Supplemental Appendix

Exome sequencing identifies *MCM8* mutation in ovarian failure and chromosomal instability

Saleh AlAsiri, Sulman Basit, Michelle A. Wood, Svetlana A. Yatsenko, Elizabeth P. Jeffries, Urvashi Surti, Deborah M. Ketterer, Sibtain Afzal, Khushnooda Ramzan, Muhammad Faiyaz-Ul Haque, Huaiyang Jiang, Michael A. Trakselis, and Aleksandar Rajkovic

Supplemental Methods

Patient Recruitment. A consanguineous family was recruited from King Khalid University Hospital (KKUH), Riyadh, Saudi Arabia (Figure 1A). The affected daughters were born to healthy parents (III-1 and III-2) of Saudi Arabian descent with no previous significant family history of primary amenorrhea or endocrine dysfunction. The parents are first cousins. There are a total of 12 children (7 females and 5 males). All affected individuals underwent detailed clinical investigation. Three affected (IV-1, IV-6, and IV-9) and six unaffected family members (III-1, III-2, IV-2, IV-3, IV-5, and IV-8) of the family provided peripheral blood samples. Five family members (IV-1, IV-6, IV-9, III-2, and IV-3) provided skin fibroblast samples. The study was approved by the Ethical Review Committee of KKUH. Informed written consent was obtained from all participating subjects.

Fertile Control Population. Women were recruited at Magee-Womens Hospital. Samples were obtained and de-identified at the time of recruitment. All control de-identified samples were from Caucasian women who have had at least one live birth or were currently pregnant with their first child at time of consent. The study was approved by the Institutional Review Board of the University of Pittsburgh.

DNA Extraction. Genomic DNA was extracted from peripheral blood samples using NucleoSpin® blood genomic DNA extraction kit (MACHEREY-NAGEL). In a 1.5 ml microcentrifuge tube (Eppendorf AG), 25 µl proteinase K, 200 µl blood and

200 µl buffer B3 (NucleoSpin®, MACHEREY-NAGEL) were added. The mixture was vortexed thoroughly for 10-20 seconds and incubated at 70°C for 10-15 minutes followed by addition of 210 µl of ethanol (96-100%) to lysate. The entire lysate was transferred to a NucleoSpin® blood column placed in a collection tube and centrifuged at 11,000 rpm for 1 minute. The collection tube with flow through was discarded and the column was placed into a new collection tube. Buffer BW (500 µl; NucleoSpin®, MACHEREY-NAGEL) was added to the column and centrifuged at 11,000 rpm for 1 minute. Again, the collection tube with flow through was discarded and the column was placed into a new collection tube. Buffer B5 (600 µl) was added to the column and centrifuged at 11,000 rpm for 1 minute. The flow-through was discarded and column was dried and placed into a new 1.5 ml microcentrifuge tube. Preheated (70°C) buffer BE (100 µl) was added to the column and incubated at room temperature for 1 minute followed by centrifugation at 11,000 rpm for 1 minute to elute the DNA. Prior to DNA extraction, the blood collected in the EDTA tubes was equilibrated to room temperature.

SNP Arrays and Homozygosity Mapping. DNA of eight individuals (III-1, III-2, IV-1, IV-2, IV-3, IV-5, IV-6, and IV-9) was used for homozygosity mapping. Genotyping was performed using the Affymetrix GeneChip Human Mapping 250K Nsp array (Affymetrix). Data was deposited in the Gene Expression Omnibus (Accession Number GSE56043). Briefly, 250 ng of genomic DNA was digested with Digestion Master Mix containing 2 µl NE buffer 2 (10X), 0.5 µl BSA

(100X; 10 mg/ml) and 1 µl Nsp1. Digested DNA sample was ligated to Nsp1 adaptor using T4 DNA ligase and amplified by 2 µl of TITANIUM Taq DNA polymerase (50X) and 100 µM PCR primer. PCR products were purified on a Clean-Up plate (Clontech Lab) and eluted by RB buffer. Purified PCR products were fragmented using Fragmentation Reagent (0.05U/µl DNase 1) for 35 minutes at 37°C followed by labeling of fragmented samples with Labeling Master Mix (30 mM GeneChip® DNA Labeling Reagent, 30 U/µl Terminal Deoxynucleotidyl Transferase) for 4 hours at 37°C. Labeled samples were hybridized to GeneChip® Human Mapping 250K Nsp Array by mixing the sample with Hybridization Master Mix, denatured on thermoblock and loaded onto the Array. Array was then placed in a hybridization oven (GeneChip® Hybridization Oven 640) for 16-18 hours. After hybridization, array was washed and stained on an automated Fluidic Station 450 followed by scanning on GeneChip® Scanner 3000 7G using GeneChip® Operating Software (GCOS). The genotype of each SNP was determined by the BRLMM algorithm incorporated in Affymetrix Genotyping Console. A call rate (percentage of SNPs genotyped by sample) of 98% was obtained across the entire sample. Mapping order, and physical and genetic distances of SNPs were obtained from Affymetrix. Analysis of SNP data was conducted using HomozygosityMapper (http://www.homozygositymapper.org). Parametric two point and multipoint linkage analysis was carried out with the online version of Superlink software. LOD scores were calculated using a fully penetrant autosomal recessive model

with a disease allele frequency of 0.01. Regions of homozygosity with LOD scores are listed in Supplemental Table 2.

Library Construction and Exome Sequencing. Genomic DNA samples (III-1, III-2, IV-1, IV-2, IV-3, IV-5, IV-6, and IV-9) were subjected to in-solution exome enrichment via the HaloPlex Exome Kit (Agilent). Following exome capture, the samples were submitted for 2 x 100 bp paired end high-throughput sequencing on a HiSeq 2500 (Illumina, Inc.). Samples were run two samples per flow cell lane. An average of 14 GB of compressed data and an average of 153,937,273 reads were generated for each sample. Quality metrics for sequencing are shown in Supplemental Table 6. Raw data were deposited in the Sequencing Reads Archive (NCBI; ID number SRP046742).

Data Alignment and Variant Calling: 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 GRCh37/hg 19 using BWA version 0.7.3a MEM (Maximum Exactly Match)[9, 10]. Alignment statistics are presented in Supplemental Tables 6 and 7. Local realignment around insertions and deletions, reads base quality recalibration, and variant calling was conducted using GATK Tool Kits version 2.6-5. GATK HaplotypeCaller (HC) was used to call variants. The GATK Variant Recalibrator was used to count the VQSLOD (log odds ratio of variant quality score) based on variants present in the HapMap (v3.3), 1000 Genomes (Omni2.5 and Gold Standard Indels b37), and dbSNP137

databases. Variants were filtered for quality using the following parameters (Supplemental Table 7): (1) mapping quality filter equal to PASS; (2) Quality Depth (QD) >2; (3) VQSLOD>0; (4) Mapping Quality (MQ) > 40; (5) Fisher test of strand bias (FS) < 60; (6) HaplotypeScore < 13; (7) MQRankSum > -12.5; and (8) ReadPosRankSum > -8. Variants were further filtered if coverage < 10, if the alternative allele was not called in < 4 reads, if Genotype Quality (GQ) < 5, if the Polymorphism Likelihood (PL) value >30. Variants were then filtered according to the Mendelian rule for homozygous recessive inheritance: (1) affected daughters are homozygous alternative genotype, (2) the parents are heterozygous, and (3) the unaffected siblings do not have a homozygous alternative genotype. Remaining variants were annotated using ANNOVAR (Last Change Date: 2013-02-11) and focus was placed on variants causing non-synonymous variants (in exons or splice sites) with an allele frequency <5% as reported available databases (1000 Genomes or the ESP6500 from NHLBI)[11]. Variants were removed if they appeared on the NIH list of highly polymorphic and frequently mutated gene lists[12, 13]. Two variants remained.

Homology Modeling. 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[14]) using SWISS-MODEL[15]. Results are shown in Figure 1C and Supplemental Figure 3.

Sanger Sequencing. Validation of the *MCM8* c.446C>G variant was completed via PCR amplification of the region surrounding the variant using KAPA HiFi Taq Polymerase according to manufacturer's instructions. Primers were designed to amplify genomic region chr20:5,935,793-5,936,051, NCBI37/hg19 (F: 5'-CTG ACA GAA GGT GGT GAA GTA ACT A; R: 5'-CTC AGT TCT AGC CAA CAT CTT TTC G). The size of PCR products were confirmed on a 1.5% agarose gel. PCR products were then sent to Beckman Coulter Genomics for sequencing. Results were analyzed using Sequencher (Gene Codes Corporation). Samples for all family members and 200 fertile controls were analyzed. Sample chromatograms are shown in Figure 1B.

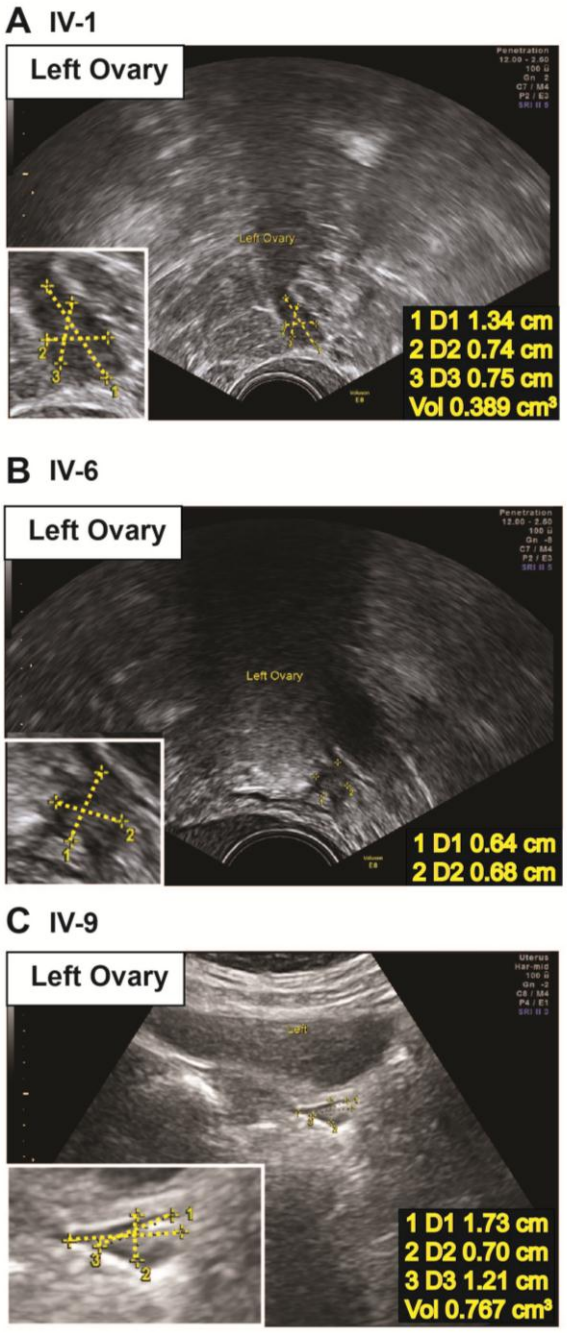
Chromosomal Instability Studies. Peripheral lymphocytes and skin fibroblasts samples were cultured in the presence of mitomycin C (MMC) or diepoxybutane (DEB) as described previously[16]. Briefly, cells were cultured in the presence of MMC at concentration 0 nM, 50 nM, 150 nM, or 300 nM or DEB. Cells were then harvested after 72 hours of incubation at 37°C. Cells were either dropped onto microscope slides (peripheral lymphocytes) or cultured on coverslips (patient fibroblasts). At least ten metaphase spreads per sample were evaluated under microscope slide for aberrations. Up to 60 aberrations per cell were counted. Results are reported in Figure 2D. ANOVA for Single Factor variation was utilized to determine both the effect of drug concentration within each cell line and the effect of cell line within each drug concentration. Two-tailed T-tests assuming

unequal variance were used to compare two cell lines at a single drug concentration.

MCM8 Foci Formation Assay. 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_2 incubator at 37°C. Human wild-type MCM8 was cloned into pEGFPC2 (Clontech) in frame with the N-terminal GFP tag using *BglII* and *BamHI*. pEGFPC2-MCM8 P149R was generated by Quikchange and KAPA HiFi polymerase introducing an *NruI* 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 or Olyumpus Provis AX70 confocal microscope. GFP foci were counted in 20 cells per condition. Statistical significance was evaluated by a two-tailed T-test (assuming unequal variance).

MCM8 DNA Binding Assays. 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^{-1} cm^{-1}. 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

194 sites of damage and MCM8 is likely to bind at these sites (1). EMSAs were
195 performed in a 15 µl reaction volume containing 20 mM sodium phosphate buffer
196 (pH 7.5), 100 mM NaCl, 20 mM β-mercaptoethanol and 5 nM DNA probe labeled
197 at the 5′-end using a standard polynucleotide kinase reaction and ^{32}P-γ-ATP, and
198 the indicated amount of MCM8. Binding reactions were allowed to equilibrate for
199 5 min followed by directly loading onto a 6% polyacrylamide/TBE gel. Gels were
200 run for 20 minutes at 13 volts cm^{-1} followed by imaging using a Storm
201 phosphorimager (GE Healthscience). Quantification of the fraction of band shift
202 was performed using the ImageQuant software (v5.0). Data were fit using
203 Kaleidagraph (Synergy) to a single site binding model defined by ΔF[MCM8]/K_d +
204 [MCM8] where F is the fraction bound and K_d is the dissociation constant.
205



Supplemental Figure 1. Representative Ultrasounds from Affected Daughters. Ultrasound images from each of the affected daughters (IV-1 – A; IV-6 – B; IV-9 – C) reveal small uteri and atrophied ovaries. (A) As visualized by transabdominal ultrasound, IV-1 had a left ovarian volume of 4.2 cm^3 but the right ovary could not be visualized. (B) IV-6 had a left ovarian volume of 2.7 cm^3 but the right ovary could also not be visualized via transvaginal ultrasound. (C) As visualized by transvaginal ultrasound, IV-9 reveals a pinpoint structures seen in both adnexae, right side with a volume of 0.7 cm^3 and left with volume of 0.4 cm^3, probably the ovaries. Insets show larger views of ovary. Only one ovary is shown in each image. Yellow dashed lines indicate length measured.

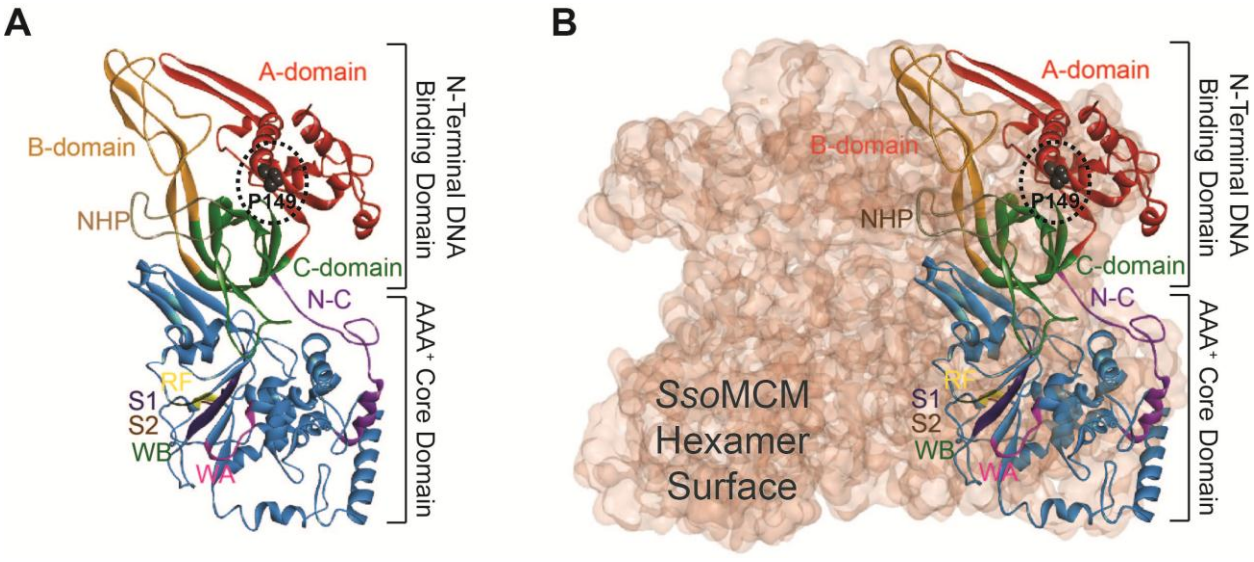
Supplemental Figure 2. Evolutionary comparison of the selected MCM8 region. CLUSTAL Omega web site (http://www.ebi.ac.uk/Tools/msa/) was used to align the sequences from MCM8 proteins in human (*Homo*, NP_001268449.1), mouse (*Mus*, NP_079952.2), frog (*Xenopus*, NP_001089437.1) and the predicted zebrafish sequence (*Danio*, XP_002665161.2). p.P149R is found at the evolutionary conserved Proline (highlighted in yellow).

```
Homo        TSEQTPQ-FLLSTKTPQSMQSTLDRFIPYKGWKLYFSEVYSDSSPLIEKIQAFEKFFTRH   106
Mus         ------------------------------------ASVYSNNSPFIEKIQAFEKFFTRH   71
Xenopus     NNGRDP--VCFAPPKPQLTQTTLDKYIPYKGWKLYFSEAYSDNSPFLEKVRAFEKFFKKQ   99
Danio       SNNNTQKGNAPQVSQARVTQATLDITCPYKGWRLYFSEGFVESSPYVEKIKVFEQYFTSQ   118
                                              :. : :.** :**::.**::*. :

Homo        IDLYDKDEIERKGSILVDFKELTEGGEVTNLIPDIATELRDAPEKTLACMGLAIHQVLTK   166
Mus         IDLYDKDEIERKGSILVDFKELTKADEITNLIPDIENALRDAPEKTLACMGLAIHQVLTK   131
Xenopus     IELYDKDEIERKGSILVDYKELLQDEDLSAAIP-LSSELKDMPEKVLECMGLAIHQVLTK   158
Danio       IDLYDKDEIERKGSILVDYKDLLSNKQVSHSLPDLARDLKEMPEKILDCLGVAIHQVLTL   178
            *:*****************:*:* .  :::  :* :   *:: *** * *:*:*******

Homo        DLERHAAELQAQEGLSNDGETMVNVPHIHARVYNYEPLTQLKNVRANYYGKYIALRGTVV   226
Mus         DLERHAAELQAQEGLSNGGETMVNVPHIYARVYNYEPLTHLKNIRATCYGKYISIRGTVV   191
Xenopus     DLETHAADLQQQEGLRTEEAPIVNVPFIHARVFNYDTLTSLKNLRASLYGKYVALRGTVV   218
Danio       DLERHAAELQGQEELPAGIRPIINIPHISARLYNYEPLTPLKSLRANLYGKFVVIRGTVV   238
            *** ***:** ** *      ::*:*.* **::**: ** **.:**. ***:: :*****
```

Supplemental Figure 3. 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 creating by threading the adjusted alignment (WT or P149R) onto the structure of *Sulfolobus solfataricus* MCM (PDBID: 3F9V)(2) using SWISS-MODEL (3). 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 1C.



Supplemental Table 1. Clinical Laboratory Investigations of Affected Daughters (IV-1, IV-6, and IV-9).

	Normal Range[a]	IV-1	IV-6	IV-9
FSH (IU/ml)	1.8-22.5[b]	73.25	95.09	42.13
AMH (ng/ml)	<6.9[c]	<0.6	<0.6	<0.6
LH (IU/ml)	1.2-100[b]	22.10	28.93	12.05
Estradiol (pg/ml)	30 to 300[b]	1.11	10.37	10.82
TSH (mIU/ml)	0.3-5.0[b]	7.68	6.51	6.24
Prolactin (ng/ml)	3-27	2.5	11.93	14.67
Uterine Volume (cm^3)[d]	20-100(4)	18.5	14.6	3.7
Right Ovarian Volume (cm^3)	6.6±0.19(5)	Not visualized	Not visualized	0.7
Left Ovarian Volume (cm^3)	6.6±0.19(5)	4.3	2.7	0.4
Anti-Thyroid Antibodies	Negative	Negative	Negative	Negative
Karyotype	46,XX	46,XX	46,XX	46,XX

[a]All hormone measures provided were prior to hormone replacement therapy.
[b]Mayo Clinic, Mayo Medical Laboratories, Rochester, MN.
[c]Reference Range for Adult Females, Esoterix Lab, Calabasas, CA.
[d]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.

Supplemental Table 2: Regions of Homozygosity Based on SNP Array			
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

Supplemental Table 3. Chromosomal Breakage as Induced by DEB treatment in Patient Lymphocytes (Raw Data).			
	Patient		
Cell Number	**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
Average Number of Breaks Per Cell	0.1	0.1	0

258

Supplemental Table 4. WES Variants in Genes Known to Be Involved in Chromosomal Instability or Primary Amenorrhea.		
Gene	**Associated Syndrome**	**Variant**
ATM	Ataxia-telangiectasia	Benign SNP
BRCA1	Breast Cancer	Benign SNP
BRCA2	Fanconi Anemia, Wilms tumor, and cancers	Benign SNP
BRIP1	Fanconi anemia, Breast Cancer	Benign SNP
ERCC4	Fanconi anemia	None
FANCA	Fanconi anemia	None
FANCB	Fanconi anemia	None
FANCC	Fanconi anemia	None
FANCD2	Fanconi anemia	None
FANCE	Fanconi anemia	None
FANCF	Fanconi anemia	None
FANCM	Fanconi anemia	None
FMR1	Fanconi anemia	None
HGPS	Hutchinson-Gilford progeria	None
MLH1	Mismatch repair cancer syndrome	None
MLH3	Colorectal Cancer, Endometrial Cancer	Benign SNP
MSH2	Mismatch repair cancer syndrome	None
MSH3	Endometrial carcinoma	Benign SNP
MSH6	Mismatch repair cancer syndrome	Benign SNP
NBN	Nijmegen breakage syndrome	Benign SNP
NBS1	Nijmegen Breakage Syndrome	None
PALB2	Fanconi anemia	None
PHF9	Fanconi anemia	None
RAD50	Nijmegen breakage syndrome-like disorder	None
RAD51C	Fanconi anemia	None
SLX4	Fanconi anemia	None
TP53	Multiple Cancers	Benign SNP
WRN	Werner syndrome	Benign SNP
XPC	Xeroderma pigmentosum	Benign SNP
XRCC9	Fanconi anemia	None

259

260

261

Supplemental Table 5. Quality Metrics for Sequencing by Sample.								
Lane on Flow Cell	**Sample ID**	**Yield (Mbases)**	**Number of Reads**	**% of raw clusters per lane**	**% of Perfect Index Reads**	**% of One Mismatch Reads (Index)**	**% of >= Q30 Bases (PF)**	**Mean Quality Score (PF)**
1	IV-9	14,038	140,381,192	44.02	98.8	1.2	89.87	35.09
1	IV-6	17,497	174,966,244	54.86	98.78	1.22	89.41	34.95
2	IV-1	9,115	91,146,660	33.77	84.25	15.75	91.5	35.57
2	IV-5	17,302	173,024,000	64.1	84.3	15.7	91.34	35.52
3	IV-3	15,629	156,291,636	45.32	86.61	13.39	88.3	34.64
3	III-1	18,405	184,050,526	53.37	86.76	13.24	88.06	34.57
4	III-2	16,120	161,199,806	48.58	99.35	0.65	89.69	35.04

262

263

Supplemental Table 6. Exome Sequencing Alignment Statistics by Sample.				
Sample ID	**Reads Aligned to the Exome**	**Average Coverage Across Exome**	**Reads Aligned to Targeted Capture Region**	**Average Coverage Across Capture Region**
IV-9	119,310,892	185	135,472,892	108
IV-6	147,764,231	233	169,010,803	137
IV-1	77,587,856	122	88,188,326	71
IV-5	147,215,164	230	166,322,091	133
IV-3	132,631,196	208	151,363,205	122
III-1	155,921,867	244	177,658,272	143
III-2	136,878,377	214	155,104,064	125

Supplemental Table 7. Variant Filtration for Homozygous Recessive Inheritance Pattern.	
	Variants Remaining
Total Variants Called in the Family	120,201
Application of GATK Quality Filters, variant kept if: (1) mapping quality filter = PASS (2) Quality Depth (QD) >2 (3) VQSLOD>0[a] (4) Mapping Quality (MQ) > 40 (5) Fisher test of strand bias (FS) < 60 (6) HaplotypeScore < 13 (7) MQRankSum > -12.5 (8) ReadPosRankSum > -8 (9) coverage > 10 (10) alternative allele called in > 4 reads (11) Genotype Quality (GQ) > 5, (12) Polymorphism Likelihood (PL) value difference <30	86,135
Application of Homozygous Recessive Inheritance Model: (1) affected daughters are homozygous alternative genotype (2) parents are heterozygous (3) unaffected siblings do not have a homozygous alternative genotype	1,030
Variants in Exons or Splice Sites as annotated using ANNOVAR (Last Change Date: 2013-02-11)	630
Nonsynonymous Variants	293
Remove Known Polymorphic Genes from NIH Recommendations (6, 7)	256
Variants with a Minor Allele Frequency <5% as reported in available databases (1000 Genomes and ESP6500 from NHLBI)	2
[a]The GATK Variant Recalibrator was used to count the VQSLOD (log odds ratio of variant quality score) based on variants present in the HapMap (v3.3), 1000 Genomes (Omni2.5 and Gold Standard Indels b37), and dbSNP137 databases.	

266

REFERENCES

1. Park, J., Long, D.T., Lee, K.Y., Abbas, T., Shibata, E., Negishi, M., Luo, Y., Schimenti, J.C., Gambus, A., Walter, J.C., et al. 2013. The MCM8-MCM9 complex promotes RAD51 recruitment at DNA damage sites to facilitate homologous recombination. *Mol Cell Biol* 33:1632-1644.

2. Brewster, A.S., Wang, G., Yu, X., Greenleaf, W.B., Carazo, J.M., Tjajadi, M., Klein, M.G., and Chen, X.S. 2008. Crystal structure of a near-full-length archaeal MCM: functional insights for an AAA+ hexameric helicase. *Proc Natl Acad Sci U S A* 105:20191-20196.

3. Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22:195-201.

4. Holm, K., Laursen, E.M., Brocks, V., and Muller, J. 1995. Pubertal maturation of the internal genitalia: an ultrasound evaluation of 166 healthy girls. *Ultrasound Obstet Gynecol* 6:175-181.

5. Pavlik, E.J., DePriest, P.D., Gallion, H.H., Ueland, F.R., Reedy, M.B., Kryscio, R.J., and van Nagell, J.R., Jr. 2000. Ovarian volume related to age. *Gynecol Oncol* 77:410-412.

6. Fuentes Fajardo, K.V., Adams, D., Program, N.C.S., Mason, C.E., Sincan, M., Tifft, C., Toro, C., Boerkoel, C.F., Gahl, W., and Markello, T. 2012. Detecting false-positive signals in exome sequencing. *Hum Mutat* 33:609-613.

7. Adams, D.R., Sincan, M., Fuentes Fajardo, K., Mullikin, J.C., Pierson, T.M., Toro, C., Boerkoel, C.F., Tifft, C.J., Gahl, W.A., and Markello, T.C. 2012. Analysis of DNA sequence variants detected by high-throughput sequencing. *Hum Mutat* 33:599-608.