Characterization of candidate genes in unexplained polyposis and colorectal cancer

Abayzeed Elsayed Osman, F.

Chapter 4

Use of sanger and next-generation sequencing to screen for mosaic and intronic *APC* variants in unexplained colorectal polyposis patients

Fadwa A. Elsayed, Carli M. J. Tops, Maartje Nielsen, Hans Morreau, Frederik J. Hes, Tom van Wezel

Familial Cancer, 2022; 21(1): 79-83

Abstract

In addition to classic germline *APC* gene variants, *APC* mosaicism and deep intronic germline *APC* variants have also been reported to be causes of adenomatous polyposis. In this study, we investigated 80 unexplained colorectal polyposis patients without germline pathogenic variants in known polyposis predisposing genes to detect mosaic and deep intronic *APC* variants. All patients developed more than 50 colorectal polyps, with adenomas being predominantly observed. To detect *APC* mosaicism, we performed next-generation sequencing (NGS) in leukocyte DNA. Furthermore, using Sanger sequencing, the cohort was screened for the following previously reported deep intronic pathogenic germline *APC* variants: c.1408+731C>T, p.(Gly471Serfs*55), c.1408+735A>T, p.(Gly471Serfs*55), c.1408+729A>G, p.(Gly471Serfs*55) and c.532-941G>A, p.(Phe178Argfs*22). We did not detect mosaic or intronic *APC* variants in the screened unexplained colorectal polyposis patients. The results of this study indicate that the deep intronic *APC* variants investigated in this study are not a cause of colorectal polyposis in this Dutch population. In addition, NGS did not detect any further mosaic variants in our cohort.

Introduction

Pathogenic germline variants in *APC* (MIM# 611731) cause familial adenomatous polyposis syndrome (FAP; MIM# 175100), a rare autosomal dominant-inherited syndrome characterized by the development of multiple colorectal adenomas and a very high risk of colorectal cancer [1-4]. In classic FAP, patients develop hundreds to thousands of colorectal adenomatous polyps, while in attenuated FAP (AFAP), patients develop fewer adenomas (< 100) at a later age than those with classical FAP [5-8]. A subset of patients with multiple colorectal adenomas and no *APC* germline variants have been found to carry biallelic variants in the base excision repair gene *MUTYH* (MIM# 604933), causing *MUTYH*-associated polyposis (MAP; MIM# 608456) [9]. In addition, a number of other genes associated with adenomatous polyposis, such as *POLE*, *POLD1*, *NTHL1*, *MSH3* and *MLH3*, have recently been reported [10-13]. The detection rate of *APC* variants in FAP patients depends on phenotype and methods. In classic FAP, *APC* germline variants can be detected in up to 85% of patients [14, 15]; however, the detection rates of *APC* germline variants in patients with fewer colorectal adenomatous polyps (AFAP patients) are lower, ranging from 10% to 30% of patients [14, 16], suggesting that a proportion of pathogenic variants remain undetected by routine methods [17-19]. Mosaic *APC* variants and deep intronic variants localized in regions not covered by PCR-based diagnostics were previously identified as additional causal factors. Using RNA-based assays and next-generation sequencing (NGS), it has been shown that a proportion of variant-negative FAP patients harbor molecular changes in deep intronic regions of *APC* [19, 20]. These studies identified deep intronic *APC* variants that result in pseudoexon formation [19, 20]. Through the use of sensitive techniques, somatic *APC* mosaicism has been demonstrated in a minority of adenomatous polyposis patients [21-26]. In addition, using deep sequence analysis of *APC* in DNA isolated from multiple adenomas, mosaic variants were identified in 9 of 18 patients with 21 to 100 adenomas; in some of these cases, NGS also detected the variants in leukocyte DNA at low frequency [27]. In this study, we investigate the role of deep intronic germline *APC* variants and mosaic *APC* variants in leukocyte DNA as possible genetic causes of colorectal polyposis in a Dutch cohort of unexplained patients with more than 50 polyps.

Materials and methods

Patients

A total of 80 index patients with more than 50 colorectal polyps (Table 1) were selected from a previously described cohort [28-31]. The cohort included patients previously screened for germline mosaic *APC* variants by denaturing gradient gel electrophoresis (DGGE) [17], the protein truncation test (PTT) [17] and high resolution melting analysis (HRMA) [21]. All cases tested negative for pathogenic germline variants in *APC*, *MUTYH*, *POLE*, and *POLD1* and for *NTHL1* hotspot variants. Clinicopathological data included date of birth, gender, age at diagnosis of colorectal polyps/adenomas, cumulative number of polyps, location and histology of polyps/adenomas, information on CRC and presence of polyps/CRC in first-degree family members. Since the term serrated adenomas is currently preferred over hyperplastic polyps, we lumped together polyps described as such under the term sessile serrated lesions with or without dysplasia. Three controls were included in this study. Leukocyte DNA from this cohort was available for the study. The study was approved by the medical ethics committee of Leiden University Medical Center, protocol P01-019.

APC intronic variant screening

Leukocyte DNA of the patients was screened for the intronic *APC* variants in Table 2 using Sanger sequencing. Primers were designed using Primer3 software http://primer3.ut.ee/ and were obtained from Eurofins Genomics (Ebersberg, Germany). The following primers with universal M13 tails were used: c.1408+731C>T, c.1408+735A>T and c.1408+729A>G; forward: 5'-TGTAAAACGACGGCCAGTATCATGCTGAACCATCTCAT-3' and reverse: 5' CAGGAAACAGCTATGACCAAATGACGAATGAAACGATG-3'. For c.532-941G>A; forward: 5' TGTAAAACGACGGCCAGTAGAGGGTTTGGGAAGTGGAG-3' and reverse: 5' CAGGAAACAGCTATGACCTCTGTGTGCCCTTAGAAAACTG-3'. Sanger sequencing of the PCR amplified fragments was performed by Macrogen (Amsterdam, Netherlands). The sequencing results were analyzed using Mutation Surveyor software (Sofgenetics, State College PA, USA).

Table 1. Clinical characteristics of the colorectal polyposis patients (n=80)

Patient characteristics	Individuals %
Number of polyps	
>100	29 (36.2%)
50-100	51 (63.8%)
Type of polyps	
Adenomas	36 (45%)
Mixed (Adenomas + Serrated*)	38 (47.5%)
Serrated	5 (6.2%)
Unknown	1 (1.3%)
Age at diagnosis with polyposis	
≥50 years	49 (61.3%)
<50 years	31 (38.7%)
Diagnosed with CRC	
Yes	27 (33.8%)
No	53 (66.2%)
Age at diagnosis with CRC	
>50	19 (70.4%)
≤48	8 (29.6%)
Sex	
Male	53 (66.2%)
Female	27 (33.8%)
Polyposis family	
Polyposis family	29
No polyposis family	37
Unknown	14
CRC family	
CRC family	33
No CRC family	34
Unknown	13

* Sessile serrated lesions with or without dysplasia

Table 2. Summary of the germline pathogenic *APC* intronic variants

Intron	Alteration in genomic DNA	Insertion length (bp)	RNA alteration	Predicted protein alteration	Publication
4	c.532-941G>A	Insertion of 167 bp	r.531_532ins532-1106_532-940	p.Phe178Argfs*22	[19]
10	c.1408+731C>T	Insertion of 83 bp	r.1408_1409ins1408+647_1408+729	p.Gly471Serfs*55	[19, 20]
10	c.1408+735A>T	Insertion of 83 bp	r.1408_1409ins1408+647_1408+729	p.Gly471Serfs*55	[19]
10	c.1408+729A>G	Insertion of 83 bp	r.1408_1409ins1408+647_1408+729	p.Gly471Serfs*55	[20]

Next-generation sequencing and data analysis

Deep *APC* sequencing was performed using a previously described custom *APC* panel [27]. The complete sequencing panel consisted of 115 amplicons (11,216 bp), covering 99.3% of the coding regions of *APC*. Libraries were prepared with Ion Ampliseq™ 2.0 Kit (Thermo Fisher Scientific, Bleiswijk, The Netherlands) according to the manufacturer's instructions and were sequenced on the Ion Torrent Proton Platform (Thermo Fisher Scientific, Bleiswijk, The Netherlands). Sequence data were analyzed as described previously [27]. Variants were annotated to the GenBank reference sequence NM_000038.4. The Integrative Genomics Viewer (IGV) (https://www.broadinstitute.org/igv/) was used to visualize the read alignment and the presence of variants against the reference genome.

Results and discussion

In this study, we attempt to identify the genetic causes of colorectal polyposis in unexplained patients with colorectal polyposis. Deep NGS of *APC* was performed to identify possible undetected pathogenic mosaic variants. Furthermore, *APC* intronic germline variants described previously [19, 20] were studied to evaluate their role. A high-risk cohort was selected for this study, consisting of 80 index patients with ≥ 50 colorectal polyps (Table 1), of whom many had a relatively early onset, which increases the probability of finding undiscovered mosaic or intronic variants. The mean age at diagnosis of colorectal polyps was 49 years (range 12-80). The majority of patients (n= 51, 63.8% with a mean age of 51 years at diagnosis) had a cumulative polyp count between 50 and 100, while 29 patients (36.2% with a mean age of 46 years at diagnosis) showed more than 100 polyps. Forty-five percent of the patients displayed only adenomatous polyps, while 47.5% of the patients displayed a mixed phenotype with adenomas and sessile serrated lesions with or without dysplasia. CRC was found in 27 patients (33.8%, with a mean age of 56 years, range 37-80). The clinical characteristics of the patients are summarized in Table 1.

First, we screened the leukocyte DNA of 80 patients for the following deep intronic heterozygous germline variants in *APC:* c.1408+731C>T, p.(Gly471Serfs*55), c.1408+735A>T, p.(Gly471Serfs*55), c.1408+729A>G, p.(Gly471Serfs*55) and c.532-941G>A, p.(Phe178Argfs*22). We did not detect any of these variants in our cohort. The study by Spier et al. [19] was the first to describe *APC*-related pseudoexons in FAP patients from Germany. These pseudoexons were caused by three heterozygous germline variants with a combined frequency of 6.4%

(8/125); *APC* c.532-941G>A was identified in five patients, *APC* c.1408+731C>T was identified in two patients, and *APC* c.1408+735A>T was identified in one patient [19]. In a second study by Nieminen et al. [20], two additional intronic variants were identified in a cohort of 54 patients from Finland: *APC* c.1408+729A>G and *APC* c.646-1806T>G and the variant identified previously by Spier et al., *APC* c.1408+731C>T. The overall reported frequency of these variants from the study by Nieminen et al. was 5.5% (3/54). The reported frequency of these intronic variants from both studies is approximately 6%. Nevertheless, we could not detect these variants in our cohort, possibly because either the frequency of intronic variants is lower in the Dutch population and the sample size of our cohort is not large enough or because these variants are local founder variants.

Subsequently, we performed deep *APC* sequencing of leukocyte DNA from 80 colorectal polyposis patients. Our positive controls were two previously described cases with mosaic *APC* variants [27]; *APC* c.4110_4111delAA was reported to be present in leukocyte DNA with 4% variant allele frequency (VAF), and *APC* c.2493dupA was reported with a VAF of 3% in leukocyte DNA. The *APC* mosaic variant c.4057G>T served as a negative control, as the variant was detected previously [27] in normal colonic mucosa and was absent in leukocyte DNA. Both positive controls, *APC* c.4110_4111delAA (Figure 1) and *APC* c.2493dupA, were clearly present, while *APC* c.4057G>T was absent in the negative control. No additional *APC* mosaic variants were detected in our cohort. A limitation of this study is that we used only leukocyte DNA for mosaicism screening due to the scarcity of available DNA from patient adenomas. Mosaicism might remain undetectable or be overlooked if the molecular analysis is limited to blood, even when sensitive techniques are applied, due to very low or even absent presentation of the mutated allele [23, 27]. Peripheral blood cells arise from the mesoderm, and when the variant occurs after mesoderm and endoderm specification (early postzygotic mutation), the mosaicism is likely restricted to the colon and is difficult to detect the variant in leukocyte DNA [23, 27, 32, 33]. In a previous study, it was recommended to test at least two or more adenomas to detect mosaic variants [27].

A recent systematic review of current *APC* mosaicism studies recommends testing adenomas from the polyposis patients without *APC* germline variant to allow the detection of low allele frequency mosaicism as well as mosaicism confined to colon [33]. Consequently, in our study, *APC* mosaic variants confined to the colon could have been missed because we could not screen the DNA from the adenomas of the patients.

Figure 1. Integrative Genomics Viewer (IGV) images of next-generation sequencing (NGS) data of mosaic *APC* c.4110_4111delAA variant detected in the leukocyte DNA of the positive control sample.

In conclusion, we did not detect any of the four previously reported *APC* intronic variants in our cohort. We also did not detect mosaic *APC* variants in our cohort using deep sequencing analysis in blood. This finding suggests that the benefit of using targeted amplicon-based NGS to further scrutinize the *APC* gene in unexplained cases of polyposis is limited. Analyzing DNA from adenomas in addition to leukocyte DNA is recommended to detect a possible underlying mosaicism. Also, other approaches, such as whole genome sequencing or transcriptome sequencing, could be employed to detect undiscovered intronic or promoter variants or other regulatory variants.

References

1. Bodmer WF, Bailey CJ, Bodmer J, et al. Localization of the gene for familial adenomatous polyposis on chromosome 5. Nature 1987;328:614-6.
2. Bisgaard ML, Fenger K, Bulow S, et al. Familial adenomatous polyposis (FAP): frequency, penetrance, and mutation rate. Hum Mutat 1994;3:121-5.
3. Fearnhead NS, Britton MP, Bodmer WF. The ABC of APC. Hum Mol Genet 2001;10:721-33.
4. Yurgelun MB, Kulke MH, Fuchs CS, et al. Cancer Susceptibility Gene Mutations in Individuals With Colorectal Cancer. J Clin Oncol 2017;35:1086-1095.
5. Knudsen AL, Bisgaard ML, Bulow S. Attenuated familial adenomatous polyposis (AFAP). A review of the literature. Fam Cancer 2003;2:43-55.
6. Jasperson KW, Tuohy TM, Neklason DW, et al. Hereditary and familial colon cancer. Gastroenterology 2010;138:2044-58.
7. Nielsen M, Hes FJ, Nagengast FM, et al. Germline mutations in APC and MUTYH are responsible for the majority of families with attenuated familial adenomatous polyposis. Clin Genet 2007;71:427-33.
8. Nieuwenhuis MH, Vasen HF. Correlations between mutation site in APC and phenotype of familial adenomatous polyposis (FAP): a review of the literature. Crit Rev Oncol Hematol 2007;61:153-61.
9. Al-Tassan N, Chmiel NH, Maynard J, et al. Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors. Nat Genet 2002;30:227-32.
10. Palles C, Cazier JB, Howarth KM, et al. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. Nat Genet 2013;45:136-44.
11. Weren RD, Ligtenberg MJ, Kets CM, et al. A germline homozygous mutation in the base-excision repair gene NTHL1 causes adenomatous polyposis and colorectal cancer. Nat Genet 2015;47:668-71.
12. Adam R, Spier I, Zhao B, et al. Exome Sequencing Identifies Biallelic MSH3 Germline Mutations as a Recessive Subtype of Colorectal Adenomatous Polyposis. Am J Hum Genet 2016;99:337-51.
13. Olkinuora A, Nieminen TT, Mårtensson E, et al. Biallelic germline nonsense variant of MLH3 underlies polyposis predisposition. Genet Med 2019;21:1868-1873.
14. Friedl W, Aretz S. Familial adenomatous polyposis: experience from a study of 1164 unrelated german polyposis patients. Hered Cancer Clin Pract 2005;3:95-114.
15. Aretz S, Stienen D, Uhlhaas S, et al. Large submicroscopic genomic APC deletions are a common cause of typical familial adenomatous polyposis. J Med Genet 2005;42:185-92.
16. Terlouw D, Suerink M, Singh SS, et al. Declining detection rates for APC and biallelic MUTYH variants in polyposis patients, implications for DNA testing policy. Eur J Hum Genet 2020;28:222-230.
17. Hes FJ, Nielsen M, Bik EC, et al. Somatic APC mosaicism: an underestimated cause of polyposis coli. Gut 2008;57:71-6.
18. Rohlin A, Wernersson J, Engwall Y, et al. Parallel sequencing used in detection of mosaic mutations: comparison with four diagnostic DNA screening techniques. Hum Mutat 2009;30:1012-20.
19. Spier I, Horpaopan S, Vogt S, et al. Deep intronic APC mutations explain a substantial proportion of patients with familial or early-onset adenomatous polyposis. Hum Mutat 2012;33:1045-50.

20. Nieminen TT, Pavicic W, Porkka N, et al. Pseudoexons provide a mechanism for allele-specific expression of APC in familial adenomatous polyposis. Oncotarget 2016;7:70685-70698.
21. Out AA, van Minderhout IJ, van der Stoep N, et al. High-resolution melting (HRM) re-analysis of a polyposis patients cohort reveals previously undetected heterozygous and mosaic APC gene mutations. Fam Cancer 2015;14:247-57.
22. Yamaguchi K, Komura M, Yamaguchi R, et al. Detection of APC mosaicism by next-generation sequencing in an FAP patient. J Hum Genet 2015;60:227-31.
23. Spier I, Drichel D, Kerick M, et al. Low-level APC mutational mosaicism is the underlying cause in a substantial fraction of unexplained colorectal adenomatous polyposis cases. J Med Genet 2016;53:172-9.
24. Ciavarella M, Miccoli S, Prossomariti A, et al. Somatic APC mosaicism and oligogenic inheritance in genetically unsolved colorectal adenomatous polyposis patients. Eur J Hum Genet 2018;26:387-395.
25. Kim B, Won D, Jang M, et al. Next-generation sequencing with comprehensive bioinformatics analysis facilitates somatic mosaic APC gene mutation detection in patients with familial adenomatous polyposis. BMC Med Genomics 2019;12:103.
26. Urbanova M, Hirschfeldova K, Obeidova L, et al. Two Czech patients with familial adenomatous polyposis presenting mosaicism in APC gene. Neoplasma 2019;66:294-300.
27. Jansen AM, Crobach S, Geurts-Giele WR, et al. Distinct Patterns of Somatic Mosaicism in the APC Gene in Neoplasms From Patients With Unexplained Adenomatous Polyposis. Gastroenterology 2017;152:546-549.e3.
28. Hes FJ, Ruano D, Nieuwenhuis M, et al. Colorectal cancer risk variants on 11q23 and 15q13 are associated with unexplained adenomatous polyposis. J Med Genet 2014;51:55-60.
29. Elsayed FA, Kets CM, Ruano D, et al. Germline variants in POLE are associated with early onset mismatch repair deficient colorectal cancer. Eur J Hum Genet 2015;23:1080-4.
30. Elsayed FA, Tops CMJ, Nielsen M, et al. Low frequency of POLD1 and POLE exonuclease domain variants in patients with multiple colorectal polyps. Mol Genet Genomic Med 2019:e603.
31. Grolleman JE, de Voer RM, Elsayed FA, et al. Mutational Signature Analysis Reveals NTHL1 Deficiency to Cause a Multi-tumor Phenotype. Cancer Cell 2019;35:256-266.e5.
32. Tuohy TM, Burt RW. Somatic mosaicism: a cause for unexplained cases of FAP? Gut 2008;57:10-2.
33. Jansen AML, Goel A. Mosaicism in Patients With Colorectal Cancer or Polyposis Syndromes: A Systematic Review. Clin Gastroenterol Hepatol 2020;18:1949-1960.