

# SCIENTIFIC REPORTS



OPEN

## Reduced expression of *APC-1B* but not *APC-1A* by the deletion of promoter 1B is responsible for familial adenomatous polyposis

Received: 10 February 2016

Accepted: 22 April 2016

Published: 24 May 2016

Kiyoshi Yamaguchi<sup>1,\*</sup>, Satoshi Nagayama<sup>2,\*</sup>, Eigo Shimizu<sup>3,\*</sup>, Mitsuhiro Komura<sup>3</sup>, Rui Yamaguchi<sup>3</sup>, Tetsuo Shibuya<sup>4</sup>, Masami Arai<sup>5</sup>, Seira Hatakeyama<sup>1</sup>, Tsuneo Ikenoue<sup>1</sup>, Masashi Ueno<sup>2</sup>, Satoru Miyano<sup>3,4,6</sup>, Seiya Imoto<sup>7</sup> & Yoichi Furukawa<sup>1</sup>

Germline mutations in the tumor suppressor gene *APC* are associated with familial adenomatous polyposis (FAP). Here we applied whole-genome sequencing (WGS) to the DNA of a sporadic FAP patient in which we did not find any pathological *APC* mutations by direct sequencing. WGS identified a promoter deletion of approximately 10 kb encompassing promoter 1B and exon 1B of *APC*. Additional allele-specific expression analysis by deep cDNA sequencing revealed that the deletion reduced the expression of the mutated *APC* allele to as low as 11.2% in the total *APC* transcripts, suggesting that the residual mutant transcripts were driven by other promoter(s). Furthermore, cap analysis of gene expression (CAGE) demonstrated that the deleted promoter 1B region is responsible for the great majority of *APC* transcription in many tissues except the brain. The deletion decreased the transcripts of *APC-1B* to 39–45% in the patient compared to the healthy controls, but it did not decrease those of *APC-1A*. Different deletions including promoter 1B have been reported in FAP patients. Taken together, our results strengthen the evidence that analysis of structural variations in promoter 1B should be considered for the FAP patients whose pathological mutations are not identified by conventional direct sequencing.

Familial adenomatous polyposis (FAP) is an autosomal dominant inherited disorder characterized by the development of hundreds to thousands of colorectal adenomas and subsequent progression to adenocarcinoma. FAP patients also have increased risk of extra-colonic tumors, such as desmoids, osteomas, and tumors in the upper gastrointestinal tract, thyroid, and brain<sup>1</sup>. FAP is caused by a germ line mutation in the tumor suppressor adenomatous polyposis coli (*APC*) gene. The *APC* gene is transcribed into multiple mRNAs of various lengths in a tissue-dependent manner, and NCBI Reference Sequence repository includes three *APC* transcript variants; variant1 (NM\_001127511.2), variant2 (NM\_001127510.2), and variant3 (NM\_000038.5). All transcripts are regulated by either promoter 1A or promoter 1B<sup>2</sup>. These two distinct promoters have been suggested to contribute to the tissue-dependent levels of the transcript variants, but which promoter preferentially functions in each tissue still remains controversial<sup>3–5</sup>.

To date, over 3,800 *APC* variants have been reported on the Leiden Open Variation Database (LOVD)<sup>6</sup>. Evidence from accumulated studies indicates that most of the mutations in *APC* are located in the 5'-half of the

<sup>1</sup>Division of Clinical Genome Research, Advanced Clinical Research Center, Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan. <sup>2</sup>Gastroenterology Center, Cancer Institute Hospital, Japanese Foundation for Cancer Research, Tokyo 135-8550, Japan. <sup>3</sup>Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan. <sup>4</sup>Laboratory of Sequence Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan. <sup>5</sup>Department of Clinical Genetic Oncology, Cancer Institute Hospital, Japanese Foundation for Cancer Research, Tokyo 135-8550, Japan. <sup>6</sup>Division of Health Medical Computational Science, Health Intelligence Center, Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan. <sup>7</sup>Division of Health Medical Data Science, Health Intelligence Center, Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to Y.F. (email: furukawa@ims.u-tokyo.ac.jp)

gene, and lead to premature truncated protein products due to nonsense mutations or small insertions/deletions. A study of genotype-phenotype correlation demonstrated that the patients with an *APC* mutation in the codon between 1 and 178 or between 312 and 412 show an attenuated form or milder phenotype with later onset and fewer polyps, than the classical polyposis with thousands of polyps in the colon<sup>7</sup>. Although a large number of *APC* germline mutations have been identified by PCR-direct sequencing, there remain a significant number of undiagnosed cases. In these cases, it is considered either that *APC* carries an alteration that is undetectable by the direct sequencing, or alternatively that a pathological genetic change in other genes is involved in the development of tumors reminiscent of FAP. The former cases include patients carrying somatic mosaicism of the *APC* gene with a very low frequency of the mutant allele<sup>8</sup>, and those with structural changes such as large deletions or duplications of *APC*<sup>9</sup>. Recent studies have demonstrated that large deletions of the transcription regulatory region of *APC* result in reduced *APC* transcripts, and that they are involved in a small subset of FAP<sup>4,10–14</sup>. The latter cases include patients with biallelic germ line mutations in the base excision repair (BER) gene *MUTYH*, an autosomal-recessively inherited disease termed *MUTYH*-associated polyposis (MAP)<sup>15</sup>. In addition, a 40 kb germline duplication upstream of the gremlin 1 (*GREMI*) was identified in patient with hereditary mixed polyposis syndrome<sup>16</sup>. Further studies have uncovered germ line mutations in the proofreading domains of DNA polymerases such as DNA polymerase  $\epsilon$  (*POLE*) and  $\delta$  (*POLD1*) that are associated with adenomatous polyposis and adenocarcinomas in the intestine, which is called polymerase proofreading-associated polyposis (PPAP)<sup>17</sup>. Recently, *NTHL1*, another gene involved in BER, was also shown to be responsible for a recessive type of adenomatous polyposis and carcinoma in the large bowel<sup>18</sup>. It is noteworthy that next-generation sequencing (NGS) has played an important role in the discoveries of new causative gene alterations.

The application of NGS is rapidly expanding from research to clinical practice. We have applied whole-genome sequencing (WGS) for a patient with both multiple adenomatous polyps in the colon and advanced rectal cancer, and without family history of polyposis. This analysis identified a *de novo* deletion in the promoter region of *APC*, which markedly decreased the transcription of the mutant allele. We further disclosed that promoter 1B plays a major role in the expression of *APC*. These data indicate that NGS improves genetic diagnosis of hereditary diseases whose mutations have been overlooked by routine genetic testing, and that analysis of allele specific expression may help the determination of pathological changes in the promoter region of responsible genes.

## Results

**Identification of a large deletion in the promoter region of the *APC* gene.** For the genetic diagnosis of the patient, we screened germ line mutations in the *APC* gene by the conventional PCR-direct sequencing method. Although seven variants including six synonymous (rs2229992, rs351771, rs41115, rs42427, rs866006, and rs465899) and one nonsynonymous (rs459552) variants were discovered in the 5'-half of the coding region where most of the *APC* mutations localize, all the seven variants turned out to be non-pathological SNPs by a search in two public databases, LOVD<sup>6</sup> and ClinVar<sup>19</sup>.

Although multiplex ligation-dependent probe amplification (MLPA) and amplicon sequencing using a panel of disease-associated genes were available for the detection of structural alteration of *APC* and genetic alterations in other responsible genes, respectively, we took advantage of WGS to test its potential in clinical genetic testing. Thus, WGS was performed using DNA extracted from the patient's peripheral blood to explore pathological mutation(s) associated with the disease. An average coverage depth of nearly 30 $\times$  was achieved, and 19,660 variants and 299 short insertion-deletion variants were detected in exons or splice sites. First, we searched for deleterious variants in the genes associated with adenomatous polyposis of the colon (*APC*, *MUTYH*, *NTHL1*, *POLD1*, and *POLE*). In complete agreement with the data obtained by the conventional screening, we confirmed the seven non-pathological mutations by WGS and failed to identify additional mutations or small insertions/deletions in *APC*. Evaluation of the non-synonymous SNVs disclosed that *MUTYH* p.Q338H (rs3219489) is a non-pathological variant, and that *NTHL1* p.L301V (rs200420874) is a variant of uncertain significance (VUS) (Table 1). Although we additionally searched for germ line mutations in other genes that predispose to colorectal polyps such as *SMAD4* and *BMPRIA* (juvenile polyposis syndrome), *STK11* (Peutz-Jeghers syndrome), *GREMI* (hereditary mixed polyposis syndrome), and *PTEN* (Cowden syndrome), all variants in these genes were not deleterious but non-pathological SNPs (Supplementary Table S1).

To determine structural alterations such as large insertions/deletions that account for the disease, we analyzed copy number alteration in genes and the flanking regions associated with adenomatous polyposis. Consequently, a significant copy number loss was found in the *APC* gene (Supplementary Fig. S1a). A visualization of the aligned reads exhibited the 5'-flanking region of *APC* with decreased sequence reads as shown in Fig. 1a, suggesting a large deletion encompassing the regulatory region (known as promoter 1B) and exon1 of the *APC* variant1 (*APC*-1B). Subsequent PCR analysis using primers across the deleted region successfully amplified the mutated allele, and the breakpoint of deletion was determined by Sanger sequencing of the PCR products (Fig. 1b). The size of the deletion was 9,588 bp (chr5:112,035,925–112,045,512; GRCh37/hg19). A homology search revealed that the flanking sequences of the breakpoints shared 77.5% identity across 200 bp (Supplementary Fig. S1b). Interestingly, the breakpoints are located within Alu repeat elements, suggesting that the deletion was likely caused by Alu-mediated recombination. The deletion of promoter 1B was confirmed by MLPA including probes for detecting two *APC* promoters (Supplementary Fig. S2a). Although deletions of promoter 1B have been reported by other groups<sup>4,11–13,20</sup>, this is the first report of a promoter deletion of *APC* in Japanese patients. To examine the inheritance of disease, we investigated the deletion in the first-grade relatives of the patient. PCR analysis using a set of primers across the deletion successfully amplified the genomic region encompassing the deletion in the patient. However, we failed to amplify the region with genomic DNA of the relatives (Fig. 1c, upper panel). In addition, semi-quantitative PCR analysis using a set of primers designed within the deletion showed that the PCR products of the patient were reduced compared with the relatives (Fig. 1c, lower panel). On the other hand, the amount of PCR products with a set of primers outside the deletion was unchanged between the

Gene	Type	Mutation	Protein alteration	dbSNP
APC	synonymous	c.T1458C	p.Y486Y	rs2229992
APC	synonymous	c.G1635A	p.A545A	rs351771
APC	synonymous	c.G4479A	p.T1493T	rs41115
APC	synonymous	c.G5034A	p.G1678G	rs42427
APC	synonymous	c.T5268G	p.S1756S	rs866006
APC	nonsynonymous	c.T5465A	p.V1822D	rs459552
APC	synonymous	c.G5880A	p.P1960P	rs465899
MUTYH	nonsynonymous	c.G1014C	p.Q338H	rs3219489
NTHL1	nonsynonymous	c.C901G	p.L301V	rs200420874
POLE	synonymous	c.G1323A	p.P441P	rs116573514
POLE	synonymous	c.C2935T	p.L979L	rs56081968
POLE	synonymous	c.G3156A	p.T1052T	rs5744857
POLE	synonymous	c.A4530G	p.A1510A	rs5744944
POLE	synonymous	c.C5334T	p.A1778A	rs11146986
POLE	synonymous	c.A6252G	p.S2084S	rs5745022

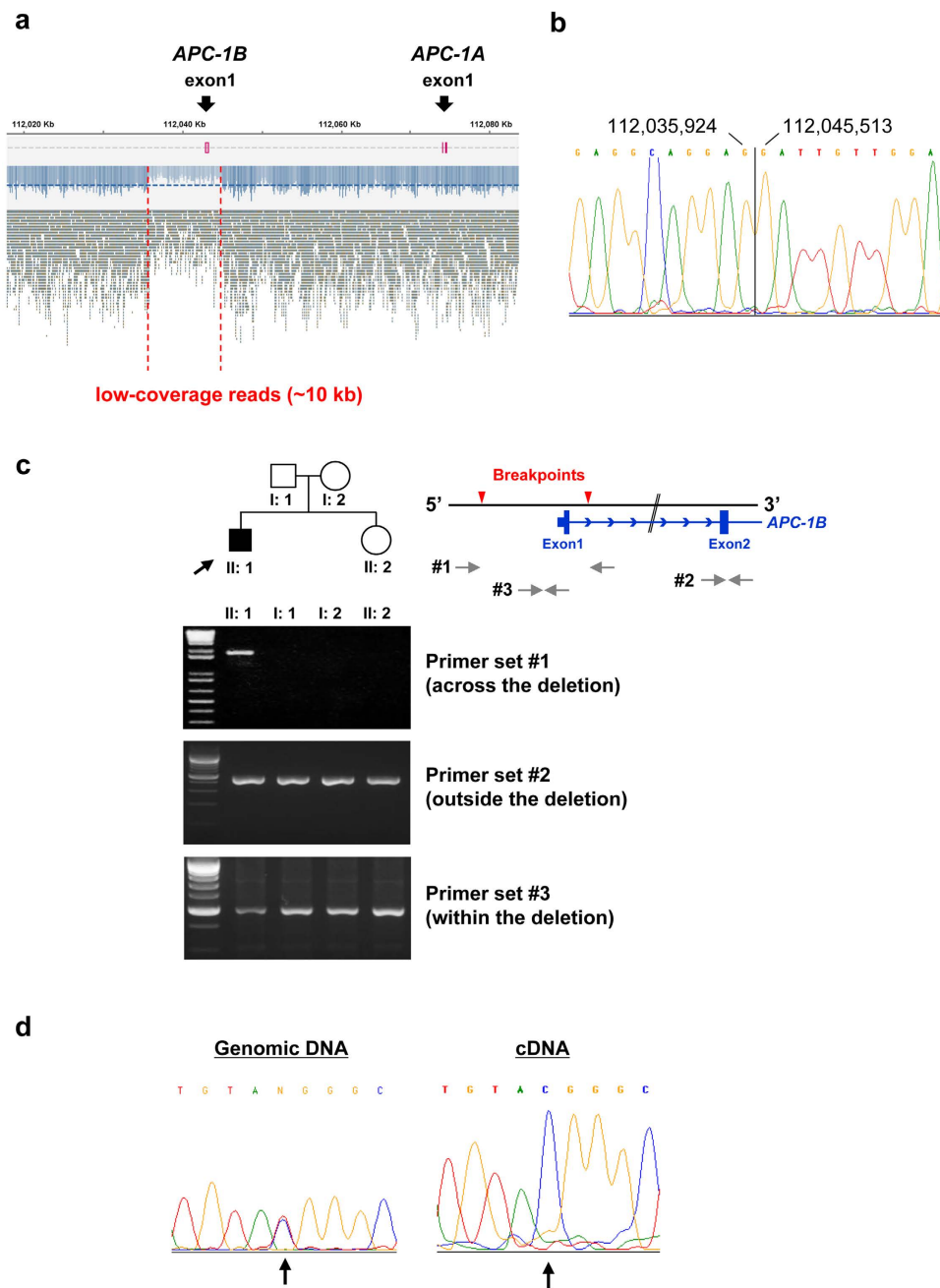
**Table 1. Summary of variants in genes associated with adenomatous polyposis of the colon (*APC*, *MUTYH*, *NTHL1*, *POLD1*, and *POLE*) detected by WGS.**

patient and the relatives (Fig. 1c, middle panel). These data suggest that the deletion is not carried in the relatives, indicating a *de novo* mutation in the patient. It is of note that this result is compatible with the patient's family history because none of the relatives were diagnosed as colonic polyposis, and no polyposis was detected in the first-grade relatives by colonoscopy.

**Transcriptional repression of the mutant *APC* allele.** Since the patient carried a large deletion which included promoter 1B, the mutant *APC* allele should have reduced transcriptional activity. Previous reports demonstrated that deletion of promoter 1B substantially reduced *APC* transcripts<sup>4,10–14</sup>. Based on this view, we speculated that one of the two alleles is down-regulated in the peripheral leucocytes as well as the epithelial cells of the colonic mucosa. To analyze allele-specific expression (ASE) of *APC* in the patient, we took the advantage of the heterozygous SNP (rs2229992) in exon12 (NM\_000038.5), and performed Sanger sequencing of the PCR products encompassing the SNP. This analysis indicated that both T and C alleles were retained in the genomic DNA (Fig. 1d). However, the T allele had almost disappeared in the cDNA, suggesting that the mutant transcripts were completely lost in the peripheral leucocytes. Due to the technical limitations of detection sensitivity in Sanger sequencing (less than 15%)<sup>8,21</sup>, we further performed deep sequencing of the cDNA, which enabled us to estimate the ratio of the two alleles in the transcripts by counting the sequence reads of the PCR products. The number of reads on the T allele was extremely low compared to the C allele (3,191 vs 25,246 reads). It is rational to consider that the T and C alleles correspond to mutant and wild type alleles, respectively. This view lead us to speculate that the 3,191 sequence reads of T allele may mainly consist of residual *APC-1A* transcript, and that the 25,246 reads of C allele include both *APC-1A* and *APC-1B* transcripts. We estimated that the mutant transcripts make up only 11.2% (3,191/28,437) of the total *APC* transcripts.

**Transcription of the mutant allele starts from the second TSS by promoter 1A.** The *APC* gene has two promoter regions (Fig. 2a), and promoter 1B is located approximately 30kb upstream of promoter 1A. Although the mutant allele lacks the region encompassing promoter 1B and exon1B, it retains promoter 1A. To investigate the activity of promoter 1A and/or others on the mutant allele, we employed cap analysis of gene expression (CAGE) analysis that allows transcriptome profiling with the simultaneous identification of tissue/cell/condition-dependent transcriptional start site (TSS). This analysis facilitates the estimation of the *APC* promoter activity by the measurement of peak height at the TSS. RNAs used for this analysis were extracted from the peripheral blood of two healthy volunteers as well as from the patient. Expectedly, we identified two tag clusters (Fig. 2b); the first cluster located at the 3' site of promoter 1B, and the second cluster at the 3' site of promoter 1A. The RNA from the two healthy volunteers showed both these peaks, and the first peak at 1B was approximately 44.5-fold higher than the second peak at 1A. Importantly, the peak at 1B was reduced in the patient to approximately 50% (39–45%) compared to the peak in the healthy volunteers (control 1 and control 2), while the peaks in *ACTB* showed no significant difference between the two (Table 2). Considering the expression of the wild type allele in the patient, the CAGE data suggested a complete loss of promoter 1B activity in the mutant allele. Of note, the CAGE data detected the peak at promoter 1A in the patient, but no new substantial peaks appeared in the flanking region, suggesting that the activity of promoter 1A was not abrogated by the deletion. In addition, the peak was slightly increased in the patient compared to the healthy controls (Fig. 2b).

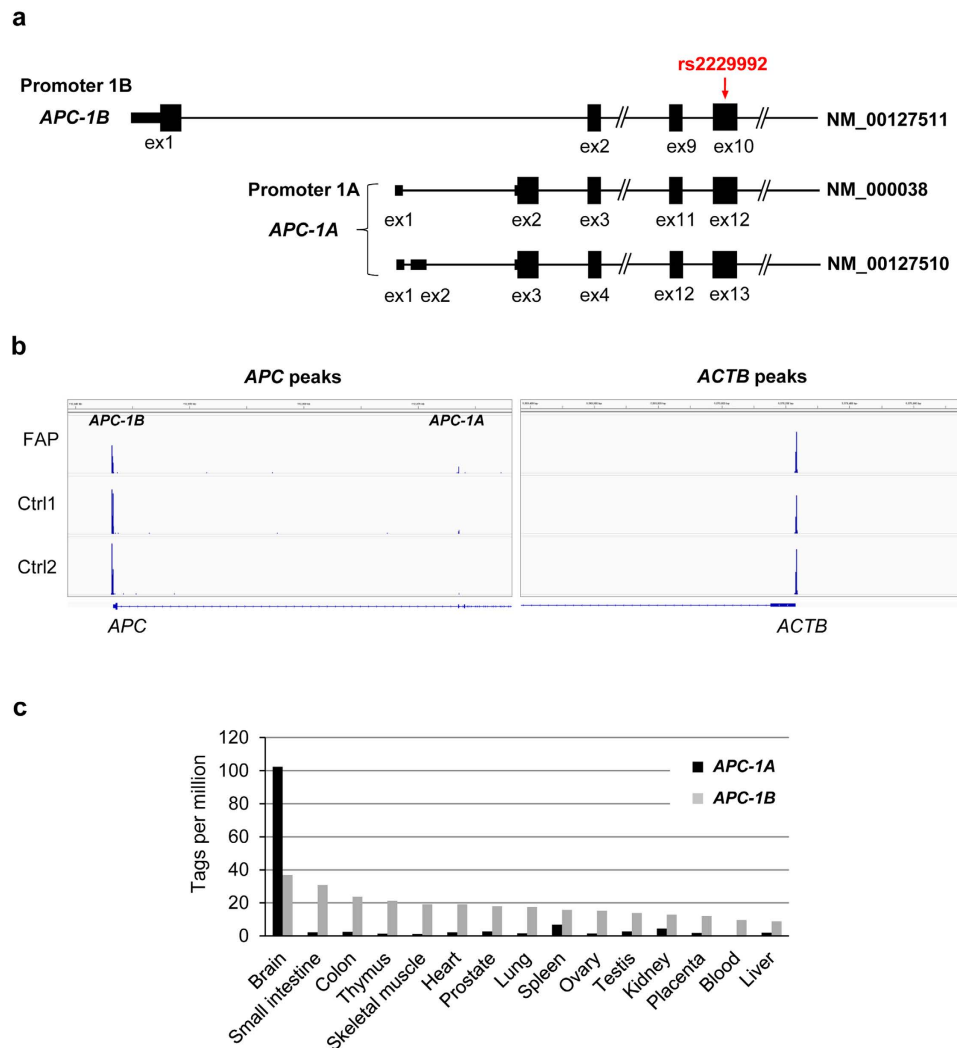
Degradation of cytosolic  $\beta$ -catenin by the *APC* destruction complex represents the key step of the canonical Wnt signaling pathway. The increased stability of  $\beta$ -catenin protein following *APC* mutation leads to the transcriptional activation of target genes in the nucleus<sup>22</sup>. Although we estimated approximately a 50% reduction of *APC* transcripts in the patient, we did not find any significant up-regulation of Wnt-target genes such as *DKK1*, *MYC*, and *CCND1* (Table 3). Since *APC* is a tumor suppressor gene, complete inactivation of the protein may be necessary for the transcriptional upregulation of these downstream genes. Consistently, motif discovery analysis



**Figure 1.** (a) A snapshot of the Tablet, a graphical viewer for NGS assemblies and alignments, showing a large deletion of ~10 kb encompassing promoter 1B and exon1B in the *APC* gene. (b) Direct sequencing of the breakpoint region. The PCR amplification was performed using primers across the deleted region (Supplementary Table S3), followed by the Sanger sequencing of the amplified fragment. (c) The PCR analysis for the deletion in the first-degree relatives. Three sets of primers were used for the amplification of a region across the deletion (upper panel), *APC-1B* exon2 (NM\_001127511.2, middle panel), and a region within the deletion (lower panel). The PCR amplification of *APC-1B* exon2 was used as an internal control. The primer sequences are shown in Supplementary Table S3. The proband is indicated by an arrow. Unaffected and affected individuals are represented by open and closed symbols (square: male and circle: female), respectively. (d) The sequence analysis of the heterozygous SNP rs2229992 using genomic DNA and cDNA. The arrow indicates the position of rs2229992 (c.1458 T/C).

showed that T-cell factor/lymphoid enhancer factor (TCF/LEF) motifs were not significantly enriched in the up-regulated gene list (Supplementary Data S1).

**Transcription start sites of *APC* in various normal tissues.** To elucidate the levels of *APC* expression in different tissues, we leveraged a CAGE-based gene expression database ([http://fantom.gsc.riken.jp/5/sstar/Main\\_Page](http://fantom.gsc.riken.jp/5/sstar/Main_Page)). The data is useful for the measurement of relative levels of transcripts initiated at each TSS among



**Figure 2.** (a) Schematic diagram of human *APC* transcript variants listed in NCBI. *APC* exhibits two forms of transcripts from promoter 1A and 1B, defined as *APC-1A* and *APC-1B*, respectively. The arrow indicates the position of heterozygous SNP rs2229992 (c.1458T/C). (b) The Integrative Genomics Viewer (IGV) snapshot of the three CAGE experiments, showing the *APC-1A*, *APC-1B*, and *ACTB* peaks. The *APC-1B* peak in the FAP patient was much lower than those in controls. The numbers of mapped reads within each peak are shown in Table 2. (c) Organ/tissue-dependent expression of *APC-1A* and *APC-1B* were identified using a CAGE-based expression data from the FANTOM5 SSTAR ([http://fantom.gsc.riken.jp/5/sstar/Main\\_Page](http://fantom.gsc.riken.jp/5/sstar/Main_Page)).

Gene	Patient	Control 1	Control 2
<i>APC-1B</i>	11.70	30.28	26.18
<i>APC-1A</i>	1.72	0.68	<0.1
<i>ACTB</i>	2526.07	2208.45	2742.55

**Table 2.** The expression of *APC* and *ACTB* in the patient and controls analyzed by CAGE. Each value represents CAGE tags that are normalized as TPM (tags per million). *ACTB* served as a control.

different tissues because CAGE does not require any intervening amplification steps and uses simple normalization. The levels of *APC-1A* and *APC-1B* expression in the tissues are listed in Supplementary Table S2. Consistent with our results and a previous report<sup>4</sup>, the expression of *APC-1B* was much higher than that of *APC-1A* in the blood, implying that promoter 1B plays a major role in the regulation of *APC* expression in most normal tissues (Fig. 2c).

## Discussion

Not only point mutations and small insertions/deletions but also structural alterations in the genome play a crucial role in human diseases. Large deletions/amplifications have been shown to be responsible for familial cancer syndromes such as hereditary breast and ovarian cancer<sup>23,24</sup>, FAP<sup>25</sup>, and Lynch syndrome<sup>26</sup>. MLPA is available

Gene	Patient	Control 1	Control 2
<i>DKK1</i>	0.57	0.68	0.61
<i>MYC</i>	31.31	59.99	46.75
<i>CCND1</i>	0.34	0.91	0.97
<i>PPARD</i>	44.38	40.37	40.54
<i>AXIN2</i>	4.36	7.14	7.55
<i>LGR5</i>	<0.10	<0.10	0.12
<i>CD44</i>	780.90	573.60	784.35

**Table 3. The expression of Wnt target genes in the patient and controls analyzed by CAGE.** Each value represents CAGE tags that are normalized as TPM (tags per million).

to capture large genomic deletions/amplifications in the responsible genes. Regarding FAP, Aretz *et al.* screened deletions/amplifications of *APC* in 174 unsolved cases by MLPA, and identified 14 different types of deletions in 26 cases, ranging from a single exon to the whole gene including the regulatory region<sup>9</sup>. Genetic testing by direct sequencing followed by MLPA analysis has increased the detection rate of pathogenic mutations in *APC* by 7–10%<sup>11</sup>. However, the identification of genetic alterations in the 5'-flanking regions, the intronic regions, and the 3'-untranslated regions of the responsible genes remains a challenge in genetic testing<sup>27</sup>.

A limited number of articles report FAP cases with promoter 1B deletion<sup>4,10–14,20</sup>. These cases included deletions with different sizes ranging from 11 kb to 132 kb (Supplementary Fig. S2b). The patient in this study carried a deletion of ~10 kb, which is the smallest deletion of promoter 1B, known to date. It is of note that these cases including ours did not show the phenotype of attenuated FAP but that of classical FAP. Since the germ line deletion encompassing the promoter 1B resulted in a marked decrease of *APC* transcripts in spite of the remaining *APC-1A* transcript, deletions or point mutations within this ~10 kb region may be involved in FAP without mutations in the coding region.

Although deletions of promoter 1B have been shown to reduce the expression of mutant *APC* allele, the degree of reduction caused by the deletions is different among previous reports. Pavicic *et al.* estimated the ASE by single nucleotide primer extension analysis, and found that a large deletion of 132 kb resulted in a reduction of the deleterious allele expression to 40–60% of the wild-type allele<sup>13</sup>. Another report revealed that a deletion of ~61 kb led to approximately a 90% reduction<sup>4</sup>. In this study, we used two approaches, ASE analysis with deep sequencing and CAGE, and disclosed that the deletion caused 87% reduction of mutant allele expression. Interestingly, the residual transcription was maintained by promoter 1A although the activity of promoter 1B was completely lost. Of note, the expression of *APC-1A* was slightly higher in the mutation carrier than that in controls. This may be explained as a compensatory mechanism against the impaired activity of promoter 1B. A similar observation has been reported by Rohlin *et al.*<sup>4</sup>.

To our knowledge, this is the first report showing a tissue-specific regulation of the *APC* transcripts at the TSS levels. We here revealed in our CAGE analysis that *APC-1B* expression is much higher than *APC-1A* expression in leukocytes, which is in complete accordance with the CAGE-based expression data in FANTOM5 SSTAR (Fig. 2b,c). Since CAGE does not require any intervening amplification step and uses simple normalization, it is suitable for measuring relative levels of transcripts initiated at each TSS among different tissues. Of note, the brain expresses an extremely high amount of *APC-1A* in addition to the high levels of *APC-1B* expression. This finding may suggest an important role of *APC* in the brain. It has been reported that canonical Wnt signaling pathway is one of the essential regulators of adult neurogenesis<sup>28–30</sup>. Strikingly, the expression of *APC* plays a key role in regulating neuronal differentiation of newly generated cells and maintaining neural stem cells in the adult neurogenic niche<sup>31</sup>. Although the role and mechanisms(s) of abundant *APC-1A* expression in the brain remain unclear, we speculate that patients lacking promoter 1B may not suffer from brain tumors associated with FAP.

In conclusion, using next-generation sequencing we identified a new causative promoter deletion of *APC* at the genome and transcript levels in a Japanese FAP patient. Of note, whole-genome sequencing in combination with ASE analysis by deep sequencing may be a useful strategy to identify deleterious genetic alterations in the regulatory regions of responsible genes for hereditary diseases.

## Materials and Methods

**Patient and DNA/RNA extraction.** A 29-year-old man was diagnosed to suffer from multiple colon adenomas and concomitant rectal adenocarcinoma by colonoscopy (Supplementary Fig. S2c,d). Additional upper gastrointestinal endoscopy revealed multiple fundic gland polyps in the stomach, and adenomas in the duodenum. These clinicopathological findings strongly suggested classical FAP. The patient's family tree is shown in Supplementary Fig. S3. None of the other family members were diagnosed as colonic polyposis. Individuals, I-1 and I-2 suffered gastric cancer (GS) and colorectal cancer (CRC), respectively. Individual, I-4 suffered CRC and thyroid cancer (TC). However, their offspring has had no malignancies except for the patient. The first-grade relatives of the patient (II-3, 4, and III-5) underwent screening of the colon by colonoscopy, but no polyposis was detected at the age indicated in Supplementary Fig. S3.

Genomic DNA was extracted from peripheral blood of the patient according to the standard phenol extraction/purification procedure<sup>32</sup>. Peripheral blood mononuclear cells were isolated from peripheral blood using Ficoll stratification. Total RNA was extracted using RNeasy Mini Kit (Qiagen, Hilden, Germany).

This project was approved by the ethical committee of Institute of Medical Science, the University of Tokyo (IMSUT-IRB, 23-18-0929). All experiments were carried out in accordance with the approved guidelines. Written informed consent was obtained from the patient and the first-degree relatives in this study.

**PCR-direct sequencing.** The coding exons in *APC* 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.

**Whole-genome sequencing.** WGS was performed using the HiSeq 2500 platform with paired-end reads of 101 bp according to the manufacturer's instructions (Illumina, San Diego, CA). A sequence library containing inserts of 250–350 bp was prepared using 1.0 µg of genomic DNA extracted from lymphocytes. A fastq file was aligned to human reference sequence (hg19) using Burrows-Wheeler Aligner (ver. 0.5.10)<sup>33</sup> and a bam file was created for data processing. The detection of single nucleotide variants and short insertions/deletions was performed by comparing the aligned bases with the reference sequence using a Bayesian approach. At each candidate position, the posterior probability of existence of variant was computed by updating a beta non-informative prior with the information of observed reads. We supposed that a random variable 'X' follows the posterior distribution representing the probability of variant. We used  $\text{Pr}(X \geq 0.05)$  as the score, and regarded the candidate positions whose scores were greater than 0.9 as statistically significant. Pathogenicity of the variants was first evaluated by two public databases, Leiden Open Source Variation Database (LOVD) and ClinVar in National Center for Biotechnology Information (NCBI). Uncharacterized non-synonymous variants were secondary evaluated by the frequencies of mutant alleles using dbSNP in NCBI, ESP-6500 in NHLBI Exome Sequencing Project, Human Genetic Variation Database (HGVD) in Kyoto University, and Integrative Japanese Genome Variation Database (iJGVD) in Tohoku University. Variants in *APC*, *BMPRIA*, or *STK11* with an allele frequency greater than 0.01, and those in *MUTYH* or *NTHL1* greater than 0.05 were considered as non-pathologic variants. The remaining non-synonymous variants were categorized in VUS. Additionally, 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<sup>34</sup>.

#### Determination of the breakpoint and analysis of allele-specific expression of the *APC* gene.

The breakpoint of the deletion was determined by PCR amplification using primers across the deleted region, followed by the Sanger sequencing of the amplified fragment. The primer sequences used for the amplification are shown in Supplementary Table S3.

To examine allele-specific expression (ASE) of *APC*, deep cDNA sequencing was performed using IonPGM Sequencing 200 kit and Sequencing 200 kit (Thermo Fisher Scientific) with libraries of PCR products prepared using Ion Plus Fragment Library Kit (Thermo Fisher Scientific). One microgram of total RNA was reverse transcribed for single-stranded cDNA using oligo(dT)<sub>18</sub> primer with Transcriptor reverse transcriptase (Roche Diagnostics, Indianapolis, IN). The cDNA was amplified with a set of primers encompassing a heterozygous single nucleotide polymorphism (SNP) at rs2229992 in the coding region of *APC*. The primer sequences are shown in Supplementary Table S3. Variants were identified using the Variant Caller deployed with Torrent Suite (Thermo Fisher Scientific), and the ASE was estimated using the allele-specific read depth.

**Identification of a *de novo* *APC* mutation in the patient.** PCR amplification was performed using DNA from the patient and the first-degree relatives with a set of primers across the deletion. Additionally, semi-quantitative PCR was performed with a set of primers within the deleted region and those outside the deletion (KOD-Plus kit, TOYOBO, Osaka, Japan). The primer sequences are shown in Supplementary Table S3.

**Cap analysis of gene expression (CAGE).** To elucidate relative levels of transcripts initiated at each TSS, we employed CAGE, a promoter-based expression analysis. The CAGE cDNA libraries were prepared using a Library Preparation kit according to the manufacturer's protocol (DNAFORM, Yokohama, Japan). Briefly, two micrograms of total RNA extracted from peripheral blood were used for synthesis of the first strand cDNA. After biotinylation of 5'-caps of RNAs, these biotinylated RNA/cDNA hybrids were captured by streptavidin-conjugated magnetic beads. The cDNAs were released from RNAs, and ligated with a linker including barcode sequences. Sequencing of the CAGE libraries were performed on the Illumina HiSeq 2500 (50 bp single-end). The CAGE sequencing data were handled using Maser (Management and Analysis System for Enormous Reads), a pipeline execution system, provided by the National Institute of Genetics ([http://cell-innovation.nig.ac.jp/index\\_en.html](http://cell-innovation.nig.ac.jp/index_en.html)). CAGE tags were normalized by the total number of tags per sample mapped to the human genome (hg19), and were indicated as tags per million (TPM). The clusters with <0.1 TPM per base were omitted. To compare the levels of *APC* expression in different normal human tissues, we used FANTOM5 SSTAR (semantic catalog of samples, transcription initiation and regulators), a CAGE-based gene expression database, provided by Riken ([http://fantom.gsc.riken.jp/5/sstar/Main\\_Page](http://fantom.gsc.riken.jp/5/sstar/Main_Page)).

#### References

1. Burt, R. W., DiSario, J. A. & Cannon-Albright, L. Genetics of colon cancer: impact of inheritance on colon cancer risk. *Annu Rev Med* **46**, 371–379 (1995).
2. Lambertz, S. & Ballhausen, W. G. Identification of an alternative 5' untranslated region of the adenomatous polyposis coli gene. *Hum Genet* **90**, 650–652 (1993).
3. Esteller, M. *et al.* Analysis of adenomatous polyposis coli promoter hypermethylation in human cancer. *Cancer Res* **60**, 4366–4371 (2000).
4. Rohlin, A. *et al.* Inactivation of promoter 1B of *APC* causes partial gene silencing: evidence for a significant role of the promoter in regulation and causative of familial adenomatous polyposis. *Oncogene* **30**, 4977–4989 (2011).

5. Segditsas, S. *et al.* Promoter hypermethylation leads to decreased APC mRNA expression in familial polyposis and sporadic colorectal tumours, but does not substitute for truncating mutations. *Exp Mol Pathol* **85**, 201–206 (2008).
6. Fokkema, I. F. *et al.* LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat* **32**, 557–563 (2011).
7. Newton, K. F. *et al.* Genotype-phenotype correlation in colorectal polyposis. *Clin Genet* **81**, 521–531 (2012).
8. Yamaguchi, K. *et al.* Detection of APC mosaicism by next-generation sequencing in an FAP patient. *J Hum Genet* **60**, 227–231 (2015).
9. Aretz, S. *et al.* Large submicroscopic genomic APC deletions are a common cause of typical familial adenomatous polyposis. *J Med Genet* **42**, 185–192 (2005).
10. Charames, G. S. *et al.* A large novel deletion in the APC promoter region causes gene silencing and leads to classical familial adenomatous polyposis in a Manitoba Mennonite kindred. *Hum Genet* **124**, 535–541 (2008).
11. Kadiyska, T. K. *et al.* APC promoter 1B deletion in familial polyposis—implications for mutation-negative families. *Clin Genet* **85**, 452–457 (2014).
12. Lin, Y. *et al.* Novel APC promoter and exon 1B deletion and allelic silencing in three mutation-negative classic familial adenomatous polyposis families. *Genome medicine* **7**, 42 (2015).
13. Pavicic, W. *et al.* Promoter-specific alterations of APC are a rare cause for mutation-negative familial adenomatous polyposis. *Genes, chromosomes & cancer* **53**, 857–864 (2014).
14. Snow, A. K. *et al.* APC promoter 1B deletion in seven American families with familial adenomatous polyposis. *Clin Genet* **88**, 360–365 (2015).
15. Al-Tassan, N. *et al.* Inherited variants of MYH associated with somatic G:C->T:A mutations in colorectal tumors. *Nat Genet* **30**, 227–232 (2002).
16. Jaeger, E. *et al.* Hereditary mixed polyposis syndrome is caused by a 40-kb upstream duplication that leads to increased and ectopic expression of the BMP antagonist GREM1. *Nat Genet* **44**, 699–703 (2012).
17. Palles, C. *et al.* Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat Genet* **45**, 136–144 (2013).
18. Weren, R. D. *et al.* A germline homozygous mutation in the base-excision repair gene NTHL1 causes adenomatous polyposis and colorectal cancer. *Nat Genet* **47**, 668–671 (2015).
19. Landrum, M. J. *et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* **42**, D980–985 (2014).
20. Kalbfleisch, T. *et al.* Characterization of an APC Promoter 1B deletion in a Patient Diagnosed with Familial Adenomatous Polyposis via Whole Genome Shotgun Sequencing. *F1000Research* **4**, 170 (2015).
21. Rohlin, A. *et al.* Parallel sequencing used in detection of mosaic mutations: comparison with four diagnostic DNA screening techniques. *Hum Mutat* **30**, 1012–1020 (2009).
22. Clevers, H. & Nusse, R. Wnt/beta-catenin signaling and disease. *Cell* **149**, 1192–1205 (2012).
23. Petrij-Bosch, A. *et al.* BRCA1 genomic deletions are major founder mutations in Dutch breast cancer patients. *Nat Genet* **17**, 341–345 (1997).
24. Swensen, J., Hoffman, M., Skolnick, M. H. & Neuhausen, S. L. Identification of a 14 kb deletion involving the promoter region of BRCA1 in a breast cancer family. *Hum Mol Genet* **6**, 1513–1517 (1997).
25. Raedle, J. *et al.* A de novo deletion of chromosome 5q causing familial adenomatous polyposis, dysmorphic features, and mild mental retardation. *The American journal of gastroenterology* **96**, 3016–3020 (2001).
26. Wijnen, J. *et al.* MSH2 genomic deletions are a frequent cause of HNPCC. *Nat Genet* **20**, 326–328 (1998).
27. Spier, I. *et al.* Deep intronic APC mutations explain a substantial proportion of patients with familial or early-onset adenomatous polyposis. *Hum Mutat* **33**, 1045–1050 (2012).
28. Kalani, M. Y. *et al.* Wnt-mediated self-renewal of neural stem/progenitor cells. *Proc Natl Acad Sci USA* **105**, 16970–16975 (2008).
29. Lie, D. C. *et al.* Wnt signalling regulates adult hippocampal neurogenesis. *Nature* **437**, 1370–1375 (2005).
30. Wexler, E. M., Paucer, A., Kornblum, H. I., Palmer, T. D. & Geschwind, D. H. Endogenous Wnt signaling maintains neural progenitor cell potency. *Stem Cells* **27**, 1130–1141 (2009).
31. Imura, T., Wang, X., Noda, T., Sofroniew, M. V. & Fushiki, S. Adenomatous polyposis coli is essential for both neuronal differentiation and maintenance of adult neural stem cells in subventricular zone and hippocampus. *Stem Cells* **28**, 2053–2064 (2010).
32. Sato, T. *et al.* Allelotype of breast cancer: cumulative allele losses promote tumor progression in primary breast cancer. *Cancer Res* **50**, 7184–7189 (1990).
33. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
34. Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5**, 557–572 (2004).

## Acknowledgements

This work was supported in part by the Grant-in-Aid for Young Scientists #26860206 (K.Y.) from the Japan Society for the Promotion of Science, and Center of Innovation program (S.M.) from the Japan Science and Technology Agency.

## Author Contributions

Y.F., S.M., S.N., S.I., R.Y. and K.Y. designed studies. K.Y. and S.H. performed experiments. E.S., M.K., R.Y., T.S., S.M. and S.I. provided data analysis and interpretation. S.N., M.A., T.I. and M.U. assisted with sample collection and data interpretation. K.Y. and Y.F. provided manuscript writing. K.Y., S.N. and E.S. contributed equally to this work. All authors contributed to and approved the final manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Yamaguchi, K. *et al.* Reduced expression of APC-1B but not APC-1A by the deletion of promoter 1B is responsible for familial adenomatous polyposis. *Sci. Rep.* **6**, 26011; doi: 10.1038/srep26011 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>