Identification of a novel pathogenic MLH1 mutation and recommended genetic screening strategy: An investigation of three Chinese Lynch syndrome pedigrees

Fan Li | Yunwei Xia | Guoguang Wang | Chaoyang Tang | Tian Zhan | Jian Shen | Jianping Zhang

Department of General Surgery, The Second Affiliated Hospital, Nanjing Medical University, Nanjing, China

Correspondence
Jianping Zhang and Jian Shen, Department of General Surgery, The Second Affiliated Hospital, Nanjing Medical University, 121 Jiangjiayuan Road, Nanjing 210011, China. Email: zjpnnjmu@hotmail.com (J. Z.); drshenjian@njmu.edu.cn (J. S.)

Funding information
This work was supported by the Natural Science Foundation of China (81874058).

Abstract

Background: Lynch syndrome (LS) is an autosomal-dominant disorder that increases the risk of many cancers. The genetic basis of LS is germline mutations in DNA mismatch repair genes.

Methods: We performed next-generation sequencing on blood cells obtained from the members of three unrelated LS pedigrees. Immunohistochemistry staining was performed to analyze protein expression.

Results: Multigene panel screening revealed three mutL homolog 1 (MLH1) pathogenic mutations (c.199G>A, c.790 + 1G>A, and c.1557_1558 + 8delGGGTACG-TAA, unreported) confirmed by Sanger sequencing. Immunohistochemistry showed a loss of MLH1 protein expression. We also confirmed that the unreported mutant allele was inherited for at least three generations.

Conclusion: These results provide new insights into the molecular mechanisms underlying the pathogenicity of MLH1 mutations and reaffirm the importance of genetic screening for the early diagnosis of LS.

KEYWORDS
Lynch syndrome, mutL protein homolog 1, next-generation sequencing

1 | INTRODUCTION

Lynch syndrome (LS), an autosomal-dominant inherited disorder, is the most frequent cause of hereditary colorectal cancer (CRC), accounting for 1%–3% of CRC cases (Bonadona et al., 2011; Hampel et al., 2005). The clinical features of families with LS include an earlier average onset age for cancer, multiple primary cancers, increased lifetime risk of CRC, and increased risk of extracolonic epithelial malignancies (Cohen & Leininger, 2014). Currently, in the absence of LS-specific symptoms, studies that identify consistent molecular markers for early diagnosis and prognosis are urgently needed.

LS is caused by the pathogenic mutant alleles of the human mismatch repair (MMR) gene mutL homolog 1 (MLH1, OMIM*120, 436), MSH2, MSH6, PMS2, and EPCAM, which trigger genomic instability and thereby lead to various cancers (Cini et al., 2015). MMR proteins are involved in repairing incorrect bases that are inserted during DNA replication. Nonsense, missense, frameshift, and splicing variants, as well as deletions of one or more exons, have been identified in MMR alleles. Indeed, some exonic and intronic variants create and/or disrupt splice sites, leading to aberrantly spliced mRNAs (Tamura et al., 2019). Therefore, it remains critical to identify new alleles to elucidate the...
complete genetic architecture of CRC to more fully understand its etiology (Peters, Bien, & Zubair, 2015).

Traditionally, high-risk patients with LS have been identified through clinical diagnostic criteria (Amsterdam criteria II and revised Bethesda guidelines; Laghi, Bianchi, Roncalli, & Malesci, 2004; Vasen, Watson, Mecklin, & Lynch, 1999) and universal tumor screening, including immunohistochemistry (IHC) and microsatellite instability (MSI; Giardiello et al., 2014). MSI is a hallmark of cancers associated with LS; it is detected by the size fractionation of several mononucleotide repeat sequences (Tomiali et al., 2014). IHC testing of tumor tissues to detect loss of MMR gene expression is used to select genes for testing. However, current clinical diagnostic criteria have significant drawbacks for the diagnosis of LS. Consequently, molecular diagnosis with next-generation sequencing (NGS) as the primary detection method is currently offered to families to compensate for the above limitations. This may become the most effective and accurate method for the diagnosis of LS (Kyrochristos & Roukos, 2019).

Here, we report the identification and characterization of three different MLH1 mutations in three unrelated Chinese families with LS, including a missense mutation (c.199G>A) and two splice site mutations (c.790 + 1G>A and c.1557_1558 + 8delGGGTACGTAA, unreported). In addition, we recommend a screening strategy suitable for the Han Chinese population (Giardiello et al., 2014).

2 MATERIALS AND METHODS

2.1 Ethical compliance

All procedures performed in this study involving human participants were in accordance with the ethical standards of the Medical Ethics Committee of Nanjing Medical University and the Declaration of Helsinki and its later amendments or comparable ethical standards (2014). Written informed consent was obtained from all individual participants included in the study.

2.2 Patients and pedigrees

Three probands (proband 1: generation III, No. 7; proband 2: generation III, No. 3; proband 3: generation IV, No. 11) were diagnosed with CRC and treated at The Second Affiliated Hospital of Nanjing Medical University. Three-generation pedigree 1 with 11 members, four-generation pedigree 2 with 10 members, and six-generation pedigree 3 with 10 members were diagnosed with cancer and enrolled in the study. The diagnostic criteria for patients with LS were based on the Amsterdam II criteria.

2.3 Immunohistochemistry

MMR protein IHC was performed on formalin-fixed and paraffin-embedded sections dewaxed in xylene, dehydrated in ethanol, boiled in 0.01 M citrate buffer (pH 6.0) for 20 min in a microwave oven, and incubated with 3% hydrogen peroxide for 5 min. After washing with PBS, the sections were incubated in 10% normal bovine serum albumin for 5 min, followed by incubation with two different types of rabbit anti-MLH1 (1:50, Abcam and MBX) antibody at 4°C overnight. The slides were incubated with anti-rabbit horseradish peroxidase-conjugated secondary antibody (1:300, Beyotime Co. Ltd) at room temperature for an additional 30 min. Staining was visualized using diaminobenzidine. Sections were counterstained with hematoxylin, dehydrated, cleared, mounted, and photographed using a panoramic-scan digital slice scanning system (3DHISTECH Co. Ltd). Quantitation of immunostaining was performed by two independent researchers who were blinded to patient details.

2.4 NGS-based clinical cancer gene test

NGS with a multigene panel of germline variants in 26 cancer predisposition genes, including ATM, BARD1, BLM, BRCA1, BRCA2, BRI1, CDH1, CHEK2, EPCAM, FAM175A, MEN1, MLH1, MRE11a, MSH2, MSH6, MUTYH, NBN, PALB2, PMS2, PTEN, RAD50, RAD51C, RAD51D, STK11, TP53, and XRCC2, was performed by Simcere Co. Ltd. Briefly, tumor and genomic DNA were isolated from tumor specimens and peripheral blood, respectively. NGS was performed on an Illumina HiSeq-2500 platform (Illumina). Data were analyzed with the 1,000 Genomes browser (https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes), ClinVar (https://www.ncbi.nlm.nih.gov/clinvar), and ExAC (http://exac.broadinstitute.org). The pathogenicity of the mutations was classified according to the recommendations of the American College of Medical Genetics and Genomics (ACMG; http://www.acmg.net) and InSiGHT variant databases (https://www.insight-group.org/variants/databases).

2.5 Sanger sequencing

To validate mutations identified by NGS, Sanger sequencing was performed by Simcere Co. Ltd. Primers were designed based on the reference sequence of the human genome downloaded from GenBank and synthesized by Invitrogen. Polymerase chain reaction (PCR) was amplified with an ABI 9700 Thermal Cycler, and the products were sequenced on an ABI PRISM 3730 automated sequencer (Applied
Biosystems). The reference sequence NM_000249.3 (*MLH1*) was used to report aberrant transcripts.

### 2.6 | In silico prediction

All three *MLH1* variants were examined through in silico splicing prediction using Alamut® Visual version 2.12.0 (Interactive Biosoftware), which included multiple prediction algorithms.

### 2.7 | Structure prediction

The amino acid sequences of the *MLH1* protein (GenBank accession number NP000240.1) were obtained from the GenBank database. The homology modeling program, Swiss-Model (http://swissmodel.expasy.org), was used to create a model of the structure of the mutated region.

### 3 | RESULTS

#### 3.1 | Clinical findings in three pedigrees

According to the sequencing results, variants classified as pathogenic in ClinVar were evaluated for sequencing depth and visually inspected using the Integrative Genomic Viewer. After filtering strategies followed by Sanger validation, three *MLH1* variants on 26 genes were detected in the probands of three unrelated Chinese pedigrees that involved c.199G>A, c.790+1G>A, and c.1557_1558 + 8delGGGTACGTAA mutations (Table 1; Figure 1).

| Pedigree | cDNA level | gDNA | Protein level | Type | Coding effect |
|----------|------------|------|---------------|------|--------------|
| 1        | c.199G>A   | g.36996701 G>A | p.(Gly67Arg)  | Substitution | Missense     |
| 2        | c.790+1G>A | g.37014545 G>A | p.(Glu227_Ser295del) | Substitution | Splicing variant |
| 3        | c.1557_1558 + 8del | g.37028931_37028940 del | p.(Glu519Aspfs*8) | Deletion | Splicing variant |

**TABLE 1** Information of mutations in three pedigrees and in silico prediction results

**FIGURE 1** Examination of the mutation in three pedigrees. First-generation sequencing (Sanger) of the blood sample of the probands result showing at the position
3.2 | Potential mechanisms underlying the loss of MLH1 expression

We then performed in silico predictions of the potential effects of the mutations on MLH1 protein structure. As shown in Figure 2a, the variant is denoted (c.199G>A) at the cDNA level and results in a G67R substitution (GGG>AGG). This should result in the partial loss of the C-terminal portion of the α-helix (Thompson et al., 2013). As shown in Figure 2b, the second variant (c.790 + 1G>A) causes the 9th and 10th exons, codons 227–295, to be skipped during mRNA splicing, leading to faulty functional domain formation of the protein (Auclair et al., 2006). Similarly, the third variant (c.1557_1558 + 8delGGTACGTAA), located at an exon-intron boundary of MLH1 exon 13, likely creates a frameshift starting at residue Glu519, with the new reading frame ending in a stop codon at position 526, as per the prediction of the Alamut software suite. Therefore, nonsense-mediated decay (NMD) might be involved in degrading mRNA (Hentze & Kulozik, 1999; Sjursen, McPhillips, Scott, & Talseth-Palmer, 2016).

3.3 | Identification of variant functions at the protein level

Furthermore, we examined the effect of the MLH1 mutation on protein expression through IHC. As shown in Figure 3, the sporadic CRC sample was used as a positive control and showed positive nuclear staining in tumor cells, whereas IHC results for two reported variants (c.199G>A and c.790 + 1G>A) showed MLH1 nuclear expression loss in tumor and stromal cells. Interestingly, the expression of mutant MLH1 (c.1557_1558 + 8delGGTACGTAA) protein was absent, which was consistent with the sequencing results. Therefore, we hypothesized that all three variants result in a loss of MLH1 expression.
3.4 | Significance of the novel MLH1 mutation site

The three families harboring these mutations all have a strong history of cancer. Considering the families’ clinical characteristics and Amsterdam criteria II, which have been widely applied to aid the diagnosis of LS, the three families were diagnosed with LS. As shown in Figure 4a, in pedigree 1, 11 members in three consecutive generations suffered malignancies and in pedigree 2, 10 members in four consecutive generations suffered malignancies. Next, we focused on the novel mutation genetics in pedigree 3. As shown in Figure 4b, of the five-generation Chinese pedigree with 31 members, six were affected by CRC and two were carriers—the proband (IV-11) and the proband’s nephew (V-2). The son of the proband’s nephew (VI-1), a 9-year-old male without cancer, also carried this mutant allele, demonstrating inheritance over three generations.

4 | DISCUSSION

In this study, through in silico prediction, we identified three germline variants, including the novel mutation c.1557_1558 + 8delGGGTACGTAA, which may lead to a frameshift mutation. The other two mutations were c.199G>A and c.790 + 1G>A, which have previously been reported in InSiGHT variant databases (https://www.insight-group.org/variants/databases) as pathogenic variants. Our IHC data are in concordance with molecular data, supporting the conclusion that protein loss can be explained by pathogenic mutations.

The first variant is denoted as MLH1 c.199G>A at the cDNA level and p.Gly67Arg (G67R) at the protein level, and it results in the change of a glycine to an arginine (GGG>AGG). This variant has been reported in many individuals with LS, isolated from LS-associated cancers, and found to be absent from healthy controls. The sequence change of c.790 + 1G>A affects a donor splice site in intron 9 of the MLH1. It is expected to disrupt mRNA splicing and result in an absent or disrupted protein product. Studies utilizing patient-derived RNA have shown that this variant causes the significant skipping of exons 9 and 10 (Auclair et al., 2006). This alternate splicing results in the loss of amino acids 227–295 in the MLH1 protein, which has been shown functionally to render MLH1 defective in MMR activity. For these reasons, this variant was classified as pathogenic.

The novel variant (c.1557_1558 + 8delGGGTACGTAA) is located in the donor splice site of exon 13 in MLH1, which may lead to the formation of a premature stop codon in in
silico analysis. However, such aberrantly spliced transcripts may increase the probability that NMD is involved in degrading aberrant transcripts. Consequently, the novel mutations that may cause aberrant transcripts are likely to form truncated proteins and thus not be functional (Lynch et al., 2009).

Numerous pathogenic \textit{MLH1} mutations have been reported to be associated with LS in multiple countries and ethnic groups. The human \textit{MLH1} is composed of 19 exons and the cDNA is predicted to encode a protein of 756 amino acid residues. Several variants are linked with ethnicity. Three mutations (c.1453G \textgreater C, c.1742C \textgreater T, and c.1758dupC) in \textit{MLH1} have only been found in Southeast Asian populations, and it is suggested that specific mutations in this group require greater attention for \textit{MLH1} genetic screening. Asia is the most populous continent, and the number of studies involving \textit{MLH1} deleterious mutations in Asian countries is increasing (Jia et al., 2018; Momma et al., 2019; Pandey & Shrestha, 2018). According to various studies, some clinical differences between western and eastern countries exist, including age at CRC diagnosis, differentiation state of the tumors, lifetime cancer risk, and frequent primary extracolonic tumors in the stomach.

**FIGURE 4** Pedigree structure of the three Chinese families. (a) Three-generation pedigree exhibiting the c.199G>A mutation and 11 members diagnosed with cancers. Four-generation pedigree exhibiting the c.790 + 1G>A mutation and 10 members diagnosed as cancers. (b) Five generations with 31 members, of which six of them were diagnosed with CRC; the novel mutation is a three-generation heritable mutation. BC, breast cancer; EC, esophageal cancer; GBM, glioblastoma multiforme; OC, ovarian cancer; PC, pancreatic cancer; SC, skin cancer.
(Liu et al., 2014). Furthermore, the use of universal clinical criteria to diagnose LS in China likely underestimates the number of cases due to the small family sizes, which are a result of the national one-child policy.

Therefore, more attention should be paid to the public and medical education of LS (Biller, Syngal, & Yurgelun, 2019; Giardiello et al., 2014). We propose that if patients fulfill any of the following conditions, they should be regarded as being at a high risk for LS: (a) presence of CRC; (b) more than two metachronous CRCs; (c) simultaneous or heterochronous tumors in other regions (stomach, small intestine, female reproductive system, urinary system, or hepatobiliary system); (d) more than two cases of CRC or ovarian cancer in the pedigree; and (e) failure to detect MMR protein by IHC in patients who are CRC-negative. Further gene sequencing should be performed to identify the mutation site and guide the family survey. We further recommend the following germline mutation testing for high-risk family members: (a) for the proband, NGS to screen for LS mutations, informed by pathological and IHC results; (b) NGS to screen other family members; and (c) providing risk management for at-risk relatives and control of the disease (Figure 5).

In conclusion, we identified three germline variants in three unrelated pedigrees from China, including a novel mutation, $MLH1$ c.1557_1558 + 8delGGGTACGTAA, which was inherited for three consecutive generations. Based on our clinical experience, we have also proposed screening and risk management strategies for the Han Chinese population.

**CONFLICT OF INTEREST**

The authors declare that they have no competing interests.

**AUTHOR CONTRIBUTIONS**

F. Li and Y. Xia analyzed the data and wrote the manuscript. G. Wang and C. Tang provided the samples. T. Zhan and F. Li performed genomic DNA extraction and sequencing. J. Zhang and J. Shen conceived the research. All the authors have read and approved the final version of the manuscript.

**DATA AVAILABILITY STATEMENT**

Not applicable.

**ORCID**

Jianping Zhang https://orcid.org/0000-0002-1929-792X

**REFERENCES**

Auclair, J., Busine, M. P., Navarro, C., Ruano, E., Montmain, G., Deseigne, F., … Wang, Q. (2006). Systematic mRNA analysis for
the effect of MLH1 and MSH2 nonsense mutations on aberrant splicing. *Human Mutation*, 27(2), 145–154. https://doi.org/10.1002/humu.20280

Biller, L. H., Syngal, S., & Yurgelun, M. B. (2019). Recent advances in Lynch syndrome. *Familial Cancer*, 18(2), 211–219. https://doi.org/10.1007/s10689-018-0017-1

Bonadona, V., Bonaiti, B., Olschwang, S., Grandjouan, S., Huiart, L., Longy, M., … Bonaiti-Pellie, C. (2011). Cancer risks associated with germline mutations in MLH1, MSH2, and MSH6 genes in Lynch syndrome. *JAMA*, 305(22), 2304–2310. https://doi.org/10.1001/jama.2011.743

Cini, G., Carnevali, I., Quaia, M., Chiaravalli, A. M., Sala, P., Giacomini, E., … Viel, A. (2015). Concomitant mutation and epimutation of the MLH1 gene in a Lynch syndrome family. *Carcinogenesis*, 36(4), 452–458. https://doi.org/10.1093/carcin/bgv015

Cohen, S. A., & Leinninger, A. (2014). The genetic basis of Lynch syndrome and its implications for clinical practice and risk management. *The Application of Clinical Genetics*, 7, 147–158. https://doi.org/10.2147/hacg.s51483

Dinjens, W. N., Dubbink, H. J., & Wagner, A. (2015). Guidelines on genetic evaluation and management of Lynch syndrome. *American Journal of Gastroenterology*, 110(1), 192–193. https://doi.org/10.1038/ajg.2014.329

Giardello, F. M., Allen, J. I., Axilbund, J. E., Boland, R. C., Burke, C. A., Burt, R. W., … Rex, D. K. (2014). Guidelines on genetic evaluation and management of Lynch syndrome: A consensus statement by the US Multi-society Task Force on colorectal cancer. *American Journal of Gastroenterology*, 109(8), 1159–1179. https://doi.org/10.1038/ajg.2014.186

Hampel, H., Frankel, W. L., Martin, E., Arnold, M., Kanduja, K., Kuebler, P., … de la Chapelle, A. (2005). Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). *New England Journal of Medicine*, 352(18), 1851–1860. https://doi.org/10.1056/nejmoa043146

Hentze, M. W., & Kulozik, A. E. (1999). A perfect message: RNA surveillance and nonsense-mediated decay. *Cell*, 96(3), 307–310. https://doi.org/10.1016/s0092-8674(00)80542-5

Jia, S., Zhang, M., Sun, Y., Yan, H., Zhao, F., Li, Z., & Ji, J. (2018). A Chinese family affected by Lynch syndrome caused by MLH1 mutation. *BMC Medical Genetics*, 19(1), 106. https://doi.org/10.1186/s12881-018-0605-x

Kyrochristos, I. D., & Roukos, D. H. (2019). Comprehensive intra-individual genomic and transcriptional heterogeneity: Evidence-based Colorectal Cancer Precision Medicine. *Cancer Treatment Reviews*, 80, 101894. https://doi.org/10.1016/j.ctrv.2019.101894

Laghi, L., Bianchi, P., Roncalli, M., & Malesci, A. (2004). Re: Revised Bethesda guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *JNCI Journal of the National Cancer Institute*, 96(18), 1402–1403; author reply 1403-1404. https://doi.org/10.1093/jnci/djh280

Liu, F., Yang, L. I., Zhou, X., Sheng, W., Cai, S., Liu, L., … Xu, Y. E. (2014). Clinicopathological and genetic features of Chinese hereditary nonpolyposis colorectal cancer (HNPPC). *Medical Oncology*, 31(10), 223. https://doi.org/10.1007/s12032-014-0223-1

Lynch, H. T., Lynch, P. M., Lanspa, S. J., Snyder, C. L., Lynch, J. F., & Boland, C. R. (2009). Review of the Lynch syndrome: History, molecular genetics, screening, differential diagnosis, and medico-legal ramifications. *Clinical Genetics*, 76(1), 1–18. https://doi.org/10.1111/j.1399-0004.2009.01230.x

Momma, T., Gonda, K., Akama, Y., Endo, E., Ujiie, D., Fujita, S., … Nomizu, T. (2019). MLH1 germline mutation associated with Lynch syndrome in a family followed for more than 45 years. *BMC Medical Genetics*, 20(1), 67. https://doi.org/10.1186/s12881-019-0792-0

Pandey, A. S., & Shrestha, S. (2018). A novel frameshift mutation in the MLH1 gene in a patient with Lynch syndrome. *Indian Journal of Cancer*, 55(4), 410–412. https://doi.org/10.4103/ijc.IJC_349_18

Peters, U., Bien, S., & Zubair, N. (2015). Genetic architecture of colorectal cancer. *Gut*, 64(10), 1623–1636. https://doi.org/10.1136/gutjnl-2013-306705

Sjursen, W., McPhillips, M., Scott, R. J., & Talseth-Palmer, B. A. (2016). Lynch syndrome mutation spectrum in New South Wales, Australia, including 55 novel mutations. *Molecular Genetics & Genomic Medicine*, 4(2), 223–231. https://doi.org/10.1002/mgg3.198

Tamura, K., Kaneda, M., Futagawa, M., Takeshita, M., Kim, S., Nakama, M., … Tatsumi-Miyajima, J. (2019). Genetic and genomic basis of the mismatch repair system involved in Lynch syndrome. *International Journal of Clinical Oncology*, 24(9), 999–1011. https://doi.org/10.1016/j.ijco.2019.04.048

Thompson, B. A., Goldgar, D. E., Paterson, C., Clendenning, M., Walters, R., Arnold, S., … Spurdle, A. B. (2013). A multifactorial likelihood model for MMR gene variant classification incorporating probabilities based on sequence bioinformatics and tumor characteristics: A report from the Colon Cancer Family Registry. *Human Mutation*, 34(1), 200–209. https://doi.org/10.1002/humu.22213

Tomiak, E., Samson, A., Spector, N., Mackey, M., Gilpin, C., Smith, E., … Asmis, T. (2014). Reflex testing for Lynch syndrome: If we build it, will they come? Lessons learned from the uptake of clinical genetics services by individuals with newly diagnosed colorectal cancer (CRC). *Familial Cancer*, 13(1), 75–82. https://doi.org/10.1007/s10068-013-9677-0

Vasen, H. F., Watson, P., Mecklin, J. P., & Lynch, H. T. (1999). New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPPC, Lynch syndrome) proposed by the International Collaborative group on HNPPC. *Gastroenterology*, 116(6), 1453–1456. https://doi.org/10.1016/S0016-5085(99)70510-X

**How to cite this article:** Li F, Xia Y, Wang G, et al. Identification of a novel pathogenic MLH1 mutation and recommended genetic screening strategy: An investigation of three Chinese Lynch syndrome pedigrees. *Mol Genet Genomic Med*. 2020;8:e1295. https://doi.org/10.1002/mgg3.1295