Recurrent APC splice variant c.835-8A > G in patients with unexplained colorectal polyposis fulfilling the colibactin mutational signature

Terlouw, D.; Suerink, M.; Boot, A.; Wezel, T. van; Nielsen, M.; Morreau, H.

Recurrent *APC* Splice Variant c.835-8A>G in Patients With Unexplained Colorectal Polyposis Fulfilling the Colibactin Mutational Signature

Diantha Terlouw,[1] Manon Suerink,[2] Arnoud Boot,[1,3] Tom van Wezel,[1] Maartje Nielsen,[2] and Hans Morreau[1]

[1]Department of Pathology, Leiden University Medical Centre, Leiden, The Netherlands; [2]Department of Clinical Genetics, Leiden University Medical Centre, Leiden, The Netherlands; and [3]Department of Cancer and Stem Cell Biology, Duke-National University of Singapore Medical School, Singapore

Keywords: Unexplained Adenomatous Polyposis Coli; Bacterial Infection; Hotspot *APC* Variant; *pks* + *E coli*..

Despite the clear autosomal dominant inheritance of germline *APC* variants causing familial adenomatous polyposis, carriers can still present with a negative family history suggesting a de novo variant. Depending on the exact temporal occurrence of the de novo variant, all or only a subset of cells in the body will be affected. Presence of a de novo variant in only a subset of cells is called mosaicism.

Jansen et al[1] reported that *APC* mosaicism can be detected using next-generation sequencing in DNA isolated from formalin-fixed paraffin-embedded adenoma tissue. These variants were often not found in leukocyte DNA. *APC* analysis in adenomas is part of our regular diagnostics for unexplained polyposis patients.

The identification of possible hotspot variants in *APC* will help to interpret findings suggestive of mosaicism. Does a finding of 2 colonic lesions sharing the same variant indicate mosaicism or is it coincidental? This question is considered in Jansen et al[1] with a patient carrying the same *APC* variant in 10 of 16 lesions.

Methods

Formalin-fixed paraffin-embedded tissue blocks from colorectal adenomas and carcinomas were collected from 201 unexplained polyposis patients. In total, 872 colorectal lesions were sequenced using next-generation sequencing. The detected variants were categorized by pathogenicity and loss of heterozygosity was determined. A more detailed description is provided in the Supplementary Methods.

Results

In 11.9% (24 of 201) of patients, true *APC* mosaicism was identified, meaning the same *APC* variant present in all analyzed lesions. After excluding the lesions of true mosaic cases, 763 lesions remained, consisting of 61 carcinomas and 702 adenomas. In 72% of these lesions at least 1 pathogenic *APC* variant was detected. In carcinomas, the frequency of *APC* variants was 69% and in adenomas was 72%.

In total, 108 *APC* variants occurred more than once in nonmosaic colorectal lesions. The most frequently observed *APC* variant, occurring in 7% of lesions, was a splice variant located in intron 8; NM_000038.5: c.835-8A>G. Two patients showed the c.835-8A>G in a true mosaic pattern. However, it was not observed in any of the normal tissues tested (n = 7; Supplementary Table 1). Moreover, in 44% of patients (16 of 36) with the c.835-8A>G variant, a subset (more than 1, ranging from 2 of 9 to 6 of 10, but not all) of lesions harbored this specific variant, a so-called hybrid mosaic pattern. Also in these patients, none of the normal tissues tested positive for the variant (n = 16).

The c.835-8A>G variant was observed in both adenomas (n = 61) and carcinomas (n = 6). The majority (46 of 67 [69%]) of lesions containing the variant were located in the distal colon. Furthermore, in 54% (36 of 67) of lesions, 1 or more other pathogenic variant was detected in the *APC* gene, in 26 (72%) of these lesions the c.835-8A>G variant showed the highest variant allele frequency. In 28% (19 of 67), loss of heterozygosity was observed, the remaining lesions (18%) did not show any second hit.

Recently, a mutational signature caused by *pks*+ *Escherichia coli* was identified.[2,3] This signature is characterized by single base substitutions T>N mostly in ATN and TTT context with strong enrichment of adenines 3 and 4 base pairs 5′ of the mutation site and a strong transcriptional strand bias.[3] Interestingly, the c.835-8A>G variant has a sequence context of TTAATTTTT (Figure 1*A*), where the underlined adenine is substituted by a guanine. Transformed in a T>N orientation (Figure 1*B*), the context perfectly fulfills the mutational signature caused by *pks*+ *E coli* with the hexanucleotide AAAATT as predominant sequencing context (Figure 1*C*). Furthermore, fulfilling the signature means that the c.835-8A>G variant is suggested to arise from adducts on the untranscribed strand and is therefore not removed by transcription-coupled nucleotide excision repair.[3]

Most current article

0016-5085
https://doi.org/10.1053/j.gastro.2020.06.055

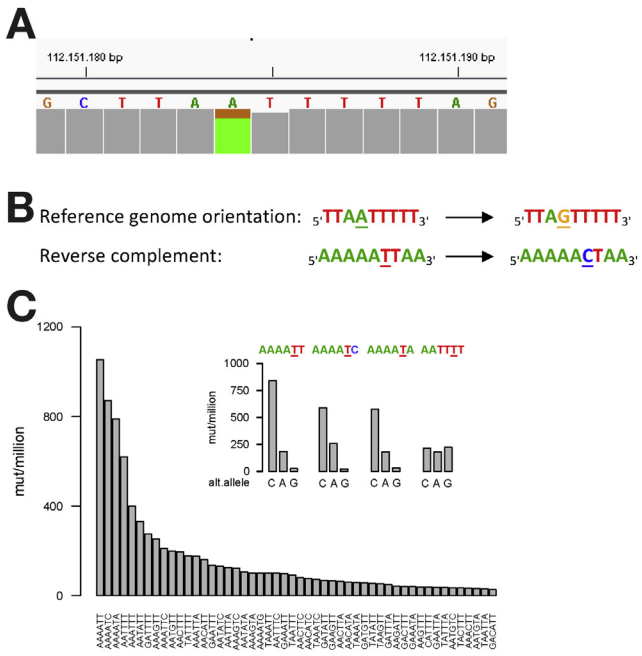

Figure 1. Comparison of the sequence context of the c.835-8A>G variant with that of the mutational signature associated with *pks*+ *E coli*. (*A*) Visualization of the variant in Integrative Genomics Viewer. (*B*) The T>N oriented sequence context of the c.835-8A>G variant. (*C*) Top 50 hexanucleotides mostly affected by the *pks*+ *E coli* mutational signature normalized to the frequency of the hexanucleotide in the human genome based on data from Boot et al.[3] For the 4 most commonly affected hexanucleotides, a breakdown of the alternative alleles is shown. AAAATT is most likely to be affected by this mutational signature, and the most common alternative allele is C.

Of the other recurrent variants, 7 fulfill the *pks*+ *E coli* mutational signature (Supplementary Table 2). Remarkably, in 13 patients, ≥50% (up to 100%) of lesions carried an *APC* variant fulfilling the *pks*+ *E coli* mutational signature (Supplementary Table 1). In total, the majority (54 of 79 [68%]) of lesions with such a variant was located distally.

Discussion

Performing *APC* mosaicism analysis in patients with unexplained polyposis provided an opportunity to study the occurrence and frequency of pathogenic *APC* variants in colorectal lesions.

The most frequently observed *APC* variant, c.835-8A>G, has been described as a germline variant twice.[4,5] Complementary DNA analysis showed that the variant creates a new splice acceptor site causing a frameshift leading to a premature stop codon. Although the variant is located in a region associated with classical familial adenomatous polyposis, the patient presented with a medium polyp burden, suggesting the variant to have a mild impact on the *APC* gene or the original splice site is still partly active, leading to some normal protein. Jarry et al[5] predicted the protein change to be p.Gly279Phefs*11. The c.835-8A>G variant has also been described somatically in 3% of sporadic colorectal cancers.[6] Interestingly, 45% of c.835-8A>G carcinomas did not carry a second hit. Moreover, the carcinomas exhibit nuclear β-catenin staining,[6] this might suggest that the splice variant provides a growth advantage to the colon crypt cell even with an intact second allele.

In 2 patients in our cohort, the c.835-8A>G variant was identified in a true mosaic pattern. One mosaic patient developed adenomas at the age of 24 years and was diagnosed with ulcerative colitis. Ulcerative colitis is known to be associated with "field cancerization" in which premalignant areas in the colon share the same dysplastic changes simultaneously through repopulation of destroyed crypts.[7] This phenomenon might be an explanation for the detected mosaicism and development of adenomas at a young age in patients with inflammatory bowel disease.

The presence of *pks*+ *E coli*, causing a specific mutational signature (Figure 1), might be an additional explanation for unexplained polyposis patients. This especially applies to the large proportion of patients carrying the c.835-8A>G variant and other *pks*+ *E coli* variants in multiple lesions. Remarkably, the *pks*+ *E coli* mutational signature seems to predominantly affect the distal colon,[8] as confirmed by the location of lesions with *pks*+ *E coli* variants in our cohort. These findings show the relevance of further research into the presence and influence of *pks*+ *E coli* in our cohort and other unexplained polyposis patients.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at http://dxdoi.org/10.1053/j.gastro.2020.06.055.

References

1. Jansen AM, et al. Gastroenterology 2017;152:546–549 e3.
2. Pleguezuelos-Manzano C, et al. Nature 2020;580:269–273.
3. Boot A, et al. Genome Res 2020;30:803–813.
4. Fostira F, et al. BMC Cancer 2010;10:389.
5. Jarry J, et al. Fam Cancer 2011;10:659–665.
6. Yaeger R, et al. Cancer Cell 2018;33:125–136.e3.
7. Baker KT, et al. Carcinogenesis 2018;39:11–20.
8. Lee-Six H, et al. Nature 2019;574:532–537.

Received May 4, 2020. Accepted June 25, 2020.

Correspondence
Address correspondence to: Prof. Hans Morreau, Department of Pathology, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, The Netherlands. e-mail: j.morreau@lumc.nl; fax: +31 71 526 8285.

Acknowledgments
Collaborators: Dr Alexandra M.J. Langers, Department of Gastroenterology, Leiden University Medical Center, Leiden; Dr Carli M. Tops, Department of Clinical Genetics, Leiden University Medical Center, Leiden; Dr Dina Ruano, Department of Pathology, Leiden University Medical Center, Leiden.

The authors acknowledge the work of Dr Dina Ruano in the bioinformatical support, Dr Carli M. Tops in assisting in the interpretation of the variants, and Dr Alexandra M.J. Langers in acquistion of the patients.

CRediT Authorship Contributions
Diantha Terlouw, MSc (Data curation: Lead; Formal analysis: Lead; Writing – original draft: Lead). Manon Suerink, MSc (Writing – review & editing: Supporting). Arnoud Boot, PhD (Writing – review & editing: Supporting). Tom van Wezel, PhD (Conceptualization: Supporting; Funding acquisition: Supporting; Supervision: Supporting; Writing – review & editing: Supporting). Maartje Nielsen, PhD (Conceptualization: Supporting; Funding acquisition: Supporting; Supervision: Supporting; Writing – review & editing: Supporting). Hans Morreau, PhD (Conceptualization: Lead; Funding acquisition: Lead; Supervision: Lead; Writing – review & editing: Lead).

Conflicts of interest
The authors disclose no conflicts.

Funding
This study is supported by the Dutch Cancer Society.

Supplementary Methods

Cohort

As part of diagnostics, patients with multiple colorectal adenomas and/or carcinomas were sent in for *APC* mosaicism testing either via their clinical geneticist or gastroenterologist at the Leiden University Medical Center. Of 201 patients without a germline explanation for the polyposis coli, colonic adenoma and/or carcinoma material was collected. The study protocol was approved by the local ethics committee (LUMC B18.042).

DNA Extraction

The material collected comprised formalin-fixed paraffin-embedded tissue blocks and H&E slides. The H&E slides were used to examine the region of interest and determine tumor percentages (preferably >30%). When possible, formalin-fixed paraffin-embedded tissue blocks were punched to collect tumor cells. Otherwise, whenever enough lesional cells were present, complete 10-μm hematoxylin-stained sections were taken (whole section). If tumor cell percentage was too low, an inverted microscope was used to scratch the tumor cells from the sections (microdissection). DNA from collected tumor cells was isolated using the automated Tissue Preparation System, as described previously.[1]

Next-Generation Sequencing

Next-generation sequencing of the *APC* gene was performed. Most (n = 538) of the lesions were sequenced for *APC* only, and 334 lesions were sequenced using a custom-designed mosaic colorectal cancer panel. This panel includes 20 colorectal cancer and polyposis-associated genes, but also hotspots of the *CTNNB1* gene. The AmpliSeq (ThermoFisher Scientific, Waltham, MA) NGS libraries were prepared following manufacturer's instructions. In short, 2 AmpliSeq primer pools were prepared, isolated DNA was added, and a first amplification polymerase chain reaction was performed. Next, the 3 primer pools were combined, after which the primers were partly digested during a second polymerase chain reaction. To ligate the barcodes, another polymerase chain reaction run was performed. Lastly, the libraries were purified with AMPureXP beads (Beckman Coulter Life Sciences, Indianapolis, IN) and the 2 pools were combined. After loading the samples on the chip using the Ion Chef (ThermoFisher Scientific), sequencing was performed in an Ion GeneStudio S5 Series sequencer (ThermoFisher Scientific).

Data Analysis

The sequencer's output of unaligned reads was mapped against the human reference genome (hg19) using Burrows-Wheeler aligner. Multiple different softwares (VarScan, ANNOVAR, and Integrative Genomics Viewer) were used for variant calling, annotation, and visualization of the alignment and variants, respectively. Whenever variant interpretation was desired, Alamut software (Interactive Biosoftware, Louen, France) was used.

Data Interpretation

Detected variants were categorized by pathogenicity: 1 = benign, 2 = likely benign, 3 = uncertain significance, 4 = likely pathogenic and 5 = pathogenic using the Leiden Open Variant Database and ClinVar. Variants were reported whenever the read count was at least 100 and variant allele frequency at least 10%. Loss of heterozygosity was determined based on allele frequencies of heterozygous single nucleotide polymorphisms and, when present, the (pathogenic) variant.

Reference

1. van Eijk R, et al. Exp Mol Pathol 2013;94:121–125.

Supplementary Table 1. Patients With the c.835-8A>G Variant in at Least 1 Lesion

Variable	Patient	CRC age, *y*	Ad[a]	Age, *y*[b]	c.835-8A>G[c]	n[d]	Other *pks*+ *E coli* variants, n[e]	Total *pks*+ *E coli* variants, n [f] (%)
Mosaic	10	69	25	69	3/3; 1x CRC	0/3 (Leu, Ur and BS)	—	3/3 (100)
	165	—	3	25	3/3	0/4 (C, Leu, Ur and BS)	—	3/3 (100)
Hybrid	3	42[g]	22	69	2/7	—	—	2/7 (28.6)
	11	—	15	50	3/7	—	3/7	6/7 (85.7)
	12	66[h]	13	65	2/9	0/4 (C)	—	2/9 (22.2)
	15	—	12	59	3/12	—	—	3/12 (25)
	17	66	70	66	3/10; 1x CRC	0/2 (Leu)	—	3/10 (30)
	21	—	27	63	2/10	—	1/10	3/10 (30)
	96	2x 65[i]	4	65	3/5	—	—	3/5 (60)
	122	55	10	55	2/4; 1x CRC	—	1/4	3/4 (75)
	125	4x 55	2	55	2/4; 2x CRC	0/4 (C)	—	2/4 (50)
	128	—	10	52	6/10	0/5 (2x C, Leu, Ur and BS)	2/10	8/10 (80)
	152	—	28	81	2/6	0/1 (C)	—	2/6 (33.3)
	156	—	36	74	2/4	—	—	2/4 (50)
	168	—	17	55	4/7	—	—	4/7 (57.1)
	174	—	10	54	2/6	—	—	2/6 (33.3)
	194	—	22	69	3/7	—	—	3/7 (42.9)
	198	69[g]	30	69	2/4	—	—	2/4 (50)
No mosaic	41	2x 62[h]	4	62	1/6	—	1/6	2/6 (33.3)
	56	—	22	75	1/3	—	—	1/3 (33.3)
	65	72[g]	24	71	1/3	—	—	1/3 (33.3)
	88	—	21	59	1/6	—	—	1/6 (16.7)
	120	—	30	59	1/4	—	—	1/4 (25)
	127	73	30	74	1/4; 1x CRC	—	—	1/4 (25)
	132	—	15	65	1/8	—	1/8	2/8 (25)

Supplementary Table 1. Continued

Variable	Patient	CRC age, *y*	Ad[a]	Age, *y*[b]	c.835-8A>G[c]	n[d]	Other *pks*+ *E coli* variants, n[e]	Total *pks*+ *E coli* variants, n [f] (%)
	149	2x 64, 70	8	64	1/3; 1x CRC	0/1 (C)	—	1/3 (33.3)
	158	65[h]	28	65	1/3	—	1/3	2/3 (66.7)
	177	—	16	65	1/4	—	—	1/4 (25)
	178	2x 69[h]	>20	69	1/4	—	—	1/4 (25)
	182	—	11	55	1/3	—	—	1/3 (33.3)
	184	—	10	37	1/4	—	—	1/4 (25)
	188	—	14	42	1/7	0/3 (Leu, Ur, and BS) (c.4348C>T as hybrid mutation)	—	1/7 (14.3)
	190	—	29	83	1/4	—	—	1/4 (25)
	191	—	23	67	1/4	—	1/4	2/4 (50)
	202	—	14	63	1/6	—	—	1/6 (16.7)
	204	—	9	59	1/3	—	1/3	2/3 (66.7)

BS, buccal swab; C, normal colon mucosa; CRC, colorectal cancer; Leu, leukocyte; Ur, urine.
[a]Cumulative number of adenomas.
[b]Age at first adenoma.
[c]Number of lesions with c.835-8A>G variant/total number of lesions tested; number of CRC with c.835-8A>G.
[d]Number of normal tissue with c.835-8A>G variant/total number of normal tissue tested.
[e]Number of lesions with 1 of the other 7 recurrent variants fulfilling the *pks*+ *E coli* mutational signature/total number of lesions tested (Supplementary Table 2).
[f]Number of lesions with a *pks*+ *E coli* variant/total number of lesions tested.
[g]CRC not sequenced for *APC*.
[h]CRC sequenced for *APC* but negative for the c.835-8A>G variant.
[i]Only 1 CRC sequenced for *APC* but negative for the c.835-8A>G variant.

Supplementary Table 2. Recurring *APC* Variants in Nonmosaic Colorectal Lesions

Variant (NM_000038.5:)	n	%	*pks+ E coli* mutational signature?
c.835-8A>G	61	7.0	Yes
c.4348C>T	51	5.9	No
c.2626C>T	34	3.9	No
c.646C>T	22	2.5	No
c.637C>T	21	2.4	No
c.694C>T	19	2.2	No
c.847C>T	15	1.7	No
c.1495C>T	12	1.4	No
c.4393_4394delAG	12	1.4	No
c.4099C>T	11	1.3	No
c.3964G>T	9	1.0	No
c.3927_3931delAAAGA	8	0.9	No
c.1690C>T	8	0.9	No
c.2413C>T	8	0.9	No
c.3340C>T	8	0.9	No
c.994C>T	6	0.7	No
c.4391_4394delAGAG	6	0.7	No
c.4630G>T	6	0.7	No
c.3493A>T	6	0.7	Yes
c.4067C>G	6	0.7	No
c.1660C>T	5	0.6	No
c.2804dupA	5	0.6	No
c.1213C>T	5	0.6	No
c.2805C>A	5	0.6	No
c.4245delT	5	0.6	No
c.4285C>T	5	0.6	No
c.6363_6365dupTGC	4	0.5	No
c.4063delT	4	0.5	Yes
c.1409-5A>G	4	0.5	No
c.1168A>G	4	0.5	No
c.1600A>T	4	0.5	Yes
c.3862G>T	4	0.5	No
c.3991A>T	4	0.5	No
c.4108A>T	4	0.5	No
c.4501delT	4	0.5	Yes
c.4747A>G	4	0.5	No
c.7490C>T	4	0.5	No
c.904C>T	4	0.5	No
c.4390G>T	3	0.3	No
c.645+1G>A	3	0.3	No
c.1312+1G>A	3	0.3	No
c.1548G>T	3	0.3	No
c.2008A>T	3	0.3	Yes
c.2821G>T	3	0.3	No
c.3441C>G	3	0.3	No
c.3916G>T	3	0.3	No
c.3934G>T	3	0.3	No
c.4053dupT	3	0.3	No
c.4057G>T	3	0.3	No
c.4063dupT	3	0.3	No
c.4233delT	3	0.3	No
c.4463delT	3	0.3	No
c.4666dupA	3	0.3	No
c.646-1G>A	2	0.2	No
c.6934C>A	2	0.2	No
c.4067C>A	2	0.2	No
c.3982C>T	2	0.2	No
c.1307delA	2	0.2	Yes
c.1548G>A	2	0.2	No
c.1968_1969delAA	2	0.2	No
c.423-6A>G	2	0.2	Yes
c.4460_4464delCTTTA	2	0.2	No
c.1548+2T>C	2	0.2	No
c.1960C>T	2	0.2	No
c.3175G>T	2	0.2	No
c.4128T>A	2	0.2	No
c.4189G>T	2	0.2	No
c.4459dupA	2	0.2	No
c.4611_4612delAG	2	0.2	No
c.4634C>A	2	0.2	No
c.509_512delATAG	2	0.2	No
c.591_592delAG	2	0.2	No
c.6868T>C	2	0.2	No
c.1234C>T	2	0.2	No
c.1312+5G>A	2	0.2	No
c.1333C>T	2	0.2	No
c.1411G>A	2	0.2	No

Supplementary Table 2. Continued

Variant (NM_000038.5:)	n	%	*pks+ E coli* mutational signature?
c.1744-1G>A	2	0.2	No
c.1958+1_1958+2dupGT	2	0.2	No
c.2205delG	2	0.2	No
c.2222delA	2	0.2	No
c.2364_2365delGCinsAT	2	0.2	No
c.2741_2742delGTinsAG	2	0.2	No
c.2932C>T	2	0.2	No
c.3030delT	2	0.2	No
c.3193C>T	2	0.2	No
c.3289G>T	2	0.2	No
c.3471_3474delGAGA	2	0.2	No
c.3682C>T	2	0.2	No
c.3907C>T	2	0.2	No
c.3956delC	2	0.2	No
c.4135G>T	2	0.2	No
c.4216C>T	2	0.2	No
c.4260_4261delCA	2	0.2	No
c.4271delC	2	0.2	No
c.4300delA	2	0.2	No
c.4429C>T	2	0.2	No
c.4485delT	2	0.2	No
c.4549C>T	2	0.2	No
c.4612_4613delGA	2	0.2	No
c.4655_4656delAG	2	0.2	No
c.5626A>G	2	0.2	No
c.5937delC	2	0.2	No
c.70C>T	2	0.2	No
c.757_761dupGGCTC	2	0.2	No
c.7835G>A	2	0.2	No
c.8416C>G	2	0.2	No
c.933+1G>A	2	0.2	No