ARTICLE





Development of an MSI-positive colon tumor with aberrant DNA methylation in a PPAP patient

Kiyoshi Yamaguchi ¹ · Eigo Shimizu² · Rui Yamaguchi² · Seiya Imoto ³ · Mitsuhiro Komura² · Seira Hatakeyama¹ · Rei Noguchi¹ · Kiyoko Takane¹ · Tsuneo Ikenoue¹ · Yoshimasa Gohda⁴ · Hideaki Yano⁴ · Satoru Miyano^{2,5} · Yoichi Furukawa ¹

Received: 24 October 2018 / Revised: 18 March 2019 / Accepted: 19 April 2019 / Published online: 14 May 2019 © The Author(s), under exclusive licence to The Japan Society of Human Genetics 2019

Abstract

Polymerase proofreading-associated polyposis (PPAP) is a disease caused by germline variations in the *POLE* and *POLD1* genes that encode catalytic subunits of DNA polymerases. Studies of cancer genomes have identified somatic mutations in these genes, suggesting the importance of polymerase proofreading of DNA replication in suppressing tumorigenesis. Here, we identified a germline frameshift variation in the *POLE* gene (c.4191_4192delCT, p.Tyr1398*) in a case with multiple adenomatous polyps and three synchronous colon cancers. Interestingly, one of the colon cancers showed microsatellite instability-high (MSI-H) and another microsatellite stable. Immunohistochemical staining revealed that the MSI-H tumor cells lost the expression of MLH1 protein. Whole genome sequencing of the MSI-H tumor did not find pathogenic somatic mutations in mismatch repair genes but found frameshift mutations in the *TET* genes that catalyze 5-methylcytosine hydroxylation. Bisulfite sequencing of the tumor corroborated an increase in the number of hypermethylated regions including the *MLH1* promoter. These data indicate that PPAP patients might develop MSI-positive tumors through epigenetic silencing of *MLH1*. These findings will contribute to comprehensive understanding of the molecular basis of tumors that involve deficiency of proofreading activity of DNA polymerases.

These authors contributed equally: Kiyoshi Yamaguchi and Eigo Shimizu

Supplementary information The online version of this article (https://doi.org/10.1038/s10038-019-0611-7) contains supplementary material, which is available to authorized users.

Yoichi Furukawa furukawa@ims.u-tokyo.ac.jp

- ¹ Division of Clinical Genome Research, Advanced Clinical Research Center, Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan
- ² Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan
- ³ Division of Health Medical Data Science, Health Intelligence Center, Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan
- ⁴ Department of Surgery, National Center for Global Health and Medicine, Tokyo 162-8655, Japan
- ⁵ Division of Health Medical Computational Science, Health Intelligence Center, Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan

Introduction

A recent study of familial polyposis patients without germline variations in the tumor suppressors APC or MUTYH identified germline variations in genes associated with proofreading of DNA replication, POLE (p.Leu424-Val, OMIM ID: 174762), and POLD1 (p.Ser478Asn, OMIM ID: 174761), and defined a new disease entity polymerase proofreading-associated polyposis termed (PPAP, OMIM ID: 615083) [1]. Although the great majority of pathogenic variations in POLE or POLD1 are missense variations within or close to their exonuclease domain [1-3], there is one reported case of *POLE* variations outside of the exonuclease domain [4]. Patients with PPAP are predisposed to various cancers including colorectal cancer (CRC). Endometrial and breast cancer seem to be selectively found in the patients with POLD1 variants [2]. The clinical phenotype of colorectal tumors in PPAP patients is similar to that observed in MUTYH-associated polyposis, attenuated familial polyposis, or Lynch syndrome. Notably, tumors in PPAP patients accumulate a large number of somatic mutations because POLE and POLD1 have exonuclease proofreading activity that plays a vital role in replication fidelity by the recognition and excision of mispaired bases [3]. Unlike tumors in Lynch syndrome, tumors in PPAP patients are often microsatellite stable (MSS), and some of them exhibit chromosomal instability [1, 5]. However, a recent report demonstrated that CRCs in three *POLE* (p.Leu424Val) variant carriers from two families were microsatellite instability-high (MSI-H), and one tumor displayed a hypermutator phenotype [6]. Thus, it is a matter of controversy whether tumors carrying germline *POLE* or *POLD1* variant exhibit MSS or MSI.

Next generation sequencing has enabled us to analyze the human and cancer genomes comprehensively and facilitated the identification of pathogenic variants for predisposition to diseases and driver changes for the development and progression of neoplasms. International collaborations in the ICGC and TCGA projects revealed that CRC shows a various spectrum and different frequencies of somatic mutations [7]. It is well known that defects in either mismatch repair (MMR) or replicative repair result in the increase of somatic mutations and render genetic instability which is involved in colorectal tumorigenesis. Spontaneous mutagenesis caused by MMR deficiency has been extensively studied in bacteria and yeast. For example, the absence of MSH2 or MSH6 increased the reversion rates of some mutations by up to 6,0000-fold in yeast [8]. Development of several types of tumors has been observed in mouse models of defective MMR [9]. In addition, pathogenic germline variants of MMR have been found in families with Lynch syndrome (also known as hereditary nonpolyposis CRC), the most common hereditary CRC syndrome. Moreover, recent studies of cancer genomes and PPAP patients have underscored the importance of DNA replication errors in human carcinogenesis. Tomasetti et al. [10] investigated the correlation between stem cell divisions and cancer incidence, and estimated that errors in DNA replication were responsible for two-thirds of the mutations in human cancers. According to the mutational analysis using mutant Pol ε and Pol δ alleles in yeast, Pol ε is responsible for leading-strand DNA replication, whereas Pol δ for lagging strand synthesis [11, 12]. Unlike Pol α , these two polymerases have intrinsic proofreading exonuclease activity and contribute to accurate replication [13]. Indeed, spontaneous mutation rates of cells lacking proofreading activity of Pol ε were more than 70 times higher at some genetic loci compared with the wild-type cells [14]. Inactivation of the proofreading activity of DNA polymerases accumulates somatic mutations in the cells and leads to the development of tumors. In agreement with these views, mice with a Pole variant tended to develop intestinal adenomas and adenocarcinomas, histiocytic sarcomas, and non-thymic lymphomas [14], and mice with a *Pold1* variant preferentially developed lymphomas, skin, and lung tumors [14, 15]. These data suggest that proofreading activity of Pol ε and Pol δ might play a tissue-dependent role in carcinogenesis [14].

Whole genome sequencing (WGS) has enabled us to comprehensively analyze the human genome, improving the chance of identifying disease-causative variants. In the present study, we identified a germline frameshift variation outside the exonuclease domain of the POLE gene in a case with multiple adenomatous polyps and synchronous CRCs using WGS. Furthermore, comprehensive genetic, methylome, and transcriptome analyses of the two synchronous tumors revealed that the one in ascending colon showed MSI-H, hypermethylation of the MLH1 promoter, BRAF (p.Val600Glu) mutation and hypermutator phenotype, and the other in sigmoid colon showed MSS, a nonsense mutation in APC (p.Glu1309*) and a missense mutation in KRAS (p.Gly12Asp). Our data indicated that deleterious variation was not restricted within the exonuclease domain of the POLE gene, and that carriers of a POLE variant might develop MSI-positive tumors through the deregulation of DNA methylation.

Materials and methods

Clinical information

A 68-year-old woman patient suffered from multiple colorectal tumors including two advanced cancers in the ascending and sigmoid colon, a relatively early cancer in the transverse colon, and multiple polyps underwent surgical operation in Department of Surgery, National Center for Global Health and Medicine, Tokyo. Family history of the patient showed that her brother (II-4) and niece (III-10) suffered from CRC, and another brother and a sister from gastric cancer (II-1) and pancreatic cancer (II-2), respectively (Fig. 1a). Histological examination of the ascending colon cancer of 8.2×7.0 cm in size disclosed that tumor was mucinous adenocarcinoma with invasion to subserosal layer (ss), and without lymphatic (ly0) or vessel (v0) invasion, or the involvement of regional lymph nodes (n0). The sigmoid colon cancer was 1.7×1.4 cm in size, was diagnosed as a tubular adenocarcinoma with moderately differentiation (tub2), and had invasion to proper muscle (mp) with moderate lymphatic invasion (ly2), no vessel invasion (v0), or the involvement of lymph nodes (n0). The transverse colon cancer was 1.0×1.0 cm in size, was diagnosed as a tubular adenocarcinoma with moderately differentiation (tub2), and had invasion to submucosal layer (sm) with severe lymphatic invasion (ly3), slight vessel invasion (v0), and no involvement of lymph nodes (n0).

Fig. 1 Family history of the patient and a pathogenic germline variant of POLE. a Pedigree of the patient. The proband is indicated by an arrow. Males are illustrated by squares, and females by circles. Unaffected and affected individuals are represented by open and closed symbols, respectively. Current age and histories of malignancy are described in the vicinity of symbols. A diagonal slash through the symbol indicates a deceased person. b A two-base deletion in exon 33 of the POLE gene was confirmed by direct PCR sequencing (left). Schematic protein structure of POLE (right). The exonuclease and polymerase activity domains reside in the N-terminal half of the protein. c Diagram showing tumor location and types of sequence analysis. B peripheral blood sample, Ta tumor arising from ascending colon, Ts tumor arising from sigmoid colon, N1, N2, and N3 normal colonic mucosa in ascending, transverse, and sigmoid colon, respectively



PCR-direct sequencing

The coding exons in *MSH2*, *MLH1*, and *MSH6* were amplified with M13-tailed target-specific primers, and the PCR products were sequenced on the Applied Biosystems 3730xl DNA Analyzer using the BigDye Direct Cycle Sequencing kit (Thermo Fisher Scientific, Waltham, MA). The primer sequences used for sequencing are available on request.

Mutations in the *POLE* and *TET* genes identified by the WGS were confirmed by PCR-direct sequencing. The primer sequences used for the amplification and sequencing are shown in Supplementary Table 1.

Microsatellite instability analysis

The NCI panel, composed of BAT25, BAT26, D5S346, D2S123, and D17S250 markers, was used for MSI analysis. Genomic DNA (30 ng) extracted from the normal mucosa and CRCs was amplified by PCR using *Ex Taq* DNA polymerase (Takara Bio, Shiga, Japan). The fluorescently labeled PCR products mixed with Hi-Di formamide and

GeneScan 500 LIZ size standard (Thermo Fisher Scientific) were subjected to fragment analysis (3130xl Genetic Analyzer, Thermo Fisher Scientific). Data analysis was performed using the GeneMapper Software 5 (Thermo Fisher Scientific).

Immunohistochemical staining

The expression of MMR proteins in CRCs was evaluated by immunohistochemical staining using an anti-MLH1 (ES05, Leica Biosystems, Wetzlar, Germany), anti-PMS2 (556415, BD Biosciences, San Jose, CA), and anti-MSH2 antibodies (NA27, Merck, Darmstadt, Germany). The sections were deparaffinized with xylene. After antigen retrieval (10 mM sodium citrate buffer for 10 min at 115 °C), nonspecific binding was blocked with goat serum, followed by overnight incubation in the antibodies (dilution: MLH1, 1:100; PMS2, 1:50; MSH2, 1:50) at 4 °C. After washing, the tissue–antibody reaction was visualized using the EnVision+ System (Agilent Technologies, Santa Clara, CA). Hematoxylin was used for nuclear counterstaining.

Whole genome sequencing

Genomic DNA was extracted from peripheral blood and two synchronous CRCs in the patient according to the standard phenol extraction/purification procedure. Sequencing was performed with paired-end reads of 126 bp on an Illumina HiSeq 2500 platform using the manufacturer's instructions (Illumina, San Diego, CA). A sequence library containing inserts of 250-350 bp was prepared using 200 ng of genomic DNA and TruSeq DNA Sample Prep kit (Illumina). The fastq files were aligned to human reference sequence (hg19) using Burrows-Wheeler Aligner (ver. 0.5.10) and bam files were created for data processing. Genomon (ver. 2.5), an in-house pipeline constructed at the Human Genome Center, The University of Tokyo, was used for the detection of single-nucleotide variants and short insertions/deletions (https://github.com/Genomon-Project). For identifying somatic mutations, EBcall (Empirical Bayesian mutation calling) algorism was used [16]. This algorism discriminates somatic mutations from sequencing errors based on an empirical Bayesian method. In addition, a copy number analysis was performed using WGS data. Abnormal copy number regions were detected using the circular binary segmentation algorithm with the R package DNAcopy.

Bisulfite sequencing

After fragmentation of DNA using *Eco*RI, 4 µg of the DNA was denatured in 0.3 M NaOH, and treated with 3.6 M sodium bisulfite and 10 mM hydroquinone at 55 °C for 16 h. Desulfonation was accomplished by the treatment with NaOH at the final concentration of 0.3 M. The converted DNA was amplified using *Ex taq* (Takara Bio) and primers (sense, 5'-GAGTAGTTTTTTTTTAGGAGTGAA-3' and antisense, 5'-AACTCCTCCTCTCCCCTTAC-3'). The PCR products were cloned into pCR2.1 using a topoisomerase TA Cloning Kit (Thermo Fisher Scientific) and transformed into DH10B competent cells (Thermo Fisher Scientific). Colonies were selected by blue/white color selection, and purified individual plasmids were sequenced by the Sanger method.

For the methylation analysis by whole genome bisulfite sequencing (WGBS), DNA bisulfite conversion from two normal colon tissues and two synchronous CRCs samples was performed using the EZ DNA Methylation-Lightning kit (Zymo, Irvine, CA). Libraries were prepared using TrueSeq DNA Methylation kit (Illumina), and were subsequently sequenced with paired-end reads of 126 bp on a HiSeq 2500. Global and unbiased DNA methylation was determined from uniquely mappable reads aligned to hg19 using Bismark [17]. To characterize differentially methylated regions (DMRs) between normal and tumor tissues, we employed the Metilene, a binary segmentation algorithm combined with a two dimensional statistical test [18]. Statistical significance was determined by false discovery rate (FDR) *q*-value (q < 0.05). In addition, hypermethylated or hypomethylated DMRs were defined as DNA methylation density with more than 20% difference in average, compared with those in the matched colonic mucosa.

Transcriptome analysis

Total RNA was extracted from three normal colonic mucosa and two CRCs (Ta and Ts) of the patient using the RNeasy Mini Kit (Qiagen, Hilden, Germany). The percentages of RNA fragments larger than 200 nucleotides (DV_{200}) were calculated using the Agilent 2100 Bioanalyzer (Agilent Technologies). RNA sequencing libraries were prepared using TruSeq RNA Access Library Prep kit according to the manufacturer's protocol (Illumina). Sequencing was performed with paired-end reads of 101 bp on an Illumina HiSeq 2500 platform. Differentially expressed genes between the normal and CRCs were identified using DESeq2. The biological significance of the expression data were assessed by Gene Set Enrichment Analysis (GSEA). Gene sets with FDR *q*-value < 0.05 were considered significant.

Results

Identification of a germline *POLE* variant in a patient with multiple CRC and family history of cancer

We first suspected Lynch syndrome because of multiple colon cancer with the patient's family history. Pathogenic variants were not identified by the screening of three genes associated with MMR (MLH1, MSH2, and MSH6). Therefore, we performed WGS and searched for deleterious variants. In the WGS data, an average coverage depth of nearly 43.3 × was achieved, and a total of 4,679,569 variants were detected. Consistent with the data of the initial screening, we detected non-pathogenic variants of MLH1, MSH2, and MSH6 (Table 1). No pathogenic variant was identified in PMS2. Furthermore, copy number analysis revealed no copy or structural alterations in the forementioned genes or the EPCAM gene (data not shown) [19]. Hence, we explored other variants associated with familial colon cancer such as APC (familial adenomatous polyposis), MUTYH (MUTYH-associated polyposis), POLE and POLD1 (PPAP), MSH3 (MSH3-associated polyposis), SMAD4 and BMPR1A (juvenile polyposis syndrome), STK11 (Peutz-Jeghers syndrome), PTEN (Cowden syndrome), NTHL1 (NTHL1-associated polyposis), and RNF43

 Table 1
 Variants of genes

 responsible for hereditary CRC
 syndromes that were detected by

 whole genome sequencing
 by

Gene	Туре	Mutation	Protein alteration	RefSeq accession number	dbSNP131
MUTYH	Nonsynonymous	c.1005G>C	p.Q335H	NM_012222	rs3219489
MUTYH	Nonsynonymous	c.74 G>A	p.G25D	NM_012222	rs75321043
MUTYH	Nonsynonymous	c.53C>T	p.P18L	NM_012222	rs79777494
MSH6	Nonsynonymous	c.116G>A	p.G39E	NM_000179	rs1042821
MLH1	Nonsynonymous	c.655A>G	p.I219V	NM_000249	rs1799977
MLH1	Nonsynonymous	c.1151T>A	p.V384D	NM_000249	rs63750447
MSH3	Nonsynonymous	c.235A>G	p.I79V	NM_002439	rs1650697
MSH3	Nonsynonymous	c.2846A>G	p.Q949R	NM_002439	rs184967
MSH3	Nonsynonymous	c.3133G>A	p.A1045T	NM_002439	rs26279
APC	Nonsynonymous	c.5465T>A	p.V1822D	NM_000038	rs459552
PMS2	Nonsynonymous	c.1621A>G	p.K541E	NM_000535	rs2228006
PMS2	Nonsynonymous	c.1454C>A	p.T485K	NM_000535	rs1805323
PMS2	Nonsynonymous	c.1408C>T	p.P470S	NM_000535	rs1805321
BMPR1A	Nonsynonymous	c.4C>A	p.P2T	NM_004329	rs11528010
PTEN	Nonsynonymous	c.10G>A	p.G4R	NM_001304717	rs12573787
PTEN	Nonsynonymous	c.194G>C	p.C65S	NM_001304717	rs2943772
POLE	Frameshift deletion	c.4191_4192del	p.Y1398*	NM_006231	-

Variants that do not change the amino acid sequence such as intronic and synonymous variants were excluded

(serrated polyposis syndrome). Consequently, we identified a two-base deletion in exon 33 of the *POLE* gene (c.4191_4192del) and confirmed the deletion by Sanger sequencing (Fig. 1b).

To confirm the variant in DNA and transcripts, we performed deep sequencing of both DNA and transcripts in the patient's leukocytes. As expected, we found that the *POLE* variant (c.4191_4192del) showed 50.4% (1,001 reads) of the total reads (1,987 reads) in the DNA. However, variant reads (827 reads) were detected in 41.9% of total reads (1,976 reads) in the transcripts. Since the frameshift variant results in a premature termination codon, the transcripts carrying the deletion are assumed to be degraded through nonsense-mediated RNA decay.

Microsatellite instability and loss of MLH1 expression

In parallel with the search for germline variants, MSI analysis of the ascending (Ta), and sigmoid (Ts) colon cancers was performed. Interestingly, all Bethesda's five markers were positive in Ta but were negative in Ts (Fig. 2a), indicating that Ta and Ts were MSI-H and MSS, respectively. MSI-H has been found in over 90% of CRC from Lynch syndrome and 10–15% of sporadic cases [20–22]. Typically, absence of DNA MMR proteins were found in these cases [23]. As expected, immunohistochemical staining of MMR proteins showed the absence of MLH1 expression in conjunction with PMS2 loss in Ta (Fig. 2b).

These patterns of loss have been often described elsewhere [24, 25]. These data suggested that the two CRCs developed through different molecular mechanisms, even though the impairment of polymerase proofreading was commonly involved in their tumorigenesis. This view prompted us to investigate genomic, epigenetic, and expression changes in the two CRCs.

Mutation profiling of the two advanced cancers (Ta and Ts)

WGS of the Ta and Ts detected a total of 488,915 and 16,687 mutations, respectively. Mutation rate of Ta was 81.5 per 10^6 bases, whereas that of Ts was 2.8 per 10^6 bases, indicating that Ta could be classified as a hypermutated tumor [7]. In addition, an increase in the number of short insertion/deletion (indel) was observed in Ta because of the MSI-H state (Fig. 3a). We further found that Ta was accompanied by BRAF (p.Val600Glu) mutation, as is often observed in MSI-H tumors [26]. Ta also carried somatic mutations in TP53 (c.764T>G, p.Ile255Ser), RNF43 (c.1976delG, p.Gly659fs), and EGFR (c.1859G>A, p.Cys620Tyr). On the other hand, Ts had deleterious mutations in APC (c.3925G>T, p.Glu1309*), KRAS (c.35G>A, p.Gly12Asp), and PIK3CA (c.241G>A p.Glu81Lys). Regarding the second hit in the POLE gene, no genetic alterations including SNV or copy number changes were found in the gene locus in Ta and Ts (Supplementary Fig. 1). Since Ta showed an MSI-H phenotype,



Fig. 2 Loss of MLH1 and PMS2 in Ta. a MSI analysis of CRCs (Ta and Ts) using the Bethesda panel of microsatellite markers.

b Immunohistochemical staining of mismatch repair proteins, MLH1, MSH2, and PMS2 in Ta and Ts

Fig. 3 Whole genome analysis of Ta and Ts. **a** The bar plots show the number (left panel) and proportion (right panel) of somatic mutations in Ta and Ts. The number of short insertion/ deletion (indel) and singlenucleotide variants (SNV) in protein-coding regions and splice sites were counted. **b** Mutation spectra in whole genome sequence data in Ta and Ts



we additionally searched for somatic mutations in *MLH1*, *MSH2*, *MSH6*, and *PMS2*; however, no pathogenic mutations were found in the four genes.

As shown in Fig. 3b, mainly C:G>T:A transitions occurred in both Ta and Ts, due to the under defective

proofreading activity [1, 27]. In addition, increased nucleotide change of A:T>G:C could be observed in Ta, which is one of the dominant signatures of MSI state [27]. To compare the mutation burden and signature of Ta and Ts with the exome sequencing data of CRC in TCGA database

[7], we first extracted mutation data of entire exonic regions from the WGS data by computational approach. Regarding the frequency of mutations, Ta was classified as a "hypermutated" tumor, but Ts as a "non-hypermutated" tumor (Supplementary Fig. 2). To analyze mutational signature, we extracted datasets of 17 CRCs with POLE mutation from the TCGA, and compared their mutational signature with that of Ta and Ts. This signature analysis has uncovered high frequency of "Signature 10" that is associated with POLE somatic mutations and "Signature 6" that is associated with defective DNA MMR (Supplementary Fig. 3a). Subsequent similarity analysis largely divided the tumors into two groups: one showing strong similarity to Signature 10 (POLE mutation) and weak similarity to Signature 6, 15, and 20 (defective DNA MMR), and the other exhibited strong similarity to Signature 6, 15, and 20 and weak similarity to Signature 10 (Supplementary Fig. 3b). It is likely that tumors with strong similarity to Signature 10 exhibit ultra-hypermutated (more than 100 mutations per Mb) and tumors with strong similarity to Signature 6, 15, and 20 exhibit hypermutated (10-100 mutations per Mb) or non-hypermutated. Intriguingly, we found that hypermethylation of MLH1 promoter or mutations in MMR genes might not impact on mutational signatures in colorectal tumors with POLE mutation.

Hypermethylation of the *MLH1* promoter and mutations in the *TET* genes

Since no mutation in MLH1 was identified in Ta, we analyzed methylation status of the MLH1 promoter. Bisulfite sequencing of the A region, which has been most commonly tested [26, 28], revealed hypermethylation of the promoter in Ta, but not in Ts (Fig. 4a). To investigate the mechanisms of aberrant methylation of the MLH1 promoter, we searched for mutations in genes associated with DNA methylation including DNA methyltransferases (DNMTs) that are essential for establishment (DNMT3a and DNMT3b) and maintenance (DNMT1) of DNA methylation patterns. Although WGS of Ta detected 26 mutations in the intronic and intergenic regions of DNMT1, all were variants of uncertain significance (VUS, data not shown). RNA-seq analysis did not show any difference in the expression of DNMT1 between normal colonic mucosa (N) and Ta (log2 ratio Ta/N: -0.67). We further analyzed the other DNA methyltransferases and ten eleven translocation (TET) enzymes, which convert 5-methylcytosine to 5hydroxymethylcytosine and promote locus-specific reversal of DNA methylation [29]. As a result, a total of 157 mutations in DNMTs and TETs were detected in Ta, and one mutation in DNMT3B in Ts. Although the mutation located downstream of the DNMT3B gene was VUS, the 157 mutations include frameshift mutations in TET1 and TET3, and nonsynonymous mutations in *TET1*, *TET2*, and *TET3* (Table 2). These mutations were confirmed by the direct PCR method (Supplementary Fig. 4).

Global features of DNA methylation in the synchronous CRCs

To obtain an unbiased insight into genome-wide changes in DNA methylation, we performed WGBS of the two cancers and matched normal colon tissues (Fig. 1c). This analysis provided a $44.7 \times -77.5 \times$ average depth of coverage across the entire genome and a $26.8 \times -52.2 \times$ average depth in CpG sites. In agreement with the previous report [30], normal colon tissues had a CpG methylation level of ~70% on average, whereas the two CRCs showed a substantial hypomethylation in this context (Ta: -5.9%and Ts: -6.3%) (Supplementary Table 2). In contrast, the majority of CHGs or CHHs was poorly methylated (normal: 1.2-2.0%, tumor: 6.3-6.8%), and average CHG or CHH methylation was higher in the CRCs compared with that observed in normal tissues (CHG, Ta: +4.7% and Ts: +4.8%/CHH, Ta: +5.1% and Ts: +5.6%, Supplementary Table 2), suggesting the different roles of CpG and non-CpG methylation in epigenetic regulation and tumorigenesis.

Since the location in the genome defines the impact of DNA methylation on transcriptional activity [31], we classified CpGs into three groups based on their genomic location: promoters, gene bodies, and intergenic regions. DNA methylation in the region of the first exon is more tightly linked to transcriptional silencing than that in the upstream promoter region [31]. Considering that the majority of the first exons in mammalians is shorter than 1 kb [32], we defined the promoter regions as 1 kb upstream of the transcription start site (TSS) to 1 kb downstream of TSS, and gene bodies as 1 kb downstream of TSS to the end of genes. The levels of global CpG methylation within the promoters were low, but significantly higher in Ta and Ts than in the normal tissues (Tukey's HSD, p < 0.001) (Fig. 4b). Since Ta carries somatic mutations in TET1 and TET3 and hypermethylation in MLH1, we interrogated whether Ta showed different patterns of global DNA methylation from Ts. As expected, Ta harbored a significantly increased number of methylated CpG within the promoters compared with Ts (Tukey's HSD, p < 0.001). However, methylation levels within the gene bodies or intergenic regions showed marginal difference between Ta and Ts.

To characterize different epigenetic states between normal and cancerous tissues, we determined DMRs in the entire genome using the Metilene software [18]. Each region with more than a 20% difference on average was defined as a hypermethylated DMR (hyper-DMR)



(Supplementary Data 1 and 2) or hypomethylated DMR (hypo-DMR) (Supplementary Data 3 and 4) when comparing the tumors to the normal tissues. As a result, we

corroborated that the *MLH1* promoter was in the list of Taspecific hyper-DMRs (FDR q-value = 2.0×10^{-44}) (Supplementary Data 1 and Fig. 4c). In addition, the number of ✓ Fig. 4 Genomic features of differentially methylated regions (DMRs) in Ta and Ts. a Methylation status of the CpG dinucleotides in the A region of the MLH1 promoter. The ratios of methylated CpGs in Ta and Ts are shown based on sequences collected from 19 and 16 colonies, respectively. Methylated CpG (●); unmethylated CpG (○). **b** Box plots of the levels of methylated CpG in each genomic region. c A snapshot of the Integrative Genomics Viewer, showing hypermethylation of the MLH1 promoter in Ta. d Venn diagrams showing the number of overlapping hypermethylated DMRs (Hyper-DMRs) or hypomethylated DMRs (Hypo-DMRs) in the genomic regions including promoters (from 1 kb upstream to 1 kb downstream of TSS), gene bodies (entire transcribed regions without 1 kb downstream of TSS), intergenic, and super enhancers (within the intergenic). Statistical significance was determined by FDR *q*-value (q < 0.05). DMRs were defined as DNA methylation density with more than a 20% difference in average, compared with those in the matched normal colonic mucosa. e Heatmap of Hyper-DMRs in promoters by WGBS concurrently with decrease in differentially expressed genes (DEGs) by RNA-seq analysis. Hyper-DMRs were defined as higher density of DNA methylation with >20% in average compared with those in the matched normal colonic mucosa. For the identification of DEGs, genes whose expression was $\log 2 < -1$ were used. f Pathways associated with the promoter hypermethylation and decreased expression were analyzed by GSEA with the KEGG pathway gene set collection. Statistical significance was determined by FDR q-value (q < 0.05). The chart shows the top ten significant gene sets of Ta (upper panel) and Ts (lower panel). ARVC arrhythmogenic right ventricular cardiomyopathy

Table 2 Variants of the TET genes in the assending CRC (Ta)

Gene	Туре	Mutation	Protein alteration
TET1	Frameshift deletion	c.106delA	p.K36fs
TET1	Nonsynonymous SNV	c.3856G>A	p.V1286I
TET2	Nonsynonymous SNV	c.1817C>T	p.T606I
TET3	Frameshift deletion	c.205_206del	p.C69fs
TET3	Nonsynonymous SNV	c.2188G>A	p.V730M

hyper-DMRs was approximately two-times greater in Ta than Ts (Fig. 4d), which may result from the decrease of 5mC hydroxylase activity by the mutations in *TET* family genes.

We further investigated alteration in DNA methylation levels within the super enhancers by integrating the methylation and super enhancer data (Super-Enhancer Archive, http://sea.edbc.org). However, the levels of DNA methylation within the super enhancers appear not to be affected much during tumorigenesis (Fig. 4d).

Identification of DMRs actively contributing to gene regulation

Unbiased genome-wide analyses have shown an inverse correlation between gene expression and DNA methylation especially in the regions near TSS [33, 34]. The mechanism by which hypomethylated regions are associated with the accessibility of the transcriptional machinery including

transcription factors is the underlying cause for this phenomenon. To identify DMRs actively contributing to gene regulation, we performed RNA sequencing using the synchronous CRCs and three adjacent normal colonic tissues (Fig. 1c), and subsequently integrated the transcriptome data into the DMRs data. As a result, 182 and 143 genes, which show hyper-DMRs within their promoters and a decreased expression (log2 T/N < -1), were identified in Ta and Ts, respectively (Fig. 4e and Supplementary Data 5). The 182 genes with Ta-specific hyper-DMRs/decreased expression included the MLH1 gene (DMR: +74.1% and expression (log2 Ta/N): -2.64) (Fig. 4e). Interestingly, the promoter region of phosphatase and tensin homolog (PTEN) gene was hypermethylated and its expression was downregulated in Ta, suggesting that activation of the PI3K/ Akt pathway by the suppression of PTEN may play a crucial role in tumor development (Supplementary Data 5).

To elucidate possible biological significance of the gene expressions affected by hypermethylation, we screened key pathways in Ta and Ts by the GSEA. As shown in Fig. 4f, the pathway analysis identified 23 and 4 gene sets associated with Ta and Ts, respectively. Although two pathways, "Calcium signaling pathway" and "Neuroactive ligand-receptor interaction", were overlapped between Ta and Ts, the remaining 21 deregulated pathways in Ta included "Cell adhesion molecules", "ECM-receptor interaction", and "Pathways in cancer" (Supplementary Data 6). These results suggest that DNA methylation events show active participation in gene expression, and that the aberrant DNA methylation contributes to tumor heterogeneity.

Discussion

Germline variants of POLE and POLD1 have been shown to predispose to colorectal adenomas and carcinomas. These missense variants seem to be concentrated within or adjacent to the exonuclease domains that are essential for proofreading activity [1-3]. In this study, we identified a novel POLE germline variant that locates outside the exonuclease domain in a case lacking pathogenic variants in known high penetrance familial CRC susceptibility genes. Considering that variants are not limited to known 'hot spots', sequencing of the entire coding regions of POLE and POLD1 is required for a better characterization of the syndrome. On the other hand, POLE is a large gene with a protein-coding region of ~7 kb, and could thus acquire a large number of mutations in cancer. However, many mutations in *POLE* were mere passengers [35]. We believe that these variants should be carefully interpreted.

Large-scale genomic studies have disclosed that most tumors with ultra-hypermutated state (>100 mutations/Mb) and a part of tumors with hypermutated state (10–100

mutations/Mb) harbor a combination of MMR deficiency and polymerase proofreading deficiency [7, 35]. However, this case carrying a POLE germline variant synchronously developed hypermutated and non-hypermutated tumors, suggesting that this variant may not solely cause an increase in mutation frequency. To our knowledge, there is one case that carried germline frameshift variation in POLE (c.5621 5622delGT), leading to the development of earlyonset CRC (diagnosed at 26 years old) [4]. However, the report included no information about tumor phenotype. The mutator effect of pathogenic variants of vPol2 (veast POLE) was recently examined using haploid yeast strains [36]. Surprisingly, substitutions (Pro286Arg, Ser459Phe, Phe367Ser, Pro286His, Pro436Arg, and Leu424Val) in the exonuclease domain of POLE increased the frequency of mutation more than a variant (Exo⁻: yPol2-Asp290Ala, Glu292Ala) that completely eliminates exonuclease activity. Furthermore, another study showed that the effect of variants in the exonuclease domain on the mutation frequency was not correlated with their effects on proofreading activity [37]. These results provide the possibility that loss-of-function mutation in POLE contributes to tumorigenesis through mechanism(s) other than loss of exonuclease activity.

It is estimated that synchronous CRC accounts for ~1.1–8.1% of all diagnosed cases of CRCs [38]. Although strong predictive parameters for developing synchronous CRC have not been fully identified yet, accumulated evidence indicated that some of those patients are associated with inflammatory bowel diseases [39–41] or familial colon cancer syndromes such as FAP [39, 40] and Lynch syndrome [41]. A recent study implied that a high proportion of germline *POLE* variants (Leu424Val or Pro436Ser) carriers had multiple synchronous or metachronous carcinomas [42]. Taken together with our findings, patients with PPAP have high risk of the synchronous cancer.

Approximately 15% of CRCs show MSI due to hypermethylation of the MLH1 promoter [43]. MLH1 hypermethylation is associated with tumorigenesis as well as resistance to chemotherapeutic agents such as Fluorouracil [44, 45]. A DNA methyltransferase DNMT1 leads to de novo methylation of cytosine residues in the MLH1 promoter under certain conditions [46]; however, the mechanism underlying hypermethylation remains ambiguous. By comparing the whole genome data from the synchronous CRCs with (Ta) and without (Ts) MLH1 hypermethylation, we identified that Ta exclusively had the frameshift mutations in TET1 and TET3 (Table 2). Both the frameshift mutations resulted in premature stop codons, leading to loss of the C-terminal catalytic domains (Cys-rich and double-stranded β helix) of TET1 and TET3. Although somatic alterations in TET2 are frequently observed in a wide range of hematological diseases including myeloid and lymphoid malignancies [29], a recent study using

TET-knockout cells revealed that TET1 preferentially catalyzed oxidization of 5mC in human embryonic stem cells rather than TET2 [47]. It seems that TET1 acts synergistically with TET2 and TET3. They also demonstrated that loss of the TET genes resulted in locus-specific hypermethylation rather than a global gain of methylation [47]. Therefore, Ta-specific hyper-DMRs including the MLH1 promoter might, in part, rely on the loss of the TET1 and TET3 genes. On the other hand, we cannot exclude the possibility that local environmental conditions in the large intestine might affect DNA methylation. Increasing evidence suggests that bacteria can alter the chromatin structure and transcriptional program of host cells by influencing diverse epigenetic factors including DNA methylation [48]. A well-known example is Helicobacter pylori infection that induces aberrant DNA methylation in the gastric mucosa. Indeed, hypermethylation of the MLH1 promoter and reduced expression of MLH1 was observed in the gastric cancer cell line co-cultured with H. pylori [49]. Furthermore, Fusobacterium-enriched CRCs were associated with CIMP-positive, TP53 wild type, and hypermethylation of the *MLH1* promoter [50]. Understanding of the role of gut microbiota in epigenetic modulation will help the elucidation of precise molecular mechanisms underlying CRC.

Our comprehensive analysis of Ta and Ts suggested that PI3K/AKT signaling pathway was deregulated through different mechanisms, namely hypermethylation/reduced expression of *PTEN* (Ta) or pathogenic mutation in *PIK3CA* (Ts). Inhibitors targeting PI3K could be an effective therapy for the patient.

Our data disclosed that PPAP-associated CRCs have a variable phenotype. Ta, one of the two tumors, showed MSI-H and loss of MLH1 expression by hypermethylation of the *MLH1* promoter, and carried mutations in *BRAF*, *TP53*, *RNF43*, *EGFR*, and the genes encoding TET family members. On the other hand, Ts was MSS and had mutations in *APC*, *KRAS*, and *PIK3CA*. WGBS demonstrated significantly higher promoter methylation in Ta than Ts, suggesting that somatic *TET* mutations may have rendered new properties to Ta through the change of global methylation. These data will be helpful for the understanding of molecular basis of tumors that involve deficiency of proofreading activity of DNA polymerases.

Acknowledgements We thank Rika Koubo and Nozomi Yusa (The University of Tokyo) for their technical assistance. This work was supported in part by the Grant-in-Aid #16H01569 (YF) from the Japan Society for the Promotion of Science, and Center of Innovation program (SM) from the Japan Science and Technology Agency.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethics This project was approved by the ethical committee of Institute of Medical Science, the University of Tokyo (IMSUT-IRB: 23-18-0929, 28-27-1017, 26-67-1118, and 27-37-1015). All experiments were carried out in accordance with the approved guidelines. Written informed consent was obtained from the patient in this study.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Palles C, Cazier JB, Howarth KM, Domingo E, Jones AM, Broderick P, et al. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. Nat Genet. 2013;45:136–44.
- Bellido F, Pineda M, Aiza G, Valdes-Mas R, Navarro M, Puente DA, et al. POLE and POLD1 mutations in 529 kindred with familial colorectal cancer and/or polyposis: review of reported cases and recommendations for genetic testing and surveillance. Genet Med. 2016;18:325–32.
- Rayner E, van Gool IC, Palles C, Kearsey SE, Bosse T, Tomlinson I, et al. A panoply of errors: polymerase proofreading domain mutations in cancer. Nat Rev Cancer. 2016;16:71–81.
- Smith CG, Naven M, Harris R, Colley J, West H, Li N, et al. Exome resequencing identifies potential tumor-suppressor genes that predispose to colorectal cancer. Hum Mutat. 2013;34: 1026–34.
- Seshagiri S. The burden of faulty proofreading in colon cancer. Nat Genet. 2013;45:121–2.
- Elsayed FA, Kets CM, Ruano D, van den Akker B, Mensenkamp AR, Schrumpf M, et al. Germline variants in POLE are associated with early onset mismatch repair deficient colorectal cancer. Eur J Hum Genet. 2015;23:1080–4.
- 7. TCGA. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012;487:330–7.
- Earley MC, Crouse GF. The role of mismatch repair in the prevention of base pair mutations in *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA. 1998;95:15487–91.
- Reitmair AH, Redston M, Cai JC, Chuang TC, Bjerknes M, Cheng H, et al. Spontaneous intestinal carcinomas and skin neoplasms in Msh2-deficient mice. Cancer Res. 1996;56:3842–9.
- Tomasetti C, Li L, Vogelstein B. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. Science. 2017; 355:1330–4.
- Miyabe I, Kunkel TA, Carr AM. The major roles of DNA polymerases epsilon and delta at the eukaryotic replication fork are evolutionarily conserved. PLoS Genet. 2011;7:e1002407.
- Pursell ZF, Isoz I, Lundstrom EB, Johansson E, Kunkel TA. Yeast DNA polymerase epsilon participates in leading-strand DNA replication. Science. 2007;317:127–30.
- Ganai RA, Johansson E. DNA replication-a matter of fidelity. Mol Cell. 2016;62:745–55.
- Albertson TM, Ogawa M, Bugni JM, Hays LE, Chen Y, Wang Y, et al. DNA polymerase epsilon and delta proofreading suppress discrete mutator and cancer phenotypes in mice. Proc Natl Acad Sci USA. 2009;106:17101–4.
- Goldsby RE, Hays LE, Chen X, Olmsted EA, Slayton WB, Spangrude GJ, et al. High incidence of epithelial cancers in mice deficient for DNA polymerase delta proofreading. Proc Natl Acad Sci USA. 2002;99:15560–5.
- Shiraishi Y, Sato Y, Chiba K, Okuno Y, Nagata Y, Yoshida K, et al. An empirical Bayesian framework for somatic mutation detection from cancer genome sequencing data. Nucleic Acids Res. 2013;41:e89.

- Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011; 27:1571–2.
- Juhling F, Kretzmer H, Bernhart SH, Otto C, Stadler PF, Hoffmann S. metilene: fast and sensitive calling of differentially methylated regions from bisulfite sequencing data. Genome Res. 2016;26:256–62.
- Ligtenberg MJ, Kuiper RP, Chan TL, Goossens M, Hebeda KM, Voorendt M, et al. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. Nat Genet. 2009;41:112–7.
- Aaltonen LA, Peltomaki P, Leach FS, Sistonen P, Pylkkanen L, Mecklin JP, et al. Clues to the pathogenesis of familial colorectal cancer. Science. 1993;260:812–6.
- Lawes DA, SenGupta S, Boulos PB. The clinical importance and prognostic implications of microsatellite instability in sporadic cancer. Eur J Surg Oncol. 2003;29:201–12.
- Liu B, Parsons R, Papadopoulos N, Nicolaides NC, Lynch HT, Watson P, et al. Analysis of mismatch repair genes in hereditary non-polyposis colorectal cancer patients. Nat Med. 1996;2: 169–74.
- Lindor NM, Burgart LJ, Leontovich O, Goldberg RM, Cunningham JM, Sargent DJ, et al. Immunohistochemistry versus microsatellite instability testing in phenotyping colorectal tumors. J Clin Oncol. 2002;20:1043–8.
- Leung WK, Kim JJ, Wu L, Sepulveda JL, Sepulveda AR. Identification of a second MutL DNA mismatch repair complex (hPMS1 and hMLH1) in human epithelial cells. J Biol Chem. 2000;275:15728–32.
- Ruszkiewicz A, Bennett G, Moore J, Manavis J, Rudzki B, Shen L, et al. Correlation of mismatch repair genes immunohistochemistry and microsatellite instability status in HNPCCassociated tumours. Pathology. 2002;34:541–7.
- 26. Parsons MT, Buchanan DD, Thompson B, Young JP, Spurdle AB. Correlation of tumour BRAF mutations and MLH1 methylation with germline mismatch repair (MMR) gene mutation status: a literature review assessing utility of tumour features for MMR variant classification. J Med Genet. 2012;49:151–7.
- 27. Heitzer E, Tomlinson I. Replicative DNA polymerase mutations in cancer. Curr Opin Genet Dev. 2014;24:107–13.
- Deng G, Chen A, Hong J, Chae HS, Kim YS. Methylation of CpG in a small region of the hMLH1 promoter invariably correlates with the absence of gene expression. Cancer Res. 1999;59:2029–33.
- Rasmussen KD, Helin K. Role of TET enzymes in DNA methylation, development, and cancer. Genes Dev. 2016;30: 733–50.
- Vidal E, Sayols S, Moran S, Guillaumet-Adkins A, Schroeder MP, Royo R, et al. A DNA methylation map of human cancer at single base-pair resolution. Oncogene. 2017;36:5648–57.
- Brenet F, Moh M, Funk P, Feierstein E, Viale AJ, Socci ND, et al. DNA methylation of the first exon is tightly linked to transcriptional silencing. PLoS ONE. 2011;6:e14524.
- Bieberstein NI, Carrillo Oesterreich F, Straube K, Neugebauer KM. First exon length controls active chromatin signatures and transcription. Cell Rep. 2012;2:62–8.
- Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, et al. Increased methylation variation in epigenetic domains across cancer types. Nat Genet. 2011;43:768–75.
- Li Y, Zhu J, Tian G, Li N, Li Q, Ye M, et al. The DNA methylome of human peripheral blood mononuclear cells. PLoS Biol. 2010;8:e1000533.
- Campbell BB, Light N, Fabrizio D, Zatzman M, Fuligni F, de Borja R, et al. Comprehensive analysis of hypermutation in human. Cancer Cell. 2017;171:1042–56 e10.

- Barbari SR, Kane DP, Moore EA, Shcherbakova PV. Functional analysis of cancer-associated DNA polymerase epsilon variants in *Saccharomyces cerevisiae*. G3. 2018;8:1019–29.
- 37. Shinbrot E, Henninger EE, Weinhold N, Covington KR, Goksenin AY, Schultz N, et al. Exonuclease mutations in DNA polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication. Genome Res. 2014;24:1740–50.
- Lam AK, Chan SS, Leung M. Synchronous colorectal cancer: clinical, pathological and molecular implications. World J Gastroenterol. 2014;20:6815–20.
- 39. Greenstein AJ, Slater G, Heimann TM, Sachar DB, Aufses AH Jr. A comparison of multiple synchronous colorectal cancer in ulcerative colitis, familial polyposis coli, and de novo cancer. Ann Surg. 1986;203:123–8.
- Lam AK, Carmichael R, Gertraud Buettner P, Gopalan V, Ho YH, Siu S. Clinicopathological significance of synchronous carcinoma in colorectal cancer. Am J Surg. 2011;202:39–44.
- 41. Liu X, Goldblum JR, Zhao Z, Landau M, Heald B, Pai R, et al. Distinct clinicohistologic features of inflammatory bowel diseaseassociated colorectal adenocarcinoma: in comparison with sporadic microsatellite-stable and Lynch syndrome-related colorectal adenocarcinoma. Am J Surg Pathol. 2012;36:1228–33.
- 42. Spier I, Holzapfel S, Altmuller J, Zhao B, Horpaopan S, Vogt S, et al. Frequency and phenotypic spectrum of germline mutations in POLE and seven other polymerase genes in 266 patients with colorectal adenomas and carcinomas. Int J Cancer. 2015;137:320–31.

- Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP. CpG island methylator phenotype in colorectal cancer. Proc Natl Acad Sci USA. 1999;96:8681–6.
- Carethers JM, Chauhan DP, Fink D, Nebel S, Bresalier RS, Howell SB, et al. Mismatch repair proficiency and in vitro response to 5-fluorouracil. Gastroenterology. 1999;117:123–31.
- 45. Plumb JA, Strathdee G, Sludden J, Kaye SB, Brown R. Reversal of drug resistance in human tumor xenografts by 2'-deoxy-5azacytidine-induced demethylation of the hMLH1 gene promoter. Cancer Res. 2000;60:6039–44.
- 46. Jair KW, Bachman KE, Suzuki H, Ting AH, Rhee I, Yen RW, et al. De novo CpG island methylation in human cancer cells. Cancer Res. 2006;66:682–92.
- 47. Verma N, Pan H, Dore LC, Shukla A, Li QV, Pelham-Webb B, et al. TET proteins safeguard bivalent promoters from de novo methylation in human embryonic stem cells. Nat Genet. 2018;50: 83–95.
- Bierne H, Hamon M, Cossart P. Epigenetics and bacterial infections. Cold Spring Harb Perspect Med. 2012;2:a010272.
- Yao Y, Tao H, Park DI, Sepulveda JL, Sepulveda AR. Demonstration and characterization of mutations induced by *Helicobacter pylori* organisms in gastric epithelial cells. Helicobacter. 2006; 11:272–86.
- Tahara T, Yamamoto E, Suzuki H, Maruyama R, Chung W, Garriga J, et al. *Fusobacterium* in colonic flora and molecular features of colorectal carcinoma. Cancer Res. 2014;74:1311–8.