Exploring APC mosaicism: prevalence, clinical consequences and underlying causes

Terlouw, D.

Exploring *APC* mosaicism; prevalence, clinical consequences and underlying causes

Diantha Terlouw

ISBN: 978-94-6483-441-3

Artwork, cover design and lay-out: Wouter Terlouw

Printed by: Ridderprint, the Netherlands

The research presented in this dissertation was performed at the departments of Pathology and Clinical Genetics of the Leiden University Medical Center and was financially supported by the Dutch Cancer Society (project number 11292)

Exploring *APC* mosaicism; prevalence, clinical consequences and underlying causes

Proefschrift

ter verkrijging van

de graad van doctor aan de Universiteit Leiden,

op gezag van rector magnificus prof.dr.ir. H. Bijl,

volgens besluit van het college voor promoties

te verdedigen op dinsdag 21 november 2023

klokke 13.45 uur

door

Diantha Terlouw

geboren te Zwijndrecht

in 1995

Promotor

Prof. Dr. H. Morreau

Co-promotores

Dr. M. Nielsen

Dr. J.T. van Wezel

Leden promotiecommissie

Prof. Dr. M.E. van Leerdam

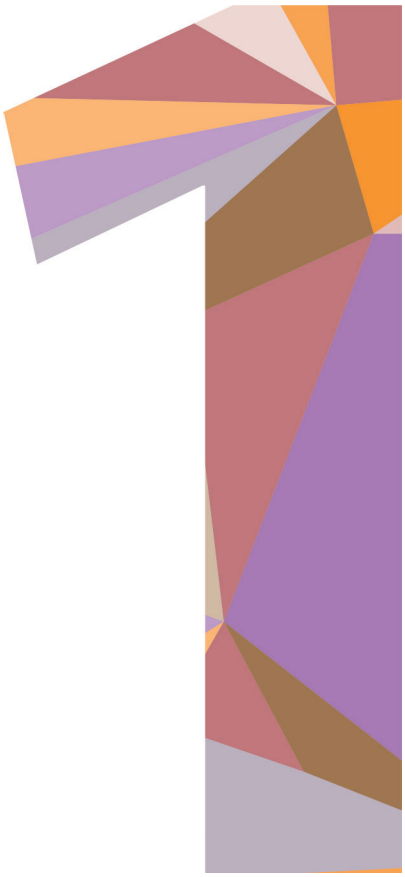
Prof. Dr. F.J. Hes (Universitair Ziekenhuis Brussel)

Prof. Dr. R.H. Sijmons (Groningen Universitair Medisch Centrum)

Dr. R.M. de Voer (Radboud Universitair Medisch Centrum)

Table of contents

Chapter 1	General introduction	7
I.	**Germline *APC* and biallelic *MUTYH* variants**	
Chapter 2	Declining detection rates for *APC* and biallelic *MUTYH* variants in polyposis patients, implications for DNA testing policy *European Journal of Human Genetics (2020)*	23
II.	***APC* mosaicism**	
Chapter 3	Prevalence and consequences of *APC* mosaicism in patients with colorectal adenomas *Manuscript in preparation*	41
Chapter 4	*APC* mosaicism, not always isolated: two first-degree relatives with apparently distinct *APC* mosaicism *Gut (2022)*	65
Chapter 5	Exploring *APC* mosaicism in upper intestinal tract adenomas *Submitted at Clinical Gastroenterology and Hepatology*	71
III.	**Colibactin**	
Chapter 6	Recurrent *APC* splice variant c.835-8A>G in patients with unexplained colorectal polyposis fulfilling the colibactin mutational signature *Gastroenterology (2020)*	87
Chapter 7	Enrichment of colibactin-associated signatures in unexplained colorectal polyposis patients *Under review at BMC Cancer*	95
Chapter 8	Colibactin mutational singatures in *NTHL1* Tumor Syndrome and *MUTYH* Associated Polyposis patients *Accepted at Genes, Chromosomes and Cancer*	109
Chapter 9	General discussion	123
Chapter 10	Nederlandse samenvatting	135
	List of publications	141
	Dankwoord	143
	Curriculum Vitae	145



General introduction

Colorectal cancer (CRC) is the third most commonly diagnosed cancer and second leading cause in cancer-associated death according to the GLOBOCAN database of the World Health Organization.[1] The CRC incidence was 1.88 million cases worldwide in 2020 and is estimated to increase to more than 3 million cases in the following 20 years.[2]

The vast majority of CRCs develop from pre-cancerous polyps, which is referred to as the adenoma-to-carcinoma sequence. Colorectal polyps are overgrowths of colorectal mucosa and are histologically classified in multiple categories.[3] Adenomatous polyps (adenomas) are the most common colorectal polyps. In the general population, approximately 25% of people aged 50 and 50% of people aged 70 are diagnosed with at least one colorectal adenoma.[4] Adenomas are further classified in three microscopical subtypes; tubular, tubulovillous and villous adenomas. This classification is based on the growth pattern, where tubulovillous is a mixture of a tubular and villous growth pattern. Villous components, high grade dysplasia or any type of adenoma bigger than 10 mm classifies advanced adenomas which influences screening and treatment decisions.[5] As suggested by the adenoma-to-carcinoma sequence, a simplification depicted in figure 1, adenomas have potential to develop into a carcinoma.[6] The transformation from normal colorectal mucosa to colorectal carcinoma often starts with two hits in the *Adenomatous Polyposis Coli (APC)* gene, as more than 70% of colorectal adenomas harbor pathogenic *APC* variants.[7]

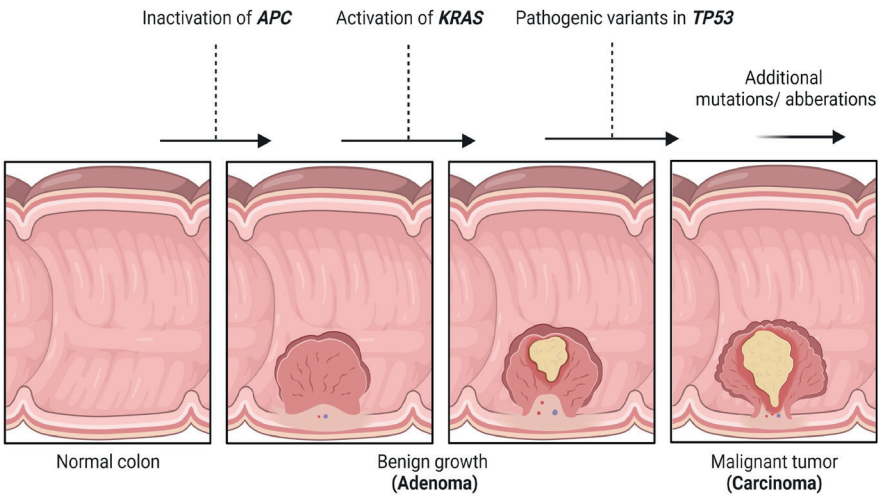


Figure 1. Adenoma-to-carcinoma sequence. Transformation from normal colon to carcinoma often starts with inactivation of *APC*, leading to activation of the Wnt pathway. Usually, the sequence continues with activating variants in *KRAS* and pathogenic variants in *TP53* which is commonly followed by additional variants and chromosomal instability. Created with Biorender.com.

As depicted in figure 2, in a normal situation, APC contributes to the canonical Wnt signaling pathway by forming a β-catenin destruction complex.[8] In absence of a Wnt ligand, this complex consisting of APC, Axin, GSK3β and kinase CK1, phosphorylates β-catenin and tags it to be degraded. In presence of Wnt ligand binding, the destruction complex is inhibited by Dishevelled. β-catenin then accumulates in the cytoplasm, translocates into the nucleus and transcribes genes promoting cell proliferation and survival.

Inactivation of APC inhibits the destruction complex to form, causing β-catenin to transcribe genes even in absence of Wnt ligand leading to uncontrolled cell proliferation and survival.[9] Subsequently, the adenoma-to-carcinoma sequence usually continues with pathogenic variants in the *KRAS* and *TP53* genes which is commonly followed by chromosomal instability.[10]

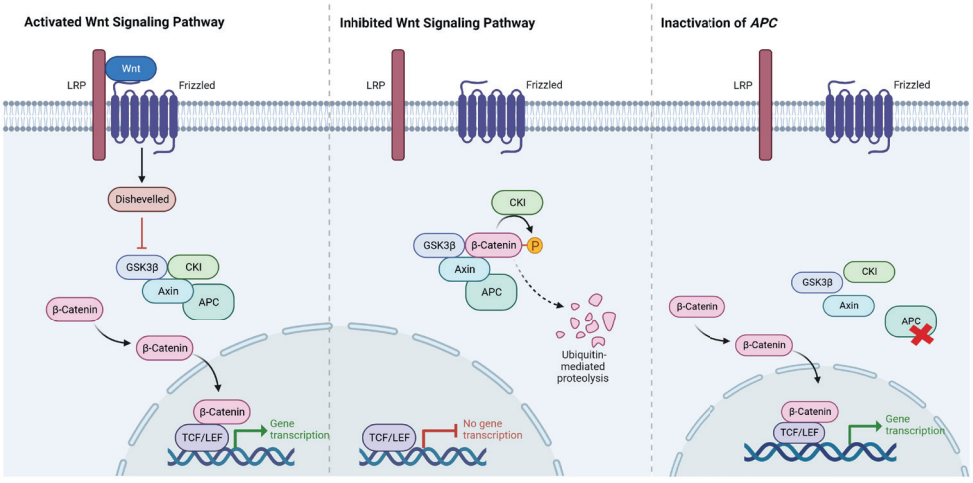


Figure 2. Canonical Wnt pathway. Left. The Wnt pathway is activated when Wnt ligand binds to the Frizzled and LRP (co-)receptor. This binding leads to recruitment of Dishevelled, inhibiting the β-catenin destruction complex. Now, β-catenin accumulates in the cytoplasm and translocates into the nucleus where it transcribes target genes. Middle. In absence of Wnt ligand, the β-catenin destruction complex tags β-catenin for ubiquitination, regulating β-catenin and gene transcription. Right. Inactivation of *APC* leads to an activated Wnt pathway in absence of Wnt ligand. Created with Biorender.com.

The second most common colorectal polyps are serrated lesions, consisting of three microscopical subtypes; hyperplastic, sessile serrated polyps and traditional serrated adenomas.[11] Although serrated lesions were for a long time thought to be benign, especially large sessile serrated lesions can develop into colorectal cancer.[12] Most sporadic left sided serrated lesions have, just like most adenomas, a *KRAS* variant activating the MAPK pathway leading to uncontrolled cell growth, proliferation, survival and migration. On the other hand, most sporadic right sided serrated lesions have a *BRAF* variant activating the MAPK pathway. This BRAF activation commonly occurs combined with an aberrant methylation of CpG islands (CIMP+) and *MLH1* promoter hypermethylation resulting in a higher potential of developing

into a colorectal carcinoma.[13] Other categories of colorectal polyps, like juvenile, hamartomatous polyps, are less common.[3]

Genetic factors, including high risk susceptibility genes and CRC-associated single nucleotide polymorphisms (SNPs), as well as environmental factors, like modifiable lifestyle factors, are of etiological significance in the development of colorectal cancer and (adenomatous) polyps.[14-16]

Genetic predisposition

The genetic significance is suggested by the estimation that in about 30% of all colorectal cancer patients a positive family history for CRC has been reported.[17] Besides the high risk susceptibility genes described below, a genetical background increasing the risk for colorectal cancer might also be explained through clustering of low-risk variants.[18]

(Attenuated) Familial Adenomatous Polyposis ((A)FAP; OMIM 175100)

With nearly 70% of all polyposis patients, the most common inherited polyposis syndrome is Familial Adenomatous Polyposis (FAP). FAP accounts for about 1% of all colorectal cancers. FAP patients harbor a pathogenic germline variant in the *APC* gene and classically develop hundreds to even thousands of colorectal polyps.[19] The polyp development in FAP patients starts from early adolescence and leads, with a penetrance of 100%, to colorectal cancer with a mean age of 39 when left untreated. Attenuated FAP (AFAP) is characterized with the development of tens to hundreds of adenomas. These patients typically develop colorectal cancer at the age of 50 to 55 with a cumulative risk of about 70%.[20] Moreover, extracolonic manifestations as stomach and duodenal polyps and carcinomas, osteomas and desmoid tumors are associated with (A)FAP.[21 22]

MUTYH-Associated Polyposis (MAP; OMIM 608456)

Approximately 15% of polyposis patients negative for a pathogenic *APC* variant have pathogenic bi-allelic *MUTYH* germline variants.[23] The MUTYH enzyme plays a key role in base excision repair (BER), involved in repairing oxidative DNA damage. As a result of oxidative DNA damage, guanine adducts 8-oxo-7,8-dihydroxy-2'-deoxyguanosine (8-oxo-dG) are formed which tends to pair with adenine instead of cytosine leading to G:C>T:A transversions. BER prevents mutagenesis by the 8-oxo-dG in which MUTYH is important for the recognition and excision of the mis-incorporated adenine opposite to the 8-oxo-dG.[24] Malfunctioning of MUTYH leads to accumulation of GAA>TAA transversions, characterized as single base substitutions (SBS) signatures 18 and 36 in the Catalogue of Somatic Mutations in Can-

cer (COSMIC) database.[23, 25, 26] More specifically 40% to 100% colorectal adenomas and 60% to 90% of colorectal carcinomas of MAP patients harbor the *KRAS* pathogenic variant NM_004985.5:c.34G>T p.Gly12Cys, fitting these signatures.[26, 27]

MUTYH-associated polyposis (MAP) patients develop tens to a hundred adenomas. In the absence of adequate surveillance colonoscopies, the lifetime risk of developing CRC is 43% to 63%.[28] Furthermore, MAP patients are at risk of developing stomach and duodenal polys and carcinomas.[29-31]

NTHL1 Tumor Syndrome (NTS; OMIM 616415)

With just 33 individuals from 20 families described, a much more rare colorectal polyposis related syndrome is caused by pathogenic biallelic *NTHL1* germline variants. NTHL1 is, like MUTYH, involved in the base excision repair pathway as a DNA glycolase removing damaged nucleotides. NTHL1 is specifically important in oxidized pyrimidines. Due to the lack of a functioning NTHL1 protein, C>T transitions will accumulate primarily at non-CpG sites, characterized in SBS30.[32, 33]

Most *NTHL1* Tumor Syndrome (NTS) patients develop adenomatous polyposis, ranging from 1 to 100 adenomas. Some patients also developed sessile serrated lesions, showing a mixed polyposis phenotype.[34] Furthermore, patients with NTS are described with other malignancies like various types of breast cancer and multiple duodenal polyps and cancer.[34, 35] The prevalence of NTS is unknown, but has been estimated to be 1 in 115,000.[35]

Other colorectal polyposis or cancer syndromes

The most prevalent heritable syndrome increasing the lifetime risk of CRC is Lynch Syndrome caused by a pathogenic variant in one of the mismatch repair (MMR) genes (*MLH1, MSH2, MSH6* or *PMS2*) or *EPCAM*.[36] Lynch syndrome is characterized by an increased risk for mainly colorectal and endometrial carcinomas and no or a small number of colorectal polyps.[37]

Other, more rare, adenomatous polyposis syndromes are caused by pathogenic variants in *POLE, POLD1*, *MSH3* and *MBD4*.[38-40] Patients with pathogenic germline variants in *RNF43, SMAD4, BMPR1A, ENG, STK11* and *PTEN* are characterized with other types of colorectal polyps.[41] Multiple other genes, like *TP53*, are linked to colorectal cancer in a multiple tumor syndrome manner.[42]

Mosaicism

Besides pathogenic germline variants, mosaicism is described as an additional explanation for the development of multiple colorectal adenomas.[43] Mosaicism means a variant present

in only a subset of body cells and arises whenever a *de novo* variant occurs after the first few embryonic cell divisions. Somatic mosaicism refers to genetically distinct cell populations presenting in only somatic tissue whereas in germline mosaicism these populations are only present in the germ cells. In case of germline or parental, the individual will be phenotypically normal, however, due to the presence of the variant in the germ cells, the variant might

be transmitted to offspring. A combination, gonosomal mosaicism; mosaic variant in both somatic and germ cells, happens whenever the variant occurs before primordial germ cells differentiation.[44, 45]

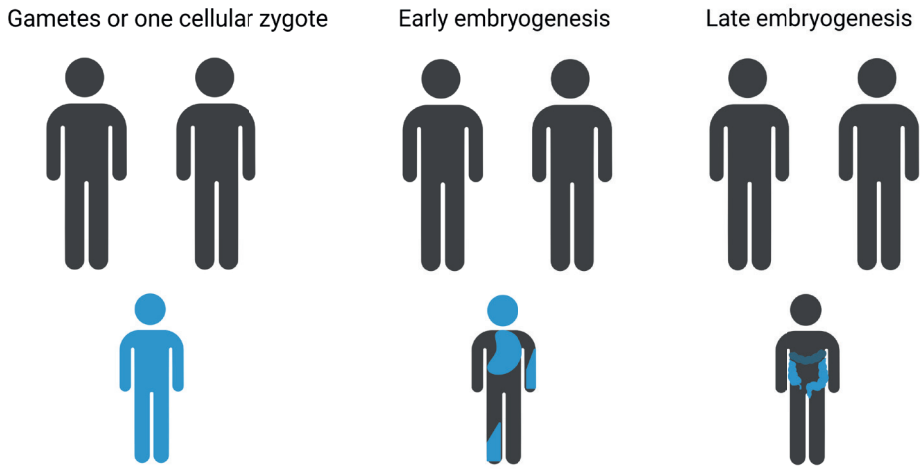


Figure 3. Timing of *de novo* variants and mosaicism. Whenever a variant occurs in the gametes of the parents or at one cellular stage of the zygote, all body cells will be affected. Whenever the variant occurs during early embryogenesis different tissues throughout the body can be affected while others are not. It is also possible that the variant occurs during late embryogenesis and causes only the large intestine affected. Created with Biorender.com and adapted from Freed et al.[44]

APC mosaicism

Despite the clear autosomal dominant inheritance, 20% to 25% of *APC* variant carriers present with a negative family history, suggesting a high proportion of *de novo* variants in FAP.[46, 47] As described above and shown in figure 3, depending on the exact temporal occurrence, all or only a subset of body cells will be affected.

As reviewed by Jansen and Goel the first *APC* mosaicism cases were already described in 1994 and 1999.[43, 48, 49] Following the description of single cases, multiple studies determined that about 20% of *de novo* patients have a mosaicism.[46, 50, 51] These studies focused on detection of low variant allele frequency variants in leukocyte DNA. Later studies used Next Generation Sequence (NGS) analysis in DNA isolated from adenomas and identified *APC*

mosaicism in 25% to 50% of unexplained polyposis coli patients.[52, 53] These mosaic variants were likely missed in regular diagnostic germline testing. Phenotypically mosaic cases are, in general, milder than germline variant carriers suggesting that *APC* mosaicism might also play a role in patients with less than 20 adenomas or developing adenomas aged above 60 or 70. These patients do not fall within the scope of the Dutch hereditary polyposis guidelines and are therefore usually not tested for pathogenic *APC* variants.[54]

Other colorectal polyposis and cancer associated genes and mosaicism

Mosaicism is also described in other colorectal cancer and polyposis associated genes like *MLH1*, *MSH2*, *PTEN* and *STK11*.[43, 55-58] Though expected less frequent than *APC* mosaicism due to the *de novo* variant frequency, prevalence of mosaicism in these and other CRC and polyposis associated genes remain largely unknown.[59]

Environmental factors

CRC is more prevalent in western industrialized countries compared to developing countries. This indicates that, besides genetic factors, environmental factors, like a western lifestyle, might play a relevant role in the development of colorectal adenomas and CRC.

Modifiable environmental risk factors

Multiple modifiable environmental risk factors are associated with increased risk for developing CRC.[60] Tobacco consumption, for example, is thought to increase the risk in a dose dependent manner while quitting smoking reduces the risk again. The risk is specifically increased for microsatellite instable tumors with CpG island methylator phenotype and BRAF variants.[61] Moreover, ethanol is metabolized to acetaldehyde which can cause DNA damage and methylation and accumulates in the colon, suggesting alcohol consumption as another modifiable risk factor.[62] It is predicted that a daily alcohol consumption increases the risk with up to 30%, mainly for microsatellite stable tumors.[63] Another association with lifestyle and CRC consumption of red and processed meat. A high intake increases the risk with 20% to 30%. It is hypothesized that this is mostly caused by the high cooking temperature promoting carcinogenesis.[64, 65] On the other hand, physical activity is thought to decrease the risk with up to 25%.[66] Physical activity furthermore decreases overweight and obesity, which in itself increase the risk of early onset colorectal cancer with 32% and 88% compared to healthy weighted individuals.[67]

Gut microbiome

Modifiable lifestyle factors are all also known to influence the gastrointestinal microbiome.[68]

The last decade the interest in the association of the gut microbiome and development of CRC has increased.[69-72] Studies comparing bacterial diversity between healthy individuals and CRC patients show significant enrichment for specific bacteria like *Fusobacterium nucleatum*, *Bacteroides Fragilis* and *Escherichia coli* while other bacteria, like *Bifidobacteria*, are decreased in CRC patients.[73-75] Also, a difference in gut microbiota between early and advanced stage CRC patients have been described.[73]

Specific bacteria are shown to drive tumorigenesis in different ways, for example causing inflammation or excreting toxins causing DNA damage. *Salmonella* and *F. nucleatum,* for example, excrete proteins, AvrA and FadA respectively, which induce the Wnt pathway and therefore enhance tumorigenesis and increase the risk of developing CRC after infection.[76-78] Another, specific *E. coli,* bacterium produce a genotoxin called colibactin. These colibactin-encoding *E. coli* are enriched in CRC and FAP patients suggesting its carcinogenic potential.[14, 79, 80]

E. coli harboring pks island

E. coli is a prevalent bacterium, colonizing the gastrointestinal tract of infants already a short time after birth.[81] These commensal bacteria are usually not pathogenic and harmless to humans. However, a few *E. coli* subtypes have acquired specific pathogenic features leading to diarrheal diseases.[82] As described above, others have acquired the ability to produce colibactin. These colibactin-encoding *E. coli* (*pks*$^+$ *E. coli*) increase colorectal tumor burden in mouse models and are detected in approximately 60% of CRC patients compared to 20% of healthy individuals.[14, 79, 80]

Pks$^+$ *E. coli* harbor the *polyketide synthase* (*pks*) gene island which encodes for the necessary equipment to produce colibactin.[83] In vitro studies showed that colibactin induces DNA crosslinks, double-strand breaks and chromosome aberrations.[79, 80, 83] Comparing intestinal organoids infected with *pks*$^+$ *E. coli* with organoids infected with *E.coli* with an impaired *pks* island characterized a specific colibactin-associated mutational signature.[84] This mutational signature is enriched for T>N mutations with an adenine 3 base pairs at the 5' side and single thymine deletions located in T homopolymers with 2 to 4 adenines to the 5' side depending on polymer length. These single base substitution and indel signatures are included in the COSMIC database as SBS88 and ID18 respectively and are detected in colorectal and oral squamous cell carcinomas.[84, 85] Remarkably, the mutational signature is also found in normal colon mucosa. Therefore, it is hypothesized that colibactin-associated DNA damage occurs early in life and acts as a driver in the CRC tumorigenesis later in life.[86] *E. coli* is not the only bacterium able to produce colibactin since the *pks* island is also detected in *Klebsiella pneumoniae*[87-89], *Enterobacter aerogenes* and *Citrobacter koseri.*[88]

Outline of this thesis

Part I of this thesis provides insights into the prevalence of *APC* and biallelic *MUTYH* germline variants. In **chapter 2** all patients tested for *APC* and *MUTYH* in the Leiden University Medical Center between 1992 and 2017 were collected. Using this cohort guidelines for *APC* and *MUTYH* testing were suggested. Moreover, this chapter shows a substantial proportion of colorectal adenomatous polyposis patients remaining unexplained using regular diagnostic genetic testing.

Part II dives into the prevalence and importance of *APC* mosaicism in this group of unexplained polyposis patients. **Chapter 3** describes the prevalence of *APC* mosaicism in our entire cohort containing various polyposis phenotypes groups. These results are used to provide *APC* mosaicism testing and surveillance guidelines. **Chapter 4** illustrates a remarkable family with two *APC* mosaicism cases underlining the importance of *APC* mosaicism testing in unexplained polyposis patients. **Chapter 5** emphasizes the importance of upper intestinal tract surveillance guidelines for *APC* mosaicism cases. It furthermore describes the possibility of *APC* mosaicism solitary to the duodenum.

Part III evaluates the influence of colibactin in the polyposis cohort. **Chapter 6** presents a commonly detected *APC* variant, c.835-8A>G, which fits the colibactin-associated mutational signature. This finding led to **chapter 7** in which additional fecal and whole genome analyses were performed to further assess the presence of colibactin and its signature in the cohort. Furthermore, **chapter 8** presents a case of *NTHL1* Tumor Syndrome with evidence of colibactin. In this report colibactin was also evaluated in a MAP cohort.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71:209-249.
2. Ervik M LF, Laversanne M, Ferlay J, Bray F. Global Cancer Observatory: Cancer Over Time. Volume 2022. Lyon, France: International Agency for Research on Cancer, 2021.
3. Miller Q, Saeed O, Mesa H. Clinical, Pathologic, and Molecular-Genetic Aspects of Colorectal Polyps. Gastrointest Endosc Clin N Am 2022;32:313-328.
4. van Hees F, Saini SD, Lansdorp-Vogelaar I, et al. Personalizing colonoscopy screening for elderly individuals based on screening history, cancer risk, and comorbidity status could increase cost effectiveness. Gastroenterology 2015;149:1425-37.
5. Rex DK, Boland CR, Dominitz JA, et al. Colorectal Cancer Screening: Recommendations for Physicians and Patients From the U.S. Multi-Society Task Force on Colorectal Cancer. Gastroenterology 2017;153:307-323.
6. Brenner H, Kloor M, Pox CP. Colorectal cancer. Lancet 2014;383:1490-1502.
7. Kinzler KW, Vogelstein B. Lessons from hereditary colorectal cancer. Cell 1996;87:159-70.
8. Clevers H. Wnt/beta-catenin signaling in development and disease. Cell 2006;127:469-80.
9. Bugter JM, Fenderico N, Maurice MM. Mutations and mechanisms of WNT pathway tumour suppressors in cancer. Nat Rev Cancer 2021;21:5-21.
10. Fearon ER. Molecular genetics of colorectal cancer. Annu Rev Pathol 2011;6:479-507.
11. Crockett SD, Nagtegaal ID. Terminology, Molecular Features, Epidemiology, and Management of Serrated Colorectal Neoplasia. Gastroenterology 2019;157:949-966 e4.
12. Choi EY, Appelman HD. A Historical Perspective and Expose on Serrated Polyps of the Colorectum. Arch Pathol Lab Med 2016;140:1079-84.
13. Kim KM, Lee EJ, Ha S, et al. Molecular features of colorectal hyperplastic polyps and sessile serrated adenoma/polyps from Korea. Am J Surg Pathol 2011;35:1274-86.
14. Dejea CM, Fathi P, Craig JM, et al. Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. Science 2018;359:592-597.
15. Peters U, Sinha R, Chatterjee N, et al. Dietary fibre and colorectal adenoma in a colorectal cancer early detection programme. Lancet 2003;361:1491-5.
16. 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.
17. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 2000;343:78-85.
18. Middeldorp A, Jagmohan-Changur S, van Eijk R, et al. Enrichment of low penetrance susceptibility loci in a Dutch familial colorectal cancer cohort. Cancer Epidemiol Biomarkers Prev 2009;18:3062-7.
19. Kinzler KW, Nilbert MC, Su LK, et al. Identification of FAP locus genes from chromosome 5q21. Science 1991;253:661-5.
20. Jasperson KW, Patel SG, Ahnen DJ. APC-Associated Polyposis Conditions. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Stephens K, Amemiya A, eds. GeneReviews((R)). Seattle (WA), 1993.
21. Bulow S, Alm T, Fausa O, et al. Duodenal adenomatosis in familial adenomatous polyposis. DAF Project Group. Int J Colorectal Dis 1995;10:43-6.
22. Nieuwenhuis MH, Mathus-Vliegen EM, Baeten CG, et al. Evaluation of management of des-

moid tumours associated with familial adenomatous polyposis in Dutch patients. Br J Cancer 2011;104:37-42.

23. 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.
24. Michaels ML, Tchou J, Grollman AP, et al. A repair system for 8-oxo-7,8-dihydrodeoxyguanine. Biochemistry 1992;31:10964-8.
25. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. Nature 2013;500:415-21.
26. Viel A, Bruselles A, Meccia E, et al. A Specific Mutational Signature Associated with DNA 8-Oxoguanine Persistence in MUTYH-defective Colorectal Cancer. EBioMedicine 2017;20:39-49.
27. Nielsen M, Morreau H, Vasen HF, et al. MUTYH-associated polyposis (MAP). Crit Rev Oncol Hematol 2011;79:1-16.
28. Lubbe SJ, Di Bernardo MC, Chandler IP, et al. Clinical implications of the colorectal cancer risk associated with MUTYH mutation. J Clin Oncol 2009;27:3975-80.
29. Nielsen M, Poley JW, Verhoef S, et al. Duodenal carcinoma in MUTYH-associated polyposis. J Clin Pathol 2006;59:1212-5.
30. Vogt S, Jones N, Christian D, et al. Expanded extracolonic tumor spectrum in MUTYH-associated polyposis. Gastroenterology 2009;137:1976-85 e1-10.
31. Collaborative Group on Duodenal Polyposis in MAP, Thomas LE, Hurley JJ, et al. Duodenal Adenomas and Cancer in MUTYH-associated Polyposis: An International Cohort Study. Gastroenterology 2021;160:952-954 e4.
32. Alexandrov LB, Jones PH, Wedge DC, et al. Clock-like mutational processes in human somatic cells. Nat Genet 2015;47:1402-7.
33. Drost J, van Boxtel R, Blokzijl F, et al. Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. Science 2017;358:234-238.
34. 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.
35. 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.
36. Kohlmann W, Gruber SB. Lynch Syndrome. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Stephens K, Amemiya A, eds. GeneReviews((R)). Seattle (WA), 1993.
37. Rijcken FE, Hollema H, Kleibeuker JH. Proximal adenomas in hereditary non-polyposis colorectal cancer are prone to rapid malignant transformation. Gut 2002;50:382-6.
38. 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.
39. 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.
40. Palles C, West HD, Chew E, et al. Germline MBD4 deficiency causes a multi-tumor predisposition syndrome. Am J Hum Genet 2022;109:953-960.
41. Ngeow J, Heald B, Rybicki LA, et al. Prevalence of germline PTEN, BMPR1A, SMAD4, STK11, and ENG mutations in patients with moderate-load colorectal polyps. Gastroenterology 2013;144:1402-9, 1409 e1-5.
42. Rengifo-Cam W, Shepherd HM, Jasperson KW, et al. Colon Pathology Characteristics in

Li-Fraumeni Syndrome. Clin Gastroenterol Hepatol 2018;16:140-141.

43. Jansen AML, Goel A. Mosaicism in Patients With Colorectal Cancer or Polyposis Syndromes: A Systematic Review. Clin Gastroenterol Hepatol 2020;18:1949-1960.

44. Freed D, Stevens EL, Pevsner J. Somatic mosaicism in the human genome. Genes (Basel) 2014;5:1064-94.

45. Rahbari R, Wuster A, Lindsay SJ, et al. Timing, rates and spectra of human germline mutation. Nat Genet 2016;48:126-133.

46. Hes FJ, Nielsen M, Bik EC, et al. Somatic APC mosaicism: an underestimated cause of polyposis coli. Gut 2008;57:71-76.

47. Bisgaard ML, Fenger K, Bulow S, et al. Familial adenomatous polyposis (FAP): frequency, penetrance, and mutation rate. Hum Mutat 1994;3:121-5.

48. Farrington SM, Dunlop MG. Mosaicism and sporadic familial adenomatous polyposis. Am J Hum Genet 1999;64:653-8.

49. Mandl M, Kadmon M, Sengteller M, et al. A somatic mutation in the adenomatous polyposis coli (APC) gene in peripheral blood cells--implications for predictive diagnosis. Hum Mol Genet 1994;3:1009-11.

50. Aretz S, Stienen D, Friedrichs N, et al. Somatic APC mosaicism: a frequent cause of familial adenomatous polyposis (FAP). Hum Mutat 2007;28:985-992.

51. Out AA, Minderhout IJHMv, Stoep Nvd, et al. High-resolution melting (HRM) re-analysis of a polyposis patients cohort reveals previously undetected heterozygous and mosaic APC gene mutations. 14 2015.

52. 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.

53. 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.

54. Syngal S, Brand RE, Church JM, et al. ACG clinical guideline: Genetic testing and management of hereditary gastrointestinal cancer syndromes. Am J Gastroenterol 2015;110:223-62; quiz 263.

55. Pastrello C, Fornasarig M, Pin E, et al. Somatic mosaicism in a patient with Lynch syndrome. Am J Med Genet A 2009;149A:212-5.

56. Sourrouille I, Coulet F, Lefevre JH, et al. Somatic mosaicism and double somatic hits can lead to MSI colorectal tumors. Fam Cancer 2013;12:27-33.

57. Gammon A, Jasperson K, Pilarski R, et al. PTEN mosaicism with features of Cowden syndrome. Clin Genet 2013;84:593-5.

58. McKay V, Cairns D, Gokhale D, et al. First report of somatic mosaicism for mutations in STK11 in four patients with Peutz-Jeghers syndrome. Fam Cancer 2016;15:57-61.

59. Win AK, Jenkins MA, Buchanan DD, et al. Determining the frequency of de novo germline mutations in DNA mismatch repair genes. J Med Genet 2011;48:530-4.

60. Katsaounou K, Nicolaou E, Vogazianos P, et al. Colon Cancer: From Epidemiology to Prevention. Metabolites 2022;12.

61. Botteri E, Borroni E, Sloan EK, et al. Smoking and Colorectal Cancer Risk, Overall and by Molecular Subtypes: A Meta-Analysis. Am J Gastroenterol 2020;115:1940-1949.

62. Salaspuro M. Microbial metabolism of ethanol and acetaldehyde and clinical consequences. Addict Biol 1997;2:35-46.

63. Amitay EL, Carr PR, Jansen L, et al. Smoking, alcohol consumption and colorectal cancer risk

by molecular pathological subtypes and pathways. Br J Cancer 2020;122:1604-1610.

64. Santarelli RL, Pierre F, Corpet DE. Processed meat and colorectal cancer: a review of epidemiologic and experimental evidence. Nutr Cancer 2008;60:131-44.

65. Aykan NF. Red Meat and Colorectal Cancer. Oncol Rev 2015;9:288.

66. Rawla P, Sunkara T, Barsouk A. Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. Prz Gastroenterol 2019;14:89-103.

67. Li H, Boakye D, Chen X, et al. Association of Body Mass Index With Risk of Early-Onset Colorectal Cancer: Systematic Review and Meta-Analysis. Am J Gastroenterol 2021;116:2173-2183.

68. Gacesa R, Kurilshikov A, Vich Vila A, et al. Environmental factors shaping the gut microbiome in a Dutch population. Nature 2022;604:732-739.

69. Sanchez-Alcoholado L, Ramos-Molina B, Otero A, et al. The Role of the Gut Microbiome in Colorectal Cancer Development and Therapy Response. Cancers (Basel) 2020;12.

70. Kim J, Lee HK. Potential Role of the Gut Microbiome In Colorectal Cancer Progression. Front Immunol 2021;12:807648.

71. Silva M, Brunner V, Tschurtschenthaler M. Microbiota and Colorectal Cancer: From Gut to Bedside. Front Pharmacol 2021;12:760280.

72. Clay SL, Fonseca-Pereira D, Garrett WS. Colorectal cancer: the facts in the case of the microbiota. J Clin Invest 2022;132.

73. Wang T, Cai G, Qiu Y, et al. Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. ISME J 2012;6:320-9.

74. Saffarian A, Mulet C, Regnault B, et al. Crypt- and Mucosa-Associated Core Microbiotas in Humans and Their Alteration in Colon Cancer Patients. mBio 2019;10.

75. Chen W, Liu F, Ling Z, et al. Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. PLoS One 2012;7:e39743.

76. Mughini-Gras L, Schaapveld M, Kramers J, et al. Increased colon cancer risk after severe Salmonella infection. PLoS One 2018;13:e0189721.

77. Wu S, Shin J, Zhang G, et al. The Bacteroides fragilis toxin binds to a specific intestinal epithelial cell receptor. Infect Immun 2006;74:5382-90.

78. Rubinstein MR, Baik JE, Lagana SM, et al. Fusobacterium nucleatum promotes colorectal cancer by inducing Wnt/beta-catenin modulator Annexin A1. EMBO Rep 2019;20.

79. Buc E, Dubois D, Sauvanet P, et al. High prevalence of mucosa-associated E. coli producing cyclomodulin and genotoxin in colon cancer. PLoS One 2013;8:e56964.

80. Arthur JC, Perez-Chanona E, Muhlbauer M, et al. Intestinal inflammation targets cancer-inducing activity of the microbiota. Science 2012;338:120-3.

81. Secher T, Brehin C, Oswald E. Early settlers: which E. coli strains do you not want at birth? Am J Physiol Gastrointest Liver Physiol 2016;311:G123-9.

82. Kaper JB. Pathogenic Escherichia coli. Int J Med Microbiol 2005;295:355-6.

83. Nougayrede JP, Homburg S, Taieb F, et al. Escherichia coli induces DNA double-strand breaks in eukaryotic cells. Science 2006;313:848-51.

84. Pleguezuelos-Manzano C, Puschhof J, Huber AR, et al. Mutational signature in colorectal cancer caused by genotoxic pks(+)E. coli. Nature 2020;580:269-+.

85. Boot A, Ng AWT, Chong FT, et al. Characterization of colibactin-associated mutational signature in an Asian oral squamous cell carcinoma and in other mucosal tumor types. Genome Res 2020;30:803-813.

86. Lee-Six H, Olafsson S, Ellis P, et al. The landscape of somatic mutation in normal colorectal

epithelial cells. Nature 2019;574:532-537.

87. Lai YC, Lin AC, Chiang MK, et al. Genotoxic Klebsiella pneumoniae in Taiwan. PLoS One 2014;9:e96292.

88. Putze J, Hennequin C, Nougayrede JP, et al. Genetic structure and distribution of the colibactin genomic island among members of the family Enterobacteriaceae. Infect Immun 2009;77:4696-703.

89. Kaur CP, Iyadorai T, Sears C, et al. Presence of Polyketide Synthase (PKS) Gene and Counterpart Virulence Determinants in Klebsiella pneumoniae Strains Enhances Colorectal Cancer Progression In-Vitro. Microorganisms 2023;11.

Part I
Germline *APC* and biallelic *MUTYH* variant

Declining detection rates for *APC* and biallelic *MUTYH* variants in polyposis patients, implications for DNA testing policy

European Journal of Human Genetics, 2020

Diantha Terlouw, Manon Suerink, Sunny S. Singh, Hans J.J.P. Gille, Frederik J. Hes, Alexandra M. J. Langers, Hans Morreau, Hans F. A. Vasen, Yvonne J. Vos, Tom van Wezel, Carli. M. Tops, Sanne W. ten Broeke, Maartje Nielsen

Abstract

This study aimed to determine the prevalence of *APC*-associated familial adenomatous polyposis (FAP) and *MUTYH* associated polyposis (MAP) in a large cohort, taking into account factors as adenoma count and year of diagnosis. All application forms used to send patients in for *APC* and *MUTYH* variant analysis between 1992 and 2017 were collected (n = 2082). Using the data provided on the application form, the *APC* and biallelic *MUTYH* prevalence was determined and possible predictive factors were examined using multivariate multinomial logistic regression analysis in SPSS. The prevalence of disease causing variants in the *APC* gene significantly increases with adenoma count while MAP shows a peak prevalence in individuals with 50–99 adenomas. Logistic regression analysis shows significant odds ratios for adenoma count, age at diagnosis, and, interestingly, a decline in the chance of finding a variant in either gene over time. Moreover, in 22% (43/200) of patients with FAP-related extracolonic manifestations a variant was identified. The overall detection rates are above 10% for patients with >10 adenomas aged <60 and >20 adenomas aged <70. Patients with variants outside these criteria had FAP-related extracolonic manifestations, colorectal cancer aged <40, somatic *KRAS* c.34G>T variant in the tumor or a first-degree relative with >10 adenomas. Therefore, *APC* and *MUTYH* testing in patients with >10 adenomas aged <60 and with >20 adenomas aged <70 is advised. Almost all FAP and MAP patients not meeting these criteria showed other characteristics that can be used as an indication to prompt genetic testing.

Introduction

Due to a combination of environmental and low penetrant risk genetic factors[1-3], a large proportion of the general population will develop one or more adenomatous polyps (25% at age 50 and 50% at age 70[4]). These polyps are possible precursors of colorectal carcinoma (CRC). The most commonly reported polyposis coli syndromes are APC-associated familial adenomatous polyposis (FAP) and MUTYH-associated polyposis (MAP).[5, 6] Variants affecting the function of these genes are found in 8–10% of all patients with polyposis, depending on age and number of adenomas. Other forms of adenomatous polyposis explaining <1% of polyposis patients include *PolE/D*-[7], *NTHL1*-[8], *MSH3*-[9], *MBD4*-[10], and *MLH3*-associated polyposis[11]. Furthermore, mosaic *APC* variants are found in a substantial proportion of the remaining unexplained polyposis cases.[12]

Identifying patients and family members with a genetic predisposition for polyposis is important due to the high CRC risk that carriers face, even at a young age. This risk can be largely circumvented through regular surveillance and adenoma removal. Since adenomas after the age of 50 are common in the general population, offering genetic testing to all

patients with adenomatous polyps is not yet cost effective. No clear guidelines for genetic testing existed until recently, and studies on the rates of variant detection have focused on patients with more than 20 adenomas lacking detailed information on the outcome of genetic testing in patients with less than 20 adenomas.

The present cohort consists of Dutch polyposis patients tested for *APC* and/or *MUTYH* variants between 1992 and 2017. The primary aim of this study was to determine the prevalence of *APC* and biallelic *MUTYH* disease causing variants in individuals referred to the clinical genetics department for DNA testing. Furthermore, we studied the relationship between the *APC* and *MUTYH* variants and several covariates. Based on these outcomes, guidelines were developed regarding the indications for referral to a clinical geneticist for DNA analyses.

Methods

Study population

This cross-sectional study was conducted amongst probands referred to a clinical geneticist (1992–2017) based on an individual's phenotype and/or family history of cancer and polyps. After consultation at centers across the Netherlands, blood samples and prespecified application forms were sent to the LUMC Laboratory of Diagnostic Genome Analysis (LDGA) for diagnostic analysis of the *APC* and *MUTYH* genes. The prespecified application form included age at testing, age at diagnosis of colorectal adenomas and/or CRC, personal history of other cancers, and a pedigree with relevant family information. The clinical information of the majority of the patients had been collected in databases developed for other studies.[3, 13, 14] These databases were merged and any required additional information was added. In total 2082 patients were included, exclusion criteria are listed in Fig. S1.

Although no clear guidelines existed, the presence of >10 adenomatous polyps was generally considered a reason for referral, as also advised by the American College of Gastroenterology (ACG).[15] Moreover, FAP-related extracolonic manifestations were considered an indication for genetic *APC* testing and a somatic NM_033360.3 (*KRAS*): c.34G>T in tumor for *MUTYH* testing.[16, 17]

Clinical genetic testing was performed with full gene Sanger sequencing and rearrangements were analyzed using multiplex ligation dependent probe amplification for the *APC* and *MUTYH* genes. *MUTYH* clinical diagnostics became available in 2004[6], individuals suspicious for MAP but tested before 2004 were analyzed retrospectively.

Missing data

Due to incompletely filled in application forms, 26 patients were included with missing val-

ues for the age at first adenoma, 9 missed age at first CRC, and 164 patients missed family history. Possible explanations for a missing or incomplete pedigree information on the application form were adoption and no contact with family members.

Both the *APC* and *MUTYH* gene were sequenced in the majority of patients. However, in 387 (19%) and 339 (16%) patients only the *MUTYH* or the *APC* gene, respectively, was tested. The reasons for not testing these genes are summarized in Table S1.

Definitions

The terms "polyp" and "adenoma" were both used to describe patient samples sent for analysis. If no histology was mentioned, "polyps" were assumed to be adenomatous. After 2004, patients with hyperplastic/serrated polyps were occasionally sent for specifically *MUTYH* analysis.[18] Patients with exclusively serrated/hyperplastic type (n = 19) were treated separately in this study. Patients with other types of polyps such as hamartomatous or juvenile polyps were excluded.

Patients with a phenotype described as "FAP" (n = 170) or "polyposis" (n = 19) were considered to have >100 adenomas, "multiple adenomas/polyps" (n = 206) and "polyps" (n = 14) were categorized as 20–49 adenomas, and "some polyps" (n = 11) as less than 10 adenomas, as described previously.3 Individuals without information on polyp history were excluded. Moreover, family members with 10 polyps or 'some' polyps were labeled as having <10 polyps, descriptions such as "FAP," "AFAP," and "multiple" were considered to have >10, and the bare description "polyps" as number unknown. When more than one first degree relative (FDR) were reported with polyps, the highest number of polyps was used. Whenever multiple family members were diagnosed with CRC, the youngest was defined as the age of CRC in that family.

An *APC* de novo variant was assumed whenever the patients parents tested negative for the *APC* variant (n = 10) or whenever the pedigree showed no relevant cancers or polyps (n = 69).

Statistical methods

Multivariable logistic regression analysis was used to assess associations between variant status (yes/no) and covariates of interest. These covariates included cumulative polyp count (<10, 10–19, 20–49, 50–99, and >100), age at diagnosis (<30, 30–39, 40–49, 50–59, and >60), history of CRC (no, <40, 40–50, and >50 (when multiple CRC, youngest age of diagnosis was used)), FDR with polys (no, yes <10, yes >10, and yes number unknown), with CRC (no, yes <50 years, yes >50 years, and yes age unknown), and year of analysis (<1995–1999,

2000–2005, 2006–2011, and 2012–2017). The patients without any adenomas were treated as a separate group.

Patients in whom *APC* or *MUTYH* was not tested were not included in the logistic regression analysis of the *APC* or *MUTYH* variant, respectively. All these patients were excluded from the analysis for overall variant detection.

Results were reported as odds ratios, with a 95% confidence interval, and a p-value <0.05 was considered statistically significant. The statistical analyses were performed using SPSS statistics 23.

Table 1. Cohort characteristics

	Total (n=2082)	***APC***	**Biallelic *MUTYH***	**Monoallelic *MUTYH***
Male - n (%)	1202 (58%)	147 (50%)	64 (54%)	27 (69%)
Adenoma count				
0	336	13 (3.9%)	3 (0.9%)	7 (2.1%)
1-9	328	1 (0.3%)	6 (1.8%)	7 (2.1%)
10-19	406	3 (0.7%)	6 (1.5%)	7 (1.7%)
20-49	590	50 (8.5%)	60 (10%)	15 (2.5%)
50-99	122	15 (12%)	22 (18%)	3 (2.5%)
>100	300	211 (70%)	22 (7.3%)	0 (0%)
Mean age at adenoma diagnosis (min-max)	53 (4-84)	36 (9-68)	49 (21-75)	54 (23-77)
CRC, yes	746 (36%)	57 (7.6%)	82 (11%)	15 (2.0%)
Mean age at (first) CRC diagnosis (min-max)	53 (12-91)	41 (21-58)	49 (21-76)	57 (28-91)
FAP extracolonic manifestations, yes	200 (10%)	43 (22%)	8 (4.0%)	4 (2.0%)
FDR with polyps	728 (38%)	156 (21%)	46 (6.3%)	18 (2.5%)
Missing	164 (8%)	12 (7.3%)	4 (2.4%)	1 (0.6%)
FDR with CRC	811 (42%)	76 (9.7%)	42 (5.1%)	26 (3.2%)
Missing	164 (8%)	12 (7.3%)	4 (2.4%)	1 (0.6%)

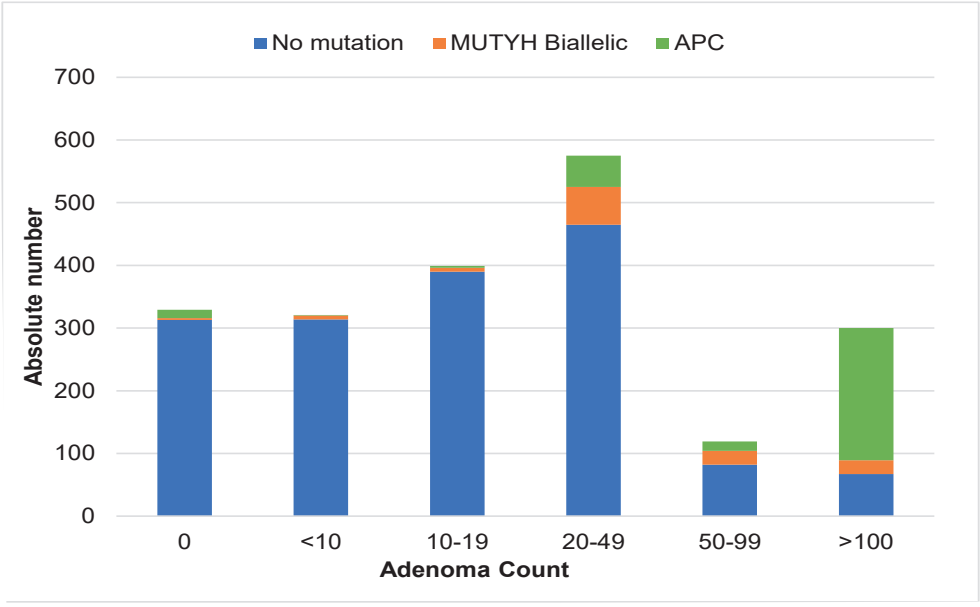
Results

Of the 2082 individuals included in the study (Table 1), in total, 14% (n = 293) carried an *APC* variant, 6% (n = 119) a biallelic *MUTYH* variant, and 2% (n = 39) a monoallelic *MUTYH* variant. Overall, a personal history of CRC was reported in 36% (n = 746) of patients. Notably, 16% (n = 336) had no history of adenomas whatsoever. In the overall cohort, variant detection rate is highest in patients with more than 20 adenomas (Fig. 1) and increases with younger ages (Fig. 2).

Association between phenotypic characteristics and a variant in *APC* and/or *MUTYH*

Multivariable logistic regression analysis (Table 2) shows that the odds of identifying a variant in either gene steadily increases with adenoma count. The odds of *APC* variant detection are highest in patients with >100 adenomas (OR 289.9; 95% CI 35.2–2385.2), while the odds ratio for biallelic *MUTYH* variants was highest for the 50–99 adenoma count (OR 10.8; 95% CI 4.0–29.1).

Fig. 1 Absolute numbers of patients sent in for genetic testing among the different adenoma count groups. APC and MUTYH variant detection depicted in green and yellow, respectively.



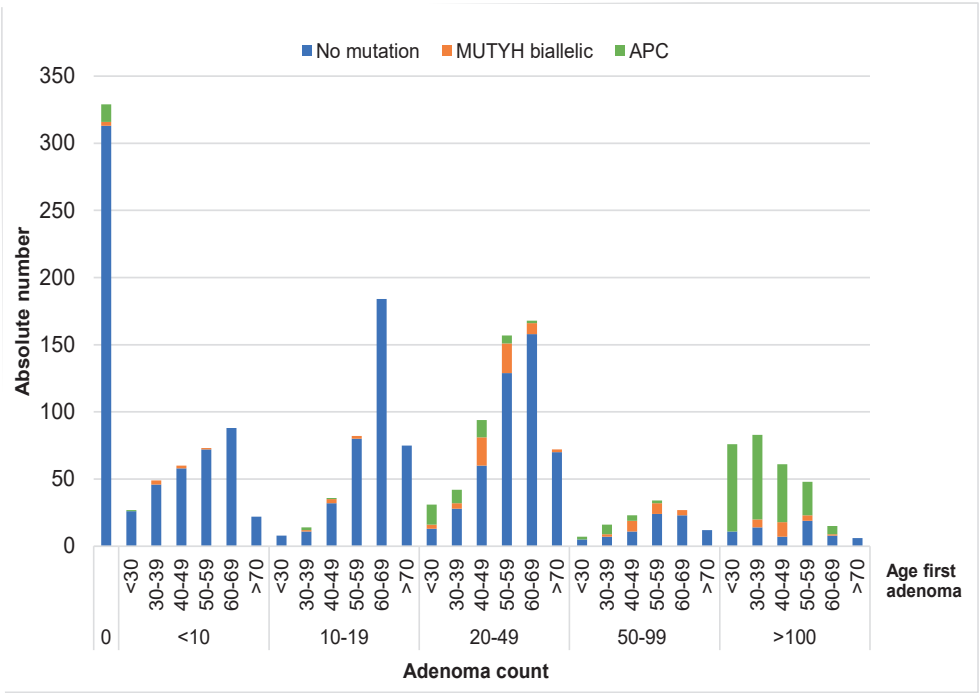
A personal history of CRC increased the likelihood of detecting biallelic *MUTYH* variants (<40: OR 3.9 [95% CI 1.5–10.0], 40–50: OR 4.5 [95% CI 2.2–9.0], and >50 OR 5.1 [95% CI 2.8–9.2]). However, no effect was found for the detection of an *APC* variant.

The chance of finding a *MUTYH* or *APC* variant was not increased in patients with a FDR with CRC. Conversely, a FDR with >10 polyps did increase the odds of detecting an *APC* variant significantly (OR 4.5 [2.5–8.4]).

Variant detection rate trends over time

Also, the chance of finding a variant decreased over the last 20 years (<1995–1999: OR 9.8

Fig. 2 Variant detection in different adenoma count groups, specified by age group



[4.7–20.3], 2000–2005: OR 3.9 [2.2–7.2], and 2006–2011: OR 1.7 [1.0–3.0]). However, the odds of finding a biallelic *MUTYH* variant were highest between 2000 and 2005. Possibly explained by the introduction of *MUTYH* diagnostics in 2004, also attributing to the increase in number of patients sent in for DNA testing in general (Fig. 3).

APC and *MUTYH* detection rates in patients with less than 20 adenomas

Since a large number of patients with less than 20 adenomas underwent genetic testing (n = 1070, 51%), these categories are described in more detail.

No adenomas

The majority of patients without adenomas underwent testing due to CRC (n = 176), FAP-related extracolonic manifestations (n = 75), or both (n = 11). Nineteen had hyperplastic polyposis, while the rest were tested based on a positive family history. *APC* was tested in 203 and *MUTYH* was tested in 259 of these patients. Thirteen FAP and three MAP patients were detected in this group (Table S2).

Nine of the *APC* variant carriers had extracolonic manifestations (mean age ~13, range 1.5–38). In addition, four had experienced CRC (two aged <40, one <50, and one >50). Of the MAP patients, all three had CRCs (<50 years old) with a *KRAS* c.34G>T transversion.

Table 2. Multivariate analysis

	APC or Biallelic *MUTYH*			*APC*			Biallelic *MUTYH*		
	N[a]	OR (95% CI)	*p-value*	N[a]	OR (95% CI)	*p-value*	N[a]	OR (95% CI)	*p-value*
Adenoma count									
<10	188	Ref	<0.001	205	Ref	<0.001	296	Ref	<0.001
10-19	292	2.6 (0.8-8.0)		296	8.5 (0.8-88.9)		367	1.5 (0.5-4.7)	
20-49	486	12.9 (5.0-33.3)		502	39.2 (4.7-324.2)		515	6.3 (2.6-15.1)	
50-99	114	15.5 (5.5-43.8)		115	32.5 (3.6-289.9)		116	10.8 (4.0-29.1)	
>100	264	59.4 (22.4-157.2)		275	289.9 (35.2-2385.2)		269	3.5 (1.3-9.5)	
Age at adenoma diagnosis									
>60	486	Ref	<0.001	502	Ref	<0.001	608	Ref	<0.001
50-59	328	5.1 (2.7-9.5)		335	3.3 (1.3-8.4)		370	3.9 (2.0-7.6)	
40-49	235	14.5 (6.8-31.0)		248	12.0 (4.6-31.4)		260	5.5 (2.3-13.5)	
30-39	164	11.4 (5.3-24.8)		172	12.6 (4.9-32.7)		188	3.9 (1.4-10.6)	
<30	131	18.7 (8.4-41.4)		136	33.1 (12.5-87.5)		137	0.9 (0.2-3.7)	
CRC (age)									
No	920	Ref	0.006	951	Ref	0.003	1065	Ref	<0.001
<40	56	1.3 (0.6-3.0)		58	0.6 (0.3-1.3)		63	3.9 (1.5-10.0)	
40-50	111	1.3 (0.7-2.5)		115	0.3 (0.1-0.6)		123	4.5 (2.2-9.0)	
>50	257	2.6 (1.5-4.5)		269	0.5 (0.2-1.1)		312	5.1 (2.8-9.2)	
FDR with polyps									
No	809	Ref	<0.001	834	Ref	<0.001	945	Ref	0.098
Yes, ≤10 polyps	140	0.5 (0.2-1.0)		153	1.5 (0.7-3.3)		191	0.3 (0.1-0.8)	
Yes, >10 polyps	184	4.5 (2.6-8.0)		191	4.5 (2.5-8.4)		194	1.2 (0.6-2.3)	
Yes, number unknown	211	0.4 (0.2-0.7)		215	0.4 (0.2-0.8)		233	0.8 (0.5-1.5)	
FDR with CRC									
No	812	Ref	0.007	842	Ref	0.155	923	Ref	0.193
Yes, ≤50y	132	1.3 (0.7-2.4)		140	1.0 (0.5-2.1)		157	1.5 (0.8-2.9)	
Yes, >50y	349	0.5 (0.3-0.8)		358	0.6 (0.3-1.0)		426	0.7 (0.4-1.3)	
Yes, age unknown	51	1.6 (0.6-4.3)		53	2.0 (0.7-6.0)		57	0.7 (0.2-2.3)	
Year of DNA testing									
2012-2017	401	Ref	<0.001	415	Ref	<0.001	511	Ref	0.006
2006-2011	511	1.7 (1.0-3.0)		521	2.0 (1.0-4.3)		591	1.1 (0.5-2.7)	
2000-2005	266	3.9 (2.2-7.2)		280	2.5 (1.2-5.4)		291	2.3 (1.2-4.7)	
<1995-1999	166	9.8 (4.7-20.3)		177	9.6 (4.1-22.2)		170	1.0 (0.5-2.0)	

a. numbers are without cases with missing information

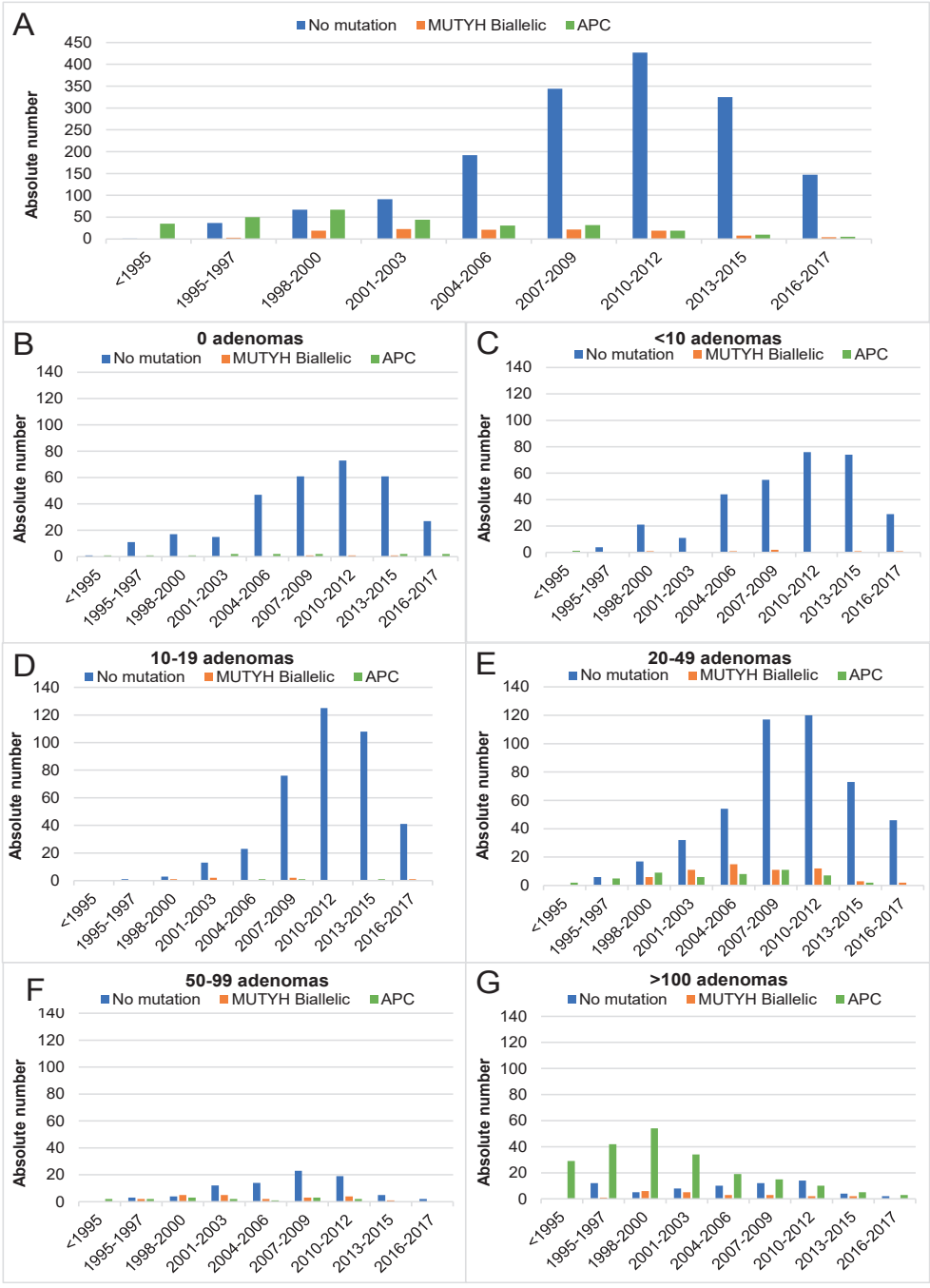


Fig. 3 Trends in variant detection. A. *APC*/*MUTYH* variant detection in all adenoma groups. B. Detection in patients without adenomas, C. 1-9 adenomas, D. 10-19 adenomas, E. 20-49 adenomas, F. 50-99 adenomas and G. more than 100 adenomas

Of the 52 patients with solely CRC aged <40, 8% (2/24) had FAP and 4% (2/50) had MAP. In patients with CRC between age 40 and 50 years this was, respectively, 4% (1/24) and 2% (1/41).

1–9 adenomas

In patients with 1–9 adenomas (n = 328; *APC* tested n = 217 and *MUTYH* tested n = 309), one *APC* and six biallelic *MUTYH* variants were identified (2% variant detection rate). In this group the *APC* variant carrier already developed adenomas by the age of 20 and had a FDR with >100 polyps. Of the MAP patients, four were affected with CRC between the ages 39 and 53. Information on *KRAS* status in tumor DNA was available for one patient, showing a somatic *KRAS* c.34G>T transversion.

10–19 adenomas

Finally, in the group with 10–19 adenomas (n = 406; *APC* tested n = 324 and *MUTYH* tested n = 401) three FAP and six MAP patients were diagnosed who all developed adenomas aged under 60.

Aged above 70

No *MUTYH* or *APC* variants were found in patients with fewer than 20 adenomas aged above 70 years (n = 82). In the patients with more than 20 adenomas aged over 70 years, one MAP patient was found (1/90, 1.1%).

The prevalence of *APC* or biallelic *MUTYH* variants in different clinical phenotypes in patients with <20 adenomas is depicted in Table S3 (as adapted from Grover et al.[19]).

***APC* de novo**

Based on family history, we surmise that a de novo variant has arisen in 24% of all *APC* variant carriers (69/292), which is comparable to the prevalence described previously.[20] Familial adenomatous polyposis (FAP This is also a plausible explanation for a negative family history in a number of FAP patients (Table S3).

Discussion

This study reports on 2082 individuals who underwent *APC* and *MUTYH* analysis at the LDGA between 1992 and 2017. The variant detection rates in patients with classic polyposis for FAP (70%) and MAP (7%) were comparable to previous studies.[5, 21-24] As expected, MAP showed a greater prevalence than FAP among individuals with 20–49 adenomas (FAP 9% vs.

MAP 10%) and 50–99 adenomas (FAP 12% vs. MAP 18%). Notably, a recent study reported lower variant rates in all adenoma groups, possibly explained by the differences in clinical background (i.e., older age) and more recent years of diagnosis (2012–2016).[25]

Although most patients undergoing DNA analysis nowadays have fewer than 20 adenomas, clinical factors associated with the presence of a germline *APC* or biallelic *MUTYH* variants in this group are still poorly understood. A study by Grover et al.[19] reported a low variant detection rate, but no clinical description of the variant carriers was provided. The study from Stanich et al.[25] analyzed a large cohort of patients with 10–20 polyps, however no patients with less than ten polyps were included. In our cohort, a large group of individuals without adenomas (n = 336) was included.

Except for four MAP patients (Table S2), all patients with *APC* or biallelic *MUTYH* variants presented with >10 adenomas aged <60, >20 adenomas aged <70, CRC below age 40, a typical *KRAS* c.34G>T variant, a FDR relative with >10 polyps, or FAP-related extracolonic manifestations explaining their referral.

Since *KRAS* was not systematically analyzed in CRC cases, no variant detection rate could be determined for this cohort. Previous studies showed in 10–25% of the CRC cases with the *KRAS* c.34G>T variant a biallelic *MUTYH* variant. *KRAS* analysis in CRC is often performed because of the prognostic and therapeutic value.[16, 17]

To analyze the impact on detection rates of several factors regression analysis was performed. While a younger age of first adenoma was associated with an increasing odds ratio of finding a variant in either gene, a personal history of CRC only increased the odds of finding a biallelic *MUTYH* variant, as also reported by Grover et al[19]. This can possibly be explained by the fact that known FAP patients undergo a (sub)total colectomy at an early age, effectively preventing the development of CRC. A family history of CRC did not influence the chance of finding either an *APC* or *MUTYH* variant. On the other hand, having a FDR with more than ten polyps clearly increased the chance of detecting an *APC* variant (OR 4.5, 2.5–8.4).

Increasing numbers of patients undergo DNA analysis while variant detection rate has steadily declined over the years. This resulted in an avoidable burden and expense for family cancer clinics and emphasizes the need for more stringent guidelines. One plausible explanation for the increase is the introduction of *MUTYH* gene testing in 2004, allowing milder phenotypes to be tested and thus increasing the number of patients with fewer than 20 adenomas. An alternative explanation is the introduction of population screening in the Netherlands in 2014 leading to increasing numbers of patients aged >55, with <10–20 ade-

nomas. However, the total number of individuals declined after 2013, possibly due to other Dutch laboratories offering *MUTYH* and *APC* testing themselves. Finally, the introduction of more sensitive techniques, such as chromoendoscopy, improvement of endoscopy equipment, implementation of adenoma detection rate as a quality measure, and better bowel preparation, has led to improved adenoma detection, particularly of low stage and small adenomas (i.e., <0.5 mm).[26-28] Moreover, a gradual incline in the percentage of de novo *APC* variants was seen over the years (<1995–1999: 14%, 2000–2005: 28%, 2006–2011: 36%, and 2012–2017: 29%), likely indicating that the majority of Dutch FAP families have been identified.

In 2015, the ACG issued guidelines for *APC* and *MUTYH* genetic testing in individuals with >10 cumulative colorectal adenomas, FAP-related extracolonic manifestations, or a family history of an adenomatous polyposis syndrome.[15] Based on our data, these guidelines may result in unnecessary testing, especially above the age of 60. On the other hand, Dutch guidelines also formulated in 2015 advise patients with either ten or more adenomas <60 years (cumulative) or 20 or more adenomas <70 years (cumulative) to be referred for genetic testing. The most recent NCCN guidelines[29] suggest genetic testing for all patients with >20 adenomas or a personal history of desmoid tumors, hepatoblastoma, cribriform-morular papillary thyroid cancer, and CHRPE, or patients with 10–20 adenomas with specific features such as age of onset influencing whether testing should be offered. Both these guidelines are supported by our data.

Stanich et al.[25] suggest testing in all patients with >10 polyps, regardless of histology or age despite their observation of declining variant rates with increasing age. Their reason is the observed detection rate in nonpolyposis related genes of around 5%. However, the 1% *CHEK2* variants reflects the prevalence in the general population[30] and does, in our opinion, not explain the polyposis phenotype. Furthermore, we excluded patients with MMR gene variants since further research is needed to draw firm conclusions about the association with polyposis.

CRC <40 years in patients without adenomas might be a reason for testing, since variants were found in 9% and 4% of our cohort in, respectively, *APC* and *MUTYH*. Testing patients with adenomas above the age of 70 should on the other hand be undertaken with caution, since the variant detection rate was 1%. Of course, other more specific circumstances might warrant testing, such as polyps below age 20 and numerous primary CRC (≥2).

One weakness of this study was that not all patients with low adenoma counts were tested for both *APC* and *MUTYH*. We detected 4% *APC* and 1% biallelic *MUTYH* variants in 0 adenoma patients, <1% *APC* and 2% *MUTYH* in 1–9 adenoma patients, and 1% *APC* and 2% *MUTYH*

in 10–19 adenomas patients. Based on the variant detection rate found in other studies, we anticipate that few or no cases were missed in our cohort.[19, 31]

Moreover, variants in other genes were not taken into account. Many of the patients were tested for *PolE/D*[32], *MSH3*, and *NTHL1* on a research basis, the proven variant carriers were excluded in this study. Possible variants in other genes such as *SMAD4*, *BMPR1A*, and *PTEN* might be present, albeit in a small percentage of our cohort. In many labs, these genes have been included in NGS panels over the recent years, but, due to their rarity and often distinct phenotype, they do not justify lowering the suggested testing threshold. Nonetheless, in the near future the NGS panels will become more extensive, including more of other polyposis and colorectal cancer related genes as already proposed by the NCCN guidelines.[29] This will increase the yield of genetic testing also for other genes than *APC* and *MUTYH*.

The 2% heterozygote *MUTYH* carriers detected in this study is higher than expected based on the 1% prevalence reported in the Exome Aggregation Consortium database but similar to what Grover et al.[19] found in patients with <20 adenomas. It is possible that some monoallelic *MUTYH* carriers have other genetic factors, which combined with *MUTYH* explains adenoma development. As illustrated by two of the *APC* variant carriers also carrying a monoallelic *MUTYH* variant.

APC mosaicism was recently identified in 25–50% of unexplained patients with >20 adenomas.[12] In most of these cases, the mosaicism was undetectable in leukocyte derived DNA and required testing of DNA isolated from >2 adenomas. Tumor testing is still logistically challenging and not performed in the current cohort. However, it might be an efficient approach in the future, especially for low adenoma count patients.

Conclusion

Adenoma count, age at adenoma diagnosis, and year of analysis are important predictive factors for *APC* and *MUTYH* variants. In view of the decline in variant detection, careful consideration for gene testing, especially in patients with lower polyp counts, is advised. Nevertheless, *APC* and *MUTYH* testing seems indicated in patients with >10 adenomas aged <60 and >20 adenomas aged <70. Other indications for referral are FAP-related extracolonic manifestations, CRC aged <40, a somatic *KRAS* c.34G>T transversion, or a FDR with >10 adenomas.

References

1. Dejea CM, Fathi P, Craig JM, et al. Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. Science 2018;359:592-597.
2. Peters U, Sinha R, Chatterjee N, et al. Dietary fibre and colorectal adenoma in a colorectal cancer early detection programme. Lancet 2003;361:1491-5.
3. 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.
4. van Hees F, Saini SD, Lansdorp-Vogelaar I, et al. Personalizing colonoscopy screening for elderly individuals based on screening history, cancer risk, and comorbidity status could increase cost effectiveness. Gastroenterology 2015;149:1425-37.
5. Kinzler KW, Nilbert MC, Su LK, et al. Identification of FAP locus genes from chromosome 5q21. Science 1991;253:661-5.
6. 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.
7. 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.
8. 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.
9. 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.
10. Sanders MA, Chew E, Flensburg C, et al. MBD4 guards against methylation damage and germ line deficiency predisposes to clonal hematopoiesis and early-onset AML. Blood 2018;132:1526-1534.
11. Olkinuora A, Nieminen TT, Mårtensson E, et al. Biallelic germline nonsense variant of MLH3 underlies polyposis predisposition. Genet Med 2019;21:1868-1873.
12. Jansen AML, Crobach S, Geurts-Giele WRR, et al. Distinct Patterns of Somatic Mosaicism in the APC Gene in Neoplasms From Patients With Unexplained Adenomatous Polyposis. Gastroenterology 2017;152:546-+.
13. Hes FJ, Nielsen M, Bik EC, et al. Somatic APC mosaicism: an underestimated cause of polyposis coli. Gut 2008;57:71-76.
14. Nielsen M, Franken PF, Reinards TH, et al. Multiplicity in polyp count and extracolonic manifestations in 40 Dutch patients with MYH associated polyposis coli (MAP). J Med Genet 2005;42:e54.
15. Syngal S, Brand RE, Church JM, et al. ACG clinical guideline: Genetic testing and management of hereditary gastrointestinal cancer syndromes. Am J Gastroenterol 2015;110:223-62; quiz 263.
16. van Puijenbroek M, Nielsen M, Tops CM, et al. Identification of patients with (atypical) MUTYH-associated polyposis by KRAS2 c.34G > T prescreening followed by MUTYH hotspot analysis in formalin-fixed paraffin-embedded tissue. Clin Cancer Res 2008;14:139-42.
17. Aimé A, Coulet F, Lefevre JH, et al. Somatic c.34G>T KRAS mutation: a new prescreening test for MUTYH-associated polyposis? Cancer Genet 2015;208:390-5.
18. Boparai KS, Dekker E, Van Eeden S, et al. Hyperplastic polyps and sessile serrated adenomas as a phenotypic expression of MYH-associated polyposis. Gastroenterology 2008;135:2014-8.

19. Grover S, Kastrinos F, Steyerberg EW, et al. Prevalence and phenotypes of APC and MUTYH mutations in patients with multiple colorectal adenomas. Jama 2012;308:485-492.
20. Bisgaard ML, Fenger K, Bülow S, et al. Familial adenomatous polyposis (FAP): frequency, penetrance, and mutation rate. Hum Mutat 1994;3:121-5.
21. Powell SM, Petersen GM, Krush AJ, et al. Molecular diagnosis of familial adenomatous polyposis. N Engl J Med 1993;329:1982-7.
22. Miyoshi Y, Ando H, Nagase H, et al. Germ-line mutations of the APC gene in 53 familial adenomatous polyposis patients. Proc Natl Acad Sci U S A 1992;89:4452-6.
23. Sieber OM, Lipton L, Crabtree M, et al. Multiple colorectal adenomas, classic adenomatous polyposis, and germ-line mutations in MYH. N Engl J Med 2003;348:791-9.
24. Wang L, Baudhuin LM, Boardman LA, et al. MYH mutations in patients with attenuated and classic polyposis and with young-onset colorectal cancer without polyps. Gastroenterology 2004;127:9-16.
25. Stanich PP, Pearlman R, Hinton A, et al. Prevalence of Germline Mutations in Polyposis and Colorectal Cancer-Associated Genes in Patients With Multiple Colorectal Polyps. Clin Gastroenterol Hepatol 2019;17:2008-2015.e3.
26. Brown SR, Baraza W, Din S, et al. Chromoscopy versus conventional endoscopy for the detection of polyps in the colon and rectum. Cochrane Database Syst Rev 2016;4:Cd006439.
27. Aranda-Hernández J, Hwang J, Kandel G. Seeing better--Evidence based recommendations on optimizing colonoscopy adenoma detection rate. World J Gastroenterol 2016;22:1767-78.
28. Brenner H, Altenhofen L, Kretschmann J, et al. Trends in Adenoma Detection Rates During the First 10 Years of the German Screening Colonoscopy Program. Gastroenterology 2015;149:356-66.e1.
29. Gupta S, Provenzale D, Regenbogen SE, et al. NCCN Guidelines Insights: Genetic/Familial High-Risk Assessment: Colorectal, Version 3.2017. J Natl Compr Canc Netw 2017;15:1465-1475.
30. Huijts PE, Hollestelle A, Balliu B, et al. CHEK2*1100delC homozygosity in the Netherlands--prevalence and risk of breast and lung cancer. Eur J Hum Genet 2014;22:46-51.
31. Knopperts AP, Nielsen M, Niessen RC, et al. Contribution of bi-allelic germline MUTYH mutations to early-onset and familial colorectal cancer and to low number of adenomatous polyps: case-series and literature review. Fam Cancer 2013;12:43-50.
32. 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;7:e00603.

Supplemental Material

Part II
APC mosaicism

Prevalence and consequences of *APC* mosaicism in patients with colorectal adenomas

Manuscript in preparation

Diantha Terlouw, Manon Suerink, Yentl Buitelaar, Monique E van Leerdam, Frederik J Hes, Demi van Egmond, Dina Ruano, Anja Wagner, Floris H Groenendijk, Isabelle C Meijssen, Eline Overwater, Sanne W Bajwa-ten Broeke, Arjen R Mensenkamp, Iris D Nagtegaal, Carli M Tops, Alexandra M J Langers, Tom van Wezel, Hans Morreau, Maartje Nielsen

Abstract

Although germline pathogenic *APC* variants are detected in most adenomatous polyposis patients, a substantial proportion still remains unexplained. Therefore, in three university medical centers in the Netherlands, *APC* mosaicism was analyzed in 458 patients with a broad spectrum of phenotypes. The mosaicism detection rate was 11.1% (51 out of 458) in the entire cohort. This rate was 17.1% (46 out of 269) in patients with ≥10 adenomas before the age of 60 or with ≥20 adenomas before the age of 70 and 2.6% (5 out of 189) in patients falling outside these Dutch hereditary polyposis testing guidelines. Overall, the odds of finding *APC* mosaicism increased significantly with adenoma count and a younger age at adenoma diagnosis. Interestingly, 28% (9 out of 32) of mosaic patients undergoing an esophagogastroduodenoscopy were diagnosed with gastroduodenal adenomas. Moreover, none of children tested inherited the mosaic variant. In one patient without children the mosaic variant was detected in semen. Hybrid mosaic cases, having a recurrent variant in some but not all lesions, were phenotypically similar to non-mosaic patients. We recommend *APC* mosaicism testing in all patients negative for germline pathogenic variants with (1) multiple adenomas before the age of 50, (2) ≥20 adenomas before the age of 60 or (3) ≥30 adenomas before the age of 70. Regular colonoscopy and at least one gastroduodenoscopy should be offered to *APC* mosaics, frequency of follow-up should depend on findings. Furthermore, offering germline testing for offspring should be considered, especially when mosaicism is exceeding the colon.

Introduction

Familial Adenomatous Polyposis (FAP), the most common polyposis syndrome, is caused by a pathogenic germline variant in the *APC* gene.[1] Familial adenomatous polyposis (FAP FAP patients classically develop hundreds to thousands colorectal adenomas. The severity depends on location of the variant in the gene; variants in the 5' or 3' end cause an attenuated phenotype with less than 100 colorectal adenomas (AFAP).[2] (A)FAP is also associated with extracolonic manifestations such as duodenal and gastric neoplasms, osteomas and desmoid tumors.[3, 4] The *APC* germline pathogenic variant detection rate is highly dependent on adenoma count, varying between 1% in patients with 10-20 adenomas to 70% in patients with more than 100 adenomas in a large Dutch cohort.[5] Still, a large proportion of polyposis patients remain unexplained after regular germline testing.

In 4% to 11% of these unexplained polyposis patients, a so-called *APC* mosaicism can be identified analyzing leukocyte DNA.[6-8] *APC* mosaicism means having a pathogenic *APC* variant in only a subset of body cells due to a *de novo* variant occurring during embryogene-

sis.[9] Depending on the timing of this occurrence, multiple or a single tissue can harbor the variant. Notable, studies performing sequencing analysis on DNA isolated from colorectal lesions identified a mosaic *APC* variant in a much larger proportion, namely 25-50%.[10-12] These results indicate a high prevalence of *APC* mosaicism in unexplained polyposis patients which is not detectable in leukocyte DNA using regular diagnostic testing.

In general, the phenotypic characteristics of patients with mosaicism are milder than observed in full germline variant carriers. This assumption suggests that *APC* mosaicism might also be present in less affected patients with between 10 and 20 colorectal adenomas aged between 60 and 70 or with more than 20 adenomas aged above 70 years. These patients do not fall within the scope of the hereditary polyposis testing guidelines and are therefore usually not tested.[13-15]

In this study, we performed targeted Next Generation Sequencing (NGS) using a broad panel on DNA isolated from colorectal lesions from unexplained polyposis patients. Moreover, patients with mild polyposis phenotypes were also included. Based on the results, we aimed to draft guidelines for *APC* mosaicism testing and surveillance.

Materials and methods

Cohort description

Up until January 2023, *APC* mosaicism testing was performed in 463 patients from three medical centers throughout the Netherlands. The majority, 379 patients, were tested in the Leiden University Medical Center (LUMC), 44 at the Erasmus Medical Center (EMC) in Rotterdam and 35 patients from the University Medical Center Groningen (UMCG) were tested at the Radboud university medical center (Radboudumc).

Leiden (LUMC)

The LUMC cohort included patients both fulfilling and not fulfilling the Dutch polyposis germline testing guidelines of either ≥10 cumulatively developed adenomas before the age of 60 years or ≥20 adenomas aged between 60 and 70. All patients fulfilling the testing guidelines were tested for germline variants at the department of Clinical Genetics. Patients not fulfilling the guidelines were included when fitting one of the following predefined groups: between 5 and 10 adenomas before the age of 50 years, between 10 and 20 adenomas aged between 60 and 70 or through population based screening aged between 55 and 75, more than 20 adenomas above the age of 70 years or multiple colorectal carcinomas before the age of 70 years.

As categorized before[5], patients with a phenotype described as 'FAP' (n=1) were considered to have >100 adenomas and 'AFAP' (n=1) to have 50-99 adenomas and a description of 'multiple adenomas' (n=3) was categorized as 30-49 adenomas and 'some polyps' (n=1) as less than 10 adenomas.

Rotterdam (EMC)

The Rotterdam cohort consists of patients that tested negative for pathogenic germline variants in adenomatous polyposis genes. In general, patients with >20 adenomas or multiple colorectal cancers were tested for *APC* mosaicism.

Groningen (UMCG)

The Groningen cohort also consists of patients negatively tested for pathogenic germline variants in adenomatous polyposis genes. *APC* mosaicism was tested whenever finding a mosaic variant influencing the surveillance guidelines for first degree relatives. Therefore, this cohort includes patients with >20 adenomas and patients with 10-20 adenomas before the age of 55.

Targeted Next-Generation Sequencing (NGS)

Leiden (LUMC)

DNA was extracted from Formalin-Fixed Paraffin Embedded (FFPE) tissue blocks of preferably 4 colorectal lesions per patient and analyzed using targeted NGS as described before.[16] In short, a custom-made panel was used consisting of: *APC*, *MUTYH*, *POLE*, *POLD1*, *NTHL1*, *MLH1*, *MSH2*, *MSH6*, *PMS2*, *MSH3*, *BMPR1A*, *RNF43*, *PTEN*, *SMAD4*, *STK11*, *ENG*, *BRCA1*, *BRCA2*, *PALB2*, *TP53*. Ampliseq NGS libraries were prepared following manufacturer's instructions. Sequencing was subsequently performed using an Ion GeneStudio S5 Series sequencer (Thermofisher Scientific). The unaligned sequence reads were mapped against the human reference genome (hg19) using TMAP software and variants were called using Torrent Variant Caller.

Detected variants were classified by pathogenicity and compared between the lesions within each patient. Mosaicism was considered when all lesions shared an identical variant. Hybrid mosaicism was considered when an identical variant was detected in a subset of lesions. Whenever mosaic, leukocytes, buccal mucosa and urine DNA was tested to determine the mosaic pattern throughout the body.

Rotterdam (EMC)

DNA of ≥2 FFPE tissues was isolated using proteinase K and 55 Chelex 100 resin. *APC* was analyzed using a custom-made targeted NGS panel consisting of *APC* and *MUTYH*. Ampliseq NGS libraries were prepared following manufacturer's instructions and sequencing was performed on an Ion s5 XL system (Thermofisher Scientific). Sequencing data was analysed using the Torrent variant caller (Thermo Fisher Scientific) or SeqPilot (JSI medical systems). Detected variants were classified by pathogenicity and compared between the lesions within each patient.

Groningen (UMCG)

Samples from the UMCG were tested for *APC* at the Radboudumc using a custom-made NGS panel based on single-molecule molecular inversion probes (smMIP) enrichment. The probes covered all regions and intron-exon boundaries, as described previously.[17, 18] Sequencing was performed on an Illumina NextSeq 500 or Novaseq 6000. Fastq files were analyzed using the SeqNext software package (JSI Medical Systems GmbH, Kippenheim, Germany). Based on the single-molecule tag, consensus reads were generated and variants in coding regions were called if present in ≥5% of all reads and ≥3 unique variant reads. Probe sequences are available upon request. Detected variants were classified by pathogenicity and compared between the lesions within each patient.

Body Mass Index (BMI) and lifestyle data

For the LUMC cohort, BMI and lifestyle data was collected using patient medical records. Additionally, a questionnaire about height, weight, smoking status, alcohol consumption and medication use was sent to 65 patients. Body Mass Index (BMI) was categorized as 'underweight' with a BMI of ≤18.4, 'healthy weight' with a BMI of 18.5-24.9, 'overweight' with 25.0-29.9 and 'obese' with a BMI of ≥30.0. Smoking status and alcohol consumption was categorized as 'never', 'former' and 'current' and smoking packyears (PY) was determined by the number of cigarette packs smoked per day multiplied by the number of smoking years. In total, BMI was determined in 136 patients, smoking status in 230 and alcohol consumption in 220 patients. Information about the medication use was gathered for about 60 patients.

Statistical analysis

Statistical analyses were performed using IBM SPSS statistics 25 and a p-value of <0.05 was considered significant. Independent T tests, Chi-square tests and Fisher's exact tests were used to assess phenotypic and lifestyle differences.

To determine associations between *APC* mosaicism and covariates of interest, a multivariable logistic regression analysis was performed. As covariates cumulative adenoma count (<10, 10-19, 20-49, 50-99, >100), age at diagnosis (<50, 50-59, 60-69 and >70), history of CRC (no, <50, 50-59 and >60) and the presence of extracolonic manifestations associated with FAP (yes, no). The results were reported as odds ratios, with a 95% confidence interval.

Results

APC mosaicism prevalence

Up until January 2023, 458 patients were tested for *APC* mosaicism in university medical centers in

Leiden, Rotterdam and Groningen using NGS on DNA isolated from colorectal adenomas or carcinomas. A large proportion of the included patients (n=189) did not fall within the scope of the Dutch colorectal cancer and polyposis testing guidelines of ≥10 adenomas before the age of 60 years or ≥20 adenomas before the age of 70 years.

In 11.1% (51 out of 458) of patients a mosaic *APC* variant was detected, as summarized in table 1. This detection rate was 17.1% (46 out of 269) and 2.6% (5 out of 189) of patients falling within and outside the scope of the Dutch testing guidelines respectively. Figure 1 shows a detection rate of more than 5% in all patients with adenomas before the age of 50 years, ≥20 adenomas before the age of 60 years or ≥30 adenomas before the age of 70 years. No mosaicism was detected in the 6 patients without any colorectal adenomas, included due to multiple CRCs.

Mosaic patients have significantly more colorectal adenomas and were younger at development of the first adenoma compared to non-mosaic patients. There was no significant difference in prevalence of colorectal carcinomas (CRC) but when suffering from CRC *APC* mosaic patients were diagnosed at a significantly younger age.

Multivariable logistic regression analysis in table 3 supports this finding by showing significant increased odds of finding a mosaic *APC* variant upon higher adenoma counts 10-19: OR 3.4 [95% CI 0.3-33.9], 20-29: OR 15.3 [95% CI 1.7-138.0], 30-49: OR 14.5 [95% CI 1.4-144.9], 50-99 OR 86.1 [95% CI 8.6-859.2], >100 OR 64.0 [95% CI 4.9-835.2] p-value <0.001) and at a younger age of adenoma diagnosis (<50: OR 34.8 [95% CI 4.0-300.2], 50-59: OR 4.5 [95% CI 0.5-41.3], 60-69 OR 2.6 [95% CI 0.3-22.7] p-value <0.001). A personal history of CRC or extracolonic manifestations associated with FAP did not affect the odds of detecting *APC* mosaicism.

Table 1 – Phenotypic characteristics comparison between cases with and without *APC* mosaicism. Gender, adenoma count age at first adenoma and CRC were significantly different between the two groups.

	Total	*APC* mosaicism	No *APC* mosaicism	*p-value*
Total	458	51	382	
Gender – n (%)				
Male	254	15 (29.4)	239 (58.7)	0.002*
Female	124	20 (39.2)	104 (25.6)	
Unknown	80	16 (31.4)	64 (15.7)	
Adenoma count – n (%)				<0.001*
0	6	0 (0)	6 (1.5)	
1-9	52	2 (3.9)	50 (12.3)	
10-19	172	7 (13.7)	165 (40.5)	
20-29	128	13 (25.5)	115 (28.3)	
30-49	63	9 (17.6)	54 (13.3)	
50-99	24	12 (23.5)	12 (2.9)	
>100	13	8 (15.7)	5 (1.2)	
Age first adenoma – mean (min-max)	57.9 (17-84)	43.5 (17-72)	59.8 (25-84)	<0.001*
CRC – n (%)	125	16 (31.4)	109 (26.8)	0.506
Age first CRC – mean (min-max)	57.2 (25-84)	47.0 (25-69)	58.7 (25-84)	0.004*
Extracolonic FAP manifestations – n (%)	26	5 (9.8)	21 (6.1)	0.163

APC mosaicism characteristics

Normal tissues tested

As summarized in supplemental table 1, leukocyte, urine and/or buccal swab DNA was analyzed for the mosaic variant in 31 patients. Distinct patterns of *APC* mosaicism were detected; from the mosaicism restricted to the colon (for example L ID 2) to extensive mosaicism throughout the entire body (for example L ID 13). Although the power is too low to reach significance, a trend was detected in higher numbers of adenomas when finding the mosaic variant in one other tissues tested (mean none: 37.2, mean at least one: 73.1). Additionally, the mosaic *APC* variant was detected in normal colon mucosa of 53% (9 out of 17 tested) of patients.

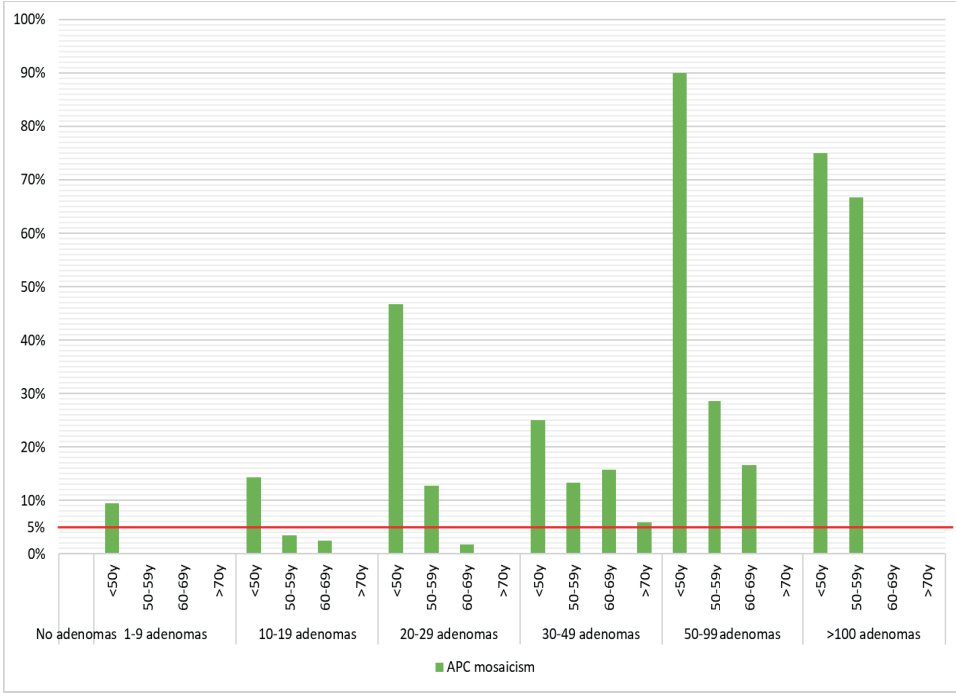


Figure 1 – Detection rates of *APC* mosaicism subdivided in adenoma count groups and stratified for age at first adenoma.

Extracolonic phenotype

An esophagogastroduodenoscopy (EGD) was performed in 32 mosaic patients of which 9 patients developed gastric or duodenum adenomas. Besides extensive mosaicism and FAP associated manifestations in L ID 13, no other patients presented with symptoms associated with FAP outside the gastrointestinal tract.

Heritability

In none of 21 children from 13 mosaic patients, the mosaic variant was detected in leukocyte DNA. Notably, in 10 of these patients leukocyte, urine and buccal swap DNA was analyzed and showed in 9 patients that the mosaicism was restricted to the colon. Interestingly, in one of the two patients (L ID 117, 281) with a desire to have children from whom semen was analyzed, the mosaic variant was detected with a variant allele frequency of 15% and 18% in duplicate testing.

Table 3 – Multivariable logistic regression analysis. The odds of finding an *APC* mosaic case increases significantly upon adenoma count and a younger age of first adenoma.

	N	OR (95% CI)	*p-value*
Adenoma count			
1-9	48	Ref	<0.001*
10-19	166	3.4 (0.3-33.9)	
20-29	116	15.3 (1.7-138.0)	
30-49	57	14.5 (1.4-144.9)	
50-99	20	86.1 (8.6-859.2)	
>100	11	64.0 (4.9-835.2)	
Age first adenoma			
>70	53	Ref	<0.001*
60-69	167	2.6 (0.3-22.7)	
50-59	121	4.5 (0.5-41.3)	
<50	77	34.8 (4.0-300.2)	
CRC (age first)			
No	309	Ref	0.919
<50	23	1.0 (0.2-4.1)	
50-59	27	1.2 (0.2-7.4)	
>60	59	1.7 (0.4-7.6)	
Extracolonic FAP manifestations			
No	394	Ref	0.812
Yes	23	0.8 (0.2-4.1)	

BMI and lifestyle (LUMC cohort only)

Table 4 shows that the non-mosaic group consists of significantly more males, smokers and are more frequently overweight or obese compared to the mosaic group. The distribution of never, former and current alcohol consumers is comparable between the two groups but the number of glasses per week is significantly higher in the non-mosaic group.

Hybrid mosaicism prevalence and phenotype (LUMC cohort only)

Seventy-nine patients (21%) were identified with a so-called hybrid mosaicism, multiple but not all lesions sharing the same *APC* variant, and are summarized in supplemental table 2. Interestingly, 31 hybrid mosaic patients have a recurrent *APC* variant fitting the colibactin-associated mutational signature as described before.[16]

Table 5 shows that hybrid mosaic cases were significantly different from the *APC* mosaic

cases in gender, adenoma count and age of first adenoma. However, hybrid mosaics were phenotypically comparable to the non-mosaic group. Therefore, hybrid mosaic cases were included as non-mosaic cases in all descriptions and analyses.

Table 4 – Comparison of BMI and lifestyle data between patients with and without *APC* mosaicism.

	Total	***APC* mosaicism**	**No *APC* mosaicism**	***p-value***
Gender – n (%)	379	36	343	
Male	254	15 (41.7)	239 (69.7)	0.002*
BMI – n (%)	135	11	124	0.042*
Underweight	2	0 (0.0)	2 (1.6)	
Healthy	47	8 (72.7)	39 (31.5)	
Overweight	61	3 (27.3)	58 (46.8)	
Obese	25	0 (0.0)	25 (20.2)	
BMI – mean (min-max)	26.8 (16.5-48.4)	23.1 (19.2-26.9)	27.2 (16.5-48.4)	0.004*
Smoking – n (%)	229	21	208	0.051
Never	68	11 (52.4)	57 (27.4)	
Former	82	6 (28.6)	76 (36.5)	
Current	79	4 (19.0)	75 (36.1)	
Smoking PY – mean (min-max)	14.5 (0-73)	4.2 (0-51)	15.6 (0-73)	0.016*
Alcohol – n (%)	219	19	200	0.152
Never	38	6 (31.6)	32 (16.0)	
Former	12	0 (0.0)	12 (6.0)	
Current	169	13 (68.4)	156 (78.0)	
Glasses/week – mean (min-max)	15.0 (0-168)	4.2 (0-14)	15.5 (0-90)	0.001*
Aspirin – n (%)	65	4	61	
Yes	9	1 (25.0)	8 (13.1)	0.458
NSAIDs – n (%)	62	3	59	
Yes	3	0 (0.0)	3 (5.1)	1.000
Antihypertensive – n (%)	64	3	61	
Yes	28	3 (100.0)	25 (41.0)	0.079

Discussion

This study reports on 458 hospital and population based patients of whom multiple colorectal adenomas or carcinomas were tested for *APC* mosaicism in Leiden, Rotterdam and Groningen. With 51 mosaic patients, a detection rate of 11.1% was found. This rate was 17.1%

(46 out of 269) and 2.6% (5 out of 189) for patients falling within and outside the scope of the Dutch polyposis and colorectal cancer guidelines for testing for germline variants; (≥10 adenomas before the age of 60 years and ≥20 adenomas before the age of 70 years).[15] These rates were substantially lower than the study of Jansen et al.[19] identifying *APC* mosaicism in patients with >20 adenomas (9 out of 18; 50%). However, the detection rate in the current cohort is 21.6% (41 out of 190) in patients with >20 adenomas before the age of 70, which is comparable to another previously performed study.[12]

Although the three centers used different testing criteria, the selection bias was minimized in the current study since the LUMC cohort consisted of the cohort of Jansen et al., all unexplained polyposis patients send in for germline genetic testing and patients falling outside the scope of the testing guidelines.

Table 5 – Phenotypic comparison of hybrid cases with mosaic cases and no mosaic cases. Hybrid mosaic cases are more comparable with cases without *APC* mosaicism.

	Total	Mosaicism	No mosaicism	Hybrid mosaicism	p-value (hybrid vs mos)	p-value (hybrid vs no mos)
Total	379	36	263	80		
Male – n (%)	254	15 (41.7)	179 (68.1)	60 (75.0)	0.001*	0.268
Unknown	1	1 (2.8)				
Adenoma count n (%)					<0.001*	0.478
0	6	0 (0.0)	6 (2.3)	0 (0.0)		
1-9	46	1 (2.78)	32 (12.2)	13 (16.3)		
10-19	158	4 (11.1)	117 (44.5)	37 (46.3)		
20-29	93	9 (25.0)	68 (25.9)	16 (20.0)		
30-49	52	7 (19.4)	32 (12.2)	13 (16.3)		
50-99	15	10 (28.7)	4 (1.5)	1 (1.3)		
>100	9	5 (13.9)	4 (1.5)	0 (0.0)		
Age first adenoma mean (min-max)	58.6 (17-84)	42.9 (17-72)	60.6 (26-84)	59.0 (25-81)	<0.001*	0.185
CRC – n (%)	107	11 (30.6)	66 (25.1)	30 (37.5)	0.533	0.034*
Age first CRC mean (min-max)	58.3 (25-84)	50.1 (29-69)	60.4 (27-84)	56.9 (25-75)	0.103	0.152
Extracolonic FAP manifestations mean (min-max)	20	2 (5.6)	16 (6.1)	2 (2.5)	0.587	0.264

As expected, prevalence of *APC* mosaicism increases with adenoma count and a younger age of adenoma diagnosis. Other factors such as suffering from CRC or FAP associated extracolonic did not influence the odds of finding *APC* mosaicism.

Most of mosaic patients presented with ≥10 adenomas before the age of 60 or ≥20 adenomas before the age of 70 and therefore fall within the Dutch hereditary polyposis guidelines which are comparable to the National Comprehensive Cancer Network guideline for genetic/familial colorectal cancer.[14, 15] More specifically, *APC* mosaicism rates were more than 5% in all patients with adenomas before the age of 50 years, ≥20 adenomas before the age of 60 years or ≥30 adenomas before the age of 70 years. Based on these data, together with our estimation of *APC* mosaicism occurring in 1 in 14 000[20], we would recommend *APC* mosaicism testing in all patients negatively tested for pathogenic germline variants with (1) multiple adenomas before the age of 50 years, (2) ≥20 adenomas before the age of 60 years or (3) ≥30 adenomas before the age of 70 years.

This cohort revealed different patterns of *APC* mosaicism; extensive mosaicism throughout the entire body to mosaicism restricted to the colon. Also, a spectrum of colorectal phenotypes were observed; mosaic patients with i.e. 10 adenomas at 59 years of age as well as extensive adenomatous polyposis. These broad spectra complicate universal surveillance guidelines. We suggest that the colonoscopy frequency *APC* mosaic patients should in principle be comparable to (A)FAP guidelines[21]; every one or two years. In case a patient has a mild phenotype with effective polypectomies, the frequency of colonoscopies could be re-evaluated to, for example, once every three years.

Moreover, in 9 out of 32 mosaic patients who underwent an EGD, duodenal or stomach lesions were detected. Therefore, based on our data, we would suggest an EGD at least once, with frequency of follow-up determined by the findings of the first EGD.

Besides consequences for the index patients, identifying *APC* mosaicism is furthermore relevant for family members. According to the Dutch surveillance guidelines, parents and siblings of unexplained polyposis patients are offered regular colonoscopies.[15] Since mosaicism occurs during embryogenesis and is therefore usually an isolated case in a family, asymptomatic parents and siblings can be reassured and no surveillance colonoscopies are needed.

Germline testing of 21 children from 13 mosaic patients showed that the variant was not transmitted to their offspring. This could suggest, as also described before[19], that the chances of hereditability are small. Notable, in ten of these patients, DNA from leukocytes, urine and buccal swab was analyzed and showed in 9 patients the mosaicism was restricted to the colon. Also, a mosaic variant was detected in semen DNA of another patient with a desire to have children. Therefore, transmission cannot be excluded and we would still recommend testing children, especially in cases with mosaicism exceeding the colon.

Six mosaic patients would have been missed using our suggested testing guidelines (L ID 10, 19, 31 and 361, R ID 43, G ID 40). One of these patients (L ID 361) has segmental polyposis, with most adenomas located at the left side of the colon. In this case, mosaicism is likely also restricted to the left side of the colon, explaining a milder phenotype. Another patient (R ID 43) was diagnosed with both colorectal and duodenal adenomas.

Two other patients (L ID 10 and 19) have mosaic variants fitting the mutational signature associated with colibactin.[16, 22, 23] The development of adenomas is likely due to the presence of colibactin and may not be due to a mosaicism. Therefore, these patients probably do not develop as much colorectal adenomas as 'real' mosaic patients. To gain more insight and provide suitable screening guidelines, further research into the association between colibactin, adenoma development and possible prevention or eradication is required.

Although most hybrid mosaic cases have a recurring *APC* variant fitting the colibactin mutational signature, other possible explanations are clonal relationship and contamination, for example mixing two adenomas during colonoscopy. Since there is no universal explanation that fits all hybrid cases, case by case evaluation is required. Still, phenotypically hybrid cases are comparable to the non-mosaic cases and should therefore, in general, be treated as such.

Non-mosaic group were have a significantly higher BMI, are more commonly former or current smokers and drink more glasses of alcohol per week compared to the mosaic patients. Although more extensive research is needed, these findings suggest that these modifiable risk factors have possibly contributed to the development of colorectal adenomas in the non-mosaic group, as previously proposed for colorectal cancer.[24, 25]

In conclusion, we recommend testing for *APC* mosaicism in all patients with (1) multiple adenomas before the age of 50 years, (2) ≥20 adenomas before the age of 60 years or (3) ≥30 adenomas before the age of 70 years. Comparable to FAP patients, we suggest patients with *APC* mosaicism should be offered regular colonoscopy screening and an esophagogastroduodenoscopy at least once. The frequency of additional endoscopies can be dependent on the findings. Furthermore, despite the small chance of transmission, we recommend considering germline testing in children of mosaic cases, especially when mosaicism is exceeding the colon.

References

1. Bisgaard ML, Fenger K, Bülow S, et al. Familial adenomatous polyposis (FAP): frequency, penetrance, and mutation rate. Hum Mutat 1994;3:121-5.
2. Jasperson KW, Patel SG, Ahnen DJ. APC-Associated Polyposis Conditions. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Gripp KW, Mirzaa GM, Amemiya A, eds. GeneReviews((R)). Seattle (WA), 1993.
3. Bulow S, Alm T, Fausa O, et al. Duodenal adenomatosis in familial adenomatous polyposis. DAF Project Group. Int J Colorectal Dis 1995;10:43-6.
4. Nieuwenhuis MH, Mathus-Vliegen EM, Baeten CG, et al. Evaluation of management of desmoid tumours associated with familial adenomatous polyposis in Dutch patients. Br J Cancer 2011;104:37-42.
5. 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.
6. Aretz S, Stienen D, Friedrichs N, et al. Somatic APC mosaicism: a frequent cause of familial adenomatous polyposis (FAP). Hum Mutat 2007;28:985-992.
7. Hes FJ, Nielsen M, Bik EC, et al. Somatic APC mosaicism: an underestimated cause of polyposis coli. Gut 2008;57:71-76.
8. Out AA, Minderhout IJHMv, Stoep Nvd, et al. High-resolution melting (HRM) re-analysis of a polyposis patients cohort reveals previously undetected heterozygous and mosaic APC gene mutations. 14 2015.
9. Rahbari R, Wuster A, Lindsay SJ, et al. Timing, rates and spectra of human germline mutation. Nat Genet 2016;48:126-133.
10. 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.
11. 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.
12. 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.
13. Syngal S, Brand RE, Church JM, et al. ACG clinical guideline: Genetic testing and management of hereditary gastrointestinal cancer syndromes. Am J Gastroenterol 2015;110:223-62; quiz 263.
14. Gupta S, Provenzale D, Regenbogen SE, et al. NCCN Guidelines Insights: Genetic/Familial High-Risk Assessment: Colorectal, Version 3.2017. J Natl Compr Canc Netw 2017;15:1465-1475.
15. Tumoren SOE, Nederland VKG, Oncogenetica W. Erfelijke tumoren: Richtlijnen voor diagnostiek en preventie. https://www.stoet.nl/wp-content/uploads/2017/04/STOET-Richtlijnenboekje-april2017_DEF.pdf 2017.
16. Terlouw D, Suerink M, Boot A, et al. Recurrent APC Splice Variant c.835-8A>G in Patients With Unexplained Colorectal Polyposis Fulfilling the Colibactin Mutational Signature. Gastroenterology 2020;159:1612-1614.e5.
17. Neveling K, Mensenkamp AR, Derks R, et al. BRCA Testing by Single-Molecule Molecular Inversion Probes. Clin Chem 2017;63:503-512.
18. Weren RD, Mensenkamp AR, Simons M, et al. Novel BRCA1 and BRCA2 Tumor Test as Basis

for Treatment Decisions and Referral for Genetic Counselling of Patients with Ovarian Carcinomas. Hum Mutat 2017;38:226-235.

19. Jansen AML, Crobach S, Geurts-Giele WRR, et al. Distinct Patterns of Somatic Mosaicism in the APC Gene in Neoplasms From Patients With Unexplained Adenomatous Polyposis. Gastroenterology 2017;152:546-+.
20. Terlouw D, Hes FJ, Suerink M, et al. APC mosaicism, not always isolated: two first-degree relatives with apparently distinct APC mosaicism. Gut 2022.
21. Yen T, Stanich PP, Axell L, et al. APC-Associated Polyposis Conditions. In: Adam MP, Everman DB, Mirzaa GM, Pagon RA, Wallace SE, Bean LJH, Gripp KW, Amemiya A, eds. GeneReviews((R)). Seattle (WA), 1993.
22. Boot A, Ng AWT, Chong FT, et al. Characterization of colibactin-associated mutational signature in an Asian oral squamous cell carcinoma and in other mucosal tumor types. Genome Res 2020;30:803-813.
23. Pleguezuelos-Manzano C, Puschhof J, Huber AR, et al. Mutational signature in colorectal cancer caused by genotoxic pks(+)E. coli. Nature 2020;580:269-+.
24. Botteri E, Borroni E, Sloan EK, et al. Smoking and Colorectal Cancer Risk, Overall and by Molecular Subtypes: A Meta-Analysis. Am J Gastroenterol 2020;115:1940-1949.
25. Soltani G, Poursheikhani A, Yassi M, et al. Obesity, diabetes and the risk of colorectal adenoma and cancer. BMC Endocr Disord 2019;19:113.

Supplemental table 1 – Overview of all patients with a mosaic *APC* variant.

C	ID	Age	Ad n	Guide	FU (months)	Cum Ad n	Extracolonic	Gastro	Mosaic variant	Leu	Ur	Buccal	Normal	Other	Other specified	Children
L	165	24	4	Yes	42	5		Yes	c.835-8A>G	0	0	0	0	NT		-
L	394	29	16	Yes	1	16		No	c.3688C>T	0	2	0.03	NT	NT		-
L	319	33	16	Yes	12	16		Yes	c.3107_3108del	2	0	1	NT	NT		-
L	361	64	18	No	6	18		Yes	c.1417C>T	NT	NT	NT	10	NT		-
L	10	69	19	No	67	19	BC60	Yes	c.835-8A>G	0	0	0	NT	NT		0/1
L	224	59	>20	Yes	28	>25	Fibroadenoma breast 51, dermato fibroma 57, BC60	Yes	c.1868_1883del	0.5	1.6	0.7	NT	NT		-
L	31	65	20	No	54	21		Yes	c.646C>T	0	0	0	NT	NT		-
L	14	29	21	Yes	60	30		Yes	c.3340C>T	0	0	0	NT	NT		-
L	163	55	21	Yes	41	21		Yes	c.3959_3960del	0	0	0	NT	NT		-
L	300	52	21	Yes	16	34		Yes	c.509_512del	0	0	0	NT	NT		0/1
L	219	38	23	Yes	31	29	1 duodenal ad 58	Yes	c.4393_4394del	0	0	NT	NT	NT		-
L	109	48	25	Yes				?	c.3340C>T	0	NT	NT	0	0	Ileum and lymph node	?
L	43	34	27	Yes	81	49		Yes	c.2566_2572del	0	NT	0	NT	NT		0/1
L	117	22	29	Yes	46	29		?	c.4192_4193del	0	0	0	NT	0	Semen	?
L	100	67	>30	Yes	82	>67		Yes	c.3211C>T	0	0	0	NT	NT		0/1
L	4	63	>30	Yes	72	>30	Lipoma64	Yes	c.4393_4394del	0	0	0	NT	NT		0/1
L	37	67	>30	Yes	84	65		Yes	c.1959-1G>A	0	0	0	15, 29	NT		0/1
L	244	43	30	Yes	22	38	3 stomach adenomas 59, fundic gland polyps	Yes	c.2612del	NT	NT	NT	NT	32	Stomach	-

Supplemental table 1 – Overview of all patients with a mosaic *APC* variant. Continued.

C	ID	Age	Ad n	Guide	FU (months)	Cum Ad n	Extracolonic	Gastro	Mosaic variant	Leu	Ur	Buccal	Normal	Other	Other specified	Children
L	208	40	32	Yes	33	48	BCC 61	Yes	c.1660C>T	0	0	0	NT	NT		-
L	2	18	46	Yes	135	59		No	c.4057G>T	0	0	0	14	0	Fibroblasts	0/2
L	19	72	49	No	70	49		Yes	c.2977A>T	0	0	0	NT	NT		-
L	9	22	>50	Yes	61	>50		Yes	c.637C>T	0	NT	NT	NT	0	Fibroblasts	-
L	250	34	>50	Yes	22	>56		Yes	c.2635C>T	0	0	0	NT	NT		-
L	18	26	>50	Yes	69	92	Possible adenomatous spot 5 mm, no PA	Yes	c.4393_4394del	0	0	NT	NT	NT		-
L	324	42	50	Yes	12	50		Yes	c.1960C>T	1	0	0	NT	NT		0/3
L	20	63	51	Yes	55	60	3 duodenal polyps 77, kerato-anchantoma 71, neurofibroma 78, M85	Yes	c.847C>T	0	0	0	NT	NT		0/2
L	34	49	58	Yes	83	58		Yes	c.1974_1975del	NT	NT	NT	2	NT		-
L	143	50	70	Yes	44	70		No	c.1960C>T	3	2	0	NT	NT		-
L	281	23	79	Yes	17	86		Yes	c.1879_1882del	6	9	7	NT	15, 18	Semen	-
L	16	46	80	Yes	60	80		No	c.3712_3713del	0	NT	0	NT	NT		0/4
L	27	35	80	Yes	104	123		Yes	c.2493dupA	3	NT	NT	NT	NT		-
L	98	37	>100	Yes				No	c.4110_4111del	4	NT	NT	4	NT		?
L	8	17	>100	Yes	78	>100		Yes	c.637C>T	0	NT	NT	NT	0	Fibroblasts	?
L	99	54	>100	Yes	99	>100	BC73	Yes	c.3904del	1	NT	NT	NT	NT		?
L	24	27	>100	Yes			Duodenal adenomas, fundic gland polyps	Yes	c.4666dup	NT	NT	NT	NT	NT		?

Supplemental table 1 – Overview of all patients with a mosaic *APC* variant. Continued.

C	ID	Age	Ad n	Guide	FU (months)	Cum Ad n	Extracolonic	Gastro	Mosaic variant	Leu	Ur	Buccal	Normal	Other	Other specified	Children
L	13	28	>200	Yes	34	>200	Duodenal polyposis, fundic gland polyps, osteoma, lipomas	Yes	c.4391_4394del	21	16	14	29	18	Lymph node	-
R	43	50	10	No		?	2 duodenal adenomas 50	Yes	c.1779G>A	0	?	0	41	?		?
R	26	51	20	Yes		?	?	?	c.4393_4394del	?	?	?	0	?		?
R	6	39	25	Yes		?	?	?	c.4057G>T	?	?	?	0	?		?
R	37	27	25	Yes		?	Fundic gland polypos 27	Yes	c.4393_4394del	4	?	?	NT	?		?
R	17	57	60	Yes		?	?	?	c.707_710dup	?	?	?	0	?		?
G	2	37	<10	Yes		?	?	?	c.4479_4480insCA	?	?	?	0	?		0/1
G	22	45	16	Yes		?	?	?	?	?	?	?	?	?		?
G	40	59	18	No		?	?	?	c.423-9A>G	?	?	?	5	?		-
G	20	55	>25	Yes		?	?	?	?	?	?	?	?	?		?
G	5	57	32	Yes		?	?	?	c.1987C>T	?	?	?	1	?		0/1
G	10	52	>32	Yes		?	?	?	c.3724C>T	?	?	?	?	?		?
G	21	21	>50	Yes		?	?	?	c.4012C>T	?	?	?	0	?		-
G	3	25	>100	Yes		?	?	?	c.3928_3929del	?	?	?	?	?		-
G	41	43	>100	Yes		?	Stomach polyps 43	Yes	c.3880C>T	?	?	?	0	?		0/1
G	4	55	>100	Yes		?		?	c.3286C>T	?	?	?	?	?		-

C – center of mosaicism testing, L – Leiden, R – Rotterdam, G – Groningen, Age – age at first adenoma, Ad (n) – number of adenomas, Guide – falling within the scope of the guideline suggested by this study, FU – follow up in months, Cum ad n – cumulatively developed adenomas at follow up, Extracolonic – extracolonic manifestations, Gastro – Gastroduodenoscopy performed. Leu, Ur, Buccal, Normal, Other – variant allele frequency of mosaic variant in leukocyte, urine, buccal swab, normal colon mucosa or other tissues, NT – not tested. Children – number of children with mosaic variant/ number of children tested for the variant, - – no children tested, ? – unknown

Supplemental table 2 – Overview of all patients with a hybrid *APC* variant.

ID	Age	Ad (n)	Hybrid variant	N hybrid/N tested	Explanation?	Leu	Ur	Buccal	Normal	Other
3	69	22	c.835-8A>G	2/7	Clb					
5	61	31	(c.3934G>T & c.1411G>A) & c.847C>T	2/7 & 2/7						
6	75	19	(c.2741_2742delinsAG & c.4348C>T) & c.4233del	2/7 & 2/7						
11	50	15	c.835-8A>G & (c.4549C>T & c.2008A>T)	3/7	Clb	0	0	0	NT	NT
12	65	13	c.835-8A>G & (c.1409-5A>G & c.2222del)	2/9 & 2/9	Clb	NT	NT	NT	0	NT
15	59	12	c.835-8A>G	3/12	Clb					
17	66	70	c.835-8A>G & c.1690C>T & c.637C>T	3/10 & 2/10 & 2/10	Clb	0	NT	NT	NT	NT
21	63	>30	c.835-8A>G & c.4348C>T	2/10 & 2/10	Clb					
23	58	15	(c.2205del & c.70C>T)	2/10	Clonal/Contamination					
30	67	>20	c.4053dup & c.4245del	2/6 & 2/6						
39	66	>10	c.4463del	2/5						
78	61	34	c.4348C>T & c.2626C>T	2/5 & 2/5						
84	67	24	(c.3682C>T & c.4733_4734del)	2/4	Clonal/Contamination					
91	67	13	c.3340C>T	2/6						
92	50	6	(c.1213C>T, c.4630G>T & c.2413C>T)	2/6	Clonal/Contamination					
96	65	10	c.835-8A>G	3/5	Clb					
105	54	38	c.4666dup	3/9						
110	64	18	c.2626C>T	2/4						
120	59	>30	c.637C>T	2/4						
122	55	10	c.835-8A>G	2/4	Clb					
125	55	2	c.835-8A>G	2/4	Clb	NT	NT	NT	0	NT

Supplemental table 2 – Overview of all patients with a hybrid *APC* variant. Continued.

ID	Age	Ad (n)	Hybrid variant	N hybrid/N tested	Explanation?	Leu	Ur	Buccal	Normal	Other
128	52	10	c.835-8A>G & (c.1600A>T & c.423-6A>G)	6/10 & 2/10	Clb	0	0	0	0	NT
132	65	15	c.4348C>T	2/8						
134	64	13	(c.4429C>T & c.3289G>T)	2/4	Clonal/Contamination					
135	46	30	c.4099C>T	2/4						
138	59	2	c.3493A>T	2/3	Clb					
147	67	16	c.3991A>T	2/4						
148	74	>30	c.4348C>T (& c.2364_2365delinsAT)	3/4 (2/4)						
152	81	28	c.835-8A>G	2/6	Clb	NT	NT	NT	0	NT
156	74	36	c.835-8A>G	2/4	Clb					
168	55	17	c.835-8A>G	4/7	Clb					
170	57	13	c.4348C>T	2/4						
173	71	27	c.1548G>T c.646C>T	2/10 & 2/10						
174	54	10	c.4099C>T & c.835-8A>G	2/6 & 2/6	Clb					
175	41	4	c.3441C>G	3/4		0	0	0	NT	NT
178	69	>20	c.5626A>G	2/6						
188	42	14	c.4348C>T	4/7		0	0	0	NT	NT
198	69	30	c.835-8A>G	2/4	Clb					
202	63	14	c.3916G>T & c.646C>T	2/6 & 2/6						
212	44	24	(c.757_761dup & c.1958+1_1958+2dup)	2/4	Clonal/Contamination					
217	59	10	c.835-8A>G	2/4	Clb					
218	50	25	c.835-8A>G	5/8	Clb					

Supplemental table 2 – Overview of all patients with a hybrid *APC* variant. Continued.

ID	Age	Ad (n)	Hybrid variant	N hybrid/N tested	Explanation?	Leu	Ur	Buccal	Normal	Other
227	59	10	c.835-8A>G	2/4	Clb					
232	61	16	c.835-8A>G	2/4	Clb					
239	63	25	c.847C>T	4/6						
242	60	28	c.835-8A>G	2/4	Clb					
243	63	25	c.835-8A>G	3/4	Clb					
246	40	10	c.835-8A>G	2/4	Clb					
258	55	13	c.835-8A>G	3/4	Clb					
261	66	8	c.4348C>T	2/4						
268	63	16	c.3788_3789delinsAG	2/7						
272	64	18	c.3340C>T	2/5						
278	40	10	c.835-8A>G	2/4	Clb					
280	49	10	c.3088A>T	3/4	Clb					
283	50	4	c.835-8A>G	2/3	Clb					
287	41	6	c.694C>T	2/4						
288	32	8	c.2767A>T & c.834+2T>C	2/4 & 2/4						
294	25	4	c.3934G>T & c.4116_4117dup	2/4 & 2/4						
303	61	11	c.556A>T	2/5						
304	55	13	c.4130_4137del	2/4						
305	54	5	c.835-8A>G	3/4	Clb					
311	63	9	c.3905del	2/5						
317	59	11	c.646C>T	3/6						

Supplemental table 2 – Overview of all patients with a hybrid *APC* variant. Continued.

ID	Age	Ad (n)	Hybrid variant	N hybrid/N tested	Explanation?	Leu	Ur	Buccal	Normal	Other
320	53	7	c.847C>T	2/4						
323	61	13	c.835-8A>G	2/4	Clb					
329	67	19	c.835-8A>G	2/4	Clb					
338	53	20	c.4057G>T	2/6						
340	63	13	(c.4526dup & KRAS)	6/8	Clonal					
348	59	22	c.847C>T	2/6						
349	58	6	c.893_894del & c.1948G>T	2/4 & 2/4						
353	57	21	c.4678G>T & c.4729_4735del	2/4 & 2/4						
363	65	40-60	c.1987C>T	2/4						
368	62	12	c.3927_3931del	2/4						
370	71	15	c.4348C>T	2/3						
372	50	multi-ple	c.3010_3019dup	3/4		0	NT	0	NT	NT
373	55	24	(c.3199C>T & SMAD4)	2/7	Clonal					
380	65	16	c.3907C>T	2/4						

Age – age at first adenoma, Ad (n) – number of adenomas, Hybrid variant – recurring *APC* variant in a proportion of adenomas, (..) – variants occurring together in multiple adenomas, N hybrid/N tested – number of adenomas with hybrid variant/number of adenomas tested. Explanation? – possible explanation for the hybrid variant, Clb – colibactin, Clonal – clonal relationship, Contam – contamination. Leu, Ur, Buccal, Normal, Other – variant allele frequency of hybrid variant in leukocyte, urine, buccal swab, normal colon mucosa or other tissues, NT – not tested.

APC mosaicism, not always isolated: two first-degree relatives with apparently distinct *APC* mosaicism

Gut 2022

Diantha Terlouw, Frederik J Hes, Manon Suerink, Arnoud Boot, Alexandra M J Langers, Carli M Tops, Monique E van Leerdam, Christi J van Asperen, Steve G Rozen, Emilia K Bijlsma, Tom van Wezel, Hans Morreau, Maartje Nielsen

APC mosaicism is briefly mentioned in the recently published BSG guidelines on hereditary colorectal cancer.[1] We wish to present a family that underlines its relevance.

A 26-year- old woman presented at the department of Clinical Genetics at Leiden University Medical Center with osteomas, lipomas, extra tooth and bowel problems. Using Sanger sequencing on leucocyte DNA, a pathogenic *APC* variant (NM_000038.6:c.4391_4394delAGAG) was detected with variant allele frequency of 20%–40%, indicating mosaicism. Subsequent colonoscopy and gastroduodenoscopy showed >200 colorectal adenomas in a patchy pattern and large variation in size and morphology, extended duodenal and gastric fundic gland polyposis. As summarised in table 1, targeted next-generation sequencing (NGS) on three adenomas and multiple normal tissues showed the same *APC* variant in all samples, suggesting an extensive mosaic pattern.

Seven years earlier, her father presented with >40 colorectal adenomas without germline pathogenic *APC* or *MUTYH* variants. However, NGS on colorectal adenomas showed another mosaic *APC* variant (NM_000038.6:c.3712_3713delAG) while in leucocyte and urine DNA this variant was absent, suggesting *APC m*osaicism limited to the colon. Figure 1 illustrates the family.

Previous studies on multiple *de novo* variants[2, 3] discuss two hypotheses: (1) an underlying hereditable defect in a DNA repair gene or (2) an underestimated *de novo* variant rate.

(1) To elucidate possible underlying genetic explanation for two mosaic cases in this family, whole-exome sequencing (WES) and mutational signature analysis by using whole-genome sequencing (WGS) were performed on leucocyte DNA and colorectal adenoma DNA respectively. WES data revealed seven rare truncating variants shared by both family members, however, these do not offer a relevant explanation of multiple mosaics in this family. In addition, the patients shared >300 rare non-synonymous coding or splicing variants. The three variants in genes associated with DNA damage repair were unlikely pathogenic.

As both patients have a mosaic deletion of AG, the search for mutational signature analysis could hint towards a specific underlying gene defect. Using SigProfiler[4], mutational signatures were assigned on WGS data. All analysed adenomas with single base substitutions show SBS1 and SBS5, both associated with ageing and observed across many tissue types.[5] No other signature was shared between the adenomas.

(2) Le Caignec et al[3] determined the probability of finding three de novo variants in *TSC2* in one family using a mathematical formula based on the *de novo* variant rate of *TSC2*. The *de novo* variant rate of *APC* is estimated between 4×10–6 and 9×10–6.[6] When imported into the same formula, these rates result in a probability of two *de novo APC* variants between

Table 1 – Overview of all tissues tested with NGS and variant allele frequencies of the detected mosaic variant. In all tissues tested from the daughter (patient 1) the mosaic variant c.4391_4394delAGAG was detected with various variant allele frequencies (VAF) ranging from 14 to 49%. The mosaic variant of the father (patient 2), c.3712_3713delAG, was solely detected in colonic tissue with a VAF ranging from 10 to 44%.

Tissue analysed with NGS	Mosaic APC variant	Cov	VAF (%)
Patient 1 (daughter)			
Adenoma 1	c.4391_4394delAGAG	1084	37
Adenoma 2	c.4391_4394delAGAG	1969	45
Adenoma 3	c.4391_4394delAGAG	1806	49
Normal mucosa	c.4391_4394delAGAG	1957	29
Lymph node	c.4391_4394delAGAG	1965	18
Leukocyte	c.4391_4394delAGAG	772	21
Urine	c.4391_4394delAGAG	1119	16
Buccal swab	c.4391_4394delAGAG	666	14
Patient 2 (father)			
Adenoma 1	c.3712_3713delAG	>2000	44
Adenoma 2	c.3712_3713delAG	>2000	24
Adenoma 3	c.3712_3713delAG	>2000	13
Adenoma 4	c.3712_3713delAG	>2000	33
Adenoma 5	c.3712_3713delAG	>2000	33
Adenoma 6	c.3712_3713delAG	94	16
Adenoma 7	X		
Adenoma 8	c.3712_3713delAG	1992	10
Adenoma 9	c.3712_3713delAG	1975	31
Adenoma 10	c.3712_3713delAG	1976	26
Leukocyte	X		
Urine	*		
Buccal swab	X		

Cov – Coverage, sequencing read count, VAF – variant allele frequency, percentage of reads with the variant, X – mosaic variant not found, * - Analysis not succeeded due to low amount of DNA in the sample.

1:6 250 000 000 and 1:1 230 000 000 families with five members; parents with three children in this particular case. Since we hypothesise that the prevalence of *APC* mosaics might be higher than germline *de novo* variants, we used our previously reported cohort[7, 8] to estimate an *APC* mosaicism rate. We found *APC* mosaicism in 7% of patients with>10 adenomas, which is 1.4x lower than biallelic *MUTYH* variants in this same cohort (10%). As biallelic *MUTYH* is known to have a population rate of 1:10 000, we roughly estimate the *APC* mosaic rate to be 1:14 000. Using this *APC* mosaicism rate in the formula, the estimated probability

of 2 *APC* mosaics in a family of 5 members is 1:20 400 000, which is still extremely low.

Our estimation, the discrepancy in pattern and phenotype and previous studies[9, 10] underline the need for *APC* mosaicism testing guidelines. Our finding furthermore emphasises caution of ruling out *APC* mosaicism in patients with a positive family history for adenomas. Finally, whenever *APC* mosaic patients have affected family members, *APC* mosaicism testing should be considered.

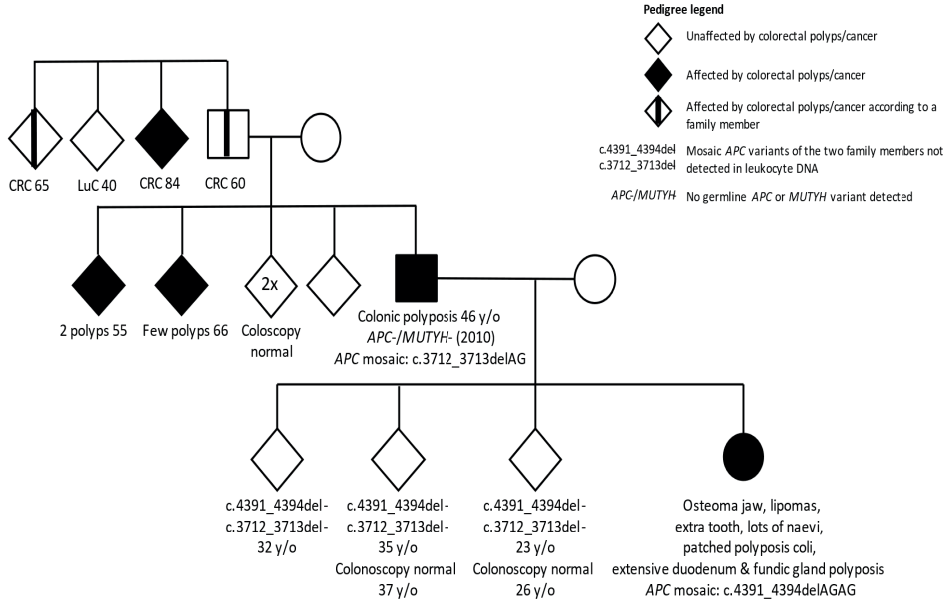


Figure 1 – Pedigree of the paternal family of both mosaic patients. The other children of the father were all genetically tested for the two mosaic *APC* variants in the family (c.4391_4394delAGAG and c.3712_3713delAG) and tested negative. Two of the three children underwent a colonoscopy without detection of polyps. Two of the father's siblings were diagnosed with a few polyps above the age of 55. Three of the siblings of the father's father suffered from colorectal cancer above the age of 60.

References

1. Monahan KJ, Bradshaw N, Dolwani S, et al. Guidelines for the management of hereditary colorectal cancer from the British Society of Gastroenterology (BSG)/Association of Coloproctology of Great Britain and Ireland (ACPGBI)/United Kingdom Cancer Genetics Group (UKCGG). Gut 2020;69:411-444.

2. Garcia B, Catasus N, Ros A, et al. Neurofibromatosis type 1 families with first-degree relatives harbouring distinct NF1 pathogenic variants. Genetic counselling and familial diagnosis: what should be offered? J Med Genet 2022.

3. Le Caignec C, Kwiatkowski DJ, Kury S, et al. Three independent mutations in the TSC2 gene in a family with tuberous sclerosis. Eur J Hum Genet 2009;17:1165-70.

4. Boot A, Ng AWT, Chong FT, et al. Characterization of colibactin-associated mutational signature in an Asian oral squamous cell carcinoma and in other mucosal tumor types. Genome Res 2020;30:803-813.

5. Alexandrov LB, Jones PH, Wedge DC, et al. Clock-like mutational processes in human somatic cells. Nat Genet 2015;47:1402-7.

6. Hes FJ, Nielsen M, Bik EC, et al. Somatic APC mosaicism: an underestimated cause of polyposis coli. Gut 2008;57:71-76.

7. 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.

8. Terlouw D, Suerink M, Boot A, et al. Recurrent APC Splice Variant c.835-8A>G in Patients With Unexplained Colorectal Polyposis Fulfilling the Colibactin Mutational Signature. Gastroenterology 2020;159:1612-1614.e5.

9. Aretz S, Stienen D, Friedrichs N, et al. Somatic APC mosaicism: a frequent cause of familial adenomatous polyposis (FAP). Hum Mutat 2007;28:985-992.

10. Jansen AML, Crobach S, Geurts-Giele WRR, et al. Distinct Patterns of Somatic Mosaicism in the APC Gene in Neoplasms From Patients With Unexplained Adenomatous Polyposis. Gas troenterology 2017;152:546-+.

Exploring *APC* mosaicism in upper intestinal tract adenomas

Manuscript submitted at Clinical Gastroenterology and Hepatology

Diantha Terlouw, Maartje Nielsen, Monique E van Leerdam, Tom van Wezel, Hans Morreau, Alexandra M J Langers

Familial Adenomatous Polyposis (FAP) is caused by germline pathogenic *APC* variants and is characterized by the development of hundreds to thousands colorectal adenomas.[1] Moreover, duodenal adenomas are observed in 50% to 90% of FAP patients, increasing the lifetime risk of small bowel carcinomas up to 12%.[2]

APC mosaicism, i.e. a pathogenic *APC* variant in only a subset of body cells, is an explanation for 25% to 50% of patients diagnosed with more than 20 colorectal adenomas in absence of a germline predisposition.[3, 4] While *APC* mosaicism offers a relevant explanation for colorectal adenomas, the extent of the risk for developing upper intestinal tract adenomas and carcinomas remains unknown. Additionally, the prevalence of *APC* mosaicism in patients with multiple upper intestinal tract adenomas without colorectal adenomas remains undetermined.

In this report, we describe the case of a 61-year-old patient who presented with unspecified upper abdominal pain, underwent a esophagogastroduodenoscopy (EGD) and was diagnosed with more than 10 duodenal adenomas. Following the initial EGD, the patient was referred to the department of Gastroenterology at the Leiden University Medical Center for the removal of an ampullary adenoma with an extensive peripapillary component. Furthermore, at least ten additional polyps were diagnosed in the descending and horizontal part of the duodenum, ranging in size from several millimeters to over two centimeters. Given the number of duodenal adenomas and the periampullary extent of the adenoma, the suspicion of FAP was raised, since periampullary adenomas are present in at least 50% of FAP patients.[5] In the colorectum, however, only one nonadvanced rectal adenoma was detected during multiple colonoscopies performed previously due to four colorectal cancer cases in the paternal family; an aunt at age 55, two uncles and grandmother aged above 80. Although no germline testing performed, targeted Next Generation Sequencing (NGS) was conducted on four duodenal adenomas and the rectum adenoma using the msCRC panel as described before.[6] Figure 1A shows that the same *APC* variant, c.4510_4513dup is present in all tested duodenal adenomas and absent in the rectum adenoma. To exclude analysis of the same lesion or a clonal relationship, two duodenal adenomas were additionally analysed using a small targeted 'Cancer Hotspot' panel. This showed unique somatic variants in *SMAD4*, *ERBB2* and *PTPN11* in T2 and in *KRAS* in T3. The presence of an identical *APC* variant in combination with different somatic variants excludes clonal relationship and confirms mosaicism. The mosaic *APC* variant was not detectable in leukocyte, urine and buccal swab DNA, suggesting a mosaicism restricted to the duodenum.

Based on the findings in this patient, targeted NGS was performed on upper intestinal tract adenomas from 13 additional patients to investigate the extent of *APC* mosaicism in these patients. This cohort included 6 patients with multiple upper intestinal adenomas, 3 pa

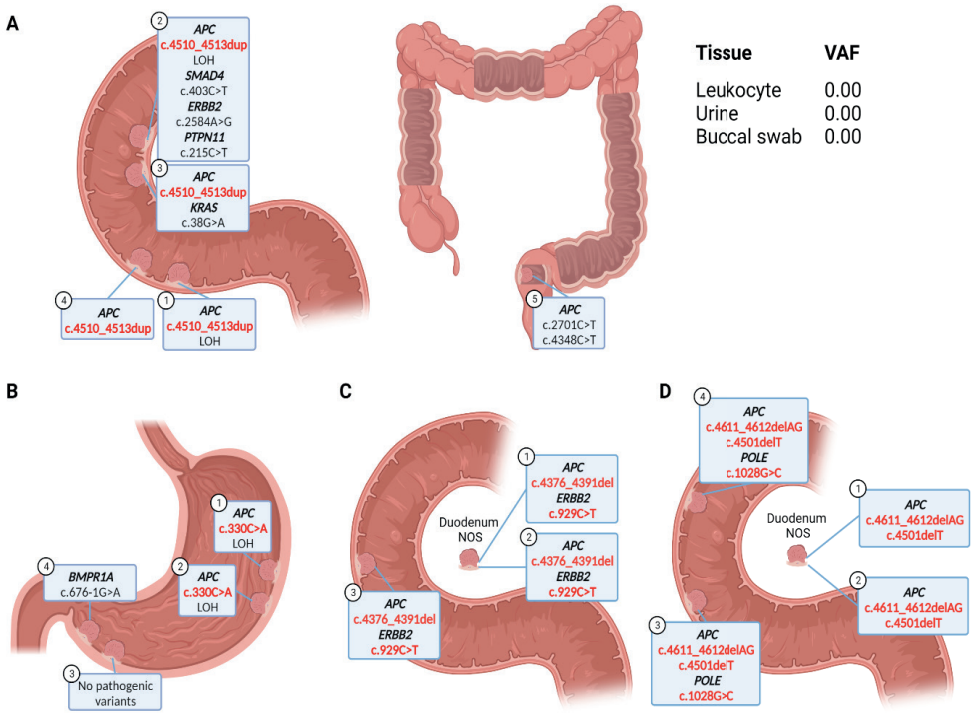


Figure 1 – Patients with at least two adenomas sharing the same *APC* variant. A. The duodenal mosaicism case presented in this report with the *APC* c.4510_4513dup variant in all duodenal adenomas but not in colorectal adenoma or other tissues tested. B. Patient with two gastric adenomas having *APC* c.330C>A (T1,T2), the location T1 and T2 and the lack of this variant in T3 and T4 suggest clonal relationship. C. Three duodenal lesions sharing the same *APC* and *ERBB2* variant which concludes clonal relationship. D. Patient with four duodenal lesions sharing two *APC* variants. T3 and T4 also harbor the same *POLE* variant. *POLE* is technically not covered in T1 and T2. Created with Biorender.com.

tients with one duodenal adenoma in combination with multiple colorectal adenomas and 4 known colorectal *APC* mosaicism patients with upper intestinal adenomas. The clinical characteristics and sequencing results are summarized in supplemental table 1.

No *APC* mosaicism was detected in the 6 patients with multiple upper intestinal adenomas and the 3 patients with one duodenal and multiple colorectal adenomas. These patients were diagnosed with 1 to 12 duodenal or gastric adenomas, with the first diagnosed at ages between 45 and 72. Three patients (patient ID 2-4) showed identical *APC* variants in at least two of the analyzed adenomas, as depicted in figure 1B-D and in supplemental table 1. In one of these three patients, ID 2 (figure 1B), two out of four gastric adenomas (T1 and T2) harbor the same *APC* variant, c.330C>A. No other somatic variants were detected. The vari-

ant was not detectable in the two other gastric adenomas (T3 and T4) and because T1 and T2 were located close to each other, we concluded that there is no convincing evidence for mosaicism and that T1 and T2 are probably clonally related. A rectum adenoma in the same patient was sequenced and the results showed two distinct *APC* variants. Another patient, ID 3 (figure 1C), harbored the *APC* variant c.4376_4391del as well as an identical *ERBB2* in all duodenal adenomas tested suggesting a clonal relationship. The third patient, ID 4 (figure 1D), had two shared *APC* variants in 4 duodenal adenomas, the first and second hit. In 2 of these lesions, T3 and T4, an identical somatic *POLE* variant was detected which was not covered in the sequencing analyses of T1 and T2. Still, the two shared *APC* variants in T1-T4 combined with the *POLE* variant in T3 and T4 suggests a clonal relationship. Furthermore, two colorectal adenomas located in the (recto)sigmoid of this patient did not show any of these *APC* or *POLE* variants. More background information on the classification and interpretation of the findings of the sequencing analysis can be found in the decision tree in supplemental figure 1.

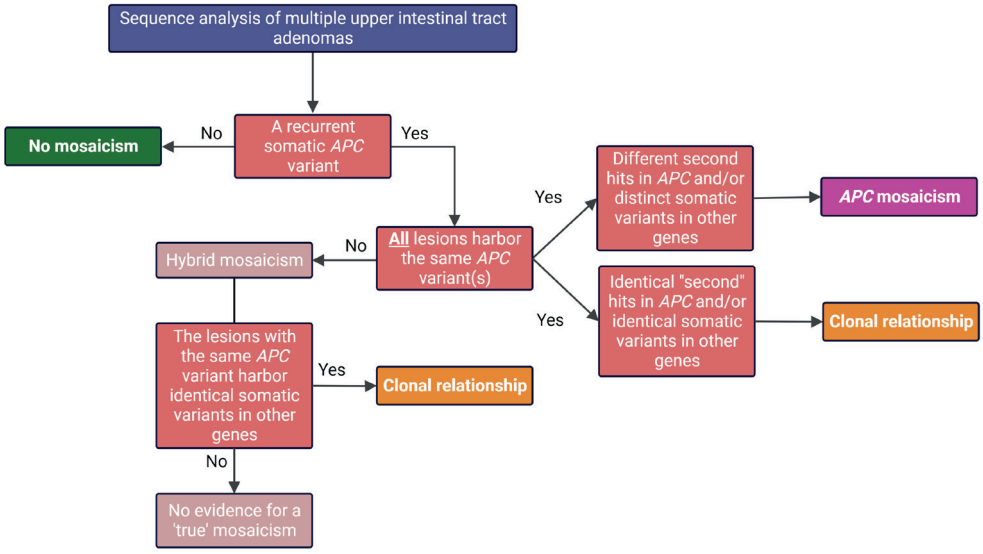
In all four colorectal mosaicism patients, patient ID 11-14, duodenal adenomas showed the mosaic variant found in colorectal adenomas. As summarized in supplemental table 1 and described before[7], in one of these patients the mosaic variant was also detected in leukocyte, urine and buccal swab DNA, showing an extensive mosaicism.

While previous studies[7-9] have described upper intestinal tract adenomas in *APC* mosaic patients, to our knowledge, this is the first reported case of solitary duodenal *APC* mosaicism and the first study to examine *APC* mosaicism in patients with duodenal adenomas regardless of the presence of colorectal adenomatous polyps. Future studies with larger cohorts could support in providing recommendations when to test for *APC* mosaicism in patients with upper intestinal tract adenomas.

The case presented in this current study demonstrates the existence of isolated duodenal *APC* mosaicism in the absence of colorectal adenomas. Furthermore, our study emphasizes the risk of developing upper intestinal tract adenomas in colorectal *APC* mosaicism cases, highlighting the relevance of advising at least one gastroduodenoscopy.

References

1. Kinzler KW, Vogelstein B. Lessons from hereditary colorectal cancer. Cell 1996;87:159-70.
2. Kadmon M, Tandara A, Herfarth C. Duodenal adenomatosis in familial adenomatous polyposis coli. A review of the literature and results from the Heidelberg Polyposis Register. Int J Colorectal Dis 2001;16:63-75.
3. Jansen AML, Crobach S, Geurts-Giele WRR, et al. Distinct Patterns of Somatic Mosaicism in the APC Gene in Neoplasms From Patients With Unexplained Adenomatous Polyposis. Gastroenterology 2017;152:546-+.
4. 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.
5. Mehta NA, Shah RS, Yoon J, et al. Risks, Benefits, and Effects on Management for Biopsy of the Papilla in Patients With Familial Adenomatous Polyposis. Clin Gastroenterol Hepatol 2021;19:760-767.
6. Terlouw D, Suerink M, Boot A, et al. Recurrent APC Splice Variant c.835-8A>G in Patients With Unexplained Colorectal Polyposis Fulfilling the Colibactin Mutational Signature. Gastroenterology 2020;159:1612-1614.e5.
7. Terlouw D, Hes FJ, Suerink M, et al. APC mosaicism, not always isolated: two first-degree relatives with apparently distinct APC mosaicism. Gut 2022.
8. 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.
9. Takao M, Yamaguchi T, Eguchi H, et al. APC germline variant analysis in the adenomatous polyposis phenotype in Japanese patients. Int J Clin Oncol 2021;26:1661-1670.



Supplemental figure 1 – Decision tree for the interpretation of sequencing results of multiple upper intestinal tract adenomas

Supplemental table 1 (part 1) - Patients characteristics and results leukocyte, urine and buccal swab DNA was tested for mosaic variants.

ID	Adenomas colon (age first)	Adenomas duodenum (age first)	Adenomas stomach (age first)	Conclusion	Leukocyte	Urine	Buccal swab
1	1 (56)	9 (60)	0	Mosaicism duodenum	0	0	0
2	1 (53)	0	4 (47)	No mosaicism (clonal relationship)	-	-	-
3	0	3 (45)	0	No mosaicism (clonal relationship)	-	-	-
4	>9 (63)	12 (63)	0	No mosaicism (clonal relationship)	-	-	-
5	32 (67)	3 (69)	0	No mosaicism	-	-	-
6	19 (64)	4 (64)	0	No mosaicism	-	-	-
7	3 (71)	4 (72)	0	No mosaicism	-	-	-
8	23 (55)	1 (57)	0	No mosaicism	-	-	-
9	8 (66)	1 (67)	0	No mosaicism	-	-	-
10	25 (49)	1 (49)	0	Hybrid mosaicism colon	-	-	-
11	30 (43)	0	4 (58)	Mosaicism colon and stomach	-	-	-
12	>200 (27)	>50 (27)	0	Extensive mosaicism throughout the body	0.21	0.16	0.14
13	51 (62)	3 (76)	0	Mosaicism colon and duodenum	0	0	0
14	23 (37)	1 (58)	0	Mosaicism colon and duodenum	0	0	-

Supplemental table 1 (part 2) - Variants detected using targeted NGS in adenomas.

ID	Location (T1)	Variants (VAF)	Location (T2)	Variants (VAF)
1	Horizontal part of duodenum	APC:c.4510_4513dup (0.31)	Duodenum	APC:c.4510_4513dup (0.33)
				SMAD4:c.403C>T (0.19)
				ERBB2:c.2584A>G (0.20)
				PTPN11:c.215C>T (0.19)
2	Stomach body	APC:c.330C>A (0.35)	Stomach body	APC:c.330C>A (0.29)
3	Duodenum	APC:c.4376_4391del (0.20)	Duodenum	APC:c.4376_4391del (0.18)
		ERBB2:c.929C>T (0.17)		ERBB2:c.929C>T (0.19)
4	Duodenum	APC:c.4611_4612del (0.16)	Duodenum	APC:c.4611_4612del (0.37)
		APC:c.4501del (0.30)		APC:c.4501del (0.38)
5	Duodenum		Duodenum	MSH2c.52_53delinsAA (0.47)
6	Duodenum		Duodenum	
7	Descending part of duodenum	BMPR1A:c.1532G>A (0.23)	Descending part of duodenum	KRAS:c.38G>A (0.24)
8	Ampulla of Vater		Ascending colon	APC:c.4348C>T (0.22)
				APC:c.1600A>T (0.24)
9	Horizontal part of duodenum		Coecum	APC:c.4348C>T (0.26)
				APC:c.2377C>T (0.23)
10	Descending part of duodenum	APC:c.4348C>T (0.30)	Sigmoid	APC:c.835-8A>G (0.59)
		APC:c.3925_3926del (0.25)		
11	Stomach	APC:c.2612del (0.32)	Ascending colon	APC:c.2612del (0.23)
				APC:c.4348C>T (0.27)
12	Descending part of duodenum	APC:c.4391_4394del (0.61)	Descending part of duodenum	APC:c.4391_4394del (0.49)
13	Ampulla of Vater	APC:c.847C>T (0.33)	Duodenum	APC:c.847C>T (0.15)
14	Descending part of duodenum	APC:c.4393_4394del (0.25)	Ascending colon	APC:c.4393_4394del (0.39)
				APC:c.2626C>T (0.20)
				APC:c.1690C>T (0.10)

Supplemental table 1 (part 2) - Variants detected using targeted NGS in adenomas. Continued.

ID	Location (T3)	Variants (VAF)	Location (T4)	Variants (VAF)
1	Ampulla of Vater	APC:c.4510_4513dup (0.19)	Horizontal part of duodenum	APC:c.4510_4513dup (0.41)
		KRAS:c.38G>A (0.21)		
2	Stomach antrum		Stomach	BMPR1A:c.676-1G>A (0.09)
3	Horizontal part of duodenum	APC:c.4376_4391del (0.18)		
		ERBB2:c.929C>T (0.24)		
4	Duodenum proximal	APC:c.4611_4612del (0.15)	Descending part of duodenum	APC:c.4611_4612del (0.10)
		APC:c.4501del (0.15)		APC:c.4501del (0.12)
		POLD1:c.1028G>C (0.18)		POLD1:c.1028G>C (0.10)
5	Coecum	APC:c.3929_3932del (0.56)	Descending colon	APC:c.646-1G>C (0.50)
				MSH2:c.53G>A (0.34)
6	Sigmoid	APC:c. 3030del (0.68)	Sigmoid	APC:c. 2991T>A (0.68)
7	Descending part of duodenum	TP53:c.524G>A (0.20)		
		BMPR1A:c.217A>T (0.23)		
8	Sigmoid	APC:c.3956del (0.60)	Rectum	APC:c.4485del (0.30)
				APC:c.3193C>T (0.30)
9	Ascending colon	APC:c.4391_4394del (0.23)	Ascending colon	APC:c.4348C>T (0.21)
		APC:c.2805C>A (0.25)		APC:c.2299C>T (0.19)
		APC:c.220G>A (0.23)		
10	Ascending colon	APC:c.4031C>A (0.24)	Sigmoid	APC:c.835-8A>G (0.18)
		POLE:c.3934G>T (0.25)		APC:c.2626C>T (0.25)
				APC:c.3916G>T (0.24)
				SMAD4:c.1082G>A (0.14)
				SMAD4:c.938C>T (0.15)
11	Transverse colon	APC:c.2612del (0.15)	Descending colon	APC:c.2612del (0.35)
		APC:c.1660C>T (0.12)		APC:c.4330C>T (0.26)
12	Descending part of duodenum	APC:c.4391_4394del (0.72)	Coecum	APC:c.4391_4394del (0.37)
		ERBB3:c.2783A>G (0.34)		APC:c.2090del (0.21)
				NRAS:c.182A>G (0.25)
13	Coecum	APC:c.847C>T (0.16)	Ascending colon	APC:c.847C>T (0.18)
		APC:c.1626+1G>T (0.16)		APC:c.4508C>A (0.17)
		APC:c.4756A>T (0.11)		
14	Transverse colon	APC:c.4393_4394del (0.21)	Transverse colon	
		APC:c.1867del (0.11)		

Supplemental table 1 (part 2) - Variants detected using targeted NGS in adenomas. Continued.

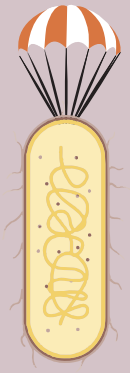
ID	Location (T5)	Variants (VAF)	Location (T6)	Variants (VAF)
1	Rectum	APC:c.2701C>T (0.34)		
		APC:c.4348C>T (0.43)		
2	Rectum	APC:c.690del (0.25)		
		APC:c.1495C>T (0.28)		
3				
4	Rectosigmoid	APC:c.847C>T (0.31)	Sigmoid	APC:c.2496del (0.40)
				APC:c.4099C>T (0.42)
5				
6				
7				
8				
9				
10	Transverse colon	APC:c.2149dup (0.19)	Transverse colon	APC:c.835-8A>G (0.23)
		APC:c.1688del (0.28)		APC:c.933+1G>T (0.10)
11				
12	Descending colon	APC:c.4391_4394del (0.45)	Sigmoid	APC:c.4391_4394del (0.49)
		APC:c.1690C>T (0.28)		APC:c.2805C>A (0.36)
		TP53:c.672+1G>C (0.12)		KRAS:c.35G>A (0.19)
13	Ascending colon	APC:c.847C>T (0.21)	Ascending colon	APC:c.847C>T (0.23)
		APC:c.4391_4394del (0.35)		APC:c.4725del (0.22)
14	Transverse colon	APC:c.4393_4394del (0.42)	Ascending colon	APC:c.4393_4394del (0.33)
		APC:c.2626C>T (0.21)		APC:c.1690C>T (0.21)

Supplemental table 1 (part 2) - Variants detected using targeted NGS in adenomas. Continued.

ID	Location (T7)	Variants (VAF)	Location (T8)	Variants (VAF)
1				
2				
3				
4				
5				
6				
7				
8				
9				
10	Sigmoid	APC:c.835-8A>G (0.37) APC:c.646C>T (0.21)	Sigmoid	APC:c.835-8A>G (0.50) APC:c.423-6A>G (0.17) CTNNB1:c.1099G>C (0.14)
11				
12				
13	Ascending colon	APC:c.847C>T (0.27) APC:c.4348C>T (0.28)	Sigmoid	APC:c.847C>T (0.22)
14	Transverse colon	APC:c.4393_4394del (0.21) APC:c.1297C>T (0.22)	Descending colon	APC:c.4393_4394del (0.35) APC:c.1902T>G (0.22) APC:c.2343del (0.20) APC:c.1213C>T (0.13) APC:c.2626C>T (0.13)

Supplemental table 1 (part 2) - Variants detected using targeted NGS in adenomas. Continued.

ID	Location (T9)	Variants (VAF)	Location (T10)	Variants (VAF)
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13	Sigmoid	APC:c.847C>T (0.25)	Rectum	APC:c.847C>T (0.59)
14				



Part III
Colibactin

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

Gastroenterology, 2022

Diantha Terlouw, Manon Suerink, Arnoud Boot, Tom van Wezel, Maartje Nielsen, Hans Morreau

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 al1 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 non-mosaic 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 1A), where the underlined adenine is substituted by a guanine. Transformed in a T>N orientation (Figure 1B), the context perfectly fulfills the mutational signature caused by *pks+ E. coli* with the hexanucleotide AAAATT as predominant sequencing context (Figure 1C). 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]

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 expla

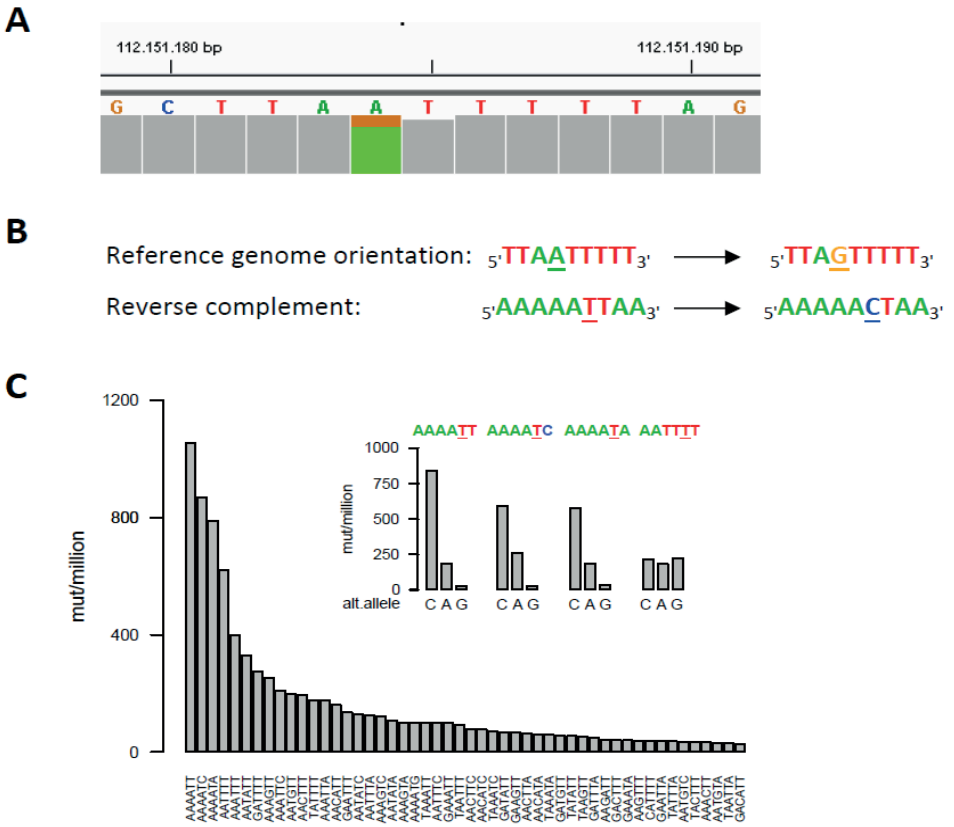


Figure 1 – Comparison of the sequence context of the c.835-8A>G variant with that of the mutational signature associated with pks+ E.coli. (A) Visualisation of the variant in Integrative Genomics Viewer. (B) The T>N oriented sequence context of the c.835-8A>G variant. (C). Top 50 hexanucleotides mostly affected by the pks+ E.coli mutational signature normalized to the frequency of the hexanucleotide in the human genome, based on data from the preprint of 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.

nation 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.

References

1. Jansen AML, Crobach S, Geurts-Giele WRR, et al. Distinct Patterns of Somatic Mosaicism in the APC Gene in Neoplasms From Patients With Unexplained Adenomatous Polyposis. Gastroenterology 2017;152:546-+.
2. Pleguezuelos-Manzano C, Puschhof J, Huber AR, et al. Mutational signature in colorectal cancer caused by genotoxic pks(+)E. coli. Nature 2020;580:269-+.
3. Boot A, Ng AWT, Chong FT, et al. Characterization of colibactin-associated mutational signature in an Asian oral squamous cell carcinoma and in other mucosal tumor types. Genome Res 2020;30:803-813.
4. Fostira F, Thodi G, Sandaltzopoulos R, et al. Mutational spectrum of APC and genotype-phenotype correlations in Greek FAP patients. BMC Cancer 2010;10:389.
5. Jarry J, Brunet JS, Laframboise R, et al. A survey of APC mutations in Quebec. Fam Cancer 2011;10:659-65.
6. Yaeger R, Chatila WK, Lipsyc MD, et al. Clinical Sequencing Defines the Genomic Landscape of Metastatic Colorectal Cancer. Cancer Cell 2018;33:125-136 e3.
7. Baker KT, Salk JJ, Brentnall TA, et al. Precancer in ulcerative colitis: the role of the field effect and its clinical implications. Carcinogenesis 2018;39:11-20.
8. Lee-Six H, Olafsson S, Ellis P, et al. The landscape of somatic mutation in normal colorectal epithelial cells. Nature 2019;574:532-537.
9. van Eijk R, Stevens L, Morreau H, et al. Assessment of a fully automated high-throughput DNA extraction method from formalin-fixed, paraffin-embedded tissue for KRAS, and BRAF somatic mutation analysis. Exp Mol Pathol 2013;94:121-5.

Supplemental Material

Enrichment of colibactin-associated mutational signatures in unexplained colorectal polyposis patients

Manuscript submitted at BMC Cancer

Diantha Terlouw, Arnoud Boot, Quinten R Ducarmon, Sam Nooij, Manon Suerink, Monique E van Leerdam, Demi van Egmond, Carli M Tops, Romy D Zwittink, Dina Ruano, Alexandra M J Langers, Maartje Nielsen*, Tom van Wezel*, Hans Morreau*

Abstract

Background Colibactin, a genotoxin produced by polyketide synthase harboring (*pks*$^+$) bacteria, induces double-strand breaks and chromosome aberrations. Consequently, enrichment of *pks*$^+$ *Escherichia coli* in colorectal cancer and polyposis suggests a possible carcinogenic effect in the large intestine. Additionally, specific colibactin-associated mutational signatures; SBS88 and ID18 in the Catalogue of Somatic Mutations in Cancer database, are detected in colorectal carcinomas. Previous research showed that a recurrent *APC* splice variant perfectly fits SBS88.

Methods In this study, we explore the presence of colibactin-associated signatures and fecal *pks* in an unexplained polyposis cohort. Somatic targeted Next-Generation Sequencing (NGS) was performed for 379 patients. Additionally, for a subset of 29 patients, metagenomics was performed on feces and mutational signature analyses using Whole-Genome Sequencing (WGS) on Formalin-Fixed Paraffin Embedded (FFPE) colorectal tissue blocks.

Results NGS showed somatic *APC* variants fitting SBS88 or ID18 in at least one colorectal adenoma or carcinoma in 29% of patients. Fecal metagenomic analyses revealed enriched presence of *pks* genes in patients with somatic variants fitting colibactin-associated signatures compared to patients without variants fitting colibactin-associated signatures. Also, mutational signature analyses showed enrichment of SBS88 and ID18 in patients with variants fitting these signatures in NGS compared to patients without.

Conclusions These findings further support colibactins ability to mutagenize colorectal mucosa and contribute to the development of colorectal adenomas and carcinomas explaining a relevant part of patients with unexplained polyposis.

Background

An enrichment of polyketide synthase (*pks*) encoding *Escherichia coli* in patients with colorectal cancer[1, 2] and polyposis[3] has implied a potential carcinogenic effect in the large intestine. These *E. coli* bacteria harbor the *pks* gene island which encodes the necessary equipment to produce the genotoxin colibactin.[4] Colibactin induces double-strand breaks and chromosome aberrations leading to a specific mutational signature that has been observed in colorectal adenocarcinomas and oral squamous cell carcinomas.[5, 6] This mutational signature is characterized by T>N mutations with an adenine 3 base pairs to the 5' side and single thymine deletions located in T homopolymers with 2 to 4 adenines to the 5' side depending on the length of the T homopolymer. These signatures are documented in the Catalogue of

Somatic Mutations in Cancer (COSMIC) database as single base substitution signature SBS88 and indel signature ID18.

E. coli is not the only bacterium able to harbor the *pks* gene island. Other bacteria mostly belonging to the Enterobacteriaceae family, such as *Klebsiella pneumoniae*, *Enterobacter aerogenes* and *Citrobacter koseri*, have also been shown to harbor *pks*.[7] Moreover, *pks* harboring bacteria are found in other organisms like bacteria in the honey bee gut or a marine sponge.[8]

We previously showed that a common *APC* splice variant c.835-8A>G and several other pathogenic *APC* variants perfectly fit the colibactin-associated mutational signatures.[9] This finding furthermore implies a possible association between colibactin and the development of colorectal neoplasms. Since a large proportion of our unexplained polyposis patient cohort showed a colibactin-associated *APC* variant in multiple adenomas, further research into the presence and impact of colibactin and its mutational signature was warranted. Therefore, for a subset of polyposis patients, metagenomics was performed on feces and Whole Genome Sequencing (WGS) with subsequent mutational signature analyses was conducted on Formalin Fixed Paraffin Embedded (FFPE) colorectal tissue blocks. Results were compared between those with and without colibactin-associated variants.

Material and methods

APC mosaicism testing

In total, 379 patients with multiple colorectal adenomas or carcinomas were tested for *APC* mosaicism. In short, DNA was isolated from Formalin Fixed Paraffin Embedded (FFPE) tissue blocks of on average 4 colorectal adenomas or carcinomas using the automated Tissue Preparation System (Siemens). Ampliseq Next Generation Sequencing (NGS) libraries (ThermoFisher Scientific) of a custom-made panel containing 20 colorectal cancer and polyposis associated genes were prepared according to manufacturer's instructions. Sequencing was performed in an Ion GeneStudio S5 Series sequencer (ThermoFisher Scientific). The raw, unaligned sequencing reads were mapped against human reference genome (hg19) using TMAP software and Torrent Variant Caller was used for variant calling. The detected variants were categorized by pathogenicity and were, when needed, visualized using Integrative Genomic Viewer[10] or interpreted using the Alamut Visual software (Sophia Genetics).

APC variants and colibactin signature

To determine whether the *APC* variants fit into the mutational signatures SBS88 and ID18, all

detected T>N and delT *APC* variants and their sequencing context were visualized using IGV. As previously described[5,6], T>N variants with the following sequencing context were labelled as fitting SBS88: 5' A-(N)-(T/A)-**T**-(T/A/G) 3'. DelT variants were labelled as fitting ID18 whenever a 2 to 4 adenine homopolymer was flanking the 5' side of a thymine homopolymer with a total of 5-6 base pairs. As illustrated in figure 1, in 269 patients no somatic variant fitting SBS88 or ID18 was found and therefore served as the control group.

Case and control group selection

A random selection of twenty-nine patients were included for fecal metagenomics and/or Whole-Genome Sequencing, as depicted in figure 1. Twenty of these patients have adenomas or carcinomas with an *APC* variant suiting SBS88 or ID18 and nine control patients do not have such a colibactin-associated *APC* variant. The patient characteristics are summarized in table 1 and somatic *APC* variants per lesion in supplemental table 1. Furthermore, the sequencing context of the *APC* variants are included in supplemental table 2.

Fecal metagenomics

Feces samples of 25 out of 29 patients were collected for deep fecal shotgun metagenomic sequencing (figure 1 and table 1). Four patients could not be included since they did not respond (N=3) or passed away (N=1). Fecal metagenomic sequencing was performed as previously described.[11] In short, stool samples were stored at -80°C, DNA was extracted and libraries were prepared according to manufacturer's protocol.

Sequencing was performed on the Novaseq6000 platform (Illumina, San Diego, CA, USA). Raw metagenomic sequences were processed, analyzed and compared to the *pks* gene island partly comparable to the method description by Nooij et al.[12] Reads mapping to the human genome (GRCh38) were removed using bowtie2 (version 2.4.2[13]) and SAMtools (version 1.11[14]) and filtered reads were quality-trimmed using fastp (version 0.20.1[15]). The pre-processing workflow is available at (https://git.lumc.nl/snooij/metagenomics-preprocessing). The quality-trimmed reads were screened for the presence of the *pks* island by mapping to the colibactin gene cluster (accession ID AM229678) using BWA-MEM (version 0.7.17[16]). Mapped reads were deduplicated using Picard MarkDuplicates (version 2.23.3[17]) to remove technical artifacts and improve quantification. The *pks* screening workflow is available at (https://git.lumc.nl/snooij/screen_pks_in_polyposis_fecal_metagenomes). As previously outlined, fecal samples positive for at least one *pks* gene were considered *pks*-positive.[12] To quantify reads per kilobase per million (RPKM) of the individual genes in the *pks* island present in the stool samples, RPKM values were calculated using the following formula: (N of mapped reads/N of base pairs of the coding sequence of the respective gene)*1,000 divided

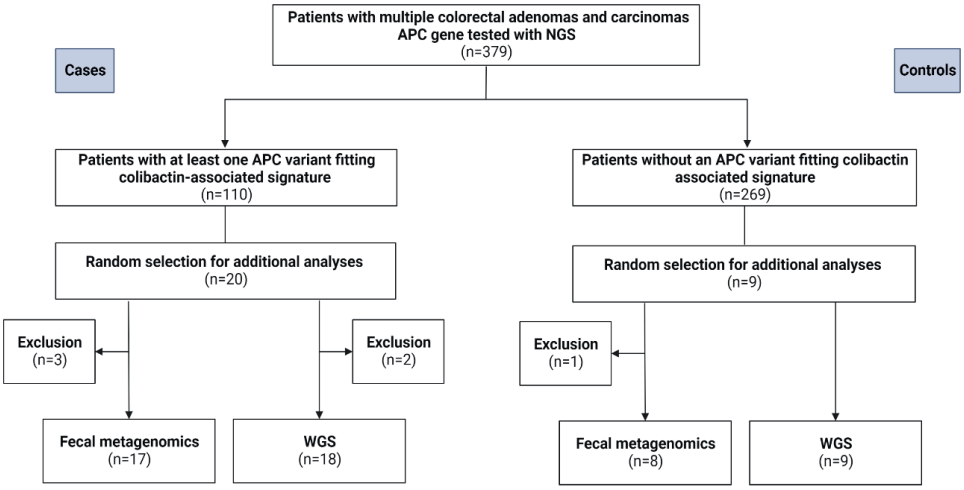


Figure 1 – **Study design and patient selection.** In total, 379 patients were tested using targeted NGS. The case group are patients with at least one *APC* variant fitting colibactin-associated mutational signature. Twenty cases are selected for additional fecal metagenomics and WGS. Patients without APC variant fitting colibactin-associated signatures serve as controls. Nine controls were selected for fecal metagenomics and WGS. Four patients could not be included for fecal metagenomics since they did not respond to sample request (N=3) or passed away (N=1). Two cases were excluded for WGS due to insufficient amount of DNA.

by N of trimmed and filtered reads*1,000,000. For mean RPKM of the entire *pks* island, RPKMs of all individual genes were summed and divided by the total number of 19 *clb* genes.

Whole-Genome Sequencing (WGS)

DNA from adenomas and carcinomas of 27 out of 29 patients was included for WGS (figure 1 and table 1). Two patients were excluded due to an insufficient amount of DNA extracted.

DNA was isolated from FFPE tissue blocks using the NucleoSpin DNA FFPE XS kit (BIOKE, Leiden, the Netherlands) according to manufacturer's instructions. WGS was performed on the BGIseq500 platform (BGI, Hong Kong, China) for 4 out of 27 patients (ID 8, 10, 12 and 13). Sequencing for the remaining 23 patients was performed on the NovaSeq6000 platform (Illumina, San Diego, USA). The raw sequencing reads were aligned to a reference genome (GRCh38). The alignment, variant calling and filtering were performed as described before.[6, 18] The mutational signature assignment using reference mutational signatures was performed using mSigAct::sparseAssignSignatures followed by mSigAct signature presence test, which provides a p-value for the null-hypothesis that a signature is not needed to explain an observed somatic mutation profile compared with the alternative hypothesis that the signature is needed, as previously described.[6]

BMI and lifestyle data

Body Mass Index (BMI) and information about lifestyle was collected using patient medical records and for some patients using a questionnaire (n=65). BMI was categorized in 4 groups: ≤18.5 'underweight', 18.5-24.9 'healthy weight', 25.0-29.9 'overweight' and ≥30.0 'obese'. Both tobacco and alcohol consumption were categorized as 'never', 'former' and 'current'. Packyears (PY) was determined as the number of packs of cigarettes smoked per day multiplied by the number of years the patient has smoked.

Statistical analysis

Statistical analyses were performed using IBM SPSS statistics 25 (Armonk, NY, USA) and a p-value of <0.05 was considered statistically significant. Independent T tests, Chi-square tests and Fisher's exact tests were used to assess the differences between the patients with and without colibactin variants based on the targeted NGS and patients with *pks* in feces with and without contribution of SBS88 and/or ID18 in the WGS data.

Results

In total, 379 unexplained polyposis patients were tested for somatic *APC* mosaicism using targeted NGS. In 110 patients, at least one colorectal adenoma or carcinoma harbored an *APC* variant that fits with one of the colibactin-associated mutational signatures. Phenotypic characteristics, like adenoma count, age at first adenoma and personal history of colorectal carcinoma did not significantly differ between the patients with (cases) and without (controls) *APC* variants fitting colibactin mutational signatures. Similarly, lifestyle factors like BMI and smoking status were not significantly different between cases and controls (supplemental table 3). The control group consisted of significantly more former alcohol consumers compared to the cases.

Fecal metagenomics

Fecal samples from seventeen patients with *APC* variants fitting SBS88 or ID18 (cases) and eight patients without *APC* variants fitting SBS88 or ID18 (controls) were used for metagenomic analysis to detect *pks* genes. As shown in table 2, 59% (10 out of 17) of the cases were *pks* positive compared to 25% (2 out of 8) of controls (p-value=0.124).

In addition, fecal metagenomics was used to quantify *pks* using RPKM values. However, no significant correlation between number of adenomas/carcinomas with *APC* variants fitting SBS88 or ID18 and the *pks* RPKM values was observed (Pearson: R=0.16, p-value=0.45).

Table 1 – Phenotypic characteristics and NGS results of patients included for fecal metagenomics and WGS.

ID	#Ad	Ad age range first	CRC age range first	#SBS88/ ID18	#tested	%	Feces	WGS
1	22	66-70	41-45	2	7	28.6	Y	Y
2	>9	66-70	66-70	3	3	100.0	Y	Y
3	15	46-50	46-50	6	7	85.7	Y	N
4	13	66-70	66-70	3	9	33.3	Y	Y
5	70	66-70	66-70	3	10	30.0	N	Y
6	4	61-65	61-65	3	6	50.0	Y	Y
7	18	56-60	-	0	4	0.0	Y	Y
8	10	51-55	51-55	3	4	75.0	Y	Y
9	2	51-55	51-55	2	4	50.0	Y	Y
10	10	51-55	-	8	10	80.0	Y	Y
11	28	81-85	-	2	6	33.3	Y	N
12	36	71-75	-	2	4	50.0	N	Y
13	27	61-65	61-65	2	3	66.7	Y	Y
14	3	21-25	-	3	3	100.0	Y	Y
15	10	51-55	-	2	6	33.3	Y	Y
16	11	51-55	-	2	3	66.7	Y	Y
17	22	66-70	-	2	4	50.0	Y	Y
18	14	61-65	-	2	6	33.3	Y	Y
19	24	41-45	-	2	4	50.0	N	Y
20	10	56-60	-	2	4	50.0	Y	Y
21	18	46-50	-	0	4	0.0	Y	Y
22	18	61-65	-	0	4	0.0	Y	Y
23	20	61-65	-	0	3	0.0	Y	Y
24	14	46-50	-	0	4	0.0	Y	Y
25	10	46-50	46-50	4	4	100.0	Y	Y
26	8	46-50	46-50	0	4	0.0	Y	Y
27	4	21-25	21-25	0	4	0.0	N	Y
28	2	61-65	61-65	0	3	0.0	Y	Y
29	9	61-65	61-65	0	5	0.0	Y	Y

Ad – numbers of colorectal adenomas developed, Ad age first – age first colorectal adenoma diagnosis, CRC age first – age of colorectal carcinoma diagnosis, #SBS88/ID18 – number of adenomas or carcinomas with a variant fitting SBS88 or ID18 based on NGS, Tested – Total number of adenomas or carcinomas tested using NGS, % - percentage of adenomas or carcinomas with a colibactin-associated variant, Feces – fecal metagenomics performed yes or no, WGS – WGS performed yes or no.

Also, no significant difference in phenotype was observed between the 10 cases with *pks* genes in feces and 7 cases without. When comparing lifestyle factors, a trend was observed towards a higher BMI in the group with *pks* in their feces (supplemental table 4).

None of the bacteria previously associated with colorectal cancer, like *Fusobacterium nucleatum*, *Bacteroides fragilis*, *Campylobacter jejuni* and *Clostridioides difficile*, or capable of producing colibactin, like *K. pneumoniae*, *E. aerogenes* and *C. koseri*, were detected in any of the stool samples (data not shown).

Mutational signature analysis

For WGS, fifty-seven colorectal adenomas or carcinomas and six normal colon mucosa samples were analyzed from eighteen patients with *APC* variants fitting SBS88 or SBS18 (cases) and nine patients without these variants (controls).

As summarized in table 2, mutational signature analysis identified SBS88 in 8 adenomas or carcinomas derived from 6 cases and ID18 in 2 lesions of 2 cases. Overall, colibactin-associated mutagenesis was detected in 38.9% (7 out of 18) cases. One adenoma of nine controls (11.1%) also showed colibactin associated mutagenesis (SBS88).

Combining fecal metagenomics and mutational signature analyses

Fifteen cases and eight controls were analyzed both using fecal metagenomics and WGS to compute mutational signature analyses. In 10 cases *pks* was found in their feces samples of which 5 patients also showed a contribution of SBS88 or ID18. Of the 5 cases without *pks* in their feces, three showed SBS88 or ID18 contribution. In 2 controls, *pks* genes were detected in feces and in one of them SBS88 was determined in colorectal lesions. Therefore, 86.7% (13 out of 15) of cases and 25% (2 out of 8) of controls showed hints of *pks* or its carcinogenic effects (p-value=0.006).

No significant differences were detected in lifestyle factors between fecal *pks+* and SBS88/ID18+ cases and fecal *pks+* and SBS88/ID18- cases (supplemental table 5).

Discussion

Using targeted NGS, 379 patients with unexplained colorectal polyposis were tested for *APC*

mosaicism. At least one somatic *APC* variant fitting one of the colibactin associated muta tional signatures (SBS88 or ID18) was found in 29% (n=110) patients. Except for the distribution of former alcohol consumption, no significant differences were observed in phenotypic characteristics or lifestyle factors between patients with and without these *APC* variants.

Table 2 – Results of fecal pks using metagenomics and mutational signatures SBS88 and ID18 using WGS.

ID	%	WGS SBS88	WGS ID18	Fecal pks
7	0	0/2	0/2	No
21	0	0/1	0/1	No
22	0	0/2	0/2	No
23	0	1/2	0/2	Yes
24	0	0/2	0/2	No
26	0	0/2	0/2	No
27	0	0/1	0/1	
28	0	0/2	0/2	Yes
29	0	0/2	0/2	No
1	28.6	1/2	0/2	Yes
5	30.0	0/2	0/2	
4	33.3	0/2	0/2	Yes
11	33.3			No
15	33.3	0/2	0/2	Yes
18	33.3	1/3	1/3	Yes
6	50.0	2/3	0/3	No
9	50.0	1/2	0/2	Yes
12	50.0	0/1	0/1	
17	50.0	0/2	0/2	Yes
19	50.0	0/3	0/3	
20	50.0	0/4	0/4	Yes
13	66.7	0/1	0/1	No
16	66.7	0/2	0/2	No
8	75.0	1/1	0/1	Yes
10	80.0	2/2	0/2	No
3	85.7			No
2	100.0	0/3	0/3	Yes
14	100.0	0/3	1/3	No
25	100.0	0/3	0/3	Yes

% - percentage of adenomas or carcinomas tested with NGS with a colibactin variant, WGS SBS88 – number of samples with SBS88 / number of samples tested, WGS ID18 – number of samples with ID18 / number of samples tested, Fecal pks – Yes for patients with and no for patients without pks in their feces sample.

Although further research is warranted, the significant difference in former alcohol consumption observed between the groups is likely attributable to the small number of patients with a former alcohol consumption status.

Fecal metagenomics revealed 59% (10 out of 17) of cases with one or more *pks* genes. This proportion is comparable to *pks*$^{+}$ *E. coli* bacteria found in colon mucosa of individuals with Familial Adenomatous Polyposis (68%) and sporadic CRC patients (55%).[1, 3] In contrast, only 25% (2 out of 8) of controls showed *pks* genes. Although numbers are small, this is comparable to the previously reported incidence of healthy individuals with *pks* genes in feces (27-29%)[12, 19, 20] and with *pks*$^{+}$ *E. coli* bacteria in colon mucosa (19-22%).[1, 3]

The current study found no significant differences in phenotypic characteristics and tobacco and alcohol consumption between patients with and without *pks* in feces. Further research is required to draw a conclusion about the correlation between BMI and *pks*$^{+}$ *E. coli*. Although not directly linked to BMI, Arima et al.[21] found that the association between the Western diet and colorectal cancer patients was only significant in patients with *pks*$^{+}$ *E. coli* in their tumor, suggesting a potential interactive carcinogenic effect between diet and *pks*$^{+}$ *E. coli*.

WGS with subsequent mutational signature analysis showed a contribution of SBS88 or ID18 in 39% (7/18) of cases, compared to 11.1% (1/9) of controls. In only one case all analyzed samples showed a contribution of SBS88. This might be explained by the variable distribution of colonic crypts with the signature within one patient.[22] Moreover, as summarized in supplemental table 1, the majority of adenomas and carcinomas (n=25) selected for WGS from cases did not harbor *APC* variants fitting SBS88 or ID18. Eighteen of these adenomas and carcinomas were located in the right colon and right sided carcinomas were less likely to have colibactin-associated signatures.[23]

Combining both fecal metagenomics and mutational signature analyses, 86.7% (13/15) of cases showed a significant enrichment towards colibactin influence compared to 25% (2/8) of controls in which both analyses were performed.

This significant enrichment of fecal *pks* and colibactin-associated mutational signatures in cases compared to controls, supports the proposition of a recent preprint that the *APC* splice variant c.835-8A>G might be used as a biomarker for *pks*$^{+}$ *E. coli* influence in the development of the adenoma or carcinoma.[23]

Despite the enrichment, no clear correlation between *pks* in feces and colibactin-associated mutational signatures in colorectal lesions was observed in individual cases. Multiple hypotheses might explain (part of) this finding, comprising both biological and technical

issues:

It was previously described that colibactin has a short-term effect, affecting the colon early in life.[22, 24, 25] Colonic mucosa of patients with a contribution of SBS88 and ID18 might therefore be affected by colibactin, but the *pks*-encoding bacteria may have been eradicated from the intestinal tract at time of feces sampling.

The other way around, in patients with *pks* detected in feces but no SBS88 or ID18 in WGS, enrichment of *pks*$^+$ bacteria after the development of adenomas but before feces sampling seems unlikely as *pks*$^+$ *E. coli* is detected in feces of newborns and therefore proposed to be transmitted during birth.[25, 26] These patients might, however, have some kind of mechanism inhibiting colibactin from entering the host cell or whenever inside the cell protects against the specific DNA damage. The protein ATG16L1 for example is described to be associated with preventing colorectal tumorigenesis in presence of *pks*$^+$ *E. coli* in cell lines and mouse models.[27] Also, colibactin production is in a recent preprint suggested to be inhibited by oxygen.[28] On the other hand, inflammation seems to promote the expansion of the colibactin-encoding *E. coli* and creates an opportunity to adhere to colon mucosa.[2] Moreover, co-localization with *B. fragilis* seems to increase DNA damage with faster tumor onset in mice.[3] These hypotheses might also play a role in whether presence of *pks*$^+$ *E. coli* in the intestinal tract actually leads to DNA damage.

Technically, the small sample set and use of shotgun metagenomics and FFPE tissue blocks are limitations of this study. Especially WGS performed on FFPE samples affects the variant and signature calling and interpretation due to fragmentation and deamination artefacts.[29-31] Moreover, shotgun fecal metagenomics is a broad analyses but a more sensitive qPCR approach performed at multiple timepoints and at time of adenoma diagnosis could give more insight into the association with adenoma development.

To conclude, in 29% of our cohort with unexplained polyposis patients a colibactin influence was suggested based on targeted NGS data. A subset of cases was included for additional analyses and showed further evidence of colibactin in fecal metagenomics and mutational signature analyses compared to controls. Further research, circumventing the complications of WGS on FFPE tissue and validating the feces analyses, should be performed to draw conclusions for individual cases. Still, these findings provide evidence that colibactin affects the colonic mucosa and plays a pivotal role in unexplained polyposis patients.

References

1. Buc E, Dubois D, Sauvanet P, et al. High prevalence of mucosa-associated E. coli producing cyclomodulin and genotoxin in colon cancer. PLoS One 2013;8:e56964.
2. Arthur JC, Perez-Chanona E, Muhlbauer M, et al. Intestinal inflammation targets cancer-inducing activity of the microbiota. Science 2012;338:120-3.
3. Dejea CM, Fathi P, Craig JM, et al. Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. Science 2018;359:592-597.
4. Nougayrede JP, Homburg S, Taieb F, et al. Escherichia coli induces DNA double-strand breaks in eukaryotic cells. Science 2006;313:848-51.
5. Pleguezuelos-Manzano C, Puschhof J, Huber AR, et al. Mutational signature in colorectal cancer caused by genotoxic pks(+)E. coli. Nature 2020;580:269-+.
6. Boot A, Ng AWT, Chong FT, et al. Characterization of colibactin-associated mutational signature in an Asian oral squamous cell carcinoma and in other mucosal tumor types. Genome Res 2020;30:803-813.
7. Wami H, Wallenstein A, Sauer D, et al. Insights into evolution and coexistence of the colibactin- and yersiniabactin secondary metabolite determinants in enterobacterial populations. Microb Genom 2021;7.
8. Silpe JE, Wong JWH, Owen SV, et al. The bacterial toxin colibactin triggers prophage induction. Nature 2022;603:315-320.
9. Terlouw D, Suerink M, Boot A, et al. Recurrent APC Splice Variant c.835-8A>G in Patients With Unexplained Colorectal Polyposis Fulfilling the Colibactin Mutational Signature. Gastroenterology 2020;159:1612-1614.e5.
10. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 2013;14:178-92.
11. Ducarmon QR, Terveer EM, Nooij S, et al. Microbiota-associated risk factors for asymptomatic gut colonisation with multi-drug-resistant organisms in a Dutch nursing home. Genome Med 2021;13:54.
12. Nooij S, Ducarmon QR, Laros JFJ, et al. Faecal microbiota transplantation influences procarcinogenic Escherichia coli in recipient recurrent Clostridioides difficile patients. Gastroenterology 2021.
13. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature Methods 2012;9:357-U54.
14. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078-2079.
15. Chen SF, Zhou YQ, Chen YR, et al. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018;34:884-890.
16. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv: Genomics. 2013.
17. https://broadinstitute.github.io/picard/index.html.
18. Boot A, Huang MN, Ng AWT, et al. In-depth characterization of the cisplatin mutational signature in human cell lines and in esophageal and liver tumors. Genome Res 2018;28:654-665.
19. Dubinsky V, Dotan I, Gophna U. Carriage of Colibactin-producing Bacteria and Colorectal Cancer Risk. Trends Microbiol 2020;28:874-876.
20. Watanabe D, Murakami H, Ohno H, et al. Stool pattern is associated with not only the prevalence of tumorigenic bacteria isolated from fecal matter but also plasma and fecal fatty acids

in healthy Japanese adults. BMC Microbiol 2021;21:196.

21. Arima K, Zhong R, Ugai T, et al. Western-Style Diet, pks Island-Carrying Escherichia coli, and Colorectal Cancer: Analyses From Two Large Prospective Cohort Studies. Gastroenterology 2022;163:862-874.
22. Lee-Six H, Olafsson S, Ellis P, et al. The landscape of somatic mutation in normal colorectal epithelial cells. Nature 2019;574:532-537.
23. Georgeson P, Steinfelder RS, Harrison TA, et al. Genotoxic colibactin mutational signature in colorectal cancer is associated with clinicopathological features, specific genomic alterations and better survival. This article is a preprint and has not been certified by peer review. MedRxiv. DOI: https://doi.org/10.1101/2023.03.10.23287127.
24. Iftekhar A, Berger H, Bouznad N, et al. Genomic aberrations after short-term exposure to colibactin-producing E. coli transform primary colon epithelial cells. Nat Commun 2021;12:1003.
25. Payros D, Secher T, Boury M, et al. Maternally acquired genotoxic Escherichia coli alters offspring's intestinal homeostasis. Gut Microbes 2014;5:313-25.
26. Tsunematsu Y, Hosomi K, Kunisawa J, et al. Mother-to-infant transmission of the carcinogenic colibactin-producing bacteria. BMC Microbiol 2021;21:235.
27. Lucas C, Salesse L, Hoang MHT, et al. Autophagy of Intestinal Epithelial Cells Inhibits Colorectal Carcinogenesis Induced by Colibactin-Producing Escherichia coli in Apc(Min/+) Mice. Gastroenterology 2020;158:1373-1388.
28. Nadège Bossuet-Greif N, Guyonnet C, Chagneau CV, et al. Oxygen concentration modulates colibactin production. This article is a preprint and has not been certified by peer review. BioRxiv. DOI: https://doi.org/10.1101/2022.06.20.496773.
29. de Schaetzen van Brienen L, Larmuseau M, Van der Eecken K, et al. Comparative analysis of somatic variant calling on matched FF and FFPE WGS samples. BMC Med Genomics 2020;13:94.
30. Robbe P, Popitsch N, Knight SJL, et al. Clinical whole-genome sequencing from routine formalin-fixed, paraffin-embedded specimens: pilot study for the 100,000 Genomes Project. Genet Med 2018;20:1196-1205.
31. Do H, Dobrovic A. Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization. Clin Chem 2015;61:64-71.

Supplemental Material

Colibactin mutational signatures in *NTHL1* Tumor Syndrome and *MUTYH* Associated Polyposis patients

Manuscript accepted at Genes, Chromosomes and Cancer

Diantha Terlouw, Arnoud Boot, Quinten R Ducarmon, Sam Nooij, Max A Jessurun, Monique E van Leerdam, Carli M Tops, Alexandra M J Langers, Hans Morreau*, Tom van Wezel*, Maartje Nielsen*

*These authors contributed equally to this work

Abstract

Polyketide synthase (*pks*) island harboring *Escherichia coli* are, under the right circumstances, able to produce the genotoxin colibactin. Colibactin is a risk factor for the development of colorectal cancer and associated with mutational signatures SBS88 and ID18. This study explores colibactin-associated mutational signatures in biallelic *NTHL1* and *MUTYH* patients. Targeted Next Generation Sequencing (NGS) was performed on colorectal adenomas and carcinomas of one biallelic *NTHL* and twelve biallelic *MUTYH* patients. Additional fecal metagenomics and genome sequencing followed by mutational signature analysis was conducted for the *NTHL1* patient. Targeted NGS of the *NTHL1* patient showed somatic *APC* variants fitting SBS88 which was confirmed using WGS. Furthermore, fecal metagenomics revealed *pks* genes. Also, in 1 out of 12 *MUTYH* patients a somatic variant was detected fitting SBS88. This report shows that colibactin may influence development of colorectal neoplasms in predisposed patients.

Introduction

Presence of colibactin is a risk factor for the development of colorectal cancer and adenomas.[1, 2] Colibactin is a genotoxin produced by specific bacteria harboring the polyketide synthase (*pks*) island, of which *Escherichia coli* (*pks*$^{+}$ *E. coli*) is one. Mutational signatures associated with colibactin are characterized and have been added to the COSMIC database as Single Base Substitution signature 88 (SBS88) and Insertion Deletion signature 18 (ID18).[1, 3] Interestingly, a specific splice variant in *APC*, c.835-8A>G, was previously described to fit SBS88 and is recently proposed to act as a possible biomarker for the colibactin-associated mutational signature in cancer.[2, 4] As 20% to 30% of the general population harbor *pks*$^{+}$ *E.coli*, colibactin may play a role in colorectal cancer patients with and without hereditary colorectal cancer syndromes.[5, 6]

Methods

Targeted Next Generation Sequencing

DNA was isolated from Formalin Fixed Paraffin Embedded (FFPE) tissue using the Tissue Preparation System (Siemens). Ampliseq Next Generation Sequencing (NGS) libraries (ThermoFisher Scientific) were prepared according to manufacturer's instructions. Sequencing was performed in an Ion GeneStudio S5 Series sequencer (ThermoFisher Scientific), raw reads were mapped against hg19 and variants called using Torrent Variant Caller. Three NGS panels were used: a limited polyposis panel including *APC, MUTYH, POLE* and *POLD1,* a custom-made panel containing 20 colorectal cancer and polyposis associated genes and an

Oncomine Comprehensive Assay (OCA) Plus (ThermoFisher) panel containing >500 genes. All T>N and delT variants were visualized using Integrative Genomic Viewer. T>N variants within sequencing context: 5' A-(N)-(T/A)-**T**-(T/A/G) 3' were determined to fit SBS88[1, 3]. DelT variants in a thymine homopolymer flanked by 2 to 4 adenine homopolymer at the 5' side with a total of 5-6 base pairs were determined to fit ID18.

Fecal metagenomics

DNA was extracted and libraries were prepared according to manufacturer's protocol and sequencing was performed on the Novaseq6000 platform (Illumina). The analyses was performed partly comparable to the method description by Nooij et al.[7] but with direct read mapping. In short, reads mapped to GRCh38 were removed and quality-trimmed. These reads were screened for the presence of the *pks* island by mapping to the colibactin gene cluster (accession ID AM229678) after which technical artifacts were removed. The pre-processing and *pks* screening workflow are available:

(https://git.lumc.nl/snooij/metagenomics-preprocessing)

(https://git.lumc.nl/snooij/screen_pks_in_polyposis_fecal_metagenomes)

Genome Sequencing

DNA was isolated from FFPE tissue blocks using the NucleoSpin DNA FFPE XS kit (BIOKE) according to manufacturer's instructions. Sequencing was performed on the NovaSeq6000 platform (Illumina). The raw sequencing reads were aligned to a reference genome (GRCh38), processed and mutational signature assignment was performed using mSigAct::sparseAssignSignatures followed by mSigAct signature presence test, as previously described.[3]

Results

Biallelic *NTHL1* patient

We describe the case of a 38 year old man with a biallelic pathogenic germline *NTHL1* variant (*NTHL1* tumor syndrome; NTS) diagnosed with two colorectal cancers: a cT3bN1M1 adenocarcinoma of the rectum and a pT1 adenocarcinoma in a pedunculated polyp in the sigmoid colon. Furthermore, a non-advanced tubular adenoma in the ascending colon was removed by snare polypectomy. The patient had a maternal aunt with breast cancer at the age of 38 and paternal grandfather with salivary duct cancer at an age above 80. Germline pathogenic variant analysis on leukocyte DNA and somatic mosaicism analysis on DNA isolated from the colorectal neoplasms were performed simultaneously. A homozygous germline pathogenic variant was identified in *NTHL1*: c.244C>T p.(Gln82*), alias p.(Gln90*).

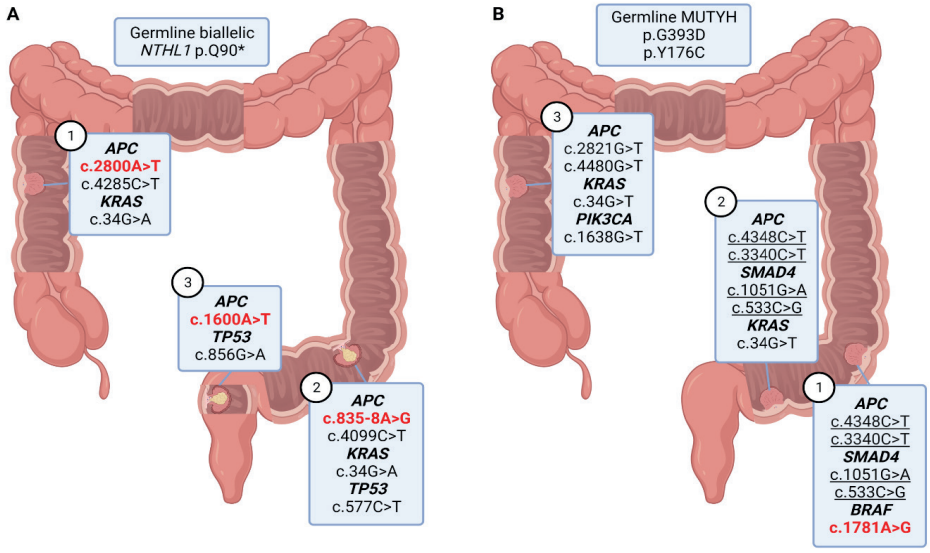


Figure 1 – A. Biallelic *NTHL1* (p.Q90*) patient with three colorectal lesions with *APC* variants fitting the colibactin mutational signature (SBS88) in red. B. The only biallelic *MUTYH* patient with (two) colorectal lesions without the MUTYH associated mutational signature (SBS36). One of these two lesions showed a *BRAF* variant suiting the colibactin mutational signature (SBS88) in red. Created with Biorender.com.

Targeted NGS on both colorectal carcinomas (T2-T3) and the tubular adenoma (T1) removed during index colonoscopy, revealed the colibactin-associated *APC* variant c.835-8A>G in T2 and two other *APC* variants in T1 and T3; c.2008A>T and c.1600A>T, depicted in figure 1a and table 1. The sequence context of c.2008A>T (ATTTT) and c.1600A>T (ATTTT) showed that these two variants also fit SBS88.

Fecal metagenomics showed presence of 6 out of 19 pks genes. Although Formalin Fixed Paraffin Embedded (FFPE) material is not optimal for genome sequencing, mutational signature analyses revealed a significant enrichment of SBS88 in one (T1) of the two analyzed colorectal lesions (T1-T2). None of these lesions showed an enrichment of ID18 or SBS30 (associated with biallelic NTHL1 variants). The distribution of mutational signatures in T1 is depicted in supplemental figure 1.

Biallelic *MUTYH* patients

Furthermore, targeted NGS was performed on 37 colorectal adenomas and 6 colorectal carcinomas from 12 biallelic *MUTYH* patients. All pathogenic variants and variants with unknown pathogenicity detected are summarized in table 1. The majority of adenomas and carcinomas (38 out of 43) showed somatic G>T variants fitting the mutational signature of

defective base excision repair due to biallelic *MUTYH* variants (SBS36). The *KRAS* c.34G>T variant was detected in 63% (25 out of 40) of the lesions in which KRAS was sequenced.

A *BRAF* variant c.1781A>G detected in patient 3 fits SBS88. As shown in figure 1B, this one variant was found in an adenoma lacking variants fitting SBS36. Moreover, the adenoma (T1) shared two *APC* and two *SMAD4* variants with another adenoma (T2), suggestive of a clonal relationship between the adenomas.

To detect additional somatic variants, an OCA Plus NGS panel was performed on T1 but the tumor mutational burden was too low to determine a mutational signature.

Discussion

In this study, NGS of a patient with a biallelic pathogenic *NTHL1* variant showed somatic variants fitting SBS88. Presence of the colibactin-associated signature is confirmed using genome sequencing and pks genes were detected in a stool sample using fecal metagenomics. Previous literature describing exome sequencing colorectal neoplasms of two biallelic NTHL1 patients showed 18 somatic T>N variants but none of these variants fit SBS88.[8] Another exome sequencing study of mono-allelic *NTHL1* patients also did not show colibactin-associated mutational signatures.[9] This is therefore the first study to present a colibactin influence in a biallelic *NTHL1* patient. Strikingly, mutational signature analysis of this same patient did not show a contribution of SBS30. Although more research is needed, the previous study investigating the *NTHL1* signature in multiple neoplasms of biallelic *NTHL1* patient showed that SBS30 did not contribute in all neoplasms to the same extent.[8]

Furthermore, colorectal carcinomas and adenomas of twelve biallelic *MUTYH* patients were analyzed using NGS. This showed, as expected, the *KRAS* variant c.34G>T in the majority of samples (63%).[10] Moreover, in one adenoma of 1 out of 12 patients a BRAF variant, c.1781A>G, was found fitting SBS88. This colibactin-associated mutational signature could unfortunately not be confirmed using a broad NGS panel. Still, this variant is recently described as one of the top 10 recurring somatic variants associated with SBS88-positive colorectal cancers.[4] Therefore, this variant hints towards colibactin mutagenesis in this adenoma.

Also, the *APC* variants c.835-8A>G and c.1600A>T are described as one of these top 10 recurring variants, supporting our findings of fitting SBS88. Both these variants were not common in the 3,916 SBS88 negative colorectal cancers included in this paper (c.835-8A>G: N=18 and c.1600A>T N=3). These findings suggest that these *APC* variants could be used as biomarkers for SBS88 lesions.

Although further research is required, since these numbers are low, this report highlights that presence of *pks*$^{+}$ *E.coli* might be considered as an additional risk factor for the development of colorectal malignancies in patients with a known predisposition to colorectal cancer or polyposis.

Table 1 – Patient characteristics and variants found using targeted Next Generation Sequencing

ID	N polyps	CRC, age first	T	NTHL1	VAF	MUTYH	VAF	KRAS	VAF	APC	VAF	Other	VAF
1	1	Yes, 37	T1 (ad)	c.244C>T	0.99	-	-	c.34G>A	0.3	c.2008A>T*	0.27	NM_000314.8:c.350A>G (3)	0.5
										c.4285C>T	0.31		
			T2 (CRC)	c.244C>T	0.98	-	-	c.38G>A	0.54	c.835-8A>G*	0.38	NM_000314.8:c.350A>G (3)	0.77
										c.4099C>T	0.32	NM_000546.5:c.577C>T	0.57
			T3 (CRC)	c.244C>T	0.94	-	-			c.1600A>T*	0.12	NM_000314.8:c.350A>G (3)	0.59
										c.4012C>T	0.26	NM_000546.5:c.856G>A	0.36
2	50-100	Yes, 44	T2 (ad)	-	-	c.1205C>T	0.46			c.4312delA	0.3		
						c.1178G>A	0.49			c.3747C>A#	0.19		
			T3 (ad)	-	-	c.1205C>T	0.55			c.2950G>T#	0.28		
						c.1178G>A	0.4			c.4630G>T#	0.54		
			T6 (ad)	-	-	c.1205C>T	0.5						
						c.1178G>A	0.46						
			T8 (ad)	-	-	c.1205C>T	0.53	c.34G>T#	0.19	c.4348C>T	0.35		
						c.1178G>A	0.45						
3	Multiple	No	T1 (ad)	-	-	c.527A>G	0.49			c.4348C>T	0.33	NM_004333.6:c.1781A>G*	0.35
						c.1178G>A	0.48			c.3340C>T	0.35	NM_005359.6:c.533C>G	0.37
												NM_005359.6:c.1051G>A	0.33
			T2 (ad)	-	-	c.527A>G	0.49	c.34G>T#	0.05	c.3340C>T	0.28	NM_005359.6:c.533C>G	0.32
						c.1178G>A	0.5			c.4348C>T	0.31	NM_005359.6:c.1051G>A	0.3
			T3 (ad)	-	-	c.527A>G	0.54	c.34G>T#	0.19	c.4480G>T#	0.25	NM_006218.4:c.1638G>T#	0.26
						c.1178G>A	0.52			c.2821G>T#	0.16		

Table 1 – Patient characteristics and variants found using targeted Next Generation Sequencing. Continued

ID	N polyps	CRC, age first	T	NTHL1	VAF	MUTYH	VAF	KRAS	VAF	APC	VAF	Other	VAF
4	100-1000	Yes, 37	T2 (CRC)	-	-	c.1205C>T	0.97	c.34G>T#	0.56	c.1897G>T#	0.22	NM_017763.5:c.1010G>A (3)	0.21
										c.4396G>T#	0.25		
			T4 (ad)	-	-	c.1205C>T	0.98	c.34G>T#	0.24	c.3466G>T#	0.24	NM_017763.5:c.1010G>A (3)	0.45
										c.4381G>T#	0.29	NM_000314.8:c.506C>A#	0.25
												NM_000059.3:c.8167G>T# (3)	0.11
												NM_006231.4:c.121A>G (3)	0.16
			T5 (ad)	-	-	c.1205C>T	0.98			c.601G>T#	0.13	NM_017763.5:c.1010G>A (3)	0.51
										c.4057G>T#	0.16	NM_002691.4:c.1840C>A# (3)	0.13
			T8 (CRC)	-	-	c.1205C>T	0.98	c.34G>T#	0.36	c.4381G>T#	0.34	NM_017763.5:c.1010G>A (3)	0.77
										c.2468C>A#	0.36	NM_000546.5:c.743G>A	0.26
												NM_000546.5:c.401T>C (3)	0.32
												NM_005359.6:c.1609G>T#	0.43
5	Multiple	Yes, 41	T1 (CRC)	-	-	c.1205C>T	0.43	c.34G>T#	0.36	c.1897G>T#	0.26	NM_000059.3:c.8072C>A# (3)	0.28
						c.527A>G	0.44					NM_002439.5:c.187C>T (3)	0.32
												NM_017763.5:c.1913G>A (3)	0.59
												NM_000546.5:c.596G>T#	0.63
			T2 (ad)	-	-	c.1205C>T	0.38	c.34G>T#	0.53	c.2602G>T#	0.33	NM_002691.4:c.2472_2473 delinsAA (3)	0.29
						c.527A>G	0.59			c.4460C>A# (3)	0.54		
			T3 (ad)	-	-	c.1205C>T	0.54	c.34G>T#	0.39	c.2602G>T#	0.33	NM_024675.4:c.1315G>A (3)	0.22
						c.527A>G	0.49			c.3856G>T#	0.52		
										c.4460C>A# (3)	0.37		

Table 1 – Patient characteristics and variants found using targeted Next Generation Sequencing. Continued

ID	N polyps	CRC, age first	T	NTHL1	VAF	MUTYH	VAF	KRAS	VAF	APC	VAF	Other	VAF
5	Multiple	Yes, 41	T4 (ad)	-	-	c.1205C>T	0.44	c.34G>T#	0.22	c.4120G>T#	0.32		
						c.527A>G	0.7						
6	Multiple	Yes, 49	T2 (ad)	-	-	c.527A>G	0.97	c.34G>T#	0.41	c.2602G>T#	0.39		
										c.4351G>T#	0.39		
			T3 (ad)	-	-	c.527A>G	0.97						
			T4 (ad)	-	-	c.527A>G	0.99						
			T5 (ad)	-	-	c.527A>G	0.97						
7	>100	Yes, 45	T1 (CRC)	-	-	c.527A>G	1	c.34G>T#	0.23			NM_005359.6:c.1088G>T# (3)	0.27
			T2 (ad)	-	-	c.527A>G	0.99			c.3769G>T#	0.68		
										c.4630G>T#	0.44		
			T3 (ad)	-	-	c.527A>G	0.99	c.34G>T#		c.4729G>T#	0.22	NM_001904.4:c.1062G>T# (3)	0.26
										c.2432C>A#	0.32		
			T4 (ad)	-	-	c.527A>G	1	c.34G>T#	0.28	c.330C>A#	0.48		
										c.4561G>T#	0.42		
8	Multiple	Yes, 58	T2 (ad)	-	-	c.1205C>T	1	c.34G>T#	0.11	c.350C>A#	0.17		
										c.3131C>A#	0.35		
										c.4120G>T#	0.12		
										c.4639G>T#	0.4		
			T3 (ad)	-	-	c.1205C>T	1	c.34G>T#	0.42	c.4120G>T#	0.35	NM_006218.4:c.1220G>T# (3)	0.18
										c.3949G>T#	0.33		

Table 1 – Patient characteristics and variants found using targeted Next Generation Sequencing. Continued

ID	N polyps	CRC, age first	T	NTHL1	VAF	MUTYH	VAF	KRAS	VAF	APC	VAF	Other	VAF
8	Multiple	Yes, 58	T5 (ad)	-	-	c.1205C>T	1	c.34G>T#	0.34	c.3139G>T#	0.31		
										c.4120G>T#	0.28		
			T6 (ad)	-	-	c.1205C>T	1	c.34G>T#	0.35	c.859G>T#	0.32	NM_005359.6:c.1609G>T#	0.33
										c.784G>T#	0.32	NM_005359.6:c.265G>T#	0.37
9	10-100	Yes, 65	T1 (CRC)	-	-	c.527A>G	0.99	c.34G>T#	0.95				
			T3 (CRC)	-	-	c.527A>G	0.99			c.4660G>T#	0.39		
			T8 (ad)	-	-	c.527A>G	0.93	c.34G>T#	0.62	c.4381G>T#	0.47		
10	30	Yes, 58	T2 (ad)	-	-	c.527A>G	0.46	c.34G>T#	0.19	c.2795C>A#	0.38		
						c.1178G>A	0.46	c.64C>A#	0.2				
			T3 (ad)	-	-	c.527A>G	0.42	c.34G>T#	0.4	c.4606G>T#	0.37		
						c.1178G>A	0.47			c.289G>A	0.39		
										c.1526_1527del	0.41		
			T4 (ad)	-	-	c.527A>G	0.57	c.34G>T#	0.37	c.4606G>T#	0.32		
						c.1178G>A	0.54			c.289G>A	0.33		
										c.1526_1527del	0.42		
			T5 (ad)	-	-	c.527A>G	0.52	c.38G>A	0.46	c.4660G>T#	0.34	NM_000059.3:c.3285G>T# (3)	0.33
						c.1178G>A	0.59			c.2821G>T#	0.31		
11	25-50	No	T3 (ad)	-	-	c.527A>G	0.99	c.34G>T#	0.31	c.2674G>T#	0.34		
										c.4660G>T#	0.28		
			T5 (ad)	-	-	c.527A>G	0.99						

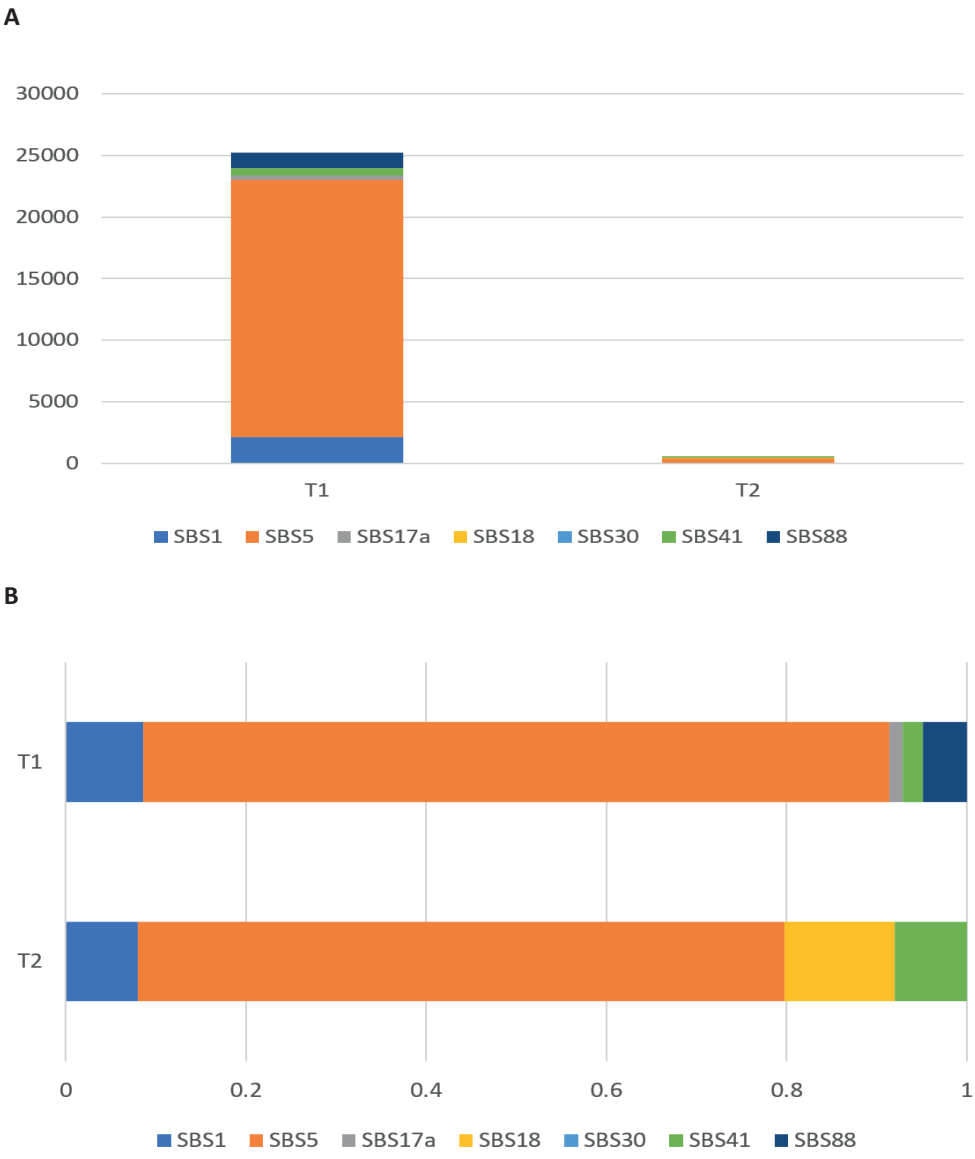
Table 1 – Patient characteristics and variants found using targeted Next Generation Sequencing. Continued

ID	N polyps	CRC, age first	T	NTHL1	VAF	MUTYH	VAF	KRAS	VAF	APC	VAF	Other	VAF
11	25-50	No	T7 (ad)	-	-	c.527A>G	0.99			c.4031C>A#	0.34		
										c.4120G>T#	0.42		
			T8 (ad)	-	-	c.527A>G	0.99			c.3862G>T#	0.47		
12	Few	Yes, 39	T2 (ad)	-	-	c.1178G>A	0.5	c.34G>T#	0.45	c.3451G>T#	0.41		
						c.527A>G	0.53						
			T3 (ad)	-	-	c.1178G>A	0.45	c.34G>T#	0.39	c.3451G>T#	0.32	NM_000546.5:c.596G>T#	0.18
						c.527A>G	0.53						
13	>50	No	T1 (ad)	-	-	c.1178G>A	1	Not covered	X	c.4390G>T#	0.19		
										c.2602G>T#	0.2		
			T2 (ad)	-	-	c.1178G>A	1	Not covered	X	c.2621C>A#	0.35		
										c.4454_4458del	0.35		
			T3 (ad)	-	-	c.1178G>A	1	Not covered	X	c.4630G>T#	0.36		
										c.2464delC	0.37		
						c.1178G>A	0.54			c.289G>A	0.33		
										c.1526_1527del	0.42		
			T5 (ad)	-	-	c.527A>G	0.52	c.38G>A	0.46	c.4660G>T#	0.34	NM_000059.3:c.3285G>T# (3)	0.33
						c.1178G>A	0.59			c.2821G>T#	0.31		

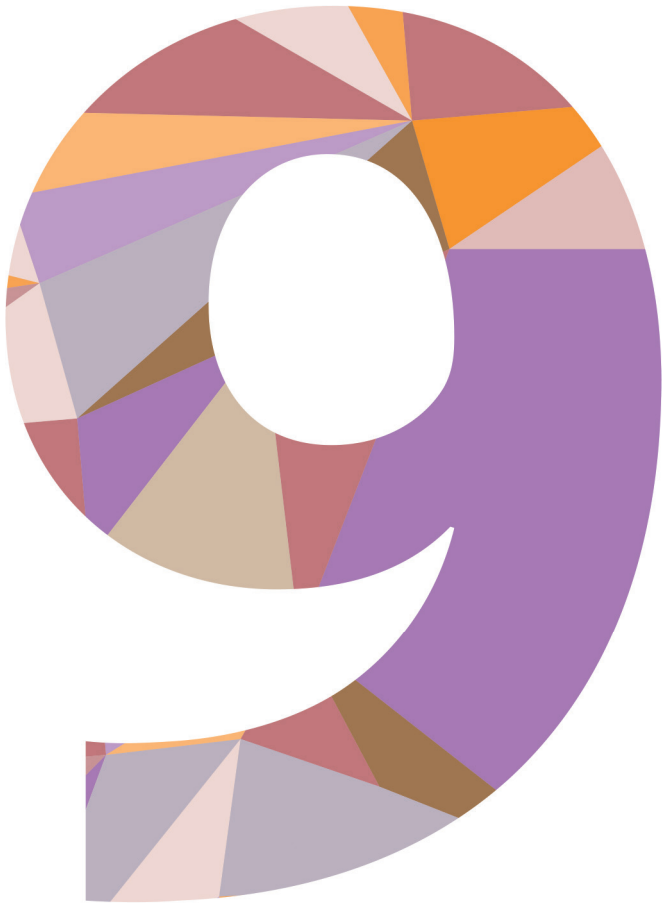
Unless otherwise specified all variants were considered to be (likely) pathogenic. N polyps – number of adenomas at time of collection, ad – adenoma, CRC – colorectal carcinoma, VAF – Variant Allele Frequency, Other – variants in other genes than NTHL1, MUTYH, KRAS or APC. (3) Variants with unknown pathogenicity, * Variants fitting with SBS88 (colibactin associated mutational signature), # Variants fitting with SBS36 (mutational signature associated with MUTYH inactivation).

References

1. Pleguezuelos-Manzano C, Puschhof J, Huber AR, et al. Mutational signature in colorectal cancer caused by genotoxic pks(+)E. coli. Nature 2020;580:269-+.
2. Terlouw D, Suerink M, Boot A, et al. Recurrent APC Splice Variant c.835-8A>G in Patients With Unexplained Colorectal Polyposis Fulfilling the Colibactin Mutational Signature. Gastroenterology 2020;159:1612-1614.e5.
3. Boot A, Ng AWT, Chong FT, et al. Characterization of colibactin-associated mutational signature in an Asian oral squamous cell carcinoma and in other mucosal tumor types. Genome Res 2020;30:803-813.
4. Georgeson P, Steinfelder RS, Harrison TA, et al. Genotoxic colibactin mutational signature in colorectal cancer is associated with clinicopathological features, specific genomic alterations and better survival. This article is a preprint and has not been certified by peer review. MedRxiv. DOI: https://doi.org/10.1101/2023.03.10.23287127.
5. Buc E, Dubois D, Sauvanet P, et al. High prevalence of mucosa-associated E. coli producing cyclomodulin and genotoxin in colon cancer. PLoS One 2013;8:e56964.
6. Dejea CM, Fathi P, Craig JM, et al. Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. Science 2018;359:592-597.
7. Nooij S, Ducarmon QR, Laros JFJ, et al. Faecal microbiota transplantation influences procarcinogenic Escherichia coli in recipient recurrent Clostridioides difficile patients. Gastroenterology 2021.
8. 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.
9. Elsayed FA, Grolleman JE, Ragunathan A, et al. Monoallelic NTHL1 Loss-of-Function Variants and Risk of Polyposis and Colorectal Cancer. Gastroenterology 2020;159:2241-2243 e6.
10. Nielsen M, Morreau H, Vasen HF, et al. MUTYH-associated polyposis (MAP). Crit Rev Oncol Hematol 2011;79:1-16.



Supplemental figure 1 – The distribution of mutational signatures found in colorectal neoplasms of the biallelic *NTHL1* patient. A. Absolute numbers of somatic variants fitting to specific mutational signatures contribution to the spectra of the neoplasms. B. The proportion of different mutational signatures found in T1 and T2.



General discussion

The aim of this thesis was, firstly, to evaluate the proportion of *APC* and *MUTYH* pathogenic variants in colorectal polyposis patients and subsequently identify the proportion of unexplained polyposis patients (Part I, **chapter 2**). Furthermore, three studies aimed to elucidate the significance of *APC* mosaicism and suggest testing and surveillance guidelines (Part II; **chapters 3-5**). Lastly, this thesis aimed to assess another explanation for the development of colorectal adenomatous polyps; the presence of *pks*$^{+}$ *E. coli* and colibactin-associated mutational signatures. (Part III, **chapters 6-8**).

Pathogenic germline variant detection rate in polyposis patients

To determine germline pathogenic *APC* and biallelic *MUTYH* variant detection rates in a Dutch cohort, we collected all patients tested in the Leiden University Medical Center between 1992 and 2017 in **chapter 2**. Comparable to most previous studies, a prevalence of 70% for FAP and 7% for MAP in patients with more than 20 adenomas was determined.[1-7] One previously performed study reported lower variant detection rates throughout the entire cohort.[6] This discrepancy could be explained by the clinical differences between the cohorts, such as age of first adenoma development. A unique aspect of our study is the large patient group with less than 20 adenomas, which could be used to evaluate testing guidelines.

Besides number of adenomas developed, the odds of finding a pathogenic germline variant in *APC* or *MUTYH* increased with a younger age of first adenoma diagnosis. A personal history of CRC only increases the odds of finding biallelic *MUTYH* variants. This can likely be explained by the (sub)total colectomy performed at an early age in FAP patients.[7] Lastly, the odds increased upon having a first-degree relative (FDR) with more than 10 adenomas only for *APC*, which is explained by the dominant and recessive inheritance pattern of FAP and MAP respectively.

Based on these findings, testing for germline pathogenic *APC* and *MUTYH* variants is indicated in patients with more than 10 adenomas before the age of 60 years and more than 20 adenomas before the age of 70 years. Other indications for testing are FAP-related extracolonic manifestations, CRC aged <40, a somatic *KRAS* c.34G>T transversion, or a FDR with >10 adenomas. These suggested guidelines are comparable to the Dutch and National Comprehensive Cancer Network (NCCN) guidelines for hereditary colorectal cancer and polyposis.[8, 9] Guidelines issued by the American College of Gastroenterology (ACG), on the other hand, might result in unnecessary testing.[10]

Our cohort also showed an increasing number of patients undergoing genetic testing for *APC* and *MUTYH* over time. This increase might, first of all, be due to the start of *MUTYH* testing in 2004, which led to more patients with milder phenotypes to be tested. Another reason for more genetic testing in polyposis patients is increased adenomas detection rates caused by more sensitive colonoscopy techniques, improved equipment and bowel preparation and introduction of population based screening in the Netherlands.[11-13] This suggests that prevalence of colorectal adenomas in the general population was possibly underestimated and we now gain relevant insight into the actual numbers. Also, modifiable risk factors like diet, alcohol and smoking, attribute to the development of about a third to half of all CRC.[14-16] This so-called Western lifestyle increases throughout both Western and non-Western countries contributing to CRC prevalence.[17] Therefore, a Western lifestyle may also contribute to the increase in colorectal adenomas in the general population.

Moreover, in **chapter 2**, a large proportion of colorectal polyposis patients remain unexplained, no germline pathogenic *APC* or biallelic *MUTYH* variants. The last decades lots of other colorectal cancer and polyposis associated genes were identified.[18-23] Due to increasing amount of genes included in Next Generation Sequencing (NGS) panels, the proportion of unexplained polyposis patients will eventually decrease. Moreover Whole Exome Sequencing (WES), analyzing the entire exosome, is used to find both newly discovered colorectal cancer or polyposis associated genes and to easily re-analyze patients in the future. Also, nowadays, the use of Whole Genome Sequencing (WGS) is more broadly introduced in the clinic, which compared to WES gives insight into possible pathogenic deep intronic variants, large genomic rearrangements or variants in the non-protein-coding sequences like regulatory sequences as promotors and enhancers, untranslated regions or Mitochondrial Iron-Regulated (MIR) genes.[24-27] Also, WES and WGS will provide data on (single nucleotide) polymorphisms which might add up to the risk of developing colorectal polyposis and cancer.[28] In the future, WGS on DNA from neoplastic tissue will provide knowledge about mutational signatures.[29] These signatures might hint towards an underlying (genetic) cause of the developed neoplasm. The broad use of these extensive sequencing techniques will eventually further decrease the prevalence of germline unexplained polyposis patients.

Prevalence of *APC* mosaicism in unexplained polyposis patients

Besides germline pathogenic *APC* and biallelic *MUTYH* variants and variants in other more rare or not yet discovered genes, a significant part of the unexplained polyposis patients are explained by *APC* mosaicism.[30-35] Especially, analysis of DNA isolated from multiple colorectal adenomatous polyps is efficient to detect *APC* mosaicism.[33] To assess the prevalence of *APC* mosaicism in patients with adenomas, we performed targeted NGS on DNA from colorectal adenomas or carcinomas of 458 patients in **chapter 3**. Moreover, this chapter

provides suggestions of *APC* mosaicism testing and surveillance guidelines. A detection rate of about 17% was found in patients falling inside the Dutch hereditary colorectal polyposis and cancer guidelines. This rate is much lower, about 3%, in patients falling outside these guidelines.

Based on the detection rates per phenotypic subgroup, we recommend *APC* mosaicism testing in all patients with (1) adenomas before the age of 50 years, (2) ≥20 adenomas before the age of 60 years or (3) ≥30 adenomas before the age of 70 years.

The broad spectrum of *APC* mosaicism phenotypes complicates an universal surveillance guideline suggestion. Still, in our opinion, *APC* mosaic patients should receive regular colonoscopies, for example every one or two years, comparable to FAP patients.[36] Re-evaluation of the follow-up could be considered in patients with effective polypectomies.

Furthermore, 28% of mosaic patients undergoing a esophagogastroduodenoscopy developed duodenal or gastric neoplasms. In **chapter 5**, we showed that the upper intestinal adenomas all harbored the mosaic variant. We therefore recommend offering at least one gastroduodenoscopy for all *APC* mosaicism patients. In chapter 5 we moreover present a case of duodenal *APC* mosaicism not affecting the colorectum. This shows the possibility of duodenal *APC* mosaicism despite colorectal adenomas and emphasizes the broad spectrum of *APC* mosaicism and its phenotype.

Moreover, children of 13 mosaic patients did not inherit the *APC* variant. Notable, of 10 patients leukocyte, urine and buccal swab was tested and nine showed a mosaicism restricted to the colorectum. Also, the mosaic variant was detected in 15% to 18% in semen DNA tested of a patient with child wish. Therefore, although chances of hereditability are small[33], we still recommend testing children especially in cases with mosaicism detected in other tissues next to the colorectum.

The family presented in **chapter 4** furthermore highlights **t**he significance of *APC* mosaicism in unexplained polyposis patients. Two first-degree relatives have different mosaic *APC* variants with distinct patterns throughout the body and distinct phenotypes. No underlying defect in DNA repair systems or mutational signatures could be identified using WES and WGS respectively.

A formula adapted from Le Caignec et al[37] determined the probability of finding two *APC* different mosaicism cases in one family to be small. Still, this family shows the value of testing for *APC* mosaicism in unexplained polyposis cases even if a FDR has a comparable phenotype.

Although important in genetic diagnostics, there are challenges in testing for (*APC*) mosaicism. In countries other than the Netherlands, in and outside Europe, *APC* mosaicism is underestimated and not regularly tested. One of the main issues are resources for sequencing multiple samples of one patient. Testing normal colorectal mucosa was a hypothesized solution. However, only in 50% of patients the mosaic variant was detected in a normal colorectal tissue sample.

Another challenge are the so-called hybrid mosaic cases were encountered. These cases have a shared variant in multiple but not all analyzed adenomas. Although this underlines necessity of analyzing more than two colorectal adenomas or carcinomas, the clinical impact remains unknown. Multiple possible explanations for hybrid mosaicism are hypothesized. We considered clonal relationship as an explanation whenever two lesions share the same precursor lesion; two adenomas or carcinomas located close to each other and share (multiple) variants.[38] Contamination, mixing two adenomas during polypectomy or mixing DNA samples during isolation or library preparation, was considered whenever multiple (*APC*) variants were shared between two adenomas or carcinomas and one of the samples also have additional (*APC*) variants. Another hypothesis was field cancerization, a mechanism in which normal tissue is replaced by tumor clones with identical *TP53* variants throughout the colon.[33, 39] This is typically described in inflammatory bowel disease in which chronic inflammation leads to crypt fission. A last explanation is that just by chance common *APC* variants occur in two adenomas of the same patient. In conclusion, no universal explanation could be found and case by case evaluation is required.

Interestingly, hybrid mosaic cases are phenotypically comparable to non-mosaic patients and significantly different from mosaic patients. Therefore, we suggest to treat hybrid mosaic cases as non-mosaic patients in surveillance and family testing guidelines for now. Although rare, an exception to this suggestion should be patients with the hybrid variant in normal colon mucosa or other tissues. In these cases, sporadic adenomas possibly developed in a background of *APC* mosaicism.

The prevalence of *APC* mosaicism might suggest a relevant role of mosaicism in other tumor syndromes. No mosaicism in any other gene included our targeted NGS panel was detected but this might be different in cohorts with other phenotypes than adenomatous polyposis. For example, mosaicism of *SMAD4* and *BMPR1A* might be present in unexplained juvenile polyposis patients.[40] Interestingly, *de novo* variant rates for genes like *BMPR1A*, *PTEN*, *SMAD4*, *STK11* and *TP53* are more than 10% of germline patients, suggestive for occurrence of mosaicism.[41-47]

Presence of colibactin as an additional explanation of colorectal adenomas

As described in **chapter 6**, a large proportion of hybrid mosaic patients shared the *APC* splice variant c.835-8A>G in multiple colorectal adenomas or carcinomas. Furthermore, this variant is the most common somatic *APC* variant detected in our cohort. The variant is predicted likely pathogenic as it leads to a premature stop codon and is detected in about 3% of sporadic colorectal carcinomas.[48-50] The c.835-8A>G variant has a (transcriptional) sequence context of AAAATT, where the underlined thymine is substituted by a cytosine, which perfectly fits the colibactin-associated mutational signature.

Colibactin is a genotoxin known to cause DNA crosslinks, double strand breaks and chromosomal aberrations.[51-53] Colibactin-associated mutational signatures are characterized as SBS88 and ID18.[54, 55]

Publicly available datasets showed that colibactin-associated mutational signatures are present in colorectal, head and neck and urinary tract cancer.[54, 55] Interestingly, the mutational signature is detected in normal colonic crypts with a variable mutational burden between individuals and even between crypts, not attributable to age. Using phylogenetics, the mutational signature was proposed to occur early in life.[56] This is supported by in vitro evidence showing genomic alterations after a short-term exposure to colibactin and the colonization of colibactin-encoding *E. coli* happening the first months after birth.[57, 58] The number of affected normal crypts was variable between patients, with some patients having a more affected left colon while in others the entire colon was affected.[56] This might explain why in our cohort, in **chapter 6 and 7**, the c.835-8A>G variant was detected both as a hybrid and 'real' mosaicism. A recent preprint supports our findings and shows that the c.835-8A>G might act as a biomarker for colibactin influence in the development of the adenoma or carcinoma.[59]

In our unexplained polyposis cohort, 110 patients had with at least one somatic *APC* variant fitting SBS88 or ID18. In **chapter 7,** fecal metagenomics and WGS of colorectal adenomas was performed to further assess the influence of colibactin. Fecal metagenomics detected *pks* genes in 25% of negative controls and 59% of patients with colibactin-associated *APC* variants. This is comparable to 19% to 29% of healthy individuals and approximately 60% of FAP and colorectal cancer patients in previous studies.[53, 60-63] Also, WGS showed an enrichment of colibactin-associated mutational signatures in 39% of cases compared to 11.1% of negative controls.

No clear correlation between presence of *pks* in feces and SBS88 and ID18 in colorectal lesions was detected. There are multiple hypotheses for this finding:

Due to the short-term effect of colibactin, affecting the colon early in life, eradication of the bacteria before feces sampling could be an explanation for patients with colibactin-associated signature without *pks* in feces.[56, 57]

Colonization of *pks*$^{+}$ bacteria after developing adenomas is unlikely in patients with *pks* in feces but no colibactin-associated signatures, as *pks*$^{+}$ *E. coli* is proposed to be transmitted during birth.[58, 64] These patients might, however, be able to inhibit colibactin from entering the host cell or protect cells against the DNA damage. For example, the autophagy-related protein ATG16L1 is associated with preventing colorectal tumorigenesis in presence of *pks*$^{+}$ *E. coli* and oxygen is associated with inhibition of colibactin production.[65, 66] On the other hand, oligosaccharides and co-colonization with enterotoxigenic *Bacteroides fragilis* are described to increase the genotoxic effect of colibactin.[60, 67] Further research should be performed to gain more knowledge about which patients are prone to the carcinogenic effect of colibactin.

Technically, especially WGS on Formalin-Fixed Paraffin Embedded tissue samples affects the performance and interpretation of mutational signature analyses due to fragmentation and deamination artefacts.[68-70] Also, complications detecting *pks* in feces due to possible low abundance of *E. coli* could be circumvented using more sensitive techniques like a specific quantitative PCR.

Chapter 8 emphasizes colibactin as a risk factor in hereditary colorectal cancer and polyposis syndromes. A biallelic *NTHL1* patient is described with *pks* in fecal metagenomics and colibactin-associated mutational signature in WGS data. A small cohort of patients with biallelic *MUTYH* variants showed one somatic *APC* variant in one lesion fitting the colibactin-associated mutational signature. Previous described WES of carcinomas of both biallelic as monoallelic *NTHL1* patients did however not show somatic variants suiting the colibactin-associated mutational signature.[71, 72] The *NTHL1* and MAP patient combined with previously described enrichment of *pks*$^{+}$ *E. coli* in FAP patients[60] and our polyposis cohort results described in **chapter 7,** suggest colibactin as an additional risk factor for development of colorectal malignancies in both sporadic colorectal neoplasms and patients with a known predisposition to CRC or polyposis. Future research should elaborate on this association but also on possible inhibition of colibactin or eradication of *pks+ E. coli*. Besides this, new research is set up to determine whether *pks*$^{+}$ *E. coli* could be used as a biomarker to neoadjuvant treatment response showing the increasing interest and implications of gut microbiome on colorectal cancer.[73]

Future perspectives

Future research into unexplained intestinal polyposis patients will be able to use faster, broader and hopefully better analysis methods. The increasing use of NGS might help in minimizing the number of unexplained polyposis patients. With the use of broad DNA sequencing, besides germline variants, somatic mutational signatures can be detected in tumor cells. These mutational signatures might hint towards the underlying known or unknown genetic cause. Additionally, whole genome sequencing can give insights into non-protein-coding sequences which can be used for finding (intronic splice site) variants in known colorectal polyposis-associated genes and possibly lead to the discovery of new candidate genes. Moreover, research into polygenic risk scores, combining pathogenic variants and single nucleotide polymorphisms in multiple genes, will help in delineating the risk of developing colorectal adenomas in individual patients or families.

Furthermore, future research in *APC* mosaicism should focus on explanations for the so-called hybrid mosaic cases. Also, more knowledge is needed about the association between variant allele frequency of the mosaic variants in different tissues or germ layers, phenotype, and risk of transmitting the variant to offspring. Based on the insights presented in this thesis, *APC* mosaicism will hopefully be recognized as an explanation for colorectal polyposis and be included as regular diagnostics in colorectal polyposis patients.

The association of current or past *pks*$^{+}$ *E. coli* (or other bacteria) derived colibactin exposure and having multiple colorectal adenomas should be more elaborately investigated in larger patient cohorts, even at a population level. Furthermore, the possible association with lifestyle factors should be studied. Moreover, future research should focus on the identification of patients carrying the colibactin-producing bacteria, for example via population-based screening programs. Possible inhibition of the DNA damaging effects of colibactin or eradication of the bacteria involved should be explored.

References

1. Kinzler KW, Nilbert MC, Su LK, et al. Identification of FAP locus genes from chromosome 5q21. Science 1991;253:661-5.
2. Powell SM, Petersen GM, Krush AJ, et al. Molecular diagnosis of familial adenomatous polyposis. N Engl J Med 1993;329:1982-7.
3. Miyoshi Y, Ando H, Nagase H, et al. Germ-line mutations of the APC gene in 53 familial adenomatous polyposis patients. Proc Natl Acad Sci U S A 1992;89:4452-6.
4. Sieber OM, Lipton L, Crabtree M, et al. Multiple colorectal adenomas, classic adenomatous polyposis, and germ-line mutations in MYH. N Engl J Med 2003;348:791-9.
5. Wang L, Baudhuin LM, Boardman LA, et al. MYH mutations in patients with attenuated and classic polyposis and with young-onset colorectal cancer without polyps. Gastroenterology 2004;127:9-16.
6. Stanich PP, Pearlman R, Hinton A, et al. Prevalence of Germline Mutations in Polyposis and Colorectal Cancer-Associated Genes in Patients With Multiple Colorectal Polyps. Clin Gastroenterol Hepatol 2019;17:2008-2015.e3.
7. Grover S, Kastrinos F, Steyerberg EW, et al. Prevalence and phenotypes of APC and MUTYH mutations in patients with multiple colorectal adenomas. Jama 2012;308:485-492.
8. Gupta S, Provenzale D, Regenbogen SE, et al. NCCN Guidelines Insights: Genetic/Familial High-Risk Assessment: Colorectal, Version 3.2017. J Natl Compr Canc Netw 2017;15:1465-1475.
9. Tumoren SOE, Nederland VKG, Oncogenetica W. Erfelijke tumoren: Richtlijnen voor diagnostiek en preventie. https://www.stoet.nl/wp-content/uploads/2017/04/STOET-Richtlijnenboekje-april2017_DEF.pdf 2017.
10. Syngal S, Brand RE, Church JM, et al. ACG clinical guideline: Genetic testing and management of hereditary gastrointestinal cancer syndromes. Am J Gastroenterol 2015;110:223-62; quiz 263.
11. Brown SR, Baraza W, Din S, et al. Chromoscopy versus conventional endoscopy for the detection of polyps in the colon and rectum. Cochrane Database Syst Rev 2016;4:Cd006439.
12. Aranda-Hernández J, Hwang J, Kandel G. Seeing better--Evidence based recommendations on optimizing colonoscopy adenoma detection rate. World J Gastroenterol 2016;22:1767-78.
13. Brenner H, Altenhofen L, Kretschmann J, et al. Trends in Adenoma Detection Rates During the First 10 Years of the German Screening Colonoscopy Program. Gastroenterology 2015;149:356-66.e1.
14. Aleksandrova K, Pischon T, Jenab M, et al. Combined impact of healthy lifestyle factors on colorectal cancer: a large European cohort study. BMC Med 2014;12:168.
15. Erdrich J, Zhang X, Giovannucci E, et al. Proportion of colon cancer attributable to lifestyle in a cohort of US women. Cancer Causes Control 2015;26:1271-1279.
16. Vajdic CM, MacInnis RJ, Canfell K, et al. The Future Colorectal Cancer Burden Attributable to Modifiable Behaviors: A Pooled Cohort Study. JNCI Cancer Spectr 2018;2:pky033.
17. Azeem S, Gillani SW, Siddiqui A, et al. Diet and Colorectal Cancer Risk in Asia--a Systematic Review. Asian Pac J Cancer Prev 2015;16:5389-96.
18. 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.
19. Valle L, Hernandez-Illan E, Bellido F, et al. New insights into POLE and POLD1 germline mutations in familial colorectal cancer and polyposis. Hum Mol Genet 2014;23:3506-12.

20. Gala MK, Mizukami Y, Le LP, et al. Germline mutations in oncogene-induced senescence pathways are associated with multiple sessile serrated adenomas. Gastroenterology 2014;146:520-9.
21. Valle L, de Voer RM, Goldberg Y, et al. Update on genetic predisposition to colorectal cancer and polyposis. Mol Aspects Med 2019;69:10-26.
22. Schubert SA, Morreau H, de Miranda N, et al. The missing heritability of familial colorectal cancer. Mutagenesis 2020;35:221-231.
23. Golubicki M, Bonjoch L, Acuna-Ochoa JG, et al. Germline biallelic Mcm8 variants are associated with early-onset Lynch-like syndrome. JCI Insight 2020;5.
24. 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.
25. Te Paske I, Mensenkamp AR, Neveling K, et al. Noncoding Aberrations in Mismatch Repair Genes Underlie a Substantial Part of the Missing Heritability in Lynch Syndrome. Gastroenterology 2022;163:1691-1694 e7.
26. Makrythanasis P, Antonarakis SE. Pathogenic variants in non-protein-coding sequences. Clin Genet 2013;84:422-8.
27. Perez-Becerril C, Evans DG, Smith MJ. Pathogenic noncoding variants in the neurofibromatosis and schwannomatosis predisposition genes. Hum Mutat 2021;42:1187-1207.
28. Middeldorp A, Jagmohan-Changur S, van Eijk R, et al. Enrichment of low penetrance susceptibility loci in a Dutch familial colorectal cancer cohort. Cancer Epidemiol Biomarkers Prev 2009;18:3062-7.
29. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. Nature 2013;500:415-21.
30. Aretz S, Stienen D, Friedrichs N, et al. Somatic APC mosaicism: a frequent cause of familial adenomatous polyposis (FAP). Hum Mutat 2007;28:985-992.
31. 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.
32. Hes FJ, Nielsen M, Bik EC, et al. Somatic APC mosaicism: an underestimated cause of polyposis coli. Gut 2008;57:71-76.
33. Jansen AML, Crobach S, Geurts-Giele WRR, et al. Distinct Patterns of Somatic Mosaicism in the APC Gene in Neoplasms From Patients With Unexplained Adenomatous Polyposis. Gastroenterology 2017;152:546-+.
34. Out AA, Minderhout IJHMv, Stoep Nvd, et al. High-resolution melting (HRM) re-analysis of a polyposis patients cohort reveals previously undetected heterozygous and mosaic APC gene mutations. 14 2015.
35. 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.
36. Yen T, Stanich PP, Axell L, et al. APC-Associated Polyposis Conditions. In: Adam MP, Everman DB, Mirzaa GM, Pagon RA, Wallace SE, Bean LJH, Gripp KW, Amemiya A, eds. GeneReviews((R)). Seattle (WA), 1993.
37. Le Caignec C, Kwiatkowski DJ, Kury S, et al. Three independent mutations in the TSC2 gene in a family with tuberous sclerosis. Eur J Hum Genet 2009;17:1165-70.
38. Greaves M, Maley CC. Clonal evolution in cancer. Nature 2012;481:306-13.
39. Galandiuk S, Rodriguez-Justo M, Jeffery R, et al. Field cancerization in the intestinal epitheli-

um of patients with Crohn's ileocolitis. Gastroenterology 2012;142:855-864 e8.

40. Calva-Cerqueira D, Chinnathambi S, Pechman B, et al. The rate of germline mutations and large deletions of SMAD4 and BMPR1A in juvenile polyposis. Clin Genet 2009;75:79-85.
41. Coburn MC, Pricolo VE, DeLuca FG, et al. Malignant potential in intestinal juvenile polyposis syndromes. Ann Surg Oncol 1995;2:386-91.
42. Mester J, Eng C. Estimate of de novo mutation frequency in probands with PTEN hamartoma tumor syndrome. Genet Med 2012;14:819-22.
43. Hernan I, Roig I, Martin B, et al. De novo germline mutation in the serine-threonine kinase STK11/LKB1 gene associated with Peutz-Jeghers syndrome. Clin Genet 2004;66:58-62.
44. Gonzalez KD, Buzin CH, Noltner KA, et al. High frequency of de novo mutations in Li-Fraumeni syndrome. J Med Genet 2009;46:689-93.
45. Gammon A, Jasperson K, Pilarski R, et al. PTEN mosaicism with features of Cowden syndrome. Clin Genet 2013;84:593-5.
46. Jansen AML, Goel A. Mosaicism in Patients With Colorectal Cancer or Polyposis Syndromes: A Systematic Review. Clin Gastroenterol Hepatol 2020;18:1949-1960.
47. McKay V, Cairns D, Gokhale D, et al. First report of somatic mosaicism for mutations in STK11 in four patients with Peutz-Jeghers syndrome. Fam Cancer 2016;15:57-61.
48. Fostira F, Thodi G, Sandaltzopoulos R, et al. Mutational spectrum of APC and genotype-phenotype correlations in Greek FAP patients. BMC Cancer 2010;10:389.
49. Jarry J, Brunet JS, Laframboise R, et al. A survey of APC mutations in Quebec. Fam Cancer 2011;10:659-65.
50. Yaeger R, Chatila WK, Lipsyc MD, et al. Clinical Sequencing Defines the Genomic Landscape of Metastatic Colorectal Cancer. Cancer Cell 2018;33:125-136 e3.
51. Nougayrede JP, Homburg S, Taieb F, et al. Escherichia coli induces DNA double-strand breaks in eukaryotic cells. Science 2006;313:848-51.
52. Arthur JC, Perez-Chanona E, Muhlbauer M, et al. Intestinal inflammation targets cancer-inducing activity of the microbiota. Science 2012;338:120-3.
53. Buc E, Dubois D, Sauvanet P, et al. High prevalence of mucosa-associated E. coli producing cyclomodulin and genotoxin in colon cancer. PLoS One 2013;8:e56964.
54. Pleguezuelos-Manzano C, Puschhof J, Huber AR, et al. Mutational signature in colorectal cancer caused by genotoxic pks(+)E. coli. Nature 2020;580:269-+.
55. Boot A, Ng AWT, Chong FT, et al. Characterization of colibactin-associated mutational signature in an Asian oral squamous cell carcinoma and in other mucosal tumor types. Genome Res 2020;30:803-813.
56. Lee-Six H, Olafsson S, Ellis P, et al. The landscape of somatic mutation in normal colorectal epithelial cells. Nature 2019;574:532-537.
57. Iftekhar A, Berger H, Bouznad N, et al. Genomic aberrations after short-term exposure to colibactin-producing E. coli transform primary colon epithelial cells. Nat Commun 2021;12:1003.
58. Tsunematsu Y, Hosomi K, Kunisawa J, et al. Mother-to-infant transmission of the carcinogenic colibactin-producing bacteria. BMC Microbiol 2021;21:235.
59. Georgeson P, Steinfelder RS, Harrison TA, et al. Genotoxic colibactin mutational signature in colorectal cancer is associated with clinicopathological features, specific genomic alterations and better survival. This article is a preprint and has not been certified by peer review. MedRxiv. DOI: https://doi.org/10.1101/2023.03.10.23287127.
60. Dejea CM, Fathi P, Craig JM, et al. Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. Science 2018;359:592-597.

61. Dubinsky V, Dotan I, Gophna U. Carriage of Colibactin-producing Bacteria and Colorectal Cancer Risk. Trends Microbiol 2020;28:874-876.

62. Putze J, Hennequin C, Nougayrede JP, et al. Genetic structure and distribution of the colibactin genomic island among members of the family Enterobacteriaceae. Infect Immun 2009;77:4696-703.

63. Nooij S, Ducarmon QR, Laros JFJ, et al. Faecal microbiota transplantation influences procarcinogenic Escherichia coli in recipient recurrent Clostridioides difficile patients. Gastroenterology 2021.

64. Payros D, Secher T, Boury M, et al. Maternally acquired genotoxic Escherichia coli alters offspring's intestinal homeostasis. Gut Microbes 2014;5:313-25.

65. Lucas C, Salesse L, Hoang MHT, et al. Autophagy of Intestinal Epithelial Cells Inhibits Colorectal Carcinogenesis Induced by Colibactin-Producing Escherichia coli in Apc(Min/+) Mice. Gastroenterology 2020;158:1373-1388.

66. Nadège Bossuet-Greif N, Guyonnet C, Chagneau CV, et al. Oxygen concentration modulates colibactin production. This article is a preprint and has not been certified by peer review. BioRxiv. DOI: https://doi.org/10.1101/2022.06.20.496773.

67. Oliero M, Calve A, Fragoso G, et al. Oligosaccharides increase the genotoxic effect of colibactin produced by pks+ Escherichia coli strains. BMC Cancer 2021;21:172.

68. de Schaetzen van Brienen L, Larmuseau M, Van der Eecken K, et al. Comparative analysis of somatic variant calling on matched FF and FFPE WGS samples. BMC Med Genomics 2020;13:94.

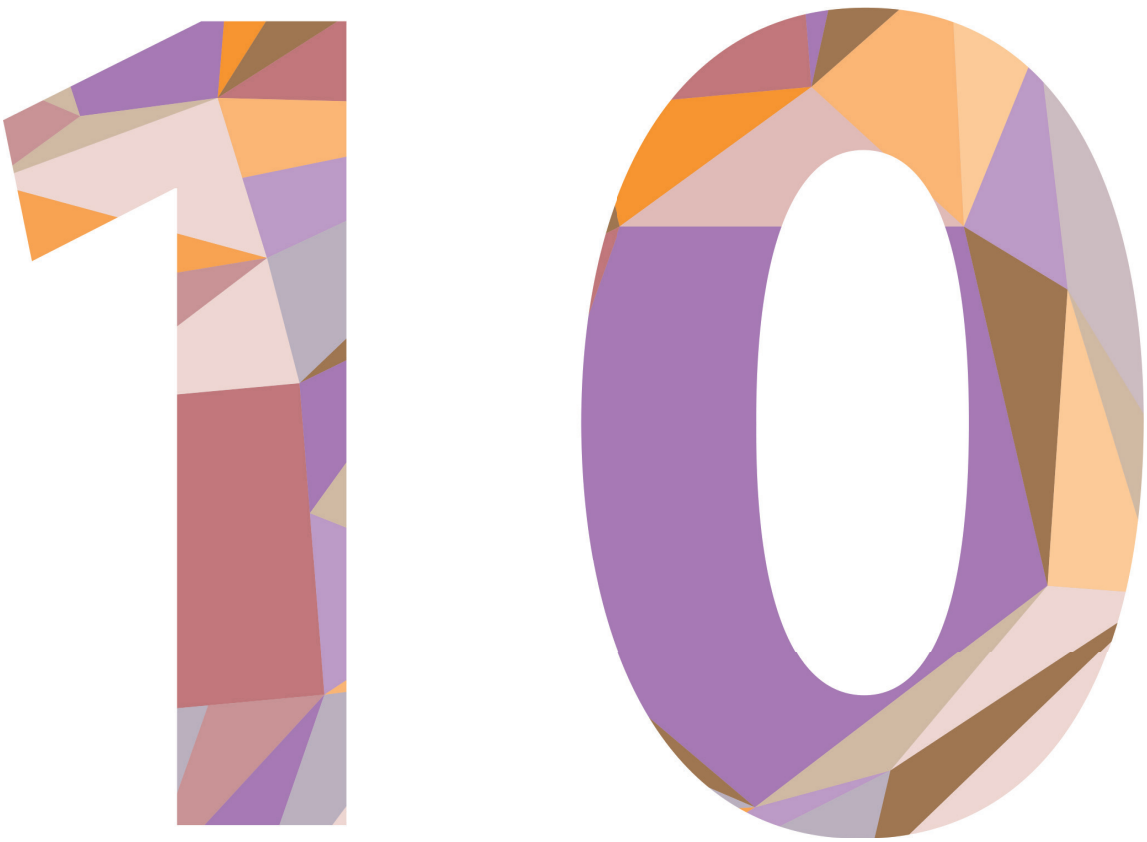
69. Robbe P, Popitsch N, Knight SJL, et al. Clinical whole-genome sequencing from routine formalin-fixed, paraffin-embedded specimens: pilot study for the 100,000 Genomes Project. Genet Med 2018;20:1196-1205.

70. Do H, Dobrovic A. Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization. Clin Chem 2015;61:64-71.

71. 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.

72. Elsayed FA, Grolleman JE, Ragunathan A, et al. Monoallelic NTHL1 Loss-of-Function Variants and Risk of Polyposis and Colorectal Cancer. Gastroenterology 2020;159:2241-2243 e6.

73. Taoum C, Carrier G, Jarlier M, et al. Determination of biomarkers associated with neoadjuvant treatment response focusing on colibactin-producing Escherichia coli in patients with mid or low rectal cancer: a prospective clinical study protocol (MICARE). BMJ Open 2022;12:e061527.



Nederlandse samenvatting

List of publications

Dankwoord

Curriculum Vitae

Onderzoek naar *APC* mozaïcisme; prevalentie, klinische consequenties en onderliggende verklaringen

Dikke darmkanker is een veel voorkomende vorm van kanker met bijna 2 miljoen gevallen wereldwijd in 2020. De verwachting is dat het aantal mensen met de diagnose dikke darmkanker blijft stijgen tot meer dan 3 miljoen in de komende 20 jaar. De meerderheid van de dikke darmkankers ontwikkelt zich vanuit een (adenomateuze) poliep en wordt daarom gezien als een voorstadium van dikke darmkanker. Ongeveer een kwart van alle mensen ontwikkelt tenminste één darmpoliep vóór de leeftijd van 50 en bij de helft van de mensen van 70 is dit het geval. Wetenschappelijk onderzoek naar de oorzaak van de ontwikkeling van dergelijke poliepen laat zien dat verschillende factoren, als voeding, gewicht en leefstijl, en ook erfelijkheid een belangrijke rol spelen.

De rol van erfelijkheid in de ontwikkeling van dikke darmkanker en poliepen wordt gesuggereerd doordat in ongeveer 30% van alle patiënten dikke darmkanker in de familie voorkomt. De twee meest voorkomende erfelijke aandoeningen die de ontwikkeling van vele dikke darmpoliepen (polyposis) veroorzaken zijn Familiaire Adenomateuze Polyposis (FAP) en *MUTYH* geassocieerde Polyposis (MAP). FAP wordt veroorzaakt door een kiembaan DNA verandering (mutatie) in het *APC* gen. FAP patiënten ontwikkelen honderden tot duizenden adenomateuze poliepen (adenomen) vanaf adolescentie en ontwikkelen dikke darmkanker op een gemiddelde leeftijd van 39 jaar. Daarnaast worden er in sommige gevallen ook poliepen of kankers buiten de dikke darm ontwikkeld, zoals in de maag of duodenum. MAP wordt veroorzaakt door twee kiembaan mutaties in het *MUTYH* gen. MAP patiënten hebben een iets milder fenotype en ontwikkelen tientallen tot honderden adenomen. Het risico op dikke darmkanker voor MAP patiënten is 43% tot 63%. Ook MAP patiënten hebben een verhoogd risico op maag en duodenum poliepen en kanker.

Naast FAP en MAP zijn er nog een aantal erfelijke aandoeningen die leiden tot polyposis, de betrokken genen zijn onder andere *NTHL1*, *POLD1 en POLE*. Bij elkaar opgeteld kunnen deze aandoeningen ongeveer 1% van alle polyposis patiënten verklaren.

In **hoofdstuk 2** van dit proefschrift worden alle patiënten die tussen 1992 en 2017 in het Leids Universitair Medisch Centrum zijn onderzocht op een kiembaan *APC* en *MUTYH* mutaties beschreven. Zoals verwacht, werd in 70% van alle patiënten met meer dan 20 adenomen een *APC* mutatie gevonden, voor *MUTYH* was dit 7%. Vernieuwend aan deze studie zijn de geïncludeerde patiënten met minder dan 20 adenomen. Deze patiënten vallen in de regel buiten de richtlijnen voor genetisch testen maar werden in deze studie meegenomen voor evaluatie van de bestaande richtlijnen. Middels verschillende statistische analyses

werd een richtlijn suggestie opgesteld waarbij de huidige Nederlandse richtlijnen worden ondersteund; genetisch testen wordt geadviseerd bij patiënten met meer dan 10 adenomen onder de leeftijd van 60 of meer dan 20 adenomen onder de leeftijd van 70.

Verder werd in **hoofdstuk 2** ook uiteengezet dat in een groot deel van de polyposis patiënten geen kiembaan verklaring gevonden kan worden in *APC* of *MUTYH*. Eerdere wetenschappelijke studies hebben aangetoond dat een zogenoemde mozaïcisme een additionele verklaring kan zijn voor het ontwikkelen van veel adenomen. Mozaïcisme wil zeggen dat een mutatie niet in alle, maar in een deel van de lichaamscellen aanwezig is. In 20% tot 25% van de FAP patiënten is de *APC* mutatie nieuw ontstaan. Afhankelijk van in welke fase van de ontwikkeling van een bevruchte eicel zo'n nieuwe mutatie ontstaat, is deze aanwezig in alle of alleen een deel van de lichaamscellen. Doordat de mutatie niet in alle lichaamscellen aanwezig is, kunnen deze patiënten gemist worden in reguliere diagnostiek waarbij DNA uit bloed wordt geanalyseerd. Verschillende studies laten zien dat 25% tot 50% van de onverklaarde polyposis patiënten met meer dan 20 adenomen verklaard kan worden door *APC* mozaïcisme.

In **hoofdstuk 3** worden meer dan 400 onverklaarde polyposis patiënten beschreven bij wie *APC* mozaïek onderzoek is verricht. Dit cohort bestaat uit patiënten die zowel binnen als buiten de Nederlandse richtlijnen voor genetisch testen vallen. Voor iedere patiënt is DNA geïsoleerd uit materiaal van gemiddeld 4 dikke darmpoliepen. De *APC* mutaties in de poliepen werden vervolgens vergeleken. *APC* mozaïcisme werd gevonden in 15% van de patiënten die binnen de Nederlandse richtlijnen voor genetisch testen vallen. De detectie graad is 3% in patiënten die buiten deze richtlijnen vallen. Op basis van deze resultaten wordt geadviseerd iedere onverklaarde polyposis patiënt die binnen de Nederlandse richtlijnen vallen te testen voor *APC* mozaïcisme.

Het belang van *APC* mozaïek testen werd verder uitgelicht in een familie beschreven in **hoofdstuk 4**. In deze familie hebben twee eerstegraads familieleden *APC* mozaïcisme maar allebei een andere *APC* mutatie. Doordat *APC* mozaïek ontstaat tijdens de ontwikkeling van een bevruchte eicel tot baby verwacht je mozaïek dus niet in meerdere familieleden. Eén van de familieleden, de dochter, heeft een uitgebreid FAP beeld, ook manifestaties buiten de dikke darm, waarbij de mozaïeke mutatie is teruggevonden in verschillende weefsels in het lichaam. De ander, de vader, heeft solitair dikke darm adenomen en de mozaïeke mutatie bevindt zich ook alleen in de dikke darm. Additionele analyses toonden geen onderliggende genetische verklaring voor de aanwezigheid van *APC* mozaïcisme in deze twee gerelateerde patiënten.

Hoofstuk 5 brengt het belang van richtlijnen voor controles van *APC* mozaïek patiënten aan

het licht. In dit hoofdstuk worden duodenum- en maagadenomen geanalyseerd van in dikke darm bekende *APC* mozaïcisme patiënten. Deze analyses lieten in al deze gevallen zien dat ook de twaalfvingerige darm en maag adenomen de mozaïeke *APC* variant hebben en mozaïcisme zich dus ook hoger in het spijsverteringskanaal bevindt. Daarom adviseren wij alle *APC* mozaïek patiënten minimaal één gastroduodenoscopie aan te bieden. Daarnaast wordt in dit hoofdstuk een patiënt beschreven met *APC* mozaïek in de twaalfvingerige darm zonder dikke darm adenomen. Dit is, naar ons weten, de eerste casus beschreven met een geïsoleerde *APC* mozaïek in de twaalfvingerige darm.

Naast erfelijkheid en *APC* mozaïcisme, zijn verschillende leefstijl factoren bekend die een rol spelen in de ontwikkeling van dikke darmpoliepen. Daarnaast is er steeds meer interesse in- en wetenschappelijk bewijs voor de rol van darmbacteriën op de ontwikkeling van dikke darmkanker. Een specifieke *polyketide synthase* bevattende *Escherichia coli* (*pks*$^{+}$ *E. coli*) bacterie wordt, met 60%, vaak teruggevonden in dikke darmkanker patiënten. Deze bacterie produceert een genotoxine genaamd colibactine. Colibactine zorgt voor DNA schade en chromosomale afwijkingen. Inmiddels is de blauwdruk (mutatie signatuur) van colibactine bekend en teruggevonden in zowel dikke darmkanker als mondkanker.

Hoofdstuk 6 van dit proefschrift beschrijft een specifieke *APC* mutatie welke vaker in ons cohort van polyposis patiënten voorkomt dan op basis van toeval of wetenschappelijke literatuur zou worden verwacht. De DNA volgorde van deze mutatie past perfect bij de blauwdruk van colibactine. De aanwezigheid van colibactine zou een verklaring kunnen zijn voor de ontwikkeling van veel adenomen bij een deel van de patiënten.

De aanwezigheid van *APC* mutaties passend bij de colibactine blauwdruk in patiënten met dikke darm adenomen is verder uiteengezet in **hoofdstuk 7**. In totaal werd in 110 patiënten in minimaal 1 van de onderzochte dikke darmlaesies een *APC* mutatie gevonden passend bij de colibactine blauwdruk. Omdat meer wetenschappelijk onderzoek naar de consequentie en invloed van colibactine nodig is, is voor een subset van de patiënten additioneel feces analyse en mutatie signatuur analyse middels een zeer moderne DNA analyse techniek (WGS) verricht. Het onderzoek in feces laat zien dat 59% van de patiënten met in minimaal 1 laesie met een *APC* mutatie passend bij de colibactine blauwdruk ook daadwerkelijk de bacterie bij zich draagt. Dit is 25% in de patiënten die geen mutatie heeft passend bij de blauwdruk. Ondanks de moeilijkheden die WGS op archiefmateriaal van de dikke darmlaesies met zich mee brengt, werd zo'n zelfde verrijking ook gezien in de mutatie signatuur analyse. Deze bevindingen leveren extra bewijs voor de invloed van *pks*$^{+}$ *E. coli* op de ontwikkeling van dikke darm adenomen en kanker.

Het laatste hoofdstuk, **hoofdstuk 8**, van dit proefschrift beschrijft een patiënt met twee

NTHL1 kiembaan mutaties waarbij *APC* mutaties passend bij de colibactine blauwdruk in dikke darmlaesies en *pks* genen in feces gevonden werden. In één van de elf MAP patiënten, bij wie middels *APC* mozaïek analyse het *APC* gen is onderzocht, werd ook een mogelijke aanwijzing voor colibactine gevonden. Deze resultaten suggereren dat de aanwezigheid van colibactine en dus de *pks*$^{+}$ *E. coli* bacterie zou moeten worden beschouwd als een risico factor voor de ontwikkeling van darmkanker ook voor patiënten met een bekende erfelijke aanleg voor dikke darm poliepen en kanker.

List of publications

APC mosaicism, not always isolated: two first-degree relatives with apparently distinct _APC_ mosaicism.

Terlouw D, Hes FJ, Suerink M, Boot A, Langers AMJ, Tops CM, van Leerdam ME, van Asperen CJ, Rozen SG, Bijlsma EK, van Wezel T, Morreau H, Nielsen M. Gut. 2022 Oct 28:gut-jnl-2022-328540

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, van Wezel T, Nielsen M, Morreau H. Gastroenterology. 2020 Oct;159(4)1612-1614

Declining detection rates for _APC_ and biallelic _MUTYH_ pathogenic variants in polyposis patients, implications for DNA testing policy.

Terlouw D, Suerink M, Singh S, Gille JJP, Hes F, Langers A, Morreau H, Vasen H, Vos Y, van Wezel T, Tops C, Ten Broeke S, Nielsen M. Eur J of Hum Genet. 2020 Feb;28(2):222-230

Performance of a RAD51-based functional HRD test on paraffin-embedded breast cancer tissue

van Wijk LM, Vermeulen S, Ter Haar NT, Kramer CJH, Terlouw D, Vrieling H, Cohen D, Vreeswijk MPG. Breast Cancer Res Treat. 2023 Sep;101007

Discordant Staining Patterns and Microsatellite Results in Tumors of MSH6 Pathogenic Variant Carriers

van der Werf-'t Lam AS, Terlouw D, Tops CM, van Kan MS, van Hest LP, Gille JJP, Duijkers FAM, Wagner A, Eikenboom EL, Letteboer TGW, de Jong MM, Bajwa-Ten Broeke SW, Bleeker FE, Gomez Garcia EB, de Wind N, van Wezel T, Morreau H, Suerink M, Nielsen M. Mod Pathol. 2023 Jun;100240

Molecular functions of MCM8 and MCM9 and their associated pathologies.

Helderman NC, Terlouw D, Bonjoch L, Golubicki M, Antelo M, Morreau H, van Wezel T, Castellví-Bel S, Goldberg Y, Nielsen M. iScience. 2023 Apr 27;26(6):106737

Molecular Profile of MSH6-Associated Colorectal Carcinomas Shows Distinct Features From Other Lynch Syndrome–Associated Colorectal Carcinomas

Helderman NC, van der Werf-'t Lam AS, Terlouw D, Bajwa-ten Broeke SW, Rodríguez-Girondo M, van Egmond D, Langers AMJ, van Leerdam ME, Rayner E, van Asperen CJ, van Hest LP, Gille HJP, Duijkers FAM, Wagner A, Eikenboom EL, Letteboer TGW, de Jong MM, Bleeker FE, Gomez Garcìa EB, Suerink M, Tops CM, de Wind N, Morreau H, Boot A, van Wezel T, Nielsen M. Gastroenterology. 2023 Mar; S0016-5085(23)00486-9

Mismatch repair deficiency and MUTYH variants in small intestine-neuroendocrine tumors.

Helderman NC, Elsayed FA, van Wezel T, Terlouw D, Langers AMJ, van Egmond D, Kilinç G, Hristova H, Farina Sarasqueta A, Morreau H, Nielsen M, Suerink M; PALGA-group collaborators. Hum Pathol. 2022 Jul;125:11-17

Prevalence of mismatch repair deficiency and Lynch syndrome in a cohort of unselected small bowel adenocarcinomas.

Suerink M, Kilinç G, Terlouw D, Hristova H, Sensuk L, van Egmond D, Farina Sarasqueta A, Langers A, van Wezel T, Morreau H, Nielsen M; PALGA-group collaborators. J Clin Pathol. 2021 Nov;74(11):724-729

The diverse molecular profiles of lynch syndrome-associated colorectal cancers are (highly) dependent on underlying germline mismatch repair mutations.

Helderman N, Bajwa-Ten Broeke S, Morreau H, Suerink M, Terlouw D, van der Werf-' T Lam AS, van Wezel T, Nielsen M. Crit Rev Oncol Hematol. 2021 Jul;163:10333

Dankwoord

Zoals eigenlijk alles in het leven, had ik de eindstreep van dit proefschrift nooit in mijn eentje kunnen behalen. Ik wil iedereen die heeft bijgedragen enorm bedanken en een aantal mensen in het bijzonder:

Allereerst mijn (co-)promotores, Hans, Tom en Maartje, hartelijk dank voor de tijd en energie die jullie in mij en dit proefschrift hebben gestoken. Dank dat ik altijd langs kon lopen voor vragen en dank voor de mogelijkheid mij te mogen ontwikkelen binnen de wetenschap. Maartje, dank voor het vertrouwen in 2017 als student en in 2019 om mij aan te dragen als PhD student. Hans en Tom, heel erg bedankt voor het vertrouwen en de bereidheid mij op te leiden tot KMBP. Ik kijk uit naar de komende jaren.

Alexandra en Monique, dank voor jullie expertise en het benaderen en includeren van de vele patiënten. Dank Manon en Sanne voor het vertrouwen wat jullie een aantal jaar geleden in mij hadden waardoor ik mijn stage bij de KG mocht beginnen. Manon, dank dat je ook tijdens mijn PhD klaar stond voor vragen en ondersteuning.

Alle MD analisten, dank voor het vele werk wat jullie voor de 'Dianthatjes' verricht hebben, zonder jullie was dit proefschrift zo goed als leeg. Collega's van de pathologie, met name KMBP(io)'s, dank voor het warme welkom en de steun het laatste jaar.

Mede-kantoortuingenoten, Anne-Sophie, Eline, Maayke, Minne, Noah en Remco dank voor jullie luisterend oor maar bovenal de gezelligheid tijdens mijn jaren in K5-187. Ook dank aan de studenten die mij hebben geholpen met de analyses, Max en Yentl. Yentl, ik vond het erg leuk om je op deze manier nog beter te leren kennen.

Lieve Ruby, Gul, Eline, Manijah en Carmen dank voor de etentjes en tripjes waar ik mijn hart kon luchten en uiteraard de interesse in mijn proefschrift.

Mijn 'Van de H' vriendinnen, zonder jullie als uitlaatklep was dit nooit gelukt. Dank voor het verzorgen van de ontspanning maar ook de vermoeide maandag en vrijdagochtenden. Dank voor de steun, liefde en warmte, ook tijdens de mindere momenten.

Lieve familie, lieve ooms, tantes, neefjes en nichtjes dank voor jullie steun en interesse. Wouter, neefie, bedankt voor je hulp met het ontwerpen van zowel de cover als binnenkant van dit proefschrift. Lieve oma's (en opa's) ook jullie bedankt voor de eeuwige steun!

Mijn sterke, stoere mama, bedankt dat jij en papa dit mogelijk hebben gemaakt; goed leren én leuke dingen doen! Dank dat jij mij hebt geleerd altijd door te gaan en het hoofd koel te houden voor wat voor hete vuren je dan ook staat.

Allerliefste paranimfen, Guido en Sofie, bedankt voor jullie hulp en dat jullie naast mij staan! Guido, mijn grote broer en steun en toeverlaat. Dank voor alles. Net zoals dat voor papa was, is niets voor jou te veel! Soof, bedankt dat jij er bent! Zo fijn dat we dichtbij elkaar wonen en je (samen met Bart) voor de nodige afleiding en gezellige tripjes zorgt.

Lieve papa, dank voor je steun en liefde wat ik nog elke dag voel!

Curriculum Vitae

Diantha Terlouw was born in Zwijndrecht, the Netherlands, on the 13th of November 1995. She graduated at her high school Walburg College in Zwijndrecht in 2014. She obtained the bachelor degree in Biomedical Sciences at the University of Leiden in 2017. During her bachelor she did an internship at the department of Clinical Genetics at the Leiden University Medical Center (LUMC) researching incidence of interval carcinomas and polyps in *PMS2* mutation carriers. After finishing this internship, she continued working at the department of Clinical Genetics as a student assistant and helped filling the *PMS2* and *MSH6* databases and continuing the study of *APC* and *MUTYH* prevalence in polyposis patients. She obtained the master degree in Biomedical Sciences with specialisation Research at the University of Leiden in 2019. She did two research internships during her master, one at the department of Parasitology at the LUMC focussing on the role of cholesterol on liver stage malaria infection and one at the department of Clinical Genetics and Pathology of the LUMC. This last internship focussed already on *APC* mosaicism and resulted in the PhD position starting in 2019. Diantha Terlouw has recently started her Clinical Scientist Molecular Pathology training at the department of Pathology at the LUMC.