 386-398.
130. Nelson LM (2009). Primary ovarian insufficiency. *N. Engl. J. Med.* **360**, 606-614.
131. Wood MA & Rajkovic A (2013). Genomic markers of ovarian reserve. *Semin. Reprod. Med.* **31**, 399-415.
132. Zhu W, Chen Y & Dutta A (2004). Rereplication by depletion of geminin is seen regardless of p53 status and activates a G2/M checkpoint. *Mol. Cell Biol.* **24**, 7140 - 7150.
133. Barber RDH, D.W.; Coleman, R.A.; Clark, B.J. (2005). GAPDH as a housekeeping gene: analysis of GAPDH mRNA expression in a panel of 72 human tissues. *Physiol. Genomics* **21**, 389-395.
134. Kapushesky MI, E.; Holloway, E.; Kurnosov, P.; Zorin, A.; Malone, J.; Rustici, G.; Williams, E.; Parkinson, H. and Brazma, A. (2010). Gene expression atlas at the European bioinformatics institute. *Nucleic Acids Res.* **38**, D690-D698.
135. Gong J, Traganos F & Darzynkiewicz Z (1995). Growth imbalance and altered expression of cyclins B1, A, E, and D3 in MOLT-4 cells synchronized in the cell cycle by inhibitors of DNA replication. *Cell Growth Differ.* **6**, 1485-1493.
136. Tsang YHH, Xianxian; Man, Wing Yu; Lee, Nelson; Poon, Randy Y.C. (2012). Novel functions of the phosphatase SHP2 in the DNA replication and damage checkpoints. *PLoS One* **7**, 1.
137. Petermann EO, M. L.; Issaeva, N.; Schultz, N.; 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.
138. Rahman FA, Ainscough JFX, Copeland N & Coverley D (2007). Cancer-associated missplicing of exon 4 influences the subnuclear distribution of the DNA replication factor CIZ1. *Hum. Mutat.* **28**, 993-1004.
139. Ast G (2004). How did alternative splicing evolve? *Nat. Rev. Genet.* **5**, 773-782.
140. Garcia-Blanco MAB, A.P.; Lasda, E.L. (2004). Alternative splicing in disease and therapy. *Nat. Biotechnol.* **22**, 535-546.
141. Zhang CH, M.L.; Krainer, A.R.; Zhang, M.Q. (2007). Dual-specificity splice sites function alternatively as 5' and 3' splice sites. *Proc. Natl. Acad. Sci. USA* **104**, 15028-15033.
142. Zhang MQ (1998). Statistical features of human exons and their flanking regions. *Hum. Molec. Genet.* **7**, 919-932.
143. Zimin A, Delcher A, Florea L, Kelley D, Schatz M, et al. (2009). A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* **10**, R42.
144. Shultz RW, Tatineni VM, Hanley-Bowdoin L & Thompson WF (2007). Genome-wide analysis of the core DNA replication machinery in the higher plants *Arabidopsis* and rice. *Plant Physiol.* **144**, 1697-1714.
145. Vijayraghavan S & Schwacha A The eukaryotic MCM2-7 replicative helicase, The Eukaryotic Replisome: a Guide to Protein Structure and Function, Vol. 62. (ed. S. MacNeill) 113-134 (Springer Netherlands, 2012).
146. Holm K, Mosfeldt Laursen E, Brocks V & Müller J (1995). Pubertal maturation of the internal genitalia: An ultrasound evaluation of 166 healthy girls. *Ultrasound Obst. Gyn.* **6**, 175-181.
147. Pavlik EJ, DePriest PD, Gallion HH, Ueland FR, Reedy MB, et al. (2000). Ovarian volume related to age. *Gynecol. Oncol.* **77**, 410-412.

148. Merz E, Miric-Tesanic D, Bahlmann F, Weber G & Wellek S (1996). Sonographic size of uterus and ovaries in pre- and postmenopausal women. *Ultrasound Obstet. Gynecol.* **7**, 38-42.
149. Jeffries EP, Denq WI, Bartko JC & Trakselis MA (2013). Identification, quantification, and evolutionary analysis of a novel isoform of MCM9. *Gene* **519**, 41-49.
150. Oostra AB, Nieuwint AWM, Joenje H & de Winter JP (2012). Diagnosis of fanconi anemia: chromosomal breakage analysis. *Anemia* **2012**, 238731.
151. Kinoshita Y, Johnson EM, Gordon RE, Negri-Bell H, Evans MT, et al. (2008). Colocalization of MCM8 and MCM7 with proteins involved in distinct aspects of DNA replication. *Microsc. Res. Tech.* **71**, 288-297.
152. Hardy CFJ, Dryga O, Seematter S, Pahl PMB & Sclafani RA (1997). *mcm5/cdc46-bob1* bypasses the requirement for the S phase activator Cdc7p. *Proc. Natl. Acad. Sci. USA* **94**, 3151-3155.
153. Fletcher RJ & Chen XS (2006). Biochemical activities of the *BOB1* mutant in *Methanobacterium thermoautotrophicum* MCM. *Biochemistry* **45**, 462-467.
154. Arnold K, Bordoli L, Kopp J & Schwede T (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195-201.
155. AlAsiri S, Basit S, Wood-Trageser MA, Yatsenko SA, Jeffries EP, et al. (2015). Exome sequencing reveals MCM8 mutation underlies ovarian failure and chromosomal instability. *J. Clin. Invest.* **125**, 258-262.
156. de Vries SS, Baart EB, Dekker M, Siezen A, de Rooij DG, et al. (1999). Mouse MutS-like protein Msh5 is required for proper chromosome synapsis in male and female meiosis. *Genes Dev.* **13**, 523-531.
157. Yoshida K, Kondoh G, Matsuda Y, Habu T, Nishimune Y, et al. (1998). The mouse RecA-like gene *Dmcl* Is required for homologous chromosome synapsis during meiosis. *Mol. Cell* **1**, 707-718.
158. Mandon-Pépin B, Touraine P, Kuttann F, Derbois C, Rouxel A, et al. (2008). Genetic investigation of four meiotic genes in women with premature ovarian failure. *Eur. J. Endocrinol.* **158**, 107-115.
159. Wong JCY, Alon N, Mckerlie C, Huang JR, Meyn MS, et al. (2003). Targeted disruption of exons 1 to 6 of the fanconi anemia group A gene leads to growth retardation, strain-specific microphthalmia, meiotic defects and primordial germ cell hypoplasia. *Hum. Molec. Genet.* **12**, 2063-2076.
160. Giri N, Batista DL, Alter BP & Stratakis CA (2007). Endocrine abnormalities in patients with Fanconi anemia. *J. Clin. Endocr. Metab.* **92**, 2624-2631.
161. Li H & Durbin R (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589-595.
162. Li H & Durbin R (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754-1760.
163. Kanemaki MT (2013). The dimeric MCM8–9 complex of *Xenopus laevis* likely has a conserved function for resistance to DNA damage. *Cell Cycle* **12**, 1338-1339.
164. Wood-Trageser Michelle A, Gurbuz F, Yatsenko Svetlana A, Jeffries Elizabeth P, Kotan LD, et al. (2014). MCM9 mutations are associated with ovarian failure, short stature, and chromosomal instability. *Am. J. Hum. Genet.* **95**, 754-762.
165. Görlich D & Mattaj IW (1996). Nucleocytoplasmic transport. *Science* **271**, 1513-1518.

166. Kalderon D, Roberts BL, Richardson WD & Smith AE (1984). A short amino acid sequence able to specify nuclear location. *Cell* **39**, 499-509.
167. Makkerh JPS, Dingwall C & Laskey RA (1996). Comparative mutagenesis of nuclear localization signals reveals the importance of neutral and acidic amino acids. *Curr. Biol.* **6**, 1025-1027.
168. Schultz N, Lopez E, Saleh-Gohari N & Helleday T (2003). Poly(ADP-ribose) polymerase (PARP-1) has a controlling role in homologous recombination. *Nucleic Acids Res.* **31**, 4959-4964.
169. Shen Y, Rehman FL, Feng Y, Boshuizen J, Bajrami I, et al. (2013). BMN 673, a novel and highly potent PARP1/2 inhibitor for the treatment of human cancers with DNA repair deficiency. *Clin. Cancer Res.* **19**, 5003-5015.
170. Shen Y, Aoyagi-Scharber M & Wang B (2015). Trapping poly(ADP-ribose) polymerase. *J. Pharmacol. Exp. Ther.* **353**, 446-457.
171. Mariano G, Ricciardi MR, Trisciuglio D, Zampieri M, Ciccarone F, et al. (2015). PARP inhibitor ABT-888 affects response of MDA-MB-231 cells to doxorubicin treatment, targeting Snail expression. *Oncotarget*.
172. Coleman RL, Sill MW, Bell-McGuinn K, Aghajanian C, Gray HJ, et al. (2015). A phase II evaluation of the potent, highly selective PARP inhibitor veliparib in the treatment of persistent or recurrent epithelial ovarian, fallopian tube, or primary peritoneal cancer in patients who carry a germline BRCA1 or BRCA2 mutation — An NRG Oncology/Gynecologic Oncology Group study. *Gynecol. Oncol.*
173. Franken NAP, Rodermond HM, Stap J, Haveman J & van Bree C (2006). Clonogenic assay of cells *in vitro*. *Nat. Protoc.* **1**, 2315-2319.
174. Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917-1920.
175. Hayashi K, Ogushi S, Kurimoto K, Shimamoto S, Ohta H, et al. (2012). Offspring from oocytes derived from *in vitro* primordial germ cell-like cells in mice. *Science* **338**, 971-975.