GERMLINE BIALLELIC MCM8 VARIANTS ARE ASSOCIATED WITH EARLY-ONSET LYNCH-LIKE SYNDROME

Mariano Golubicki,¹,²* Laia Bonjoch,³,* José G. Acuña-Ochoa,³ Marcos Díaz-Gay,³,§ Jenifer Muñoz,³ Miriam Cuatrecasas,⁴ Teresa Ocaña,³ Soledad Iseas,¹ Guillermo Mendez,¹ Daniel Cisterna,² Stephanie A. Schubert,⁵ Maartje Nielsen,⁶ Tom van Wezel,⁵ Yael Goldberg,⁷ Eli Pikarsky,⁸ Juan Robbio,¹ Enrique Roca,¹ Antoni Castells,³ Francesc Balaguer,³ Marina Antelo,¹,&o and Sergi Castellví-Bel.³&

¹Oncology Section, Hospital of Gastroenterology "Dr. C. B. Udaondo", Buenos Aires, Argentina.
²Molecular Biology Laboratory, Hospital of Gastroenterology "Dr. C. B. Udaondo", Buenos Aires, Argentina.
³Gastroenterology Department, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Hospital Clínic, Universitat de Barcelona, Barcelona, Spain.
⁴Pathology Department, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD) and tumor Bank-Biobank, Hospital Clínic, Barcelona, Spain.
⁵Department of Pathology, Leiden University Medical Center, Leiden University, Leiden, The Netherlands.
⁶Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands.
⁷Recanati Genetics Institute, Rabin Medical Center, Petah Tikva, Israel.
⁸The Lautenberg Center for Immunology and Cancer Research, IMRIC, Hebrew University-Hadassah Medical School, Jerusalem, Israel.

*Contributed equally to this work
&Joint senior authors
Present address: Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA, 92093, USA

Corresponding authors: Dr. Sergi Castellví-Bel, Institut d’Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Rosselló 153, 08036 Barcelona, Spain; 34-93-2275400 ext.4183; sbel@clinic.cat; Dr. Marina Antelo, Oncology Section, Hospital of Gastroenterology “Dr. C. B. Udaondo”, Buenos Aires, Argentina; 54 9 11 4 304 3202; machuantelo@hotmail.com

Article type: Research article

Word count: 6,603 words

Conflict of interest statement

The authors have declared that no conflict of interest exists.
ABSTRACT

Lynch syndrome is the most common colorectal cancer (CRC) hereditary form and it is characterized by DNA mismatch repair (MMR) deficiency. The term Lynch-like syndrome (LLS) is used for patients with MMR-deficient tumors and neither germline mutation in MLH1, MSH2, MSH6, PMS2, or EPCAM, nor MLH1 somatic methylation. Biallelic somatic inactivation or cryptic germline MMR variants undetected during genetic testing have been proposed to be involved. Sixteen patients with early-onset LLS CRC were selected for germline and tumor whole-exome sequencing. Two potentially pathogenic germline MCM8 variants were detected in a LLS male patient with fertility problems. A knockout cellular model for MCM8 was generated by CRISPR-Cas9 and detected genetic variants were produced by mutagenesis. DNA damage, microsatellite instability and mutational signatures were monitored. DNA damage was evident for MCM8KO cells and the analyzed genetic variants. Microsatellite instability and mutational signatures in MCM8KO cells were compatible with the involvement of MCM8 in MMR. Replication in an independent familial cancer cohort detected additional carriers. Unexplained MMR-deficient CRC cases, even showing somatic biallelic MMR inactivation, may be caused by underlying germline defects in genes different than the MMR genes. We suggest MCM8 as a new gene involved in CRC germline predisposition with a recessive pattern of inheritance.

Keywords: Colorectal cancer, genetic predisposition to disease, Lynch syndrome, Lynch-like syndrome, DNA mismatch repair, CRISPR-Cas9.
INTRODUCTION

The most frequent cause of hereditary colorectal cancer (CRC) is Lynch syndrome (LS), accounting for 3% of all stages of CRC (1). LS is an autosomal dominant condition caused by germline pathogenic variants in one of the DNA mismatch repair system (MMR) genes or the EPCAM gene (2). Although MSH2 and MLH1 account for most of the LS-associated CRC cases, PMS2 and MSH6 variants are actually more prevalent on population basis (3). This syndrome has a marked gene-dependent variable penetrance for CRC and endometrial carcinoma (12-55%), and an increased risk for various other extra-colonic tumors (4). Surveillance colonoscopies every two years starting at age 25-35 and yearly endometrial screening from age 40 have been advised to reduce morbidity and mortality related to cancer (5).

Tumors of LS patients display microsatellite instability (MSI), caused by the initial germline inactivation and a second somatic hit in the other allele of one of the MMR genes (2). However, the MSI specificity for LS is low because it also occurs in 12-15% of sporadic CRC cases, usually due to somatic MLH1 promoter region hypermethylation. Nevertheless, for a MSI-positive CRC, absence of MLH1 somatic hypermethylation or MMR germline pathogenic variants may be as common as 70% (6). MMR deficiency (MMRd) in tumors could be due to a sporadic “chance” event (evidenced by somatic mutations in BRAF, double somatic DNA MMR genes, or MLH1 promoter hypermethylation), or caused by an underlying “syndromic” predisposition to MMRd tumors by undetected germline defects either in the DNA MMR genes known to be associated with LS or in other “novel” genes. The cases in the latter option have been termed “Lynch-like syndrome” (LLS) and management decisions in these patients and their families are complicated because of a suspected but unconfirmed hereditary origin (7).

Then, three possibilities could explain a MMRd tumor: 1) cryptic/undetected germline mutation in DNA MMR gene (occult/undetected LS), 2) double somatic mutation in DNA MMR gene (sporadic MMRd) in 50-60% of cases (8-10), and 3) mutations (germline) in other genes
involved in DNA MMR (novel heritable predisposition). The last two hypotheses may overlap in some patients, and it is unclear if this is the case in early-onset LLS cases.

Keeping in mind the last possibility for a MMRd tumor and to further explore it, the aim of the present study was to investigate if early-onset LLS patients carried potentially pathogenic germline variants in new CRC predisposition genes causing a MMRd tumor phenotype. Accordingly, we performed exome sequencing in the germline and tumor DNA of 16 early-onset CRC LLS patients, and carried out an exhaustive functional evaluation for 2 rare recessive germline variants detected in a candidate gene.
RESULTS

Clinical characteristics and germline sequencing results

Sixteen patients with non-polyposis LLS CRC diagnosed before the age of 40 were selected. These patients presented tumors with MSI and/or immunohistochemical loss of MLH1, MSH2, MSH6 or PMS2, wild-type *BRAF* V600E and/or negative *MLH1* methylation, and with no germline pathogenic variants in the MMR or *EPCAM* genes.

Germline whole-exome sequencing (WES) were performed in all DNA samples. A recessive analysis was prioritized in the analyzed patients since the family pattern was more compatible with an autosomal recessive inheritance. After variant prioritization, germline WES data analysis selecting for variants located on genes with a function compatible with cancer development yielded two potentially pathogenic *MCM8* variants in a male early-onset CRC LLS patient (LLS17, Figure 1A). LLS17 presented an IIIB stage (T4N1M0), left-colon sided, well-differentiated, mucinous adenocarcinoma diagnosed at 40 years old, referred no cancer family history and reported fertility problems with his spouse suffering an advanced pregnancy miscarriage. The *MCM8* variants identified were p.(Lys118Glufs*5), and p.(Ile138Met). Manual visualization of WES data showed they were in trans and mutually exclusive when a single sequencing read passed through both genomic positions (Figure 1B). Both were validated by Sanger sequencing (Figure 1C). The p.(Lys118Glufs*5) frameshift variant was predicted pathogenic and very rare in the general population (11/276,744). On the other hand, the missense p.(Ile138Met) variant was also *in silico* predicted as deleterious and, to be destabilizing for the protein, as well as moderately rare in controls (1,604/282,692).

Tumor analysis

Previous tumor analysis in LLS17 showed an MSI-positive phenotype with loss of MLH1/PMS2 protein expression (Supplementary Figure S1), wild-type *BRAF* V600E, and no *MLH1* somatic hypermethylation. Somatic WES revealed a high tumor mutational burden (77
single nucleotide variants/Mb), an important contribution of the MMR-related mutational signature SBS15 (Figure 2) and provided a set of relevant somatic variants, which are summarized in Supplementary Table S1. Among them, two somatic MLH1 truncating variants (c.129dupA and c.1831delA) were detected and subsequently confirmed by Sanger sequencing (Figure 3A). A putative mosaicism for these variants was disregarded in the patient by manual inspection of the germline WES data at the corresponding genomic locations (Figure 3B).

**CRISPR/Cas9 MCM8\(^{\text{KO}}\) modeling**

We established CRISPR/Cas9 MCM8\(^{\text{KO}}\) knock-outs (\(\text{MCM8}^{\text{KO}}\)) on DLD-1 cells (human CRC cell line) in order to test the functional impairment of the identified variants and further validate their suspected role in CRC germline predisposition. According to bioinformatic CRISPR prediction tools, a sgRNA targeting the fifth exon was selected. The disruption of \(\text{MCM8}\) and the purity of the clones were confirmed by Sanger sequencing. \(\text{MCM8}^{\text{KO}}\) expression depletion was verified at both RNA and protein levels (Figure 4). Two \(\text{MCM8}^{\text{KO}}\) clones (5.2 and 5.3) were selected to carry out further expansion and functional characterization.

**Functional characterization of germline variants**

To evaluate the functional effect of the identified \(\text{MCM8}\) variants, site-directed mutagenesis was performed on a vector carrying the wild-type ORF of this gene. Both p.(Lys118Glufs*5) and p.(Ile138Met) variants were generated and confirmed by Sanger sequencing (Figure 5A). Both vectors carrying the selected variants were transiently transfected in \(\text{MCM8}^{\text{KO}}\) cells and their expression was detected at the mRNA level (Figure 5B). However, cells expressing the \(\text{MCM8}\) p.(Lys118Glufs*5) variant showed absence of MCM8 protein expression, suggesting its depletion (Figure 5C).

Since MCM8 is involved in repair of DNA double-strand breaks (DSB), the comet assay was used to monitor DNA damage (Figure 6A). Experimental conditions were set up by a time-course of DNA repair kinetics on \(\text{MCM8}^{\text{WT}}\) and \(\text{MCM8}^{\text{KO}}\) cells. After oxaliplatin
treatment, cells were allowed to recover during different time points. DNA repair was already noticeable in $MCM8^{WT}$ cells after a 16-hour resting period, whereas at the same time-point, $MCM8^{KO}$ 5.2 and 5.3 clones still showed sustained DNA damage (Supplementary Figure S2A). This 16-hour resting period was selected to further characterize the repair activity of $MCM8$ variants through the comet assay. Results revealed that $MCM8^{WT}$ cells almost completely recovered DNA integrity 16 h after treatment, as they showed less dispersed DNA tail comets. Both $MCM8^{KO}$ 5.2 and 5.3 cells showed an increased DNA damage grade after the same resting period, reflecting DNA-DSB repair impairment (Figure 6B, upper panel).

Once the implication of $MCM8$ on DNA DSB repair was confirmed, we proceeded to test the effect of p.(Ile138Met) and p.(Lys118Glufs*5) variants. To do so, the $MCM8^{KO}$ 5.2 clone was transiently transfected with either $MCM8^{WT}$, p.(Ile138Met) or with p.(Lys118Glufs*5) expressing vectors, exposed to oxaliplatin treatment, and then allowed to recover. We observed that cells expressing both p.(Ile138Met) and p.(Lys118Glufs*5) variants showed larger tails (Figure 6B, lower panel), and therefore, a higher DNA damage retention in comparison with the rescued phenotype ($MCM8^{WT}$). The quantitative analysis from the obtained comet images confirmed the impaired DNA repair capacity of MCM8-depleted cells and those expressing both $MCM8$ variants (Figure 6C, Supplementary Figure 2B).

Microsatellite instability and mutational signatures analyses

To determine if depletion of $MCM8$ was involved in alteration of the MMR system in our model, MSI was analyzed in $MCM8^{WT}$, $MCM8^{KO}$ 5.2 and $MCM8^{KO}$ 5.3 cells after 30, 60 and 90 days of sub-culturing. When comparing with $MCM8^{WT}$, a MSI-like profile was especially evident in the $MCM8^{KO}$ 5.2 clone, already present at 30 days of sub-culturing, whether it was more subtle for the $MCM8^{KO}$ 5.3 clone (Figure 7A).

To gain more insight about the relationship between $MCM8$ deficiency and the MSI phenotype, we performed WES and mutational signature analysis in $MCM8^{WT}$, $MCM8^{KO}$ 5.2 and $MCM8^{KO}$ 5.3 cells after 120 days of culturing. We were able to determine which genetic variants appeared during culturing and focus our signature analysis on them. Single-base
signatures (SBS) SBS1, SBS5, SBS20 and SBS44 were detected in the analyzed samples (Figure 7B). SBS1, SBS5 and SBS44 were present in all samples. SBS44 is associated with a defective DNA MMR and was also present in $MCM8^{WT}$ cells, most likely due to the fact that cells used as cellular model (DLD-1) are MSH6-defective. Interestingly, both $MCM8^{KO}$ clones acquired a distinctive significant contribution of the SBS20 not present in $MCM8^{WT}$. This signature also represents one of the seven signatures associated with a MMR system impairment and it is currently associated at the moment to concurrent $POLD1$ pathogenic variants and defective DNA MMR (11). Regarding mutational burden, $MCM8^{KO}$ 5.2 cells accumulated more variants than $MCM8^{WT}$ or $MCM8^{KO}$ 5.3 cells (Supplementary Figure 3A). At the point variant level, no relevant putative somatic mutations were detected by WES in $MCM8^{WT}$, $MCM8^{KO}$ 5.2 and $MCM8^{KO}$ 5.3 cells after culturing. However, considering length of indel substitutions, a higher frequency of indels greater than 5 base pairs was found in the $MCM8^{KO}$ 5.3 clone, suggesting that other alterations in DNA repair mechanisms, such as homologous recombination (HR) (12), could be a concomitant defect associated to $MCM8$ depletion (Supplementary Figure 3B).

**Screening of the candidate gene variants in an independent cohort**

An independent cohort of 131 Dutch unaffiliated familial cancer cases (mainly CRC) with available WES data was accessible. Both $MCM8$ and $MCM9$ were screened in this cohort, since a link between $MCM9$ and inherited predisposition to mixed polyposis and early-onset CRC has also been previously suggested and both proteins cooperate to perform their function (13). Results are summarized in Table 1. A breast cancer patient carried biallelic $MCM8$ genetic variants in the context of a breast cancer family, and 5 additional heterozygote carriers were detected. Regarding $MCM9$, two families carried biallelic genetic variants, including a LLS patient with premature ovarian failure and a familial MMR-proficient CRC patient, and heterozygote variants were found in 12 patients. Family trees for biallelic carriers are available in Supplementary Figure S4. Mostly, additional segregation for the detected genetic variants was not possible in these families with the exception of one of the families.
carrying biallelic \textit{MCM9} variants where an affected brother of the proband also carried both variants. However, according to an \textit{in silico} prediction (CADD>15), most of them could correspond to potentially pathogenic genetic variants.
DISCUSSION

We analyzed 16 patients with early-onset non-familial non-polyposis CRC and LLS to identify novel germline candidate genes in this scenario. Although biallelic somatic pathogenic variants in MMR genes may account for over half of CRC labeled as LLS (10), some LLS patients may have a hereditary origin, especially in the early-onset setting, such as biallelic MUTYH or monoallelic POLE/POLD1 carriers (14). We identified two potentially pathogenic germline biallelic MCM8 variants in one of these patients, also presenting somatic biallelic MLH1 mutations. Therefore, our main hypothesis was that germline alterations in MCM8 following a recessive pattern of inheritance could be involved in predisposing to CRC with a somatic MMRd phenotype.

The MCM8 gene (mini-chromosome maintenance 8 homologous recombination repair factor) raised interest as a candidate due to its previous association with ovarian/gonadal failure and chromosomal instability (15). Also, it forms a complex with MCM9, a gene required in the MMR system (16) and recently associated with germline predisposition to CRC also with recessive pattern of inheritance (13). MCM8 works together with MCM9 in a helicase hexameric complex involved in several functions such as genome maintenance, meiotic recombination and DSB repair via HR (17). Although germline variants in several DNA helicases have been associated to cancer (18), and recessive MCM9 variants have shown to predispose to CRC and premature ovarian failure (POF) in a recessive manner (13), there is no evidence currently of an association between MCM8 and germline CRC predisposition. Remarkably, germline alterations in this gene have been related to POF and chromosomal instability (19, 20). Very recently, the same MCM8 truncation variant found in LLS17 in this study p.(Lys118Glufs*5) was also found in homozygosis in two sisters affected with POF (21).

MCM8 participates in DSB-repair by HR interacting with MRN complex (MRE11-RAD50-NSB1), and is required for nuclease activity and stable association with DSB (22). Cells lacking MCM8 or MCM9 are viable but highly sensitive to interstrand crosslinking.
(ICL)-inducing agents, and exhibit more chromosome aberrations in the presence of mitomycin C compared with wild-type cells (15). During ICL repair, MCM8 and MCM9 form nuclear foci that partly colocalize with Rad51. MCM8-9 dimer also works downstream of the Fanconi anemia and the BRCA2/Rad51 pathways, and is required for HR that promotes sister chromatid exchanges, as a hexameric ATPase/helicase (23). Although only \textit{MCM8} variants located in the helicase domain have been tested to explore HR deficiency (23), we showed to some extent that the \textit{MCM8} p.(Ile138Met) variant, located in the DNA binding domain, can impair DSB repair capacity.

By performing \textit{MCM8} gene editing in a cellular model and ectopic reintroduction of the identified genetic variants, we were able to demonstrate its plausible effect on DNA repair efficiency. Besides, although several genes implicated in DNA repair have been spotlighted as probable causes of the underlying mechanisms of LLS CRC (24), this study differs from previous reports in the fact of performing a thorough functional characterization of the proposed candidate gene. Additionally, although HR and MMR systems are inextricably linked (25), we were capable of suggesting that both systems could be impaired by biallelic pathogenic variants in \textit{MCM8}.

However, the relationship between MCM8 and the MMR system still remains unclear. MCM9 was previously reported to be required for mammalian MMR (16), as it co-immunoprecipitates with MCM8, MSH2 and MLH1 (both key MMR system proteins) and its deficiency is linked to MSI and a MMR alteration. MCM8 and MCM9 arose early during eukaryotic evolution and are conserved among many eukaryotic organisms (not in yeast or \textit{C. elegans}). \textit{D. melanogaster} and related species only possess the MCM8 protein (i.e. REC) and MCM9 is absent (26). Additionally, MCM8 protein expression is lost in \textit{MCM9} knock-out cells, suggesting that both proteins are needed for the formation of an hexameric complex and its stabilization (16, 23). Thus, this fact is in agreement with variants in \textit{MCM8} and \textit{MCM9} being responsible for similar phenotypes as those previously reported (13). However, MCM8 or
MCM9 defects may not be always linked with a MMR deficient or MSI-positive tumor. In fact, the MMR system was preserved and MSI was negative in the biallelic MCM9 CRC patient previously reported (Goldberg Y and Pikarsky E, unpublished data). On the other hand, the developed MCM8KO model showed impairment of the MMR system. Additionally, MCM8KO cells developed the SBS20 mutational signature over time. This signature was previously associated with MSI CRC and concomitant POLD1 somatic mutations (11, 27). It represents a unique mutational profile reflecting the biological interaction between POLD1 and the MMR system, rather than the sum of both processes. It would remain to be elucidated whether the interaction between MCM8 and the MMR system could represent a similar profile. Based on these results, we hypothesize that MCM8 pathogenic variants may perturbate the function of the MCM8/MCM9 hexameric complex, impairing both the MMR and the HR-mediated DNA repair pathways (Figure 8).

Finally, the presence of MCM8/MCM9 genetic variants in the independent analyzed cohort with a putative involvement in familial cancer/CRC predisposition supports to some extent our conclusions. Besides, as highlighted by the biallelic MCM8 carrier with breast cancer, a potential pleiotropy effect for this gene could be hypothesized to include germline predisposition to breast cancer besides CRC. Further family segregation and functional characterization of these variants is warranted, as well as replication in additional familial CRC cases to confirm the implication of these genes in hereditary CRC and other neoplasms.

To conclude, with our study we provide evidence that some LLS CRC cases with a defective MMR system, showing somatic biallelic MMR inactivation, may be caused by underlying germline pathogenic variants in new CRC predisposing genes, especially in patients with an early onset. We suggest MCM8 as a potential CRC predisposing gene following a recessive inheritance pattern, and recommend this gene and MCM9 to be included in future screening of unaffiliated familial CRC cohorts in order to gain additional knowledge of its involvement in germline CRC predisposition.
METHODS

DNA extraction
Germline, tumor and cultured cells’ DNA was extracted with the QIAamp DNA Blood, QIAamp DNA FFPE and QIAamp DNA Mini kits, respectively (Qiagen, Hilden, Germany), following the manufacturer's instructions.

Exome sequencing
Germline and tumor WES were performed in DNA samples using the HiSeq2000 platform (Illumina, San Diego, CA, USA) and SureSelectXT Human All Exon v5 kit (Agilent, Santa Clara, CA, USA) for exon enrichment at CNAG (Centre Nacional d’Anàlisi Genòmica, www.cnag.org.eu). Indexed libraries were massively parallel-sequenced using a paired-end 2 × 75 bp read length protocol. Sequencing data quality control previous to its analysis was performed in all samples using the Real-Time Analysis software sequence pipeline (Illumina). The Burrows–Wheeler Aligner (BWA-MEM algorithm) was used for the human reference genome read mapping (build hs37d5, based on NCBI GRCh37) (28).

Sequencing data analysis
The GATK HaplotypeCaller tool was used for single nucleotide variants (SNV) and short indels calling in germline samples, and GATK MuTect2 and Strelka2 were applied in tumor samples using previously developed R language in-house pipelines (29-32). Several databases were evaluated for variant annotation, including SnpEff and dbNSFP. PhyloP, SIFT, PolyPhen2, MutationTaster, LRT and CADD were used for missense variants pathogenicity prediction, as previously described (30, 33). For germline DNA data, we selected those genes following a recessive pattern of inheritance with two potentially pathogenic variants per individual, both presenting an allele frequency in the Genome Aggregation Database (gnomAD, https://gnomad.broadinstitute.org/) lower than 1%. Guidelines for the interpretation of sequence variants from the American College of Medical
Genetics and Genomics and the Association for Molecular Pathology were also used (34). STRUM was used for predicting the effect of genetic variants regarding fold stability change of protein molecules (35). Concerning tumor variants, they were restrained to those having a coverage $\geq 10X$ both in germline and somatic samples, an alternative allelic frequency in the tumor $\geq 20\%$ and being truncating or missense variants fulfilling at least three of the missense pathogenicity tools criteria. SigProfiler (11, 36) was used to perform SNV mutational signature refitting analysis according to COSMIC reference signatures (37). Tumor mutational burden was assessed exploring somatic SNVs and indels.

Additional germline variant prioritization was carried out to select those actually relevant for the CRC phenotype, and manually curated using the Integrative Genomics Viewer (38). Germline and tumor candidate variants were subsequently validated by Sanger sequencing. Primer details are available in Supplementary Table S2.

**Human CRC cell line**

The DLD-1 human CRC cell line was purchased from the American Type Culture Collection (ATCC, Manassas, VA, USA) and cultured in RPMI-1640 medium supplemented with 10% fetal bovine serum (Gibco, ThermoFisher, Waltham, MA, USA), at 37°C in 5% CO$_2$.

**Plasmids**

LentiCRISPRv2-Puro (Addgene, #98290) expression vector was available. $MCM8$ ORF (NM_032485.5) cloned into the pcDNA3.1 expression vector (OHu10568D) was purchased from GenScript (Nanjing, China).

**CRISPR/Cas9-mediated MCM8 knockout generation**

The Benchling (http://benchling.com) CRISPR tool was used to design the single guide RNA (sgRNA) against the coding region of the $MCM8$ gene. The sgRNA was cloned into the LentiCRISPRv2-Puro vector and transiently transfected into the DLD-1 CRC cell line. Two days later, transfected cells were puromycin-selected (4 µg/mL) and seeded into a 96-well
plate at a density of 1 cell/well. Several clones were characterized and selected for further analysis. *MCM8* gene editing was validated by Sanger sequencing, and gene down-regulation and depletion were checked by quantitative real-time PCR and Western Blot, respectively.

**Antibodies**

Polyclonal antibody against MCM8 (ab183045) was from GeneTex (Irvine, CA). Anti-GAPDH (14C10) was purchased from Cell Signaling (Danvers, MA). Goat anti-rabbit (SA5-10036) DyLight® 800 secondary antibody was acquired from Thermo Fisher Scientific (Waltham, MA, USA).

**RNA extraction and quantitative real-time PCR**

Total RNA extraction was performed with the RNeasy Mini Kit according to manufacturer’s instructions (Qiagen, Hilden, Germany). RNA was retrotranscribed using the Applied Biosystems™ High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher, Waltham, MA, USA). Multiplex quantitative PCR was performed with the Applied Biosystems™ 7300 PCR System by using a specific TaqMan assay for *MCM8*-FAM (hs01067422_m1). The endogenous control gene was GAPDH-VIC (4326317E). Relative expression levels of each target gene were calculated for each sample as \(-\Delta \text{Ct}\) values (\(-\Delta \text{Ct}= – [\text{Ct target gene} – \text{Ct endogenous control}]\)).

**Protein extraction and Western blot**

To obtain whole-cell protein extracts, cells were detached from cell culture plates with Accutase® (Sigma Aldrich, San Luis, MO, USA) and lysed with RIPA buffer supplemented with cOmplete Protease Inhibitor Cocktail and PhosSTOP™ (Roche, Basilea, Switzerland). Sample protein concentrations were determined by using the Pierce BCA Protein Assay Kit (Thermo Fisher, Waltham, MA, USA). Equal amounts of protein lysates were resolved in NuPAGE Bis-Tris protein gel electrophoresis, followed by protein transfer onto Immobilon PVDF membranes (Millipore, MA, USA), according to manufacturer’s protocols (Thermo
Proteins were blotted with the indicated primary and secondary Dylight® antibodies and detected by using the Odyssey Imaging System (LI-COR, Lincoln, NE, USA).

**Site-directed mutagenesis**

The Q5® Site-Directed Mutagenesis Kit (NEB, Ipswich, MA, USA) was used to introduce the variants of interest in the wild-type (WT) pcDNA3.1-\textit{MCM8} expression vector. Mutagenic primers were designed using the NEBaseChanger tool and obtained from IDT (Coralville, IA, USA) (Supplementary Table S2). Mutagenesis products were verified by Sanger sequencing.

**Microsatellite instability assay by multiplex PCR**

\textit{MCM8}\textsubscript{WT} and \textit{MCM8}\textsubscript{KO} cells were maintained during 90 days. DNA from cells was extracted every 30 days to test MSI profiles in two different microsatellite markers (BAT25, BAT26) (39) by using a capillary electrophoresis genetic fragment analyzer (Applied Biosystems, Foster City, CA, USA). Primer sequences are available in Supplementary Table S2.

**Neutral single-cell gel electrophoresis (Comet assay)**

\textit{MCM8}\textsubscript{KO} clones were transiently transfected with WT or mutated p.(Lys118Glufs*5) or p.(Ile138Met) plasmids using X-tremeGENE HP DNA transfection reagent. Two days later, cells were subjected to selection with 1 mg/mL of G418 (InvivoGen, San Diego, CA, USA) for 72 h, as the \textit{MCM8} expression pcDNA3.1 vector carries a neomycin resistance cassette.

To induce DNA damage, cells were incubated in media containing 100µM oxaliplatin for 90 min. After a resting period of 16 hours, cells were collected to evaluate their DNA repair capacity by a neutral comet assay (Trevigen, Gaithersburg, MD, USA). As a control, oxaliplatin-treated cells with no resting period were also collected.

**Exome sequencing and mutational signatures in \textit{MCM8}\textsubscript{KO} cell lines**
Three cell lines (MCM8<sup>WT</sup> cells and two different MCM8<sup>ko</sup> clones) were cultured during 120 days to allow mutation accumulation. Then, DNA samples were obtained and evaluated by WES together with a DNA sample from the original MCM8<sup>WT</sup> pool (day 0) to assess the initial mutational background. Data was processed as previously described. Mutect2, VarScan and Strelka2 were used for variant calling (29, 31, 40). Only the variants that were called by at least two of these three callers were chosen for further analysis. Results from MCM8<sup>WT</sup> cells at the initial time-point (0 days) were used as parental normal control against the three cells lines cultured for 120 days to exclude the initial mutational background and only consider genetic variants that appeared during culturing. SigProfiler was used again as previously described (11, 36).

**Screening of the candidate gene variants in an independent cohort**

In order to seek for additional patients with potentially pathogenic germline variants in our candidate gene, access to WES data from an independent cohort of 131 Dutch unaffiliated familial cases (mainly CRC) was granted for replication purposes. We selected coding potentially pathogenic variants (missense, frameshift, truncating, splicing altering, CADD>15), presenting an allele frequency in gnomAD lower than 1%.

**Statistics**

For multiple comparisons, 1-way ANOVA was followed by Tukey’s multiple-comparisons test. P values were considered significant at P < 0.05.

**Study approval**

Sixteen patients with non-polyposis LLS CRC diagnosed before the age of 40 were selected from a cohort previously described by Antelo et al. (6) and from the high-risk clinic for gastrointestinal cancer at Hospital Clinic in Barcelona. These patients presented tumors with MSI and/or immunohistochemical loss of MLH1, MSH2, MSH6 or PMS2, wild-type BRAF V600E and/or negative MLH1 methylation, and with no germline pathogenic variants in the
MMR or *EPCAM* genes. This study was approved by the institutional ethics committee (2011/6440) and written informed consent was obtained in all cases.
AUTHOR CONTRIBUTIONS

Conceptualization, MG, LB, MA and SCB; Funding acquisition, MA, SCB, AC, and FB; Investigation, MG, LB, JGAO, MDG, JM, MC, TO, SAS, MN, TW, YG, EP, MA and SCB; Resources, MG, LB, JGAO, MDG, JM, MC, TO, SAS, MN, TW, YG, EP, AC, FB, MA and SCB; Supervision, LB, MA and SCB, Visualization, MG, LB, JGAO, MDG, JM, MC, TO, SAS, MN, TW, YG, EP, MA and SCB; Writing – original draft, MG, LB, MA and SCB; Writing – review & editing, all authors.
ACKNOWLEDGEMENTS

MG and MA were supported by Foundation Nelia et Amadeo Barletta contract. LB was supported by a Juan de la Cierva postdoctoral contract (FJCI-2017-32593). JM was supported by a contract from CIBEREHD. MD-G was supported by a contract from Agència de Gestió d’Ajuts Universitaris i de Recerca -AGAUR- (Generalitat de Catalunya, 2019FI_B2_00203). CIBEREHD is funded by the Instituto de Salud Carlos III. This research was supported by grants from Foundation Nelia et Amadeo Barletta FNAB and the Argentinian National Cancer Institute. This research was supported by grants from Fondo de Investigación Sanitaria/FEDER (16/00766, 17/00878, 17/01304, 19/01867), Fundación Científica de la Asociación Española contra el Cáncer (GCB13131592CAST), PERIS (SLT002/16/00398, Generalitat de Catalunya), CERCA Program (Generalitat de Catalunya) and Agència de Gestió d’Ajuts Universitaris i de Recerca (Generalitat de Catalunya, GRPRE 2017SGR21, GRC 2017SGR653). This research was funded by a research grant from the Dutch Digestive Foundation (MLDS FP13-13) awarded to TvW. This article is based upon work from COST Action CA17118, supported by COST (European Cooperation in Science and Technology, www.cost.eu).

We are sincerely grateful to the patients, Dr. Celia Badenas, Prof. Dr. Esteban Cvitkovic, CNAG and the Biobank of Hospital Clínic–IDIBAPS. The work was carried out (in part) at the Esther Koplowitz Centre, Barcelona.
REFERENCES


Figure 1. Family tree and germline sequencing results. (A) LLS17 family. Black arrow indicates CRC patient carrying the germline MCM8 variants. (B) Screenshot showing that the biallelic variants are in trans. The variants are mutually exclusive when a single read crosses both positions. (C) Sanger validation of MCM8 germline biallelic variants p.(Lys118Glufs*5) and p.(Ile138Met) (c.351_354delAAAG and c.414A>G). CRC, colorectal cancer. AML, acute myocardial infarction. MSI-H, microsatellite instability high.
Figure 2. Tumor mutational profiling analysis. (A) Mutational signatures contribution extracted with SigProfiler of the LLS17 tumor sample showing a significant contribution of the SBS15 signature, associated with defective MMR system. (B) Single base substitution, insertion and deletion, and doublet base substitution profiles.
Figure 3. Tumor *MLH1* genetic variants. (A) Sanger validation of *MLH1* somatic variants found in LLS17 tumor. For the c.192dupA variant, the reverse complementary sequence is shown. (B) Screenshot of IGV manual curation showing germline location of the position of *MLH1* somatic variants, discarding mosaicism. In the lower panel, Sanger chromatogram confirming the absence of both *MLH1* variants in germline DNA.
Figure 4. CRISPR/Cas9 MCM8KO modeling. (A) Sanger sequencing confirmed a disruption in exon 5 of the WT MCM8 sequence. The black arrow indicates the Cas9 cleavage site, which is located three nucleotides upstream the PAM sequence. The Cas9 cleavage efficiency was validated in the mixed pool of cells two days after transfection. Gene editing in clones 5.2 and 5.3 was also validated by Sanger sequencing. (B) A reduction of MCM8 RNA expression levels was detected in MCM8KO clones when compared MCM8WT cells (WT), according to real-time PCR results. Samples were assayed in triplicate (n = 1). The solid line represents the mean value. (C) Immunoblotting analysis of MCM8KO protein extracts showed a loss of expression within MCM8 weight range (94kDa). Representative blot of n = 2. The uncropped Western Blot can be found in Supplementary Material. For both real-time PCR and Western blot, GAPDH was used as an internal constitutive control. PAM: Protospacer adjacent motif.
Figure 5. *MCM8* variant generation and expression in *MCM8<sup>KO</sup>* cells. (A) Sanger sequencing of site-directed mutagenesis products confirmed the correct generation of *MCM8* p.(Lys118Glufs*5) (c.345delAGAA) and p.(Ile138Met) (c.414A>G) variants. Changes in the nucleotide sequence are marked, and the altered amino acid sequence is highlighted in bold. (B) mRNA expression and (C) protein levels of both MCM8 variants and the MCM8<sup>WT</sup> rescued phenotype in the *MCM8<sup>KO</sup>* 5.2 clone. In (B), samples were assayed in triplicate (*n* = 1). The solid line represents the mean value. In (C), Western blot analysis of MCM8 expression showed no band within the range of MCM8 weight (94 kDa) in the p.(Lys118Glufs*5) *MCM8<sup>KO</sup>* transfected cells, suggesting a knock-out-like pattern. Representative blot of *n* = 3. The uncropped Western Blot can be found in Supplementary material. GAPDH was used as internal constitutive control.
Figure 6. *MCM8* genetic variants displayed less ability to repair damaged DNA. (A) Overview of the DNA repair experiment (comet assay). (B) Representative images of neutral comet assay (*n = 3*). Upper panel, DNA damage impairment detected in *MCM8*KO cells in comparison to *MCM8*WT cells. Lower panel, *MCM8*KO 5.2 expressing p.(Ile138Met) or p.(Lys118Glufs*5*) (short format is displayed). MCM8 proteins showed lower DNA repair capacity than the rescued phenotype (*MCM8*WT). Scale bar: 100 µm. (C) Quantitative analysis of DNA damage in the three independent experiments, measured as the amount of tail DNA.
Box and whiskers represent 25–75 and 5–95 percentiles, respectively. The solid line represents the median value. **$P<0.01$, ***$P<0.001$, 1-way ANOVA with Tukey post hoc test.
Figure 7. *MCM8* deficiency can initiate microsatellite instability and mutational signature related to MSI. (A) After 30 days of sub-culturing, *MCM8<sup>KO</sup>* 5.2 cells already showed a profile shift for BAT25 and BAT26 microsatellite markers when comparing with day 0. This profile was detected throughout the different time points of the culture process. The *MCM8<sup>KO</sup>* 5.3 clone exhibited milder MSI alterations only apparent for BAT25. Profile shifts are indicated with an arrow. (B) *MCM8<sup>KO</sup>* 5.2 and 5.3 displayed a significant contribution of SBS20 mutational signature (associated with concomitant *POLD1* mutations and defective DNA mismatch repair) whereas it was not detected on *MCM8<sup>WT</sup>* cells cultured over the same period.
Figure 8. Proposed molecular mechanism of pathogenicity for MCM8 variants. The MCM8/MCM9 hexameric complex is required for the MMR system and the HR-mediated DNA repair caused by interstrand crosslinking agents. Previous work has demonstrated that MCM9 interacts with the MMR proteins and is recruited to the mismatch lesion (16). Also, the MCM8/MCM9 complex is required for the MRN protein complex at DNA damage foci to facilitate DNA resection, a key step on DNA DSB repair (22). We propose that biallelic germline mutations in MCM8 affect the MCM8/MCM9 protein complex, impairing both DNA repair pathways and leading to the accumulation of non-corrected mutations. These alterations, in some cases, can also promote somatic biallelic MMR inactivation and LLS. CRC, colorectal cancer. DSB, double-strand breaks. EO LLS, early-onset Lynch-like syndrome. HR, homologous recombination. ICL, interstrand crosslinking. MMR, mismatch repair.
Table 1. Variants in *MCM8* and *MCM9* detected in 131 Dutch cancer patients. All variants have allele frequency < 0.01 in the gnomAD database. Potentially pathogenic genetic variants (CADD>15) are highlighted in bold.

<table>
<thead>
<tr>
<th>MCM8 biallelic carrier</th>
<th>Patient</th>
<th>Cohort</th>
<th>Chr</th>
<th>Pos</th>
<th>R</th>
<th>A</th>
<th>Gen</th>
<th>DNA</th>
<th>Protein</th>
<th>GnomAD</th>
<th>CAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>SXS48</td>
<td>Breast cancer</td>
<td>20</td>
<td>593927</td>
<td>T</td>
<td>A</td>
<td>MC</td>
<td>c.692T</td>
<td>p.(Ile231Lys)</td>
<td>0</td>
<td>23.6</td>
<td></td>
</tr>
<tr>
<td>SXS48</td>
<td>Breast cancer</td>
<td>20</td>
<td>594820</td>
<td>A</td>
<td>G</td>
<td>MC</td>
<td>c.994A</td>
<td>p.(Thr332Ala)</td>
<td>0.00000398</td>
<td>23.3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MCM8 monoallelic carriers</th>
<th>Patient</th>
<th>Cohort</th>
<th>Chr</th>
<th>Pos</th>
<th>R</th>
<th>A</th>
<th>Gen</th>
<th>DNA</th>
<th>Protein</th>
<th>GnomAD</th>
<th>CAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>DDPOL6</td>
<td>Colonic polyposis</td>
<td>20</td>
<td>596676</td>
<td>A</td>
<td>G</td>
<td>MC</td>
<td>c.2149</td>
<td>p.(Ile717Val)</td>
<td>0.00757</td>
<td>22.9</td>
<td></td>
</tr>
<tr>
<td>DDPOL6</td>
<td>Colonic polyposis</td>
<td>20</td>
<td>593582</td>
<td>A</td>
<td>G</td>
<td>MC</td>
<td>c.414A</td>
<td>p.(Ile138Met)</td>
<td>0.0062</td>
<td>12.5</td>
<td></td>
</tr>
<tr>
<td>DDPOL6</td>
<td>Colonic polyposis</td>
<td>20</td>
<td>596557</td>
<td>A</td>
<td>G</td>
<td>MC</td>
<td>c.1886</td>
<td>p.(Asn629Ser)</td>
<td>0.0000398</td>
<td>0.017</td>
<td></td>
</tr>
<tr>
<td>MSS2-19</td>
<td>MMR-proficient</td>
<td>20</td>
<td>594396</td>
<td>C</td>
<td>T</td>
<td>MC</td>
<td>c.832C</td>
<td>p.(Arg278Cys)</td>
<td>0.000374</td>
<td>16.74</td>
<td></td>
</tr>
<tr>
<td>MSS23-1</td>
<td>MMR-proficient</td>
<td>20</td>
<td>596797</td>
<td>G</td>
<td>A</td>
<td>MC</td>
<td>c.2209</td>
<td>p.( Ala737Thr)</td>
<td>0.000768</td>
<td>29.1</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MCM9 biallelic carriers</th>
<th>Patient</th>
<th>Cohort</th>
<th>Chr</th>
<th>Pos</th>
<th>R</th>
<th>A</th>
<th>Gen</th>
<th>DNA</th>
<th>Protein</th>
<th>GnomAD</th>
<th>CAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>011-6929</td>
<td>Lynch-like syndrome</td>
<td>6</td>
<td>119149</td>
<td>G</td>
<td>A</td>
<td>MC</td>
<td>c.1642</td>
<td>p.(Arg548Trp)</td>
<td>0.00000665</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>011-6929</td>
<td>Lynch-like syndrome</td>
<td>6</td>
<td>119252</td>
<td>T</td>
<td>A</td>
<td>MC</td>
<td>c.152A</td>
<td>p.(Asn51Ile)</td>
<td>0.0000199</td>
<td>28.2</td>
<td></td>
</tr>
<tr>
<td>MSS13-1</td>
<td>MMR-proficient</td>
<td>6</td>
<td>119135</td>
<td>T</td>
<td>C</td>
<td>MC</td>
<td>c.3425</td>
<td>p.(Lys1142Arg)</td>
<td>0.000234</td>
<td>23.1</td>
<td></td>
</tr>
<tr>
<td>MSS13-1</td>
<td>familial CRC</td>
<td>6</td>
<td>119149</td>
<td>A</td>
<td>G</td>
<td>MC</td>
<td>c.1640</td>
<td>p.(Leu547Pro)</td>
<td>0.0000532</td>
<td>24.5</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MCM9 monoallelic carriers</th>
<th>Patient</th>
<th>Cohort</th>
<th>Chr</th>
<th>Pos</th>
<th>R</th>
<th>A</th>
<th>Gen</th>
<th>DNA</th>
<th>Protein</th>
<th>GnomAD</th>
<th>CAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>030-5374</td>
<td>Lynch-like syndrome</td>
<td>6</td>
<td>119136</td>
<td>T</td>
<td>C</td>
<td>MC</td>
<td>c.3286</td>
<td>p.(Met1096Val)</td>
<td>0.00239</td>
<td>0.078</td>
<td></td>
</tr>
<tr>
<td>015-6177</td>
<td>Lynch-like syndrome</td>
<td>6</td>
<td>119137</td>
<td>C</td>
<td>A</td>
<td>MC</td>
<td>c.1974</td>
<td>p.(Gln658His)</td>
<td>0.00801</td>
<td>22.2</td>
<td></td>
</tr>
<tr>
<td>F69245</td>
<td>CRC family</td>
<td>6</td>
<td>119137</td>
<td>C</td>
<td>A</td>
<td>MC</td>
<td>c.1974</td>
<td>p.(Gln658His)</td>
<td>0.00801</td>
<td>22.2</td>
<td></td>
</tr>
<tr>
<td>024-5894</td>
<td>Lynch-like syndrome</td>
<td>6</td>
<td>119136</td>
<td>C</td>
<td>G</td>
<td>MC</td>
<td>c.3034</td>
<td>p.(Glu1012Gln)</td>
<td>0.000107</td>
<td>11.96</td>
<td></td>
</tr>
<tr>
<td>025-5029</td>
<td>Lynch-like syndrome</td>
<td>6</td>
<td>119150</td>
<td>T</td>
<td>A</td>
<td>MC</td>
<td>c.1521</td>
<td>p.(Glu507Asp)</td>
<td>0.00992</td>
<td>24.2</td>
<td></td>
</tr>
<tr>
<td>018-6429</td>
<td>Lynch-like syndrome</td>
<td>6</td>
<td>119137</td>
<td>T</td>
<td>A</td>
<td>MC</td>
<td>c.2144</td>
<td>p.(Asp715Val)</td>
<td>0.00137</td>
<td>11.8</td>
<td></td>
</tr>
<tr>
<td>MSS2-19</td>
<td>MMR-proficient</td>
<td>6</td>
<td>119135</td>
<td>T</td>
<td>C</td>
<td>MC</td>
<td>c.3425</td>
<td>p.(Lys1142Arg)</td>
<td>0.000234</td>
<td>23.1</td>
<td></td>
</tr>
<tr>
<td>NA41-1</td>
<td>MMR-proficient</td>
<td>6</td>
<td>119137</td>
<td>-</td>
<td>T</td>
<td>MC</td>
<td>c.1987</td>
<td>p.(Ser663Phe)</td>
<td>0</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>NA96-14</td>
<td>MMR-proficient</td>
<td>6</td>
<td>119147</td>
<td>G</td>
<td>C</td>
<td>MC</td>
<td>c.1915</td>
<td>p.(Leu639Val)</td>
<td>0.000625</td>
<td>25.4</td>
<td></td>
</tr>
<tr>
<td>MSS6-19</td>
<td>MMR-proficient</td>
<td>6</td>
<td>119234</td>
<td>T</td>
<td>C</td>
<td>MC</td>
<td>c.911A</td>
<td>p.(Asn304Ser)</td>
<td>0.00357</td>
<td>26.1</td>
<td></td>
</tr>
<tr>
<td>NA209-1</td>
<td>MMR-proficient</td>
<td>6</td>
<td>119137</td>
<td>T</td>
<td>A</td>
<td>MC</td>
<td>c.2144</td>
<td>p.(Asp715Val)</td>
<td>0.00137</td>
<td>11.8</td>
<td></td>
</tr>
</tbody>
</table>