Characterization of candidate genes in unexplained polyposis and colorectal cancer

Abayzeed Elsayed Osman, F.

Characterization of candidate genes in unexplained polyposis and colorectal cancer

Fadwa Abayzeed Elsayed Osman

ISBN: 978-94-6483-467-3
Cover lay-out, lay-out and design: Daisy Zunnebeld | persoonlijkproefschrift.nl
Printing: Ridderprint | www.ridderprint.nl

Financial support for the printing of this thesis was kindly provided by the Department of Pathology of the Leiden University Medical Center and ChipSoft.

Characterization of candidate genes in unexplained polyposis and colorectal cancer

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 28 november 2023

klokke 15.00 uur

door

Fadwa Abayzeed Elsayed Osman

Promotor
Prof. dr. H. Morreau

Co-promotores
Dr. T. van Wezel
Dr. M. Nielsen

Leden promotiecommissie
Prof. dr. C.J. van Asperen
Prof. dr. F.J. Hes, Universitair Ziekenhuis Brussel
Prof. dr. M.E. van Leerdam
Dr. S.W. Bajwa-ten Broeke, Universitair Medisch Centrum Groningen

Table of contents

Chapter 1

General introduction and outline of this thesis

General introduction

Colorectal cancer

Colorectal cancer (CRC; MIM 114500) is the third most commonly diagnosed cancer and the second leading cause of cancer death worldwide [1-3], with almost 1.9 million new cases and 1 million deaths in 2020. CRC accounts for 10% of all cancer diagnoses and 9.4% of all cancer deaths [2, 4, 5]. CRC results from the progressive accumulation of genetic and epigenetic alterations that lead to the transformation of normal colonic epithelium to colon adenocarcinoma [6]. In the progression from colorectal adenoma to carcinoma, three major pathways are distinguished: chromosomal instability (CIN), microsatellite instability (MSI) and CpG island methylator phenotype (CIMP) [7]. CIN is the most common type of genomic instability observed in CRC and occurs in 80%-85% of colorectal tumors [6]. While the majority of CRCs occur sporadically, an estimated 35% of CRCs are due to heritable factors [8, 9]. Between 5% and 10% of all CRC cases are associated with well-characterized hereditary polyposis and/or CRC syndromes [9]. The etiology of up to 30% of inherited CRCs is not completely understood, and the underlying genetic factors contributing to the risk of CRC remain undefined [10]. Genome-wide association studies (GWAS) have successfully identified common, low-penetrance single nucleotide polymorphisms (SNPs) associated with the risk of CRC [11-17]. In recent years, major efforts have been made to identify the genetic causes, as the identification of germline pathogenic variants substantially facilitates the clinical management of patients and their families.

Hereditary colorectal cancer syndromes

Hereditary CRC syndromes (Table 1), characterized by dramatic increases in the risk of colorectal neoplasia, are phenotypically divided into polyposis and nonpolyposis syndromes, based largely on the number and histology of the colorectal polyps. The polyposis syndromes can be further divided into adenomatous, hamartomatous, serrated and mixed polyposis syndromes according to the predominant type of polyps, e.g., adenomatous polyps, hamartomatous polyps or serrated polyps. Polyposis is defined by the constitutive development of multiple polyps in the colon and rectum. Polyps are benign outgrowths of tissue into the lumen of the colorectum, but they have the potential to evolve into an in situ carcinoma by the accumulation of additional somatic mutations [18]. This phenomenon is known as the adenoma-to-carcinoma sequence, and it is accepted that more than 95% of colorectal cancers arise from adenomas. Syndromic nonpolyposis CRC is subdivided on the basis of molecular tumor phenotype as DNA mismatch repair-deficient (MMRD) or mismatch

repair-proficient (MMRP) CRC [19-21]. The development of polyps in patients with a nonpolyposis CRC predisposition syndrome is rare, but these polyps evolve rapidly into carcinomas since the polyp-to-carcinoma sequence appears to be accelerated in these patients [22]. Several high-penetrance genes with inherited germline variants, such as *APC* (MIM 611731), *BMPR1A* (MIM 601299), *GREM1* (MIM 603054), *MLH1* (MIM 120436), *MSH2* (MIM 609309), *MSH3* (MIM 600887), *MSH6* (MIM 600678), *MUTYH* (MIM 604933), *NTHL1* (MIM 602656), *PMS2* (MIM 600259), *POLD1* (MIM 174761), *POLE* (MIM 174762), *PTEN* (MIM 601728), *RNF43* (MIM 612482), *SMAD4* (MIM 600993) and *STK11* (MIM 602216), are known to be associated with CRC syndromes [23].

Table 1. CRC predisposition syndromes

Syndrome	Genes	Pattern of inheritance	Prevalence
Lynch syndrome	*MLH1*, *MSH2*, *MSH6*, *PMS2*, *EPCAM*	Dominant	2% - 4%
Familial adenomatous polyposis	*APC*	Dominant	< 1%
MUTYH-associated polyposis	*MUTYH*	Recessive	< 1%
Polymerase proofreading-associated polyposis	*POLE*, *POLD1*	Dominant	Unknown
NTHL1-associated polyposis	*NTHL1*	Recessive	Unknown
MSH3-associated polyposis	*MSH3*	Recessive	Unknown
Serrated polyposis syndrome	*RNF43*	Dominant	Unknown
Constitutional MMR deficiency syndrome	*MLH1*, *MSH2*, *MSH6*, *PMS2*	Recessive	Unknown
Hereditary mixed polyposis syndrome	*GREM1*	Dominant	Unknown
Juvenile polyposis	*SMAD4*, *BMPR1A*	Dominant	< 1%
Peutz-Jeghers Syndrome	*STK11*	Dominant	< 1%
PTEN hamartoma tumor syndrome	*PTEN*	Dominant	< 1%

Nonpolyposis syndromes

Lynch syndrome

Lynch syndrome (LS; MIM 120435), previously referred to as hereditary nonpolyposis colorectal cancer (HNPCC), is the most common cause of hereditary CRC, accounting for approximately 2% - 4% of all CRCs [10, 24, 25]. The lifetime CRC risk is estimated to be 50%-80% [10, 24]. This syndrome also predisposes patients to extracolonic cancers, such as cancers of the endometrium, small bowel, ureter and renal pelvis, stomach, hepatobiliary tract and ovary [26-32]. LS is inherited in an autosomal dominant pattern and is caused by germline pathogenic variants in one of the MMR genes (*MLH1*, *MSH2*, *MSH6*, *PMS2*) or 3' end deletion of the *EPCAM* gene, leading to transcriptional read-through into and subsequent epigenetic silencing of MSH2 [33-38]. For LS, the lifetime risk for CRC is highly variable and dependent on the gene involved. The risk for CRC-associated *MLH1* and *MSH2*

mutations is generally higher than the risks associated with mutations in the other LS-related genes [32, 39]. The MMR system consists of several proteins that repair DNA damage during replication and maintain genome stability mainly by correcting base-base and small insertion–deletion mismatches that are generated during DNA replication. MMR proteins function as heterodimers in two main complexes, MutS heterodimers (MSH2/MSH6 and MSH2/MSH3) and MutL heterodimers (MLH1/PMS2, MLH1/PMS1 and MLH1/MLH3). The MutS heterodimers recognize mismatches and small insertions/deletions (indels). The MutL heterodimers form a MutS/MutL/DNA complex for exonuclease activity and termination of mismatch-provoked excision [40, 41]. MMR defects lead to genomic instability and the accumulation of secondary mutations, resulting in a strong mutator phenotype. Mutations occur especially in simple repetitive DNA sequences and microsatellites, resulting in microsatellite instability (MSI). MSI is a hallmark of MMRD cancers and is found in > 90% of LS colorectal cancers [42-44]. In up to 15% of sporadic CRCs, MSI is caused by somatic hypermethylation of the *MLH1* promoter and associated silencing of *MLH1*. These patients frequently also exhibit specific mutations in *BRAF* (V600E) [32].

Familial colorectal cancer type X

In a fraction of families fulfilling the Amsterdam 1 criteria for HNPCC [45], CRCs are microsatellite stable and without MMR gene mutations. These families are defined as having familial colorectal cancer type X (FCCTX) [46, 47]. This heterogeneous group of families has an increased risk of developing CRC and other related tumors [48]. Although the clinical identification of FCCTX has improved in recent years, its genetic etiology remains unknown [47, 49]. Some genes, such as *BMPR1A* [50], *BRCA2* [51], *FAN1* [52], *OGG1* [53], *RPS20* [54], *SEMA4A* [55] and *SETD6* [56], have already been reported to be potentially associated with FCCXT. In addition, a review suggested a possible association with *BCR*, *BMP4*, *CENPE*, *CDH18*, *GABBR2*, *GALNT12*, *GREM1*, *HABP4*, *KIF24* and *ZNF367* [57]. Moreover, a review by Nejadtaghi et al. [58] identified *APC*, *BMPR1A*, *BRAF*, *BRCA2*, *KRAS*, *MGMT*, *RPS20*, *SEMA4A,* and hypermethylation of at least one gene of the MMR system as potentially related to FCCTX. Despite these studies, no defined set of genes is conclusively associated with FCCTX.

Polyposis syndromes

Familial adenomatous polyposis

Less than 1% of all CRCs occur due to familial adenomatous polyposis (FAP; MIM 175100). FAP represents the most common gastrointestinal polyposis syndrome

and the second most common cause of hereditary CRC [59, 60], with an estimated incidence varying from 1:8000 to 1:37600 [61]. FAP is an autosomal dominant precancerous condition characterized by the development of colorectal adenomas, which inevitably progress to colorectal carcinoma unless detected early [10]. In the classic form of FAP, patients develop hundreds to thousands of colorectal adenomas during adolescence or the third decade of life, and the lifetime risk of CRC is almost 100%. Attenuated FAP (AFAP) is a phenotypically distinct form of FAP in which patients have a milder manifestation than classic FAP. AFAP is characterized by fewer colorectal adenoma polyps (less than 100 polyps), a later age of adenoma development and a lower lifetime risk of CRC (70%) [62, 63]. FAP is caused by germline variants in the tumor suppressor gene *APC* [64-67]. *APC* is located on chromosome 5q21-q22 and consists of 15 exons encoding a protein of 2845 amino acids (310 kDa). *APC* plays a major role in the Wnt signaling pathway by negatively regulating the β-catenin oncoprotein [68-70]. Germline *APC* variants lead to the development of multiple adenomas as a result of inactivation of the remaining wild-type *APC* allele in the tumor, either through somatic mutations or through loss of heterozygosity of *APC* [59, 70, 71]. Correlations between the FAP phenotype and the site of mutation in the *APC* gene have been reported; patients with AFAP generally have a mutation in the 5′ or 3′ region of the *APC* gene, whereas individuals with FAP carry mutations elsewhere in this gene [72]. De novo variants are responsible for approximately 25% of FAP cases who lack a family history of the disease, and approximately 20% of these have somatic mosaicism [73-77].

MUTYH-associated polyposis

In 2002, Al Tassan et al. reported for the first time that inherited defects of the base excision repair gene *MUTYH* predispose patients to multiple colorectal adenomas and carcinoma [78], causing MUTYH-associated polyposis (MAP; MIM 608456) [78-81]. MAP is an autosomal recessive inherited syndrome caused by biallelic germline variants in the base excision repair gene *MUTYH*, characterized by a greatly increased lifetime risk of CRC (80%) [82] and accounting for less than 1% of CRC cases [60, 83]. An estimated 1 in every 20,000 European individuals have biallelic *MUTYH* variants [61]. *MUTYH* encodes a DNA glycosylase involved in oxidative DNA damage repair, is located on chromosome 1 between p32.1 and p34.3 and consists of 16 exons [84]. The enzyme excises adenine bases from the DNA backbone at sites where adenine is inappropriately paired with guanine, cytosine, or 8-oxo-7,8-dihydroguanine, a major oxidatively damaged DNA lesion [85-87]. Consequently, tumors from MAP patients with dysfunctional MUTYH protein display an excess of somatic mutations with a strong bias toward C:G > to A:T

transversions at NpCpA or NpCpT sites in multiple genes, including *APC* and *KRAS* [88-90]. A molecular hallmark of cancers caused by MAP is the presence of the somatic *KRAS* c.34G>T mutation [91, 92]. MAP patients show substantial variability in clinical features but usually present with an attenuated polyposis phenotype, showing fewer than 100 adenomas, although a few MAP patients with CRC without polyps have also been reported. The evidence that monoallelic variants confer an elevated CRC risk is somewhat controversial. In a large population-based series, biallelic *MUTYH* variant carriers showed a 28-fold increased risk for CRC, while monoallelic *MUTYH* variants were not associated with an increased CRC risk [93]. However, in other studies, a small increased risk for CRC was reported for *MUTYH* monoallelic variant carriers [94, 95]. Win et al. reported that the CRC risk for monoallelic variant carriers depends on family history and can be sufficiently high to warrant consideration of more intensive CRC screening than for the general population. CRC risk is higher for monoallelic carriers of Y179C than for G396D [96]. A previous study reported that biallelic *MUTYH* carriers have an increased risk of bladder and ovarian cancers, while *MUTYH* monoallelic carriers have an increased risk of gastric, liver, breast and endometrial cancers [97].

Polymerase proofreading-associated polyposis

Germline pathogenic variants affecting the exonuclease domain of *POLE* and *POLD1* predispose patients to multiple colorectal adenomas and carcinomas, causing so-called polymerase proofreading-associated polyposis (PPAP; MIM 615083, 612591) [98-102]. PPAP is an autosomal dominant disease with a high penetrance [98]. In addition to multiple adenomas and CRC, variant carriers also present with extra colonic cancers, such as endometrial, ovarian, brain, pancreatic, and small intestinal cancer and melanoma [103-106]. A recent study indicated that PPAP constitutes 0.1-0.4% of familial cancer cases, reaching 0.3-0.7% when only CRC and polyposis are considered [107]. Although the precise risk and mean age of CRC development are not clear, a study found patients with variants in *POLE* to have a 28% risk and patients with *POLD1* variants to have an 82% to 90% risk of CRC by the age of 70 years [108]. *POLE* and *POLD1* encode the catalytic subunits of DNA polymerases epsilon and delta, respectively. Polymerase epsilon and delta are involved in DNA replication of the leading and lagging strands and possess an accurate proofreading domain that removes incorrectly inserted nucleotides during DNA replication [109]. While the majority of CRCs from *POLE* or *POLD1* variant carriers are MMR proficient, a subset of CRCs in *POLE* variant carriers showed MMR deficiency without germline MMR gene variants [110]. De novo variants in *POLE* have been identified in several singletons [99], but the prevalence of de novo *POLE* variants remains to be determined. Tumors

from *POLE* and *POLD1* pathogenic variant carriers show an ultrahypermutated phenotype with the number of somatic mutations exceeding 100 mutations/Mb [111, 112]. *POLE* defects are associated with signature SBS10 and show an excess of C:G>A:T and C:G>T:A [113, 114]. Thus far, no clear signature has been described for *POLD1*-mutated CRCs.

NTHL1-associated tumor syndrome

In 2015, a rare recessive inherited form of polyposis and CRC syndrome that is caused by biallelic pathogenic variants in the base excision repair gene *NTHL1* was discovered [115]. After the discovery, several additional families from different ethnic groups with biallelic germline variants in *NTHL1* in a homozygous or compound heterozygous state were reported [116-122]. Different extracolonic malignancies were observed in individuals with biallelic germline *NTHL1* variants, including malignancies of the endometrium, breast and duodenum [115, 116, 119]. Based on the frequency of loss-of-function (LoF) variants in the publicly available database, the incidence of NTHL1 deficiency is estimated to be 1:114,770, approximately fivefold lower than the incidence of MAP (1:19,079) [61]. Endonuclease III-like protein 1, encoded by *NTHL1*, is a bifunctional glycosylase involved in base excision repair that recognizes and removes oxidized pyrimidines [123]. Tumors from biallelic *NTHL1* LoF variant carriers show a bias toward C>T transitions at non-CpG sites [115, 124] with a unique mutational signature referred to as signature SBS30 [124]. Signature 30 has previously been identified in one patient with breast cancer [125]. Retrospective analysis of tumor and germline sequencing data of this breast cancer patient revealed a heterozygous germline *NTHL1* variant with loss of heterozygosity in the tumor [124].

MSH3-associated polyposis

Another polyposis syndrome with a recessive inheritance pattern is referred to as MSH3-associated polyposis (MIM 617100) [126]. After whole-exome sequencing (WES) of leukocyte DNA from 102 unrelated individuals with unexplained adenomatous polyposis, two unrelated individuals with compound heterozygous LoF germline variants in *MSH3* were identified, suggesting that *MSH3* mutations represent an additional recessive subtype of colorectal adenomatous polyposis [126]. The tumors from the carriers demonstrated high microsatellite instability of di- and tetranucleotides (Elevated Microsatellite Alterations at Selected Tetranucleotide repeats (EMAST) [127]) and immunohistochemical loss of MSH3 in normal and tumor tissues [126]. The associated phenotype was characterized by the presence of colorectal and duodenal adenomas, CRC, gastric cancer and early-onset astrocytoma [126].

Constitutional MMR deficiency syndrome

Constitutional MMR deficiency (CMMRD; MIM 276300) syndrome is a rare autosomal recessive childhood cancer predisposition syndrome caused by biallelic pathogenic germline variants in one MMR gene (*MLH1*, *MSH2*, *MSH6* and *PMS2*). CMMRD is characterized by a high risk of developing a broad spectrum of malignancies during childhood and adolescence, including mainly T-cell non-Hodgkin lymphomas, high-grade gliomas and gastrointestinal tumors, mainly CRC tumors. Another characteristic of CMMRD is café-au-lait maculae (CALM) [128, 129]. Remarkably, a large proportion of CMMRD patients develop multiple synchronous adenomas ranging from a few up to > 100 polyps, mimicking attenuated familial adenomatous polyposis [130-132]. Polyps in CMMRD can also histologically resemble those in juvenile polyposis [131].

Serrated polyposis syndrome

Serrated polyposis syndrome (SPS; MIM 617108) was previously known as hyperplastic polyposis syndrome (HPS). SPS is characterized by the presence of multiple serrated polyps throughout the colon and rectum and is associated with an increased risk of CRC for affected individuals and their first-degree relatives [133-135]. The prevalence of SPS is estimated to be 1:2000 in the general population [134]. In 2014, Gala et al. reported the association between SPS and *RNF43* by identifying a novel germline variant in two individuals with multiple serrated polyps [136]. Subsequently, the role of *RNF43* germline variants as the cause of multiple serrated polyps was supported by several other studies [137-139]. The study by Yan et al. showed loss of the remaining wild-type allele from carriers through somatic mutations or loss of heterozygosity, adding the potential role of *RNF43* in the development of colonic serrated neoplasia [138]. Buchanan et al. proposed that mutations in *RNF43* might account for only a small proportion of SPS, and consequently, there is no need for routine germline testing of *RNF43* in individuals who meet the criteria for SPS [140].

Hereditary mixed polyposis syndrome

Hereditary mixed polyposis syndrome (HMPS1 MIM 601228) is a rare autosomal dominant disorder that is associated with an increased risk of CRC, characterized by polyps of multiple and mixed morphologies, including serrated lesions, Peutz–Jeghers polyps, juvenile polyps and conventional adenomas [141-144]. The genetic etiology for HMPS1 was first described in 2012, when a 40-kb duplication in the 5′ regulatory region of *GREM1* was identified as a causal mutation in families of Ashkenazi Jewish origin and was shown to lead to increased and ectopic expression of *GREM1* in the colonic mucosa [144]. Excess GREM1 protein

levels suppress bone morphogenetic protein [144], allowing epithelial cells to retain stem cell-like properties, form ectopic crypts and ultimately become neoplastic [145]. The 40-Kb duplication has been identified in 1:184 Ashkenazi Jewish individuals with a personal or familial history of polyposis or CRC [146]. In addition to the founder Ashkenazi duplication, several other *GREM1* variants were identified in families with polyposis and CRC [147-149].

Hamartomatous polyposis syndromes

Hamartomatous polyposis syndromes (HPSs) are a rare heterozygous group of disorders that are inherited in an autosomal-dominant manner and are characterized by the development of hamartomatous polyps of the gastrointestinal tract. Hamartomatous polyposis syndromes have malignant potential for the development of CRC as well as extracolonic cancers [63]. These conditions account for less than 1% of CRC cases and occur at approximately one-tenth of the frequency of adenomatous polyposis syndromes [150, 151]. The hamartomatous polyposis syndromes include juvenile polyposis syndrome (JPS), Peutz-Jegher's syndrome (PJS) and PTEN hamartoma tumor syndrome (PHTS).

Juvenile polyposis syndrome (JPS)

JPS is characterized by the development of multiple gastrointestinal polyps, the most common location of which is the colon (98%). Patients with JPS syndrome have a high risk of colon cancer, and there is also an increased risk of gastroduodenal cancer. Pathogenic germline variants in *SMAD4* or *BMPR1A* are found in approximately 20-60% of JPS patients [63].

Peutz–Jeghers syndrome (PJS)

PJS is caused by germline variants in *STK11* (previously known as *LKB1*) and is characterized by multiple characteristic hamartomatous polyps in the gastrointestinal tract associated with mucocutaneous pigmentation. Patients with PJS have an increased risk for CRC and extra colonic cancers [63].

PTEN hamartoma tumor syndrome (PHTS)

Germline variants in the tumor suppressor gene *PTEN* are responsible for a group of phenotypically diverse conditions, which have collectively been called PTEN hamartoma syndrome (PHTS) [63, 134, 152]. PHTS includes Cowden syndrome (CS) and Bannayan-Riley-Ruvalcaba syndrome (BRRS), both of which are inherited in an autosomal dominant pattern [151, 153, 154]. CS is rarely identified before adulthood and is characterized by multiple developmentally disorganized benign growths,

or hamartomas, with an increased risk of both benign and malignant tumors [155]. Individuals with CS are at risk for developing breast, thyroid, endometrial, colon, skin and renal cancers [156]. BRRS patients show gastrointestinal hamartomatous polyps, lipomas, macrocephaly and developmental delay [152].

Missing heritable factors in CRC and polyposis

The exact contribution of heritable factors to CRC and polyposis is still not fully understood. Based on Nordic twin and family studies, it has been estimated that 12-35% of all CRCs are linked to genetic factors [8, 157]. Later, estimates for heritability of CRC decreased to approximately 15% of all CRC cases [158, 159]. The currently known high-penetrant Mendelian polyposis and/or CRC syndromes can only explain 5-10% of all CRC cases [8, 60, 160, 161]. In the case of polyposis, the genetic causes remain unexplained in approximately 20% of polyposis cases [162]. In approximately 60% of MMRD CRCs without somatic *MLH1* promoter hypermethylation, no underlying germline MMR variants are known. These patients are referred to as having suspected Lynch syndrome (sLS) or Lynch-like syndrome (LLS) [163]. Studies have shown that patients with double somatic MMR pathogenic variants can still have hereditary CRC caused by genes involved in DNA repair since they can lead to acquired pathogenic variants in the MMR genes [164-166]. The genetic background is unknown for 50-60% of hereditary nonpolyposis colorectal cancer (HNPCC) families who fulfil the Amsterdam criteria [45] but do not have a mutation in one of the MMR genes (MMRP), referred to as familial colorectal cancer type X (FCCTX) [167]. In addition to the identification of rare high-penetrant risk genes contributing to the heredity of CRC, it is estimated that common variants may explain approximately 12% of the relative risk for CRC [14, 16, 161, 168]. In more than approximately one-third of CRC patients with a suspected hereditary cause, the underlying genetic factors remain unexplained [157]. It is important to resolve this issue with heritability, and the identification of genetic factors has important implications for the carriers and their families, as it helps risk assessment, directs clinical management, and guides preventive and therapeutic options [10, 169].

Novel candidate genes for CRC and polyposis

Recently, different candidate genes have been identified but require further evidence to be implemented in routine genetic testing. New candidate genes have been proposed for predisposition to hereditary CRC and polyposis, such as *BUB1* [170], *BUB3* [170], *FAN1* [52], *LRP6* [171], *RPS20* [54], *FOCAD* [172], *PTPN12* [171], *GALTN12* [173, 174], *MIA3* [175] and the constitutional epigenetic silencing of *PTPRJ* [176]. Recently, *MCM8* was proposed for predisposition to CRC with a recessive pattern of

inheritance [177]. In a systematic review performed to validate the association between variants in *RPS20*, *FANCM*, *FAN1*, *TP53*, *BUB1*, *BUB3*, *LRP6* and *PTPN12* and the development of CRC, the evidence supports the association between variants in *RPS20* and CRC but not in the other candidate genes [178].

Outline of this thesis

The aim of this thesis is to study the underlying genetic causes of unexplained polyposis and CRC. In particular, the role of *POLE*, *POLD1*, *APC* and *NTHL1* in unexplained cases was studied.

Chapter 2 describes the assessment of the prevalence of *POLE* p.(Leu424Val) and *POLD1* p.(Ser478Asn) in a Dutch series of index patients with unexplained familial early onset CRC and polyposis. In this study, we analyzed phenotypes and tumor characteristics in *POLE* variant carriers. We proposed that MMR deficiency in the tumors from *POLE* p.(Leu424Val) carriers is due to secondary MMR somatic mutation resulting from the hypermutation phenotype caused by the *POLE* variants.

In **Chapter 3**, the sequencing of the exonuclease domains of *POLE* and *POLD1* in unexplained index patients with multiple colorectal polyps is described in search for novel germline variants in these genes.

Chapter 4 focuses on screening of *APC* for mosaic and deep intronic variants in unexplained colorectal polyposis patients to study their role as predisposing factors for polyposis and CRC in this cohort.

Chapter 5 shows the molecular and clinical characterization of the tumor spectrum of individuals with biallelic LoF germline variants in *NTHL1*. To establish the disease phenotype of individuals with NTHL1 deficiency, we identified individuals with biallelic LoF germline variants in *NTHL1* and performed mutational signature analysis on different tumor types from these individuals to determine the association between NTHL1 deficiency and tumor development.

In **Chapter 6,** the role of monoallelic LoF germline variants in *NTHL1* in the risk of polyposis and/or CRC is investigated. Finally, **Chapter 7** provides a general discussion of the thesis and future perspectives.

References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394-424.
2. 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.
3. Mattiuzzi C, Sanchis-Gomar F, Lippi G. Concise update on colorectal cancer epidemiology. Ann Transl Med 2019;7:609.
4. Ferlay J, Colombet M, Soerjomataram I, et al. Cancer statistics for the year 2020: An overview. Int J Cancer 2021.
5. Xi Y, Xu P. Global colorectal cancer burden in 2020 and projections to 2040. Transl Oncol 2021;14:101174.
6. Grady WM, Carethers JM. Genomic and epigenetic instability in colorectal cancer pathogenesis. Gastroenterology 2008;135:1079-99.
7. Goel A, Nagasaka T, Arnold CN, et al. The CpG island methylator phenotype and chromosomal instability are inversely correlated in sporadic colorectal cancer. Gastroenterology 2007;132:127-38.
8. 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.
9. 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.
10. Jasperson KW, Tuohy TM, Neklason DW, et al. Hereditary and familial colon cancer. Gastroenterology 2010;138:2044-58.
11. Broderick P, Carvajal-Carmona L, Pittman AM, et al. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nat Genet 2007;39:1315-7.
12. Dunlop MG, Dobbins SE, Farrington SM, et al. Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. Nat Genet 2012;44:770-6.
13. Houlston RS, Cheadle J, Dobbins SE, et al. Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. Nat Genet 2010;42:973-7.
14. Law PJ, Timofeeva M, Fernandez-Rozadilla C, et al. Association analyses identify 31 new risk loci for colorectal cancer susceptibility. Nat Commun 2019;10:2154.
15. Orlando G, Law PJ, Palin K, et al. Variation at 2q35 (PNKD and TMBIM1) influences colorectal cancer risk and identifies a pleiotropic effect with inflammatory bowel disease. Hum Mol Genet 2016;25:2349-2359.
16. Schmit SL, Edlund CK, Schumacher FR, et al. Novel Common Genetic Susceptibility Loci for Colorectal Cancer. J Natl Cancer Inst 2019;111:146-157.
17. Tomlinson IP, Webb E, Carvajal-Carmona L, et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. Nat Genet 2008;40:623-30.
18. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell 1990;61:759-67.

19. Valle L, Vilar E, Tavtigian SV, et al. Genetic predisposition to colorectal cancer: syndromes, genes, classification of genetic variants and implications for precision medicine. J Pathol 2019;247:574-588.
20. Kastrinos F, Samadder NJ, Burt RW. Use of Family History and Genetic Testing to Determine Risk of Colorectal Cancer. Gastroenterology 2020;158:389-403.
21. Hampel H, Kalady MF, Pearlman R, et al. Hereditary Colorectal Cancer. Hematol Oncol Clin North Am 2022;36:429-447.
22. Lynch HT, Snyder CL, Shaw TG, et al. Milestones of Lynch syndrome: 1895-2015. Nat Rev Cancer 2015;15:181-94.
23. Terradas M, Munoz-Torres PM, Belhadj S, et al. Contribution to colonic polyposis of recently proposed predisposing genes and assessment of the prevalence of NTHL1- and MSH3-associated polyposes. Hum Mutat 2019;40:1910-1923.
24. Ma H, Brosens LAA, Offerhaus GJA, et al. Pathology and genetics of hereditary colorectal cancer. Pathology 2018;50:49-59.
25. Hampel H, Frankel WL, Martin E, et al. Feasibility of screening for Lynch syndrome among patients with colorectal cancer. J Clin Oncol 2008;26:5783-8.
26. Vasen HF, Offerhaus GJ, den Hartog Jager FC, et al. The tumour spectrum in hereditary non-polyposis colorectal cancer: a study of 24 kindreds in the Netherlands. Int J Cancer 1990;46:31-4.
27. Dowty JG, Win AK, Buchanan DD, et al. Cancer risks for MLH1 and MSH2 mutation carriers. Hum Mutat 2013;34:490-7.
28. Baglietto L, Lindor NM, Dowty JG, et al. Risks of Lynch syndrome cancers for MSH6 mutation carriers. J Natl Cancer Inst 2010;102:193-201.
29. Senter L, Clendenning M, Sotamaa K, et al. The clinical phenotype of Lynch syndrome due to germ-line PMS2 mutations. Gastroenterology 2008;135:419-28.
30. ten Broeke SW, Brohet RM, Tops CM, et al. Lynch syndrome caused by germline PMS2 mutations: delineating the cancer risk. J Clin Oncol 2015;33:319-25.
31. Win AK, Lindor NM, Young JP, et al. Risks of primary extracolonic cancers following colorectal cancer in lynch syndrome. J Natl Cancer Inst 2012;104:1363-72.
32. Cox VL, Saeed Bamashmos AA, Foo WC, et al. Lynch Syndrome: Genomics Update and Imaging Review. Radiographics 2018;38:483-499.
33. Ligtenberg MJ, Kuiper RP, Chan TL, et al. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3′ exons of TACSTD1. Nat Genet 2009;41:112-7.
34. Miyaki M, Konishi M, Tanaka K, et al. Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. Nat Genet 1997;17:271-2.
35. Bronner CE, Baker SM, Morrison PT, et al. Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. Nature 1994;368:258-61.
36. Nicolaides NC, Papadopoulos N, Liu B, et al. Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. Nature 1994;371:75-80.
37. Fishel R, Lescoe MK, Rao MR, et al. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell 1993;75:1027-38.
38. Leach FS, Nicolaides NC, Papadopoulos N, et al. Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. Cell 1993;75:1215-25.
39. Giardiello FM, Allen JI, Axilbund JE, et al. Guidelines on genetic evaluation and management of Lynch syndrome: a consensus statement by the US Multi-Society Task Force on colorectal cancer. Gastroenterology 2014;147:502-26.
40. Jiricny J. The multifaceted mismatch-repair system. Nat Rev Mol Cell Biol 2006;7:335-46.
41. Li GM. Mechanisms and functions of DNA mismatch repair. Cell Res 2008;18:85-98.

42. Gausachs M, Mur P, Corral J, et al. MLH1 promoter hypermethylation in the analytical algorithm of Lynch syndrome: a cost-effectiveness study. Eur J Hum Genet 2012;20:762-8.
43. Loukola A, Eklin K, Laiho P, et al. Microsatellite marker analysis in screening for hereditary nonpolyposis colorectal cancer (HNPCC). Cancer Res 2001;61:4545-9.
44. Aaltonen LA, Salovaara R, Kristo P, et al. Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. N Engl J Med 1998;338:1481-7.
45. Vasen HF, Mecklin JP, Khan PM, et al. The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). Dis Colon Rectum 1991;34:424-5.
46. Francisco I, Albuquerque C, Lage P, et al. Familial colorectal cancer type X syndrome: two distinct molecular entities? Fam Cancer 2011;10:623-31.
47. Garcia FAO, de Andrade ES, de Campos Reis Galvão H, et al. New insights on familial colorectal cancer type X syndrome. Sci Rep 2022;12:2846.
48. Martín-Morales L, Garre P, Lorca V, et al. BRIP1, a Gene Potentially Implicated in Familial Colorectal Cancer Type X. Cancer Prev Res (Phila) 2021;14:185-194.
49. Sánchez-Tomé E, Rivera B, Perea J, et al. Genome-wide linkage analysis and tumoral characterization reveal heterogeneity in familial colorectal cancer type X. J Gastroenterol 2015;50:657-66.
50. Nieminen TT, Abdel-Rahman WM, Ristimäki A, et al. BMPR1A mutations in hereditary nonpolyposis colorectal cancer without mismatch repair deficiency. Gastroenterology 2011;141:e23-6.
51. Garre P, Martín L, Sanz J, et al. BRCA2 gene: a candidate for clinical testing in familial colorectal cancer type X. Clin Genet 2015;87:582-7.
52. Seguí N, Mina LB, Lázaro C, et al. Germline Mutations in FAN1 Cause Hereditary Colorectal Cancer by Impairing DNA Repair. Gastroenterology 2015;149:563-6.
53. Kim IJ, Ku JL, Kang HC, et al. Mutational analysis of OGG1, MYH, MTH1 in FAP, HNPCC and sporadic colorectal cancer patients: R154H OGG1 polymorphism is associated with sporadic colorectal cancer patients. Hum Genet 2004;115:498-503.
54. Nieminen TT, O'Donohue MF, Wu Y, et al. Germline mutation of RPS20, encoding a ribosomal protein, causes predisposition to hereditary nonpolyposis colorectal carcinoma without DNA mismatch repair deficiency. Gastroenterology 2014;147:595-598.e5.
55. Schulz E, Klampfl P, Holzapfel S, et al. Germline variants in the SEMA4A gene predispose to familial colorectal cancer type X. Nat Commun 2014;5:5191.
56. Martín-Morales L, Feldman M, Vershinin Z, et al. SETD6 dominant negative mutation in familial colorectal cancer type X. Hum Mol Genet 2017;26:4481-4493.
57. Dominguez-Valentin M, Therkildsen C, Da Silva S, et al. Familial colorectal cancer type X: genetic profiles and phenotypic features. Mod Pathol 2015;28:30-6.
58. Nejadtaghi M, Jafari H, Farrokhi E, et al. Familial Colorectal Cancer Type X (FCCTX) and the correlation with various genes-A systematic review. Curr Probl Cancer 2017;41:388-397.
59. Leoz ML, Carballal S, Moreira L, et al. The genetic basis of familial adenomatous polyposis and its implications for clinical practice and risk management. Appl Clin Genet 2015;8:95-107.
60. Yurgelun MB, Kulke MH, Fuchs CS, et al. Cancer Susceptibility Gene Mutations in Individuals With Colorectal Cancer. J Clin Oncol 2017;35:1086-1095.
61. Weren RD, Ligtenberg MJ, Geurts van Kessel A, et al. NTHL1 and MUTYH polyposis syndromes: two sides of the same coin? J Pathol 2018;244:135-142.
62. Samadder NJ, Baffy N, Giridhar KV, et al. Hereditary Cancer Syndromes-A Primer on Diagnosis and Management, Part 2: Gastrointestinal Cancer Syndromes. Mayo Clin Proc 2019;94:1099-1116.

63. Kidambi TD, Kohli DR, Samadder NJ, et al. Hereditary Polyposis Syndromes. Curr Treat Options Gastroenterol 2019;17:650-665.
64. Bodmer WF, Bailey CJ, Bodmer J, et al. Localization of the gene for familial adenomatous polyposis on chromosome 5. Nature 1987;328:614-6.
65. Kinzler KW, Nilbert MC, Su LK, et al. Identification of FAP locus genes from chromosome 5q21. Science 1991;253:661-5.
66. Nishisho I, Nakamura Y, Miyoshi Y, et al. Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. Science 1991;253:665-9.
67. Groden J, Thliveris A, Samowitz W, et al. Identification and characterization of the familial adenomatous polyposis coli gene. Cell 1991;66:589-600.
68. Fearnhead NS, Britton MP, Bodmer WF. The ABC of APC. Hum Mol Genet 2001;10:721-33.
69. Mishra N, Hall J. Identification of patients at risk for hereditary colorectal cancer. Clin Colon Rectal Surg 2012;25:67-82.
70. Fearnhead NS, Wilding JL, Bodmer WF. Genetics of colorectal cancer: hereditary aspects and overview of colorectal tumorigenesis. Br Med Bull 2002;64:27-43.
71. 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.
72. Peters U, Bien S, Zubair N. Genetic architecture of colorectal cancer. Gut 2015;64:1623-36.
73. Bisgaard ML, Fenger K, Bülow S, et al. Familial adenomatous polyposis (FAP): frequency, penetrance, and mutation rate. Hum Mutat 1994;3:121-5.
74. Farrington SM, Dunlop MG. Mosaicism and sporadic familial adenomatous polyposis. Am J Hum Genet 1999;64:653-8.
75. Aretz S, Uhlhaas S, Caspari R, et al. Frequency and parental origin of de novo APC mutations in familial adenomatous polyposis. Eur J Hum Genet 2004;12:52-8.
76. Hes FJ, Nielsen M, Bik EC, et al. Somatic APC mosaicism: an underestimated cause of polyposis coli. Gut 2008;57:71-6.
77. Talseth-Palmer BA. The genetic basis of colonic adenomatous polyposis syndromes. Hered Cancer Clin Pract 2017;15:5.
78. 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.
79. Jones S, Emmerson P, Maynard J, et al. Biallelic germline mutations in MYH predispose to multiple colorectal adenoma and somatic G:C-->T:A mutations. Hum Mol Genet 2002;11:2961-7.
80. 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.
81. Sampson JR, Dolwani S, Jones S, et al. Autosomal recessive colorectal adenomatous polyposis due to inherited mutations of MYH. Lancet 2003;362:39-41.
82. Kastrinos F, Syngal S. Inherited colorectal cancer syndromes. Cancer J 2011;17:405-15.
83. Cleary SP, Cotterchio M, Jenkins MA, et al. Germline MutY human homologue mutations and colorectal cancer: a multisite case-control study. Gastroenterology 2009;136:1251-60.
84. Slupska MM, Baikalov C, Luther WM, et al. Cloning and sequencing a human homolog (hMYH) of the Escherichia coli mutY gene whose function is required for the repair of oxidative DNA damage. J Bacteriol 1996;178:3885-92.
85. Slupska MM, Luther WM, Chiang JH, et al. Functional expression of hMYH, a human homolog of the Escherichia coli MutY protein. J Bacteriol 1999;181:6210-3.
86. Cadet J, Davies KJA. Oxidative DNA damage & repair: An introduction. Free Radic Biol Med 2017;107:2-12.

87. Chen J, Huang Z, Wu X, et al. Oxidative stress induces different tissue dependent effects on Mutyh-deficient mice. Free Radic Biol Med 2019;143:482-493.
88. Rashid M, Fischer A, Wilson CH, et al. Adenoma development in familial adenomatous polyposis and MUTYH-associated polyposis: somatic landscape and driver genes. J Pathol 2016;238:98-108.
89. Pilati C, Shinde J, Alexandrov LB, et al. Mutational signature analysis identifies MUTYH deficiency in colorectal cancers and adrenocortical carcinomas. J Pathol 2017;242:10-15.
90. 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.
91. Lipton L, Halford SE, Johnson V, et al. Carcinogenesis in MYH-associated polyposis follows a distinct genetic pathway. Cancer Res 2003;63:7595-9.
92. 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.
93. 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.
94. Win AK, Hopper JL, Jenkins MA. Association between monoallelic MUTYH mutation and colorectal cancer risk: a meta-regression analysis. Fam Cancer 2011;10:1-9.
95. Win AK, Cleary SP, Dowty JG, et al. Cancer risks for monoallelic MUTYH mutation carriers with a family history of colorectal cancer. Int J Cancer 2011;129:2256-62.
96. Win AK, Dowty JG, Cleary SP, et al. Risk of colorectal cancer for carriers of mutations in MUTYH, with and without a family history of cancer. Gastroenterology 2014;146:1208-11. e1-5.
97. Win AK, Reece JC, Dowty JG, et al. Risk of extracolonic cancers for people with biallelic and monoallelic mutations in MUTYH. Int J Cancer 2016;139:1557-63.
98. 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.
99. 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.
100. Spier I, Holzapfel S, Altmuller J, et al. Frequency and phenotypic spectrum of germline mutations in POLE and seven other polymerase genes in 266 patients with colorectal adenomas and carcinomas. Int J Cancer 2015;137:320-31.
101. Chubb D, Broderick P, Frampton M, et al. Genetic diagnosis of high-penetrance susceptibility for colorectal cancer (CRC) is achievable for a high proportion of familial CRC by exome sequencing. J Clin Oncol 2015;33:426-32.
102. Bellido F, Pineda M, Aiza G, et al. POLE and POLD1 mutations in 529 kindred with familial colorectal cancer and/or polyposis: review of reported cases and recommendations for genetic testing and surveillance. Genet Med 2016;18:325-32.
103. Church DN, Briggs SE, Palles C, et al. DNA polymerase ε and δ exonuclease domain mutations in endometrial cancer. Hum Mol Genet 2013;22:2820-8.
104. Rohlin A, Zagoras T, Nilsson S, et al. A mutation in POLE predisposing to a multi-tumour phenotype. Int J Oncol 2014;45:77-81.
105. Hansen MF, Johansen J, Bjornevoll I, et al. A novel POLE mutation associated with cancers of colon, pancreas, ovaries and small intestine. Fam Cancer 2015;14:437-48.
106. Aoude LG, Heitzer E, Johansson P, et al. POLE mutations in families predisposed to cutaneous melanoma. Fam Cancer 2015;14:621-8.
107. Mur P, García-Mulero S, Del Valle J, et al. Role of POLE and POLD1 in familial cancer. Genet Med 2020.

108. Buchanan DD, Stewart JR, Clendenning M, et al. Risk of colorectal cancer for carriers of a germ-line mutation in POLE or POLD1. Genet Med 2018;20:890-895.
109. Nick McElhinny SA, Gordenin DA, Stith CM, et al. Division of labor at the eukaryotic replication fork. Mol Cell 2008;30:137-44.
110. Jansen AM, van Wezel T, van den Akker BE, et al. Combined mismatch repair and POLE/POLD1 defects explain unresolved suspected Lynch syndrome cancers. Eur J Hum Genet 2016;24:1089-92.
111. Campbell BB, Light N, Fabrizio D, et al. Comprehensive Analysis of Hypermutation in Human Cancer. Cell 2017;171:1042-1056.e10.
112. Kane DP, Shcherbakova PV. A common cancer-associated DNA polymerase ε mutation causes an exceptionally strong mutator phenotype, indicating fidelity defects distinct from loss of proofreading. Cancer Res 2014;74:1895-901.
113. Alexandrov LB, Kim J, Haradhvala NJ, et al. The repertoire of mutational signatures in human cancer. Nature 2020;578:94-101.
114. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. Nature 2013;500:415-21.
115. 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.
116. Rivera B, Castellsague E, Bah I, et al. Biallelic NTHL1 Mutations in a Woman with Multiple Primary Tumors. N Engl J Med 2015;373:1985-6.
117. Chubb D, Broderick P, Dobbins SE, et al. Rare disruptive mutations and their contribution to the heritable risk of colorectal cancer. Nat Commun 2016;7:11883.
118. Belhadj S, Mur P, Navarro M, et al. Delineating the Phenotypic Spectrum of the NTHL1-Associated Polyposis. Clin Gastroenterol Hepatol 2017;15:461-462.
119. Fostira F, Kontopodis E, Apostolou P, et al. Extending the clinical phenotype associated with biallelic NTHL1 germline mutations. Clin Genet 2018;94:588-589.
120. Groves A, Gleeson M, Spigelman AD. NTHL1-associate polyposis: first Australian case report. Fam Cancer 2019.
121. Belhadj S, Quintana I, Mur P, et al. NTHL1 biallelic mutations seldom cause colorectal cancer, serrated polyposis or a multi-tumor phenotype, in absence of colorectal adenomas. Sci Rep 2019;9:9020.
122. Altaraihi M, Gerdes AM, Wadt K. A new family with a homozygous nonsense variant in NTHL1 further delineated the clinical phenotype of NTHL1-associated polyposis. Hum Genome Var 2019;6:46.
123. Krokan HE, Bjørås M. Base excision repair. Cold Spring Harb Perspect Biol 2013;5:a012583.
124. 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.
125. Nik-Zainal S, Davies H, Staaf J, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature 2016;534:47-54.
126. 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.
127. Carethers JM, Koi M, Tseng-Rogenski SS. EMAST is a Form of Microsatellite Instability That is Initiated by Inflammation and Modulates Colorectal Cancer Progression. Genes (Basel) 2015;6:185-205.
128. Wimmer K, Etzler J. Constitutional mismatch repair-deficiency syndrome: have we so far seen only the tip of an iceberg? Hum Genet 2008;124:105-22.

129. Bakry D, Aronson M, Durno C, et al. Genetic and clinical determinants of constitutional mismatch repair deficiency syndrome: report from the constitutional mismatch repair deficiency consortium. Eur J Cancer 2014;50:987-96.
130. Jasperson KW, Samowitz WS, Burt RW. Constitutional mismatch repair-deficiency syndrome presenting as colonic adenomatous polyposis: clues from the skin. Clin Genet 2011;80:394-7.
131. Levi Z, Kariv R, Barnes-Kedar I, et al. The gastrointestinal manifestation of constitutional mismatch repair deficiency syndrome: from a single adenoma to polyposis-like phenotype and early onset cancer. Clin Genet 2015;88:474-8.
132. Aronson M, Gallinger S, Cohen Z, et al. Gastrointestinal Findings in the Largest Series of Patients With Hereditary Biallelic Mismatch Repair Deficiency Syndrome: Report from the International Consortium. Am J Gastroenterol 2016;111:275-84.
133. East JE, Vieth M, Rex DK. Serrated lesions in colorectal cancer screening: detection, resection, pathology and surveillance. Gut 2015;64:991-1000.
134. Patel R, Hyer W. Practical management of polyposis syndromes. Frontline Gastroenterol 2019;10:379-387.
135. Stanich PP, Pearlman R. Hereditary or Not? Understanding Serrated Polyposis Syndrome. Curr Treat Options Gastroenterol 2019;17:692-701.
136. 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.
137. Taupin D, Lam W, Rangiah D, et al. A deleterious RNF43 germline mutation in a severely affected serrated polyposis kindred. Hum Genome Var 2015;2:15013.
138. Yan HHN, Lai JCW, Ho SL, et al. RNF43 germline and somatic mutation in serrated neoplasia pathway and its association with BRAF mutation. Gut 2017;66:1645-1656.
139. Quintana I, Mejías-Luque R, Terradas M, et al. Evidence suggests that germline RNF43 mutations are a rare cause of serrated polyposis. Gut 2018;67:2230-2232.
140. Buchanan DD, Clendenning M, Zhuoer L, et al. Lack of evidence for germline RNF43 mutations in patients with serrated polyposis syndrome from a large multinational study. Gut 2017;66:1170-1172.
141. Thomas HJ, Whitelaw SC, Cottrell SE, et al. Genetic mapping of hereditary mixed polyposis syndrome to chromosome 6q. Am J Hum Genet 1996;58:770-6.
142. Whitelaw SC, Murday VA, Tomlinson IP, et al. Clinical and molecular features of the hereditary mixed polyposis syndrome. Gastroenterology 1997;112:327-34.
143. Rozen P, Samuel Z, Brazowski E. A prospective study of the clinical, genetic, screening, and pathologic features of a family with hereditary mixed polyposis syndrome. Am J Gastroenterol 2003;98:2317-20.
144. Jaeger E, Leedham S, Lewis A, et al. Hereditary mixed polyposis syndrome is caused by a 40-kb upstream duplication that leads to increased and ectopic expression of the BMP antagonist GREM1. Nat Genet 2012;44:699-703.
145. Davis H, Irshad S, Bansal M, et al. Aberrant epithelial GREM1 expression initiates colonic tumorigenesis from cells outside the stem cell niche. Nat Med 2015;21:62-70.
146. Lieberman S, Walsh T, Schechter M, et al. Features of Patients With Hereditary Mixed Polyposis Syndrome Caused by Duplication of GREM1 and Implications for Screening and Surveillance. Gastroenterology 2017;152:1876-1880.e1.
147. Rohlin A, Eiengård F, Lundstam U, et al. GREM1 and POLE variants in hereditary colorectal cancer syndromes. Genes Chromosomes Cancer 2016;55:95-106.
148. McKenna DB, Van Den Akker J, Zhou AY, et al. Identification of a novel GREM1 duplication in a patient with multiple colon polyps. Fam Cancer 2019;18:63-66.

149. Venkatachalam R, Verwiel ET, Kamping EJ, et al. Identification of candidate predisposing copy number variants in familial and early-onset colorectal cancer patients. Int J Cancer 2011;129:1635-42.
150. Gammon A, Jasperson K, Kohlmann W, et al. Hamartomatous polyposis syndromes. Best Pract Res Clin Gastroenterol 2009;23:219-31.
151. Manfredi M. Hereditary hamartomatous polyposis syndromes: understanding the disease risks as children reach adulthood. Gastroenterol Hepatol (N Y) 2010;6:185-96.
152. Macken WL, Tischkowitz M, Lachlan KL. PTEN Hamartoma tumor syndrome in childhood: A review of the clinical literature. Am J Med Genet C Semin Med Genet 2019;181:591-610.
153. Marsh DJ, Kum JB, Lunetta KL, et al. PTEN mutation spectrum and genotype-phenotype correlations in Bannayan-Riley-Ruvalcaba syndrome suggest a single entity with Cowden syndrome. Hum Mol Genet 1999;8:1461-72.
154. Smpokou P, Fox VL, Tan WH. PTEN hamartoma tumour syndrome: early tumour development in children. Arch Dis Child 2015;100:34-7.
155. Hobert JA, Eng C. PTEN hamartoma tumor syndrome: an overview. Genet Med 2009;11:687-94.
156. Gammon A, Jasperson K, Champine M. Genetic basis of Cowden syndrome and its implications for clinical practice and risk management. Appl Clin Genet 2016;9:83-92.
157. Czene K, Lichtenstein P, Hemminki K. Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. Int J Cancer 2002;99:260-6.
158. Mucci LA, Hjelmborg JB, Harris JR, et al. Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. Jama 2016;315:68-76.
159. Graff RE, Möller S, Passarelli MN, et al. Familial Risk and Heritability of Colorectal Cancer in the Nordic Twin Study of Cancer. Clin Gastroenterol Hepatol 2017;15:1256-1264.
160. Lynch HT, de la Chapelle A. Hereditary colorectal cancer. N Engl J Med 2003;348:919-32.
161. Huyghe JR, Bien SA, Harrison TA, et al. Discovery of common and rare genetic risk variants for colorectal cancer. Nat Genet 2019;51:76-87.
162. Mongin C, Coulet F, Lefevre JH, et al. Unexplained polyposis: a challenge for geneticists, pathologists and gastroenterologists. Clin Genet 2012;81:38-46.
163. Buchanan DD, Rosty C, Clendenning M, et al. Clinical problems of colorectal cancer and endometrial cancer cases with unknown cause of tumor mismatch repair deficiency (suspected Lynch syndrome). Appl Clin Genet 2014;7:183-93.
164. Morak M, Heidenreich B, Keller G, et al. Biallelic MUTYH mutations can mimic Lynch syndrome. Eur J Hum Genet 2014;22:1334-7.
165. Pearlman R, Frankel WL, Swanson B, et al. Prevalence and Spectrum of Germline Cancer Susceptibility Gene Mutations Among Patients With Early-Onset Colorectal Cancer. JAMA Oncol 2017;3:464-471.
166. Castillejo A, Vargas G, Castillejo MI, et al. Prevalence of germline MUTYH mutations among Lynch-like syndrome patients. Eur J Cancer 2014;50:2241-50.
167. Zetner DB, Bisgaard ML. Familial Colorectal Cancer Type X. Curr Genomics 2017;18:341-359.
168. Weigl K, Chang-Claude J, Hsu L, et al. Establishing a valid approach for estimating familial risk of cancer explained by common genetic variants. Int J Cancer 2020;146:68-75.
169. Patel SG, Boland CR. Colorectal Cancer in Persons Under Age 50: Seeking Causes and Solutions. Gastrointest Endosc Clin N Am 2020;30:441-455.
170. de Voer RM, Geurts van Kessel A, Weren RD, et al. Germline mutations in the spindle assembly checkpoint genes BUB1 and BUB3 are risk factors for colorectal cancer. Gastroenterology 2013;145:544-7.

171. de Voer RM, Hahn MM, Weren RD, et al. Identification of Novel Candidate Genes for Early-Onset Colorectal Cancer Susceptibility. PLoS Genet 2016;12:e1005880.
172. Weren RD, Venkatachalam R, Cazier JB, et al. Germline deletions in the tumour suppressor gene FOCAD are associated with polyposis and colorectal cancer development. J Pathol 2015;236:155-64.
173. Clarke E, Green RC, Green JS, et al. Inherited deleterious variants in GALNT12 are associated with CRC susceptibility. Hum Mutat 2012;33:1056-8.
174. Evans DR, Venkitachalam S, Revoredo L, et al. Evidence for GALNT12 as a moderate penetrance gene for colorectal cancer. Hum Mutat 2018;39:1092-1101.
175. Schubert SA, Ruano D, Elsayed FA, et al. Evidence for genetic association between chromosome 1q loci and predisposition to colorectal neoplasia. Br J Cancer 2017;117:1215-1223.
176. Venkatachalam R, Ligtenberg MJ, Hoogerbrugge N, et al. Germline epigenetic silencing of the tumor suppressor gene PTPRJ in early-onset familial colorectal cancer. Gastroenterology 2010;139:2221-4.
177. Golubicki M, Bonjoch L, Acuña-Ochoa JG, et al. Germline biallelic Mcm8 variants are associated with early-onset Lynch-like syndrome. JCI Insight 2020;5.
178. Broderick P, Dobbins SE, Chubb D, et al. Validation of Recently Proposed Colorectal Cancer Susceptibility Gene Variants in an Analysis of Families and Patients-a Systematic Review. Gastroenterology 2017;152:75-77.e4.

Chapter 2

Germline variants in *POLE* are associated with early onset mismatch repair deficient colorectal cancer

Fadwa A. Elsayed, C. Marleen Kets, Dina Ruano, Brendy van den Akker, Arjen R. Mensenkamp, Melanie Schrumpf, Maartje Nielsen, Juul T. Wijnen, Carli M. Tops, Marjolijn J. Ligtenberg, Hans F.A. Vasen, Frederik J. Hes, Hans Morreau and Tom van Wezel

European Journal of Human Genetics, 2015; 23(8): 1080-1084

Abstract

Germline variants affecting the exonuclease domains of *POLE* and *POLD1* predispose to multiple colorectal adenomas and/or colorectal cancer (CRC). The aim of this study was to estimate the prevalence of previously described heterozygous germline variants *POLE* c.1270C>G, p.(Leu424Val) and *POLD1* c.1433G>A, p.(Ser478Asn) in a Dutch series of unexplained familial, early onset CRC and polyposis index cases. We examined 1188 familial CRC and polyposis index patients for *POLE* p.(Leu424Val) and *POLD1* p.(Ser478Asn*)* variants using competitive allele-specific PCR. In addition, protein expression of the *POLE* and DNA mismatch repair genes was studied by immunohistochemistry in tumours from *POLE* carriers. Somatic mutations were screened using semiconductor sequencing. We detected three index patients (0.25%) with a *POLE* p.(Leu424Val*)* variant. In one patient, the variant was found to be de-novo. Tumours from three patients from two families were microsatellite instable, and immunohistochemistry showed MSH6/MSH2 deficiency suggestive of Lynch syndrome. Somatic mutations but no germline *MSH6* and *MSH2* variants were subsequently found, and one tumour displayed a hypermutator phenotype. None of the 1188 patients carried the *POLD1* p.(Ser478Asn) variant. *POLE* germline variant carriers are also associated with a microsatellite instable CRC. *POLE* DNA analysis now seems warranted in microsatellite instable CRC, especially in the absence of a causative DNA mismatch repair gene germline variant.

Introduction

Faithful DNA replication and the repair of errors are both essential for the maintenance of genomic stability and suppression of carcinogenesis [1]. Duplication of genomes with high accuracy is achieved through three mechanisms: the high selectivity of DNA polymerases, exonucleolytic proofreading; and post replication mismatch repair [2]. The DNA polymerases ε (POLε) and δ (POLδ) are required for efficient genome replication in the eukaryotic replication fork [3]. Their major component proteins, encoded by *POLE* and *POLD1*, respectively, possess an intrinsic 3'–5' proofreading domain that removes incorrectly inserted nucleotides during DNA synthesis [4-9]. Studies in yeast have shown that mutations in the proofreading domains of POLε or POLδ increase spontaneous mutation rates [8, 9]. In addition, somatic mutations in the proofreading domains of *POLD1* and *POLE* have been identified in microsatellite instable (MSI) and hypermutated subgroups of colorectal cancers (CRCs) [10-12].

Recently, Palles *et al* reported that heterozygous germline variants in the proofreading domain of the DNA polymerases *POLE* and *POLD1* predispose, with a high penetrance, to multiple colorectal adenomas, early onset CRC (OMIM #114500) and endometrial cancer (OMIM #608089). These variants were found by whole-genome sequencing and linkage analysis in three large families with a dominant pattern of CRC and multiple adenomas [13]. Subsequent screening of 3805 CRC patients revealed that these variants are relatively rare: *POLE* p.(Leu424Val) was found 12 times, and *POLD1* p.(Ser478Asn) only once, in patients with a positive family history of adenomas or CRC. The tumours seen in *POLE* and *POLD1* carriers were microsatellite stable and showed a hypermutator phenotype [13]. Valle *et. al* [14] detected a single *POLE* p.(Leu424Val) variant in a screen of 858 familial/early onset CRC and polyposis patients.

The goal of our study was to estimate the prevalence of germline variants in *POLE* and *POLD1* in a Dutch series of unexplained familial, early onset CRC and polyposis index cases. In addition, we analysed phenotypes and tumour characteristics in this patient series.

Materials and methods

Samples

DNA from index patients with colorectal polyposis [15] and familial CRC [16] was analysed for *POLE* NM_006231.2:c.1270C>G, p.(Leu424Val) and *POLD1*

NM_002691.1:c.1433G>A, p.(Ser478Asn). Leukocyte DNA from 485 polyposis cases was included. These patients had developed ≥10 colorectal adenomas and had been previously tested negative for *APC* and *MUTYH* germline mutations at the Laboratory for Diagnostic Genome Analysis in Leiden. Clinical data were collected from the Netherlands Foundation for the Detection of Hereditary Tumours and from clinical genetics departments in the Netherlands [17]. The familial CRC cohort comprised 703 patients, mainly from the south-western region of the Netherlands, with most cases submitted by clinical genetics departments. These patients met clinical criteria for MSI testing, which are based on early onset of disease and/or familial clustering of CRC and associated cancers, corresponding to the revised Bethesda criteria. Samples were collected between 1997 and 2013, and DNA for this cohort was available from peripheral blood (340 cases) or from formalin-fixed paraffin embedded normal mucosa (363 cases). These samples were described before, only DNA that passes quality check was included in the study [16]. The study was approved by the local medical ethical committee of the Leiden University Medical Center (P01-019).

Genotyping

POLE p.(Leu424Val) and *POLD1* p.(Ser478Asn) were tested using the competitive allele-specific PCR (KASPar) assay, following the manufacturer's protocol (LGC Genomics, Berlin, Germany). The primers were designed using Primerpicker (KBioscience, Hoddesdon, UK). The following primers were used to analyse *POLE* c.1270C>G: POLE_L424V_C1; *5'*- GGA TCA TAG CCT AGC TTG GCC TT-*3',* POLE_L424V_A2; 5'-GAA GGT CGG AGT CAA CGG ATT CCT TCC TGT GGG CAG TCA TAA TG-3' and POLE_L424V_A1; 5'-GAA GGT GAC CAA GTT CAT GCT CCT TCC TGT GGG CAG TCA TAA TC-3'. For *POLD1* c.1433G>A, we used: POLD1_S478N_C2; 5'–TCT GCT CGC CCA GGA AGT GGA A-3', POLD1_S478N_A2; 5' - GAA GGT CGG AGT CAA CGG ATT CCT ACA CGC TCA ATG CCG TGA A-3' and POLD1_S478N_A1; 5'-GAA GGT GAC CAA GTT CAT GCT ACA CGC TCA ATG CCG TGA G-3'.

Variants were identified using the CFX manager software v3.0 (Bio-Rad, Veenendaal, the Netherlands). Formalin-fixed paraffin embedded and leukocyte DNA samples were genotyped in separate experiments for accurate genotyping results. Samples positive for *POLE* c.1270C>G, p.(Leu424Val) were subsequently validated by Sanger sequencing of leukocyte DNA and of DNA extracted from formalin-fixed paraffin embedded tissues, using both normal and tumour DNA where available. Sanger sequencing was performed by Macrogen (Amsterdam, the Netherlands). The following primers, with universal M13 tails (upper case), were used for *POLE* c.1270C>G; forward: 5'-TGT AAA ACG ACG GCC AGT cca tct

gga tgc gtg cac a-3' and reverse: 5'-CAG GAA ACA GCT ATG ACC gaa tca tcc tgg ctt ctg ttc tca- 3'. For *POLD1* validation we used the oligonucleotides, forward: 5'-TGT AAA ACG ACG GCC AGT ctg tcc ttg gaa ggc cact-3' and reverse: 5'-CAG GAA ACA GCT ATG ACC gag gtc agg gag gca gca-3'. Sequencing primers were designed using Primer3 software (http://primer3.wi.mit.edu/) and all oligonucleotides were manufactured by IDT (Leuven, Belgium).

The *POLE* p.(Leu424Val) carriers were submitted to the LOVD database http://databases.lovd.nl/shared/genes/POLE, IDs 00019773 (PT1), 00019821 (PT2) 00019822 (PT3) and 00019824 (PT4).

Immunohistochemistry

Immunohistochemistry (IHC) for the DNA MMR proteins was performed as previously described [18]. POLE IHC was performed using 4 µm thick, formalin-fixed paraffin embedded tissue sections on StarFrost adhesive slides (Waldemar Knittel, Braunschweig, Germany), dried overnight at 37°C. Tissue sections were deparaffinised three times in xylene for a total of 15 min and subsequently rehydrated with 100% ethanol three times for a total of 9 min. Endogenous peroxidase activity was blocked by immersing the tissue sections in a 0.3% solution of hydrogen peroxide (H_2O_2) and methanol at room temperature for 20 min. This treatment was followed by hydration in a graded ethanol series to distilled water. Antigen retrieval was then performed by boiling the tissue sections in 0.01M citrate buffer (pH 6.0) for 10 min using a microwave oven, after which the tissue sections were allowed to cool in the same buffer to room temperature. The tissue sections were then washed twice with distilled water, for 5 min each time, followed by two 5 min washes in 1x phosphate buffered saline. This was followed by the addition of the primary antibody (ab110876, ABCAM, Cambridge UK, dilution 1:1600) and the incubation of the covered slides overnight in a humid chamber at room temperature.

After three 5 min washes in 1x phosphate buffered saline the next morning, the tissue sections were incubated for 30 min with horseradish peroxidase-conjugated multimer secondary antibody (PowerVision Poly/HRP, DAKO, Heverlee, Belgium). The slides then received three 5 min rinses in 1x phosphate buffered saline and the antigen-antibody reaction was visualized using 3,3'-diaminobenzidine in chromogenic solution, applied for 10 minutes. The tissue sections were subsequently rinsed in distilled water and then stained with haematoxylin for 20 sec. The tissue sections were rinsed in running tap water for 5 min and washed twice with distilled water, followed by dehydration in a

graded ethanol series (50, 70 and 100%) and cleared in xylene. Finally, the tissue sections were dried, mounted and covered for microscopic analysis. Tissue stroma and normal epithelium or lymph follicles served as positive internal controls when analysing POLE, MLH1, PMS2, MSH2, and MSH6 expression.

Somatic mutation screening

DNA from normal and tumour tissues of the *POLE* p.(Leu424Val) carriers PT1 and PT2 was screened for somatic mutations using a custom AmpliSeq panel (Life Technologies Europe, Bleiswijk, the Netherlands) of the DNA mismatch repair (MMR) genes *MLH1*, *MSH2*, *MSH6*, *PMS2*, supplemented with *APC* and *TP53*. The panel was sequenced using the Ion PGM system at Leiden Genome Technology Center, the Netherlands (www.lgtc.nl) and analysed with the Nextgene Software package (Softgenetics, State College, PA, USA). MMR DNA sequencing was performed as previously described in the case of PT3 [19]. Somatic *KRAS* and *BRAF* hotspot mutations (*KRAS* exon 2 and 3 and *BRAF* p.V600E were tested as described previously [20]. Somatic mutations are submitted to the COSMIC database (ID # 616).

Methylation

MSH2 methylation was studied as described earlier using the MSP1 and MSP3 primer sets for both methylated and unmethylated DNA [21]. DNA was bisulfite treated using the EZ DNA methylation KIT™ (ZYMO Research Corp, Irvine, CA, USA).

Results and discussion

We assessed the prevalence of *POLE* p.(Leu424Val) and *POLD1* p.(Ser478Asn) in a cohort of 1188 Dutch index patients with multiple polyps or familial CRC. All patients were originally referred for possible FAP (OMIM #175100), MAP (OMIM #608456) or Lynch syndrome (OMIM #120435) and no variants were found in the relevant genes. Although we did not detect *POLD1* p.(Ser478Asn), three index patients with the *POLE* p.(Leu424Val) variant were identified. These patients developed multiple colorectal adenomas, of whom two showed early onset CRC. We can therefore confirm that *POLE* variants play a minor but tangible role in familial CRC and polyposis. Despite an enrichment in our cohort for inherited CRC and polyposis, the frequency (3 in 1188; 0.25%) is also comparable to the currently reported frequency [13, 14].

The first patient (PT1) was diagnosed with a MSI caecum tumour and two adenomas at age 40. She developed, cumulatively, around 30 polyps and presented with a microsatellite stable endometrial cancer at age 50. Her son, patient 2 (PT2), is also a carrier of *POLE* p.(Leu424Val). In addition, he was clinically diagnosed with neurofibromatosis type I, the latter originating in his father's family. He developed a MSI caecum adenocarcinoma at age 30 and was diagnosed with an astrocytoma at age 15, presumably as a consequence of neurofibromatosis. Interestingly, a patient with multiple polyps, and two astrocytomas at age 26, has been described in a family with a *POLD1* variant (family SM6) [13].

As tumour tissue was available for the *POLE* p.(Leu424Val) carriers PT1 and PT2 (mother and son respectively), we performed IHC for POLE, MLH1, MSH2, MSH6 and PMS2. Considering that the *POLE* exonuclease domain is essential for the maintenance of replication fidelity, the lack of predictive value of POLE IHC is to be expected, in contrast to the predictive value for the MMR genes. However, the availability of tumour tissue from the *POLE* variant carriers allowed us to test this. Indeed, both MSI-H tumour tissues showed a positive nuclear POLE protein expression in tumour cells as compared with normal cells from the same tissue. POLE IHC therefore does not appear to be predictive for the effect of the variant.

The MSI-H caecum tumour from PT1 showed loss of only MSH6 protein expression in tumour cells (Figure 1), whereas the adenocarcinoma from PT2 showed loss of both MSH2 and MSH6 protein expression in tumour cells (Figure 1). The loss of nuclear expression of MSH6 and MSH6/MSH2 in the tumours from this family is indicative for Lynch syndrome because of an inherited *MSH6* or *MSH2* germline variant. However, no germline variants were found in *MSH6* or *MSH2*, nor in any of the other MMR genes. We therefore screened PT1 and PT2 tumours for somatic mutations in *MLH1*, *MSH2*, *MSH6*, *PMS2*, and also for somatic mutations in *APC* and *TP53* (Table 1). The mother's tumour (PT1) showed a hypermutator phenotype, with multiple somatic mutations in the MMR genes and in *APC* and *TP53,* whereas the adenocarcinoma from PT2 showed no *APC* or *TP53* mutations. Interestingly, PT1 showed three somatic mutations in *MSH6.* The first is a pathogenic truncating mutation, c.2629G>T, p.(Glu877*), the second is a possibly pathogenic missense mutation, c.4000C>T, p.(Arg1334Trp) that might affect the splice donor site in exon 9, while the third mutation is probably non-pathogenic, c.3725G>A, p.(Arg1242His). A single truncating *MSH2* mutation, c.643C>T, p.(Gln215*), was found in the son's tumour (PT2, Table 1), although a second hit in *MSH2* was not detected. No somatic

mutations were detected in *KRAS* and *BRAF* hotspots in the tumours from PT1 and PT2. These somatic mutations show an excess of G:C to T:A substitutions, similar to the reported EDM-associated mutator phenotype [13]. No other somatic mutations or previously undetected germline mutations were found. In light of the hypermutator phenotype associated with *POLE* mutations, a second point mutation was to be expected. On the other hand, promoter methylation or loss of heterozygosity at the MMR locus as the second hit could explain the loss of MSH2 or MSH6. Recently it was demonstrated that the loss of heterozygosity accounts for the second hit in over 50% of MSI tumours, albeit mostly in MLH1 deficient tumours [19].

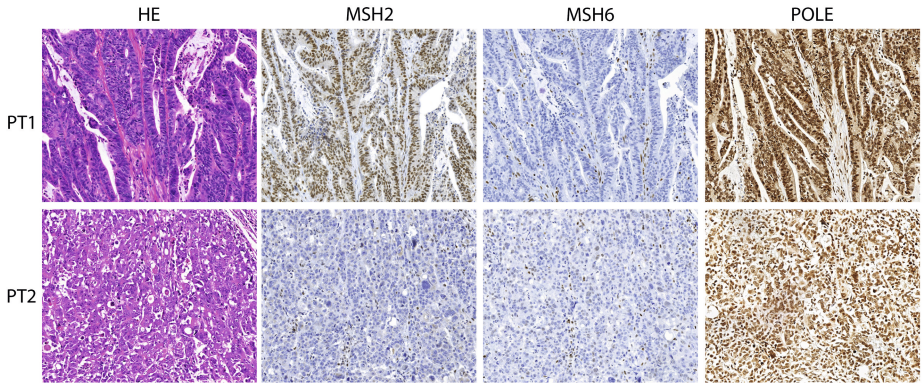


Figure 1. HE staining and MSH2, MSH6 and POLE immunohistochemistry (20x) of the MSI-H tumours from *POLE* variant carriers PT1 and PT2 (one family).

Both cases show MSH6 negative staining, with positive stromal cells (brown). PT2 also shows loss of MSH2 in tumour cells.

Table 1. Summary of somatic mutations in colorectal cancer genes in two tumours from *POLE* p.(Leu424Val) carriers (PT1 and PT2) from one family

Case	Gene.refGene	DNA Level (cDNA)	Protein	Chr	Start	End	Ref	Alt	ExonicFunc.refGene	Pathogenic
PT1	MSH6	NM_000179.2:c.2629G>T	p.Glu877*	2	48027751	48027751	G	T	Stopgain SNV	Truncating
PT1	MSH6	NM_000179.2:c.2291C>T	p.Thr764Ile	2	48027413	48027413	C	T	Non-synonymous SNV	Unlikely pathogenic
PT1	MSH6	NM_000179.2:c.3725G>A	p.Arg1242His	2	48033421	48033421	G	A	Non-synonymous SNV	Unknown pathogenicity
PT1	MSH6	NM_000179.2:c.4000C>T	p.Arg1334Trp	2	48033789	48033789	C	T	Non-synonymous SNV	Unknown pathogenicity
PT1	MSH2	NM_000251.2:c.49G>A	p.Val17Ile	2	47630379	47630379	G	A	Non-synonymous SNV	Unlikely pathogenic
PT1	MLH1	NM_000249.2:c.31C>A	p.Leu11Met	3	37035069	37035069	C	A	Non-synonymous SNV	Unknown pathogenicity
PT1	APC	NM_000038.5:c.680A>G	p.Asp227Gly	5	112128177	112128177	A	G	Non-synonymous SNV	Unknown pathogenicity
PT1	APC	NM_000038.5:c.1778G>A	p.Trp593*	5	112170682	112170682	G	A	Stopgain SNV	Truncating
PT1	APC	NM_000038.5:c.2662G>A	p.Ala888Thr	5	112173953	112173953	G	A	Non-synonymous SNV	Unlikely pathogenic
PT1	APC	NM_000038.5:c.4540C>T	p.Pro1514Ser	5	112175831	112175831	C	T	Non-synonymous SNV	Unlikely pathogenic
PT1	APC	NM_000038.5:c.5117C>T	p.Ser1706Leu	5	112176408	112176408	C	T	Non-synonymous SNV	Unlikely pathogenic
PT1	APC	NM_000038.5:c.8314T>C	p.Ser2772Pro	5	112179605	112179605	T	C	Non-synonymous SNV	Unlikely pathogenic
PT1	TP53	NM_000546.5:c.523C>T	p.Arg175Cys	17	7578407	7578407	G	A	Non-synonymous SNV	Likely pathogenic
PT1	TP53	NM_000546.5:c.742C>T	p.Arg248Trp	17	7577539	7577539	G	A	Non-synonymous SNV	Likely pathogenic
PT1	TP53	NM_000546.5:c.916C>T	p.Arg306*	17	7577022	7577022	G	A	Stopgain SNV	Truncating
PT2	MSH2	NM_000251.2:c.643C>T	p.Gln215*	2	47637509	47637509	C	T	Stopgain SNV	Truncating
PT2	APC	NM_000038.5:c.2045G>A	p.Gly682Glu	5	112173336	112173336	G	A	Non-synonymous SNV	Likely pathogenic

No mutations were found in *KRAS*, *BRAF* and *PMS2*.

Patient 3 (PT3) was diagnosed with multiple polyps at age 34. Two of the patient's brothers were affected with CRC at ages 18 and 37, while a third brother developed liver cancer at age 27. The patient's mother died aged 31 because of unknown causes, probably cancer. Tumour tissue from an adenoma with early cancer and one adenoma with high-grade dysplasia was available for examination. The adenoma with cancer (estimated tumour percentage 50%) showed microsatellite instability and immunohistochemical loss of MSH2 and MSH6 protein staining in the absence of germline variants in *MSH2* and *MSH6*. A probably non-pathogenic heterozygous missense mutation, c.1550 C>T, p.(Ala517Val), in the *MSH2* gene was detected in the adenoma with early cancer but not in normal tissue. Testing for somatic mutations in other genes could not be performed because of the limited availability of tumour DNA. Hypermethylation of the *MSH2* promoter was not detected in the tumour or normal tissue of PT3. Microsatellite instability analysis of an adenoma with high-grade dysplasia, with an estimated percentage of cells with high-grade dysplasia of 65%, showed no microsatellite instability and normal staining of the DNA mismatch repair proteins MSH2 and MSH6, as well as MLH1 and PMS2. Tumour tissue and germline DNA from relatives of PT3 were not available.

Patient 4 (PT4) was diagnosed with a microsatellite stable colon cancer and polyposis at the age of 33. No tumour tissue was available for further analysis. Although a history of CRC was reported in both paternal and maternal branches of the family, the parents were not affected by polyposis and did not carry the *POLE* variant, indicating that the *POLE* variant in PT4 was probably de-novo. There was also no evidence for gonadal mosaicism in the parents; the sibling is not known to have polyposis. In a recent study, the same *POLE* variant was also detected as a de-novo occurrence in a patient with early onset CRC and polyposis [14].

Interestingly, the germline *POLE* variant in the currently studied families (PT1, PT2 and PT3) is associated with a Lynch syndrome phenotype with MSI tumours and MSH6 or MSH2/MSH6 protein loss. This contrasts with previously identified *POLE* and *POLD1* germline variant carriers who developed microsatellite stable tumours [13]. Somatic *POLE* mutations have been reported in both microsatellite stable and MSI tumours [10, 12]. The Lynch syndrome phenotype reported here is not because of the germline variants in the MMR genes but more likely the result of somatic inactivation (PT1 and PT2). The hypermutator phenotype associated with inherited *POLE* mutations suggests a causative role for the somatic mutations although the second hit could not be identified. A similar

co-occurrence of somatic MMR and *POLE* mutations was recently shown in a sequencing effort of 147 CRC genomes. Eight of the eleven *POLE*-mutated tumours showed additional *MSH6* somatic mutations, and of these, five cases also showed *MSH2* mutations [12]. Moreover, MSI tumours with two somatic *MSH2* mutations, lacking MSH2 and MSH6 protein expression, [22] or with loss of MLH1 protein staining in the tumour, [23-25] have also been reported for patients with bi-allelic variants in the base excision repair gene *MUTYH*. Similarly to *POLE* germline variants, *MUTYH* missense variants can also induce somatic mutations in MMR genes, although the mechanism behind the co-occurrence of mutations in the different DNA repair defects remains elusive.

Conclusions

The three new families with *POLE* variants reported here bring the total number of reported families to 17. We have demonstrated that *POLE* germline variants can give rise to a Lynch syndrome-like phenotype, with MSI-H tumours displaying negative IHC for one of the MMR genes. IHC for POLE is not helpful in identifying currently known variants. We now recommend that testing for *POLE* should be considered when screening unexplained MSI-H tumours, and while clinical surveillance of *POLE* carriers appears to be indicated from a relatively young age, further conclusions regarding clinical management should be based on a larger series than the currently identified patients.

References

1. Venkatesan RN, Treuting PM, Fuller ED, et al. Mutation at the polymerase active site of mouse DNA polymerase delta increases genomic instability and accelerates tumorigenesis. Mol Cell Biol 2007;27:7669-82.
2. Schaaper RM. Base selection, proofreading, and mismatch repair during DNA replication in Escherichia coli. J Biol Chem 1993;268:23762-5.
3. Garg P, Burgers PM. DNA polymerases that propagate the eukaryotic DNA replication fork. Crit Rev Biochem Mol Biol 2005;40:115-28.
4. Seshagiri S. The burden of faulty proofreading in colon cancer. Nat Genet 2013;45:121-2.
5. Preston BD, Albertson TM, Herr AJ. DNA replication fidelity and cancer. Semin Cancer Biol 2010;20:281-93.
6. McCulloch SD, Kunkel TA. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. Cell Res 2008;18:148-61.
7. Sugino A. Yeast DNA polymerases and their role at the replication fork. Trends Biochem Sci 1995;20:319-23.
8. Morrison A, Bell JB, Kunkel TA, et al. Eukaryotic DNA polymerase amino acid sequence required for 3'----5' exonuclease activity. Proc Natl Acad Sci U S A 1991;88:9473-7.
9. Simon M, Giot L, Faye G. The 3' to 5' exonuclease activity located in the DNA polymerase delta subunit of Saccharomyces cerevisiae is required for accurate replication. Embo j 1991;10:2165-70.
10. Yoshida R, Miyashita K, Inoue M, et al. Concurrent genetic alterations in DNA polymerase proofreading and mismatch repair in human colorectal cancer. Eur J Hum Genet 2011;19:320-5.
11. Network CGA. Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012;487:330.
12. Kim TM, Laird PW, Park PJ. The landscape of microsatellite instability in colorectal and endometrial cancer genomes. Cell 2013;155:858-68.
13. 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.
14. Valle L, Hernández-Illán 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.
15. Hes FJ, Nielsen M, Bik EC, et al. Somatic APC mosaicism: an underestimated cause of polyposis coli. Gut 2008;57:71-6.
16. 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.
17. 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.
18. de Jong AE, van Puijenbroek M, Hendriks Y, et al. Microsatellite instability, immunohistochemistry, and additional PMS2 staining in suspected hereditary nonpolyposis colorectal cancer. Clin Cancer Res 2004;10:972-80.
19. Mensenkamp AR, Vogelaar IP, van Zelst-Stams WA, et al. Somatic mutations in MLH1 and MSH2 are a frequent cause of mismatch-repair deficiency in Lynch syndrome-like tumors. Gastroenterology 2014;146:643-646.e8.
20. van Eijk R, Licht J, Schrumpf M, et al. Rapid KRAS, EGFR, BRAF and PIK3CA mutation analysis of fine needle aspirates from non-small-cell lung cancer using allele-specific qPCR. PLoS One 2011;6:e17791.

21. Chan TL, Yuen ST, Kong CK, et al. Heritable germline epimutation of MSH2 in a family with hereditary nonpolyposis colorectal cancer. Nat Genet 2006;38:1178-83.
22. Morak M, Heidenreich B, Keller G, et al. Biallelic MUTYH mutations can mimic Lynch syndrome. Eur J Hum Genet 2014;22:1334-7.
23. Cleary SP, Cotterchio M, Jenkins MA, et al. Germline MutY human homologue mutations and colorectal cancer: a multisite case-control study. Gastroenterology 2009;136:1251-60.
24. Colebatch A, Hitchins M, Williams R, et al. The role of MYH and microsatellite instability in the development of sporadic colorectal cancer. Br J Cancer 2006;95:1239-43.
25. Lefevre JH, Colas C, Coulet F, et al. MYH biallelic mutation can inactivate the two genetic pathways of colorectal cancer by APC or MLH1 transversions. Fam Cancer 2010;9:589-94.

Chapter 3

Low frequency of *POLD1* and *POLE* exonuclease domain variants in patients with multiple colorectal polyps

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

Molecular Genetics & Genomic Medicine, 2019; 7(4): e00603

Abstract

Background

Germline mutations affecting the exonuclease domains of *POLE* and *POLD1* predispose to colorectal adenomas and carcinoma. Here, we aimed to screen the exonuclease domains to find the genetic causes of multiple colorectal polyps in unexplained cases.

Methods

Using a custom next-generation sequencing panel, we sequenced the exonuclease domains of *POLE* and *POLD1* in 332 index patients diagnosed with multiple colorectal polyps without germline alteration in colorectal polyposis predisposing genes.

Results

We identified two variants of unknown significance. One germline *POLD1* c.961G>A, p.(Gly321Ser) variant was found in two cases. The first patient was diagnosed with multiple polyps at age 35 and colorectal cancer (CRC) at age 37, with no known family history of CRC. The second patient was diagnosed with CRC at age 44 and cumulatively developed multiple polyps; this patient had two sisters with endometrial cancer who did not carry the variant. Furthermore, we identified a novel *POLD1* c.955T>G, p.(Cys319Gly) variant in a patient diagnosed with multiple colorectal adenomas at age 40. Co-segregation analysis showed that one sister who cumulatively developed multiple adenomas from age 34, and another sister who developed CRC at age 38, did not carry the variant. We did not identify pathogenic variants in *POLE* and *POLD1*.

Conclusion

This study confirms the low frequency of causal variants in these genes in the predisposition for multiple colorectal polyps, and also establishes that these genes are a rare cause of the disease.

Introduction

The heritable component of colorectal cancer (CRC) is approximately 35% [1], yet only up to 5% is associated with well-characterized hereditary syndromes [2], which are caused by germline mutations in known high-penetrance CRC genes [3-7]. A substantial proportion of CRC syndromes have been associated with malfunctioning DNA repair pathways [4, 8-10]. Recently, an autosomal dominant CRC syndrome, caused by monoallelic germline mutations in the exonuclease domains of *POLE* (OMIM #174762) and *POLD1* (OMIM #174761), was identified. Palles *et al.* reported heterozygous germline variants in the exonuclease domains of the DNA polymerases *POLE* and *POLD1. These variants* predispose individuals, with a high penetrance, to multiple colorectal adenomas and early-onset CRC. These mutations were found by whole-genome sequencing and linkage analysis in three large families that each possess a dominant pattern of CRC and multiple adenomas. Subsequent screening of 3,805 CRC patients revealed that these variants are relatively rare in patients with a positive family history of adenomas or CRC: *POLE* p.(Leu424Val) was found 12 times and *POLD1* p.(Ser478Asn) only once. The tumors from the carriers were microsatellite stable and showed a hypermutated phenotype [6]. Further studies have shown that *POLE* and *POLD1* mutations predispose individuals to multiple adenomas and early-onset CRC [11-17]. The germline *POLE* p.(Leu424Val) mutation was demonstrated to also be associated with a Lynch syndrome-like phenotype with microsatellite instable (MSI) tumors and somatic MSH6 or MSH2/MSH6 protein loss in the tumors [12]. Additionally, *POLE* and *POLD1* somatic mutations can give rise to a Lynch syndrome-like phenotype and microsatellite instable colorectal cancer [18]. To discover the underlying genetic causes of multiple colorectal polyps and CRC in genetically unexplained cases, we aimed to screen the exonuclease domains of *POLE* and *POLD1* in this group of patients. Generally, identification of new heritable risk factors may play a role in increasing the understanding of mechanisms underlying multiple polyp initiation and assist in the implementation of preventive strategies.

Material and methods

Samples

Ethical compliance; the study was approved by the local medical ethics committee of the Leiden University Medical Center (P01-019). A total of 332 index patients with multiple colorectal polyps were included in this study. All patients were originally referred to the Laboratory for Diagnostic Genome

Analysis in Leiden for possible FAP or MAP syndrome; no potentially pathogenic germline variants were found in the entire genes *APC* (OMIM #611731) and *MUTYH (*OMIM #604933). More recently, the patients had tested negative for *POLE NM_006231.2:c.1270C>G, p.(Leu424Val)* and *POLD1* NM_002691.3:*c.*1433G>A, p.(Ser478Asn) variants [12] and for *NTHL1* (OMIM #602656) NM_002528.6:c.268C>T, p.(Gln90*). Clinical data were collected from the Netherlands Foundation for the Detection of Hereditary Tumors (NFDHT) and from clinical genetics departments in the Netherlands; collected data included date of birth, gender, date of diagnosis with polyps, cumulative number of polyps counted at colonoscopy or in excised bowel, location and histology of polyps, presence of duodenal polyps, information on CRC, presence of polyps/CRC in first degree family members, date of last contact and status at last contact.

Targeted next-generation sequencing

Targeted next-generation sequencing of leukocyte DNA was performed using a custom M13-tailed sequencing panel on the Ion Torrent platform (Thermo Fisher, Waltham, MA, USA). The exonuclease domains of *POLE* exons 9-14 and *POLD1* exons 8-12 were screened. Primers for overlapping amplicons were designed using Primer3 (http://primer3.ut.ee/) and ordered from Integrated DNA Technologies (IDT Leuven, Belgium). Primer sequences are available upon request. Following the manufacturer's protocol - briefly, PCR amplicons were generated from 10 ng of leukocyte DNA using two primer pools. The PCR pools were subsequently combined and purified using AMPureXP beads. To add sample barcodes and Ion Torrent adapters, a second round of PCR was performed using M13 primers with A and P1 tails and sample barcodes. The PCR products were pooled, purified using AMPureXP beads and quantified using the Bioanalyzer High Sensitivity DNA kit (Agilent Technologies, Santa Clara, California). Size selection was performed, and the final concentration of the library was measured with a Bioanalyzer High Sensitivity DNA kit. Emulsion PCR was performed on an Ion One Touch 2 System (Thermo Fisher). The quality of the emulsion PCR was measured using the Qubit IonSphere Quality Control Kit, and libraries were sequenced using the Ion Personal Genome Machine (PGM).

Data analysis

The sequence data were checked for quality using the quality control tool for high throughput sequence data, FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Subsequently, data were aligned to the human genome 19 (hg19, Genome Reference Consortium GRCh37) as a reference using the Burrows-Wheeler Aligner (BWA, http://bio-bwa.sourceforge.net). Variant calling

was performed using VarScan software (http://varscan.sourceforge.net/). Subsequently, variant annotation was performed with Annovar software (http://annovar.openbioinformatics.org). Variants with a minor allele frequency (MAF) >1%, as reported in dbSNP, ExAc or Go-ESP, were also excluded. The Integrative Genomics Viewer (IGV, http://software.broadinstitute.org/software/igv/) was used to visualize the read alignment and the presence of variants against the reference genome. Alamut software (Interactive Biosoft-ware, Rouen, France), Align GVGD (http://agvgd.hci.utah.edu/agvgd_input.php), PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/) and Combined Annotation Dependent Depletion (CADD, http://cadd.gs.washington.edu/snv) were used for variant interpretation.

Validation and segregation analysis by Sanger sequencing

Sanger sequencing was performed to validate the *POLD1* NM_002691.3:c.961G>A, p.(Gly321Ser) and *POLD1* NM_002691.3:c.955T>G, p.(Cys319Gly) variants detected by the next-generation sequencing panel, followed by co-segregation analysis for available material from family members. Leukocyte DNA, in addition to both normal and tumor DNA, were used when available. Sanger sequencing was performed by Macrogen (Amsterdam, the Netherlands). Sequencing results were analyzed using Mutation Surveyor software (Sofgenetics, State College, PA).

Results

A cohort of 332 Dutch patients with multiple colorectal polyps, without known pathogenic germline mutations, were screened to identify mutations in the exonuclease domain of *POLE* and *POLD1*. The mean age at diagnosis of colorectal polyps was 55.48 years (range 13-82). Approximately 44.9% of the patients have adenomatous polyps, while 43.3% of the patients displayed a mixed phenotype, predominantly adenomas with hyperplastic or serrated type. The majority of cases (56.6%) had a cumulative polyp count of 10 to 50. CRC was found in 126 patients (38%) at a mean age of diagnosis of 53 years (range 21-80). Clinical characteristics of the index patients are summarized in Table 1. Using targeted next-generation sequencing, we screened the exonuclease domain of *POLE* and *POLD1*. Two *POLE NM_006231.2:c.1270C>G, p.(Leu424Val) mutation* carriers that we previously reported [12] were included as controls in this study. We detected the *POLE* c.1270C>G variant in the controls, but no additional *POLE* mutations were found. For *POLD1*, we identified two variants. A heterozygous germline *POLD1* NM_002691.3:c.961G>A, p.(Gly321Ser) variant located in the exonuclease domain

(EDM) was identified in patient P1. In silico analysis predicted that this variant is likely to affect the function of the protein. The amino acid is highly conserved across species, up to Baker's yeast, and highly conserved at the nucleotide level (PhyloP: 5.53). There are small physicochemical differences between glycine and serine (Grantham distance: 56 [0-215]). Although the glycine and serine differ in polarity, charge and size, this change is considered a conservative amino acid substitution. This variant is predicted to be deleterious (SIFT score: 0.0), disease-causing by Mutation Taster (p-value: 1), possibly damaging by PolyPhen-2 v2.2.2r398 (score of 0.88 [sensitivity: 0.82; specificity: 0.94]), and likely to interfere with function by Align GVGD (class C55 [GV: 0.00 - GD: 55.27]). Furthermore, the Combined Annotation Dependent Depletion (CADD Phred, v1.3) is 29.7, predicting that this is may be pathogenic variant (Table 2). Patient P1 was diagnosed with multiple colorectal polyps at age 35 (>100 polyps, mostly hyperplastic and some adenomas) and a microsatellite stable (MSS) CRC at age 37. In addition to the *POLD1* variant, the patient is heterozygous for the pathogenic NM_001128425.1:c.536A>G, p.(Tyr179Cys) variant in *MUYTH.* The patient has no known family history of CRC; furthermore, no analyzable tumor tissue is available for further study.

Another patient P2 was identified with the *POLD1* c.961G>A, p.(Gly321Ser) variant. The patient was diagnosed with CRC at age 44 and one adenomatous polyp with low-grade dysplasia at age 47 and two serrated adenomas at age 54. She had two sisters with endometrial carcinoma (EC). Both sisters did not carry the variant (Figure 1A). The tumor from this patient is mismatch repair deficient (MMRD), with microsatellite instability (MSI-H), negative MLH1/PMS2 immunohistochemistry and with *MLH1* promoter hypermethylation (OMIM #120436). No somatic mutations in *KRAS* (OMIM #190070) exon 2, codons 12/13 and *BRAF* exon 15 (OMIM #164757) were found. Due to the highly degraded nature of the formalin-fixed paraffin-embedded (FFPE) derived DNA, we were unable to determine the tumor mutation burden.

The second *POLD1* NM_002691.3:c.955T>G, p.(Cys319Gly) exonuclease domain variant was identified in a patient (P3) diagnosed with multiple colorectal polyps at ages 40 and 51 (>100 adenomas). Co-segregation was performed using available DNA from affected family members. One sister cumulatively developed multiple polyps from age 34; however, leukocyte DNA tested negative for the variant. Another sister, diagnosed with CRC at age 38, had both normal and tumor DNA available for further analysis. Using both DNA samples, this case was shown to be a noncarrier. No further DNA was available

for two other affected family members (Figure 1B). This variant could affect the function of the protein, the affected amino acid is highly conserved and the affected nucleotide is also moderately conserved (phylop: 2.87). With a large physicochemical difference between cysteine and glycine (Grantham distance: 159 [0-215]), the variant is predicted to most likely interfere with function of the protein by Align GVGD (class C65 [GV: 0.00 – GD: 158.23]) and probably damaging by PolyPhen-2 v2.2.2r398 (score of 1.00 [sensitivity: 0.00; specificity: 1.00]). The CADD Phred v1.3 score of 24.4, predicting that this is a pathogenic variant that is furthermore supported by SIFT (score: 0) and Mutation Taster (p-value: 1) (Table 2). In summary, this variant is a novel missense change that might be damaging to protein structure and function but did not show convincing co-segregation. No tumor material from the patient was available for further studies.

Table 1. Clinical characteristics of the index patients included in this study (n=332)

Clinical characterization	Individuals %
Number of polyps	
<10	53 (16.0%)
10-50	188 (56.6%)
50-100	49 (14.8%)
>100	29 (8.7%)
Unknown	13 (3.9%)
Type of polyps	
Adenomas	149 (44.9%)
Adenoma + Hyperplastic	103 (31.0%)
Adenomas + Hyperplastic + Serrated	32 (9.6%)
Adenoma + Serrated	7 (2.1%)
Hyperplastic + Serrated	2 (0.6%)
Hyperplastic	5 (1.5%)
Serrated	1 (0.3%)
Unknown	33 (9.9%)
Age at diagnosis with polyposis	
>50 years	186 (56.0%)
≤50 years	146 (44.0%)
Diagnosed with CRC	
Yes	126 (38.0%)
No	206 (62.0%)
Sex	
Male	191 (57.5%)
Female	141 (42.5%)

Table 2. *POLD1* germline variants in the exonuclease domain identified by next-generation sequencing

Patient	Alteration in genomic DNA	Protein alteration	MAF	rsID	Mutation taster	SIFT	PolyPhen-2	Grantham distance	Align GVGD	Segregation	CADD	Variant classification
P1, P2	c.961G>A	p.Gly321Ser	ExAC= 0.0005 Go-ESP=0.0002 TOPMED= 0.0003	Rs41554817	Disease causing	Deleterious	Possibly damaging	Predicted not to be deleterious	Likely to interfere with function	P1: Segregation not performed, unclear family history For P2: not segregate in tested family members	Predicted to be pathogenic	VUS
P3	c.955T>G	p.Cys319Gly	N.A	N.A	Disease causing	Deleterious	Probably damaging	Predicted to be deleterious	Highly likely to interfere with function	Not segregate in tested family members	Predicted to be pathogenic	VUS

MAF: minor allele frequency; rsID: variant identifier in dbSNP; ExAc: exome aggregation consortium; Go-ESP: exome sequencing project; TOPMED: trans-omics in precision medicine; N.A: not available; CADD: combined annotation dependent depletion; VUS: variant of uncertain significance.
GenBank reference sequence: *POLD1;* NM_002691.3

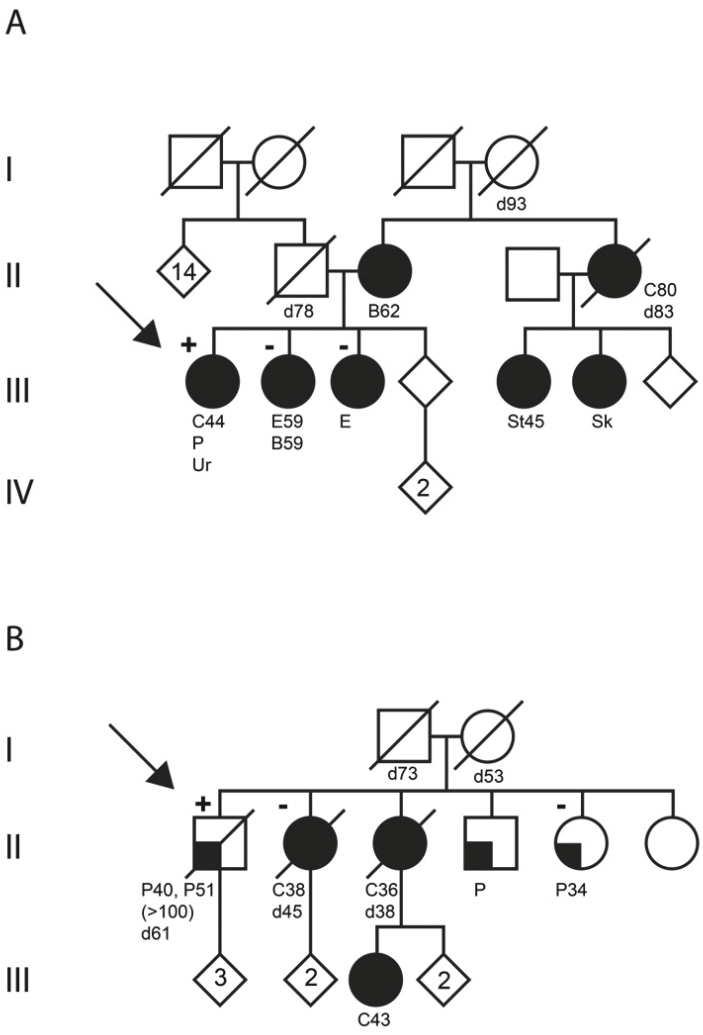


Figure 1. Pedigrees of the families with germline *POLD1* variants.

A, represents the family pedigree of the index patient P2 with *POLD1* c.961G>A, p.(Gly321Ser). B, represents the family pedigree of index patient P3 with *POLD1* c.955T>G, p.(Cys319Gly). Filled symbol, cancer; symbol filled quarter, individual with colorectal polyps. [+], *POLD1* variant carrier; [-], noncarriers. The probands are indicated by an arrow. C, colorectal cancer; P, colorectal polyps; Ur, urothelial cell cancer; E, endometrial cancer; B, breast cancer; St, stomach cancer; Sk, skin cancer; d, deceased; number next to letter, ages at diagnosis or at death.

Discussion

Pathogenic variants affecting the exonuclease domains of *POLE* and *POLD1* are associated with polyposis and colorectal cancer. Here, we screened the exonuclease domain of *POLE* and *POLD1* to detect causative variants in 332 index patients with multiple colorectal polyps. We detected two missense, heterozygous variants in *POLD1*. The *POLD1* c.961G>A, p.(Gly321Ser) variant was identified in two patients with multiple colorectal polyps and CRC. Gly321Ser is highly conserved and predicted to be damaging by in silico analysis tools. However, the available evidence is currently insufficient to evaluate the effect of this variant on the function of the protein; therefore, the variant is classified as a Variant of Unknown Significance (VUS). Co-segregation analysis was not feasible in one of the families and not supportive in the other family. In addition, the absence of available tumor tissue for functional analysis hampered further characterization of this VUS in P1. The tumor from patient P2 is MSI-H with *MLH1* promoter hypermethylation, therefore the mismatch repair deficiency phenotype is caused by somatic *MLH1* promoter hypermethylation and not due to germline defects. While tumors from *POLE* and *POLD1* pathogenic variants carriers showed hypermutated phenotype [6, 12], *POLD1* tumors with exonuclease domain mutations at highly conserved motifs (Exo1, 11, 111) were not consistently hypermutant [19]. It has been shown that mutations in *POLE* and *POLD1* do not always show a functional impact, therefore, determining the pathogenicity of mutations in these genes can be challenging [19]. The Gly321Ser variant is found in databases at a very low frequency (rs41554817, ExAc= 0.0005, GO-ESP= 0.0002 and TOPMED= 0.0003). Interestingly, patient P1 also carries a heterozygous *MUTYH* p.(Tyr179Cys) variant in addition to *POLD1* c.961G>A, p.(Gly321Ser), possibly suggesting that both genes could act cooperatively and together confer an increased CRC risk. The co-occurrence of the *MUTYH* pathogenic mutation with another mutation in *MSH2* or *MSH6* has been reported [20, 21]. Recently, a patient with the *POLD1* c.961G>A, p.(Gly321Ser) variant was reported; this patient developed CRC at age 41. No segregation analysis could be performed for the family as no DNA was available [18].

The exonuclease domain *POLD1* c.955T>G, p.(Cys319Gly) variant was identified in a patient who developed multiple colorectal polyps, with a family history of CRC and multiple polyps. The variant was only present in the patient but not in two affected siblings with CRC or multiple polyps, suggesting a possible de novo *POLD1* variant in patient P3. De novo mutations within *POLE* have been previously identified in some studies [11, 12]. While the in silico evidence suggests

a pathogenic variant, the lack of co-segregation in the family is not supportive. No tumor material from this patient was available to analyze further. However, it is still unclear whether or not the variant impaired protein function. Functional assays are required for better evaluation of these variants. Notably, the *POLD1* c.955T>G, p.(Cys319Gly) variant was not observed in population databases (ExAc, GO-ESP and TOPMED) and has not been reported in association with *POLD1*-related disease. In the present study, we did not find pathogenic variants in *POLE* and *POLD1*. These genes have a low frequency in the predisposition for multiple polyps. It is worth mentioning that we previously evaluated the prevalence of the recurrent mutations *POLE c.1270C>G,* p.(Leu424Val) and *POLD1 c.1433G>A*, p.(Ser478Asn) in a cohort of Dutch index patients with multiple polyps or familial CRC. Although we did not detect *POLD1* p.(Ser478Asn), three index patients with the *POLE* p.(Leu424Val) variant were detected [12]. Despite an enrichment in our cohort for inherited CRC and polyposis, the frequency (0.25%) is also comparable to reported frequencies [6, 11, 13]. These results confirm the low frequency of these genes as a rare cause of the disease.

Recently, both *POLE* and *POLD1* mutations were identified outside the exonuclease domains [13, 17, 19], suggesting other domains may be responsible for proofreading and should also be screened.

In conclusion, we identified no convincing pathogenic variants in exonuclease domains of *POLE* and *POLD1* in the current study. We recommend that screening of *POLE* and *POLD1* should still be considered, although pathogenic variants in *POLE* and *POLD1* probably occur at a low frequency in patients with multiple colorectal polyps. Multigene panels offer significant benefits over sequential single-gene testing by reducing costs, time and increasing the sensitivity. Moreover, making feasible the analysis of multiple low frequency genes in highly heterogenous syndromes. Indeed, including the two genes in multigene panels that are used to screen for pan-cancer mutations will allow to identify these rare mutations.

References

1. 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.
2. Jasperson KW, Tuohy TM, Neklason DW, et al. Hereditary and familial colon cancer. Gastroenterology 2010;138:2044-58.
3. Kinzler KW, Nilbert MC, Su LK, et al. Identification of FAP locus genes from chromosome 5q21. Science 1991;253:661-5.
4. 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.
5. Lynch HT, de la Chapelle A. Hereditary colorectal cancer. N Engl J Med 2003;348:919-32.
6. 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.
7. 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.
8. Nicolaides NC, Papadopoulos N, Liu B, et al. Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. Nature 1994;371:75-80.
9. Peltomaki P. Deficient DNA mismatch repair: a common etiologic factor for colon cancer. Hum Mol Genet 2001;10:735-40.
10. Hendriks YM, Jagmohan-Changur S, van der Klift HM, et al. Heterozygous mutations in PMS2 cause hereditary nonpolyposis colorectal carcinoma (Lynch syndrome). Gastroenterology 2006;130:312-22.
11. 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.
12. Elsayed FA, Kets CM, Ruano D, et al. Germline variants in POLE are associated with early onset mismatch repair deficient colorectal cancer. Eur J Hum Genet 2015;23:1080-4.
13. Spier I, Holzapfel S, Altmuller J, et al. Frequency and phenotypic spectrum of germline mutations in POLE and seven other polymerase genes in 266 patients with colorectal adenomas and carcinomas. Int J Cancer 2015;137:320-31.
14. Chubb D, Broderick P, Frampton M, et al. Genetic diagnosis of high-penetrance susceptibility for colorectal cancer (CRC) is achievable for a high proportion of familial CRC by exome sequencing. J Clin Oncol 2015;33:426-32.
15. Bellido F, Pineda M, Aiza G, et al. POLE and POLD1 mutations in 529 kindred with familial colorectal cancer and/or polyposis: review of reported cases and recommendations for genetic testing and surveillance. Genet Med 2016;18:325-32.
16. Rohlin A, Eiengard F, Lundstam U, et al. GREM1 and POLE variants in hereditary colorectal cancer syndromes. Genes Chromosomes Cancer 2016;55:95-106.
17. Esteban-Jurado C, Gimenez-Zaragoza D, Munoz J, et al. POLE and POLD1 screening in 155 patients with multiple polyps and early-onset colorectal cancer. Oncotarget 2017;8:26732-26743.
18. Jansen AM, van Wezel T, van den Akker BE, et al. Combined mismatch repair and POLE/POLD1 defects explain unresolved suspected Lynch syndrome cancers. Eur J Hum Genet 2016;24:1089-92.
19. Campbell BB, Light N, Fabrizio D, et al. Comprehensive Analysis of Hypermutation in Human Cancer. Cell 2017;171:1042-1056.e10.

20. Cohen SA, Tan CA, Bisson R. An Individual with Both MUTYH-Associated Polyposis and Lynch Syndrome Identified by Multi-Gene Hereditary Cancer Panel Testing: A Case Report. Front Genet 2016;7:36.
21. van Puijenbroek M, Nielsen M, Reinards TH, et al. The natural history of a combined defect in MSH6 and MUTYH in a HNPCC family. Fam Cancer 2007;6:43-51.

Chapter 4

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

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

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

Abstract

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

Introduction

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

Materials and methods

Patients

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

APC intronic variant screening

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

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

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

* Sessile serrated lesions with or without dysplasia

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

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

Next-generation sequencing and data analysis

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

Results and discussion

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

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

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

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

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

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

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

References

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

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

Chapter 5

Mutational signature analysis reveals NTHL1 deficiency to cause a multi-tumor phenotype

Fadwa A. Elsayed*, Judith E. Grolleman*, Richarda M. de Voer*, Maartje Nielsen*, Robbert D.A. Weren*, Claire Palles, Marjolijn J.L. Ligtenberg, Janet R. Vos, Sanne W. Ten Broeke, Noel F.C.C. de Miranda, Renske A. Kuiper, Eveline J. Kamping, Erik A.M. Jansen, M. Elisa Vink-Börger, Isabell Popp, Alois Lang, Isabel Spier, Robert Hüneburg, Paul A. James, Na Li, Marija Staninova, Helen Lindsay, David Cockburn, Olivera Spasic-Boskovic, Mark Clendenning, Kevin Sweet, Gabriel Capellá, Wenche Sjursen , Hildegunn Høberg-Vetti, Marjolijn C. Jongmans, Kornelia Neveling, Ad Geurts van Kessel, Hans Morreau, Frederik J. Hes, Rolf H. Sijmons, Hans K. Schackert, Clara Ruiz-Ponte, Dagmara Dymerska, Jan Lubinski, Barbara Rivera, William D. Foulkes, Ian P. Tomlinson, Laura Valle, Daniel D. Buchanan, Sue Kenwrick, Julian Adlard, Aleksandar J. Dimovski, Ian G. Campbell, Stefan Aretz, Detlev Schindler, Tom van Wezel, Nicoline Hoogerbrugge#, Roland P. Kuiper#

*These authors contributed equally
#Co-senior author

Cancer Cell, 2019; 35(2): 256-266.e5

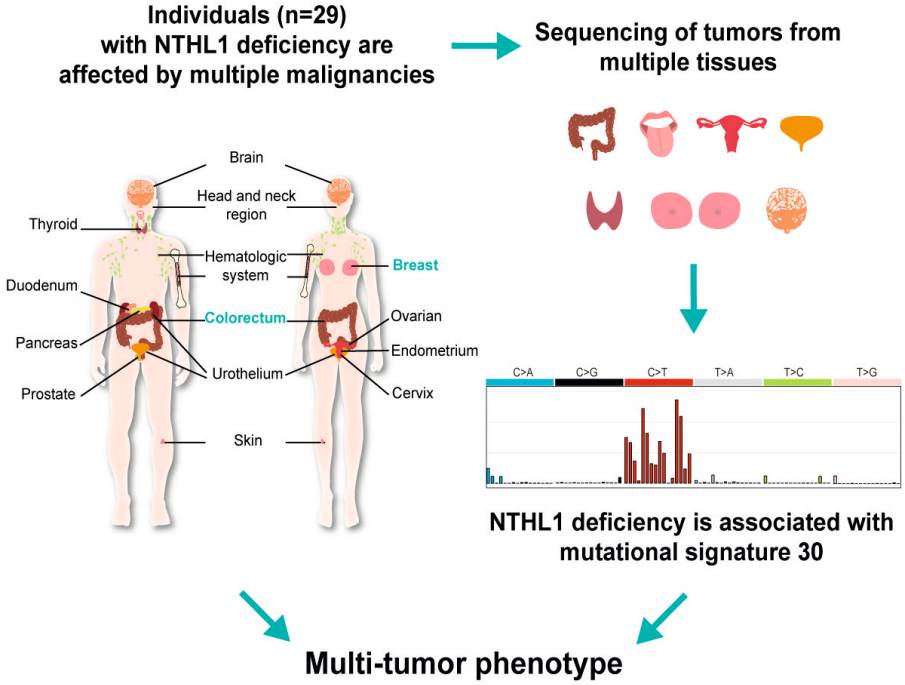
Summary

Biallelic germline mutations affecting *NTHL1* predispose carriers to adenomatous polyposis and colorectal cancer, but the complete phenotype is unknown. We describe 29 individuals carrying biallelic germline *NTHL1* mutations from 17 families, of which 26 developed one (*n* = 10) or multiple (*n* = 16) malignancies in 14 different tissues. An unexpected high breast cancer incidence was observed in female carriers (60%). Mutational signature analysis of 14 tumors from seven organs revealed that NTHL1 deficiency underlies the main mutational process in all but one of the tumors (93%). These results reveal *NTHL1* as a multi-tumor predisposition gene with a high lifetime risk for extracolonic cancers and a typical mutational signature observed across tumor types, which can assist in the recognition of this syndrome.

Significance

Individuals with a cancer predisposition syndrome benefit from customized surveillance, including screening for early-stage malignancies. However, design of an optimal surveillance program is difficult for rare cancer syndromes, particularly when the tumor spectrum is broad. This study describes the tumor phenotype observed in 17 families with NTHL1 deficiency and demonstrates that a unique *NTHL1*-associated mutational signature can be detected across tumors from 7 different organs of patients with biallelic germline *NTHL1* mutations, thereby linking a broad spectrum of cancers to this syndrome despite low patient numbers. This study illustrates the power of mutational signature analysis in defining tumor phenotypes in rare cancer predisposition syndromes and provides proof-of-principle for recognizing new patients with cancer syndromes based on tumor sequence data.

Graphical abstract



Highlights

- Biallelic germline *NTHL1* mutations predispose to a multi-tumor syndrome
- Biallelic germline *NTHL1* mutation carriers are at risk for breast cancer
- Tumors from NTHL1 deficient patients reveal a cross cancer *NTHL1*-associated signature
- Mutational signature analyses can assist to identify germline DNA repair defects

Introduction

A major proportion of known adenomatous polyposis and colorectal cancer (CRC) predisposing genes directly affects genomic maintenance. These alterations include biallelic, and thus recessively inherited, mutations in the base excision repair genes *MUTYH* and *NTHL1* [1, 2], and dominantly inherited mutations in the polymerase proofreading domains of the *POLE* and *POLD1* polymerase genes [3]. In addition to adenomatous polyposis and CRC, these syndromes appear to predispose to the development of other types of cancer [2, 4-7].

The first families described with *NTHL1* mutations were of Dutch origin, all having the same truncating germline mutation (p.Gln90*) in a homozygous state [2]. Since then, additional families of German, Spanish, British and Greek descent with p.Gln90* mutations have been reported, in two cases in compound heterozygosity with another truncating *NTHL1* mutation (c.709+1G>A and p.Gln287*, respectively) [2, 5, 8-10]. Three of these families have previously been described in detail [5, 10]. The findings underscore the major contribution of this p.Gln90* mutation in causing the *NTHL1*-associated polyposis phenotype in different demographic populations, but also emphasize the role of other pathogenic mutations in this gene.

With the limited number of families with biallelic germline *NTHL1* mutations described thus far, the phenotypic spectrum and cancer risk estimates have not been established. Consequently, diagnosis of this syndrome can easily be missed in patients that present with cancers not yet linked to NTHL1 deficiency. In this study, we aimed to define the molecular and clinical characteristics of the tumor spectrum of individuals with biallelic germline *NTHL1* mutations and provide a strategy that can assist in the recognition of DNA repair cancer syndromes even in the absence of family history or other clinical parameters.

Results

Individuals with biallelic germline *NTHL1* mutations develop multiple primary tumors

We collected 19 previously unreported individuals with biallelic germline *NTHL1* mutations from 11 unrelated families (Figure S1), which were identified by targeted mutational screening of polyposis and familial CRC patients or by individual identifications in diagnostic or research settings (Table S1). Thus far, in total 29 individuals (14 male/15 female) from 17 families have been identified. We obtained and updated detailed clinical information for all of these individuals (Table 1). All individuals that received a colonoscopy (24 out of 29 individuals) were diagnosed with adenomatous polyps and 33% were additionally diagnosed with one or more hyperplastic polyps. Twenty-six individuals were diagnosed with a (pre)malignancy (90%), of which 16 developed multiple primary tumors (range: 2-5; Figure 1 and Table 1). Only one out of 33 second tumors could potentially be considered as therapy-related (Table S2). The majority of individuals developed one or more CRCs (59%), albeit that this is likely the result of a selection bias in our study population. In addition, 66% of the encountered tumors were extracolonic. In total, 14 types of (pre)malignancies and benign

tumors were observed, of which nine were recurrently encountered (Figure 1 and Table 1). Cervical (pre)malignancies and basal cell carcinomas were diagnosed in two and three individuals, respectively. Furthermore, urothelial cell cancers (UCCs) and head and neck squamous cell carcinomas (HNSCCs) were each encountered in four individuals. Hematologic malignancies, endometrial (pre)malignancies, and brain tumors were observed in five individuals. Strikingly, nine out of 15 women (60%) were diagnosed with breast cancer.

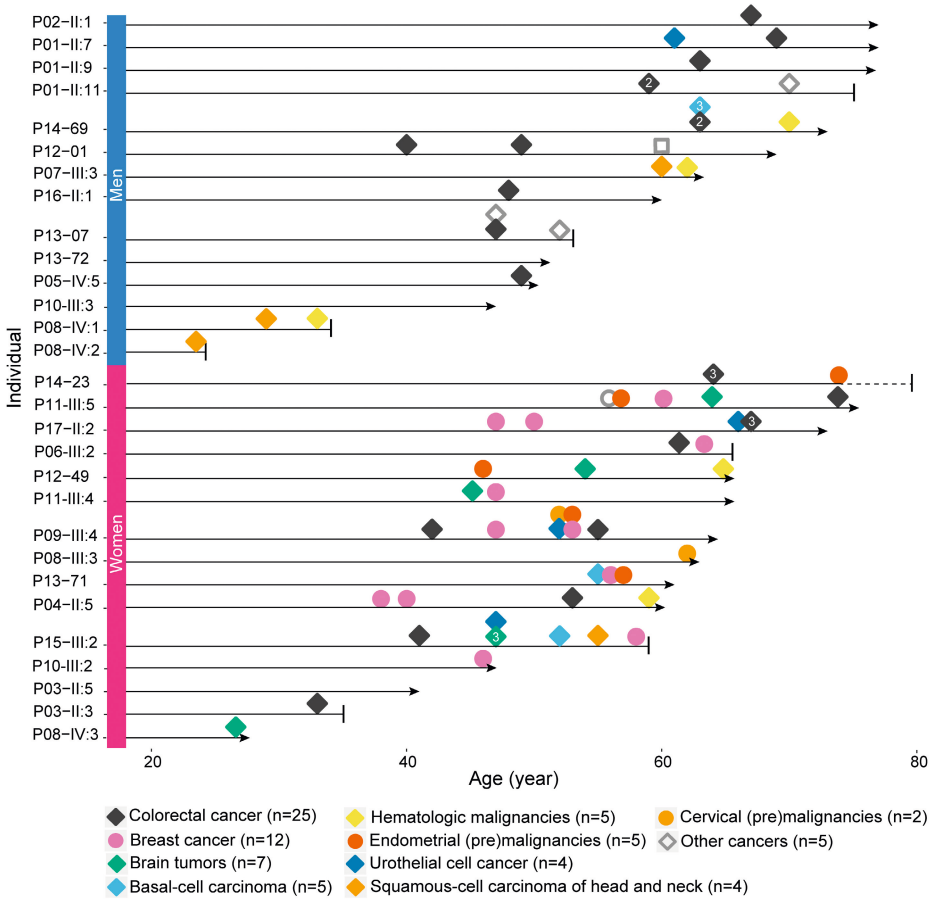


Figure 1. Age of diagnosis of benign meningiomas and (pre)malignant tumors per classification of all 29 individuals with biallelic germline *NTHL1* mutations.

Sixteen patients developed multiple malignant tumors and one patient (P11-III:4) had a benign meningioma before she developed breast cancer. Round, square, or diamond symbols indicate a female, male, or non-gender specific malignancy, respectively. Numbers indicate multiple similar malignancies at the same time. Arrowheads indicate current age, and vertical lines mark the age of death. Dashed horizontal lines indicate uncertainty about time of death. Patients are ranked based on gender (blue and pink bars represent men/women, respectively) and current age/age of death. See also Figure S1.

Table 1. Clinical phenotype of individuals with biallelic germline *NTHL1* mutations

Family	cDNA change (NM002528.6)	Amino Acid change	Patient ID[a]	M/F	Malignancies and pre-malignancies[b]	Polyps[c]	Benign lesions[b]	Publication
1	c.268C>T	p.Gln90*	**P01-II:11**	M	CRC (59), cecum CRC (59), transversum ThyC (70), follicular	Multiple A		This study
			P01-II:7	M	Renal pyelum cancer[d] (61), papillary CRC (69), ileocecal	Multiple A	Neurofibroma	This study
			P01-II:9	M	CRC (63), appendix	>30A		This study
2	c.268C>T/ c.806G>A	p.Gln90* / p.Trp269*	**P02-II:1**	M	CRC (67), rectum	50-100A		This study
3	c.268C>T	p.Gln90*	**P03-II:3**	F	CRC (33), sigmoid	1A 2H		This study
			P03-II:5	F	None (41)	6A 7H		This study
4	c.268C>T/ c.733dup	p.Gln90*/ p.Ile245Asnfs*28	**P04-II:5**	F	BC (right, 38), ductal BC (left, 40), ductal CRC (53), cecum AML[e] (59)	1A		This study
5	c.268C>T	p.Gln90*	**P05-IV:5**	M	CRC (49), rectum	200 polyps; >11A, 8H, 1S		This study
6	c.268C>T/ c.235_236insG	p.Gln90*/ p.Ala79Glyfs*2	**P06-III:2**	F	CRC (61), transversum BC (right, 63), triple negative	Multiple A >30H		This study
7	c.806G>A/ c.859C>T	p.Trp269*/ p.Gln287*	**P07-III:3**	M	SCC of the parotid gland[g] (60), AML[e] (62)	>40A		This study

Table 1. (continued)

Family	cDNA change (NM002528.6)	Amino Acid change	Patient ID[a]	M/F	Malignancies and pre-malignancies[b]	Polyps[c]	Benign lesions[b]	Publication
8	c.545G>A	p.Trp182*	**P08-IV:1**	M	SCC of the mouth base[g] (29)	No colonoscopy performed	MDS[e] (33)	This study
			P08-IV:2	M	SCC of the tongue tip[g] (24)	No colonoscopy performed		This study
			P08-IV:3	F	Brain tumor[f] (27)	No colonoscopy performed		This study
			P08-III:3	F	CC[h] (62)	No colonoscopy performed		This study
9	c.268C>T	p.Gln90*	**P09-III:4**	F	CRC (42), rectum BC (left, 47), lobular BlC[d] (52), papillary Endocervical adenocarcinoma in situ[h] (52) BC (right, 53), ductal EC[i] (53), serous CRC (55), transversum	11A >4H		This study
10	c.268C>T	p.Gln90*	**P10-III:2**	F	BC (right, 46)	13A	Skin hemangiomas (3x) Ovary cysts Liver cysts	This study
			P10-III:3	M	None (46)	2A 1H		This study
11	c.268C>T/ c.390>A	p.Gln90*/ p.Tyr130*	**P11-III:4**	F	BC (right, 47), mixed ductal/papillary	13A 2H	Meningioma[f] (45) Breast papilloma (left, 49) Uterine polyps	This study
			P11-III:5	F	OC (57), mixed endometrioid/mucinous[j] EC[i] (57), mixed endometrioid/mucinous[j] BC (left, 60), papillary and triple negative CRC (73), ascendens	No colonoscopy performed	Meningioma[f] (64), right parasellar meninges	This study

Table 1. (continued)

Family	cDNA change (NM002528.6)	Amino Acid change	Patient ID[a]	M/F	Malignancies and pre-malignancies[b]	Polyps[c]	Benign lesions[b]	Publication
12	c.268C>T	p.Gln90*	**P12-01**	M	CRC (40), rectum CRC (49), cecum PC (60)	15A[#]		(Weren et al., 2015)[2]
			P12-49	F	Endometrial complex hyperplasia[i] (46) Non-Hodgkin lymphoma[e](65)	40A	Psammomatous meningioma[f](54)	(Weren et al., 2015)[2]
13	c.268C>T	p.Gln90*	**P13-07**	M	CRC (47), rectum PaC (47) DC (52)	50A	Biliary tract hamartoma (52)	(Weren et al., 2015)[2]
			P13-71	F	BCC (55) BC (56) EC[i] (57)	50A		(Weren et al., 2015)[2]
			P13-72	M	None	10A		(Weren et al., 2015)[2]
14	c.268C>T	p.Gln90*	**P14-23**	F	CRC (64), rectum CRC (64), ascendens CRC (64), ascendens EC[i] (74)	20A		(Weren et al., 2015)[2]
			P14-69	M	CRC (63), cecum CRC (63), ascendens BCC (63), nose tip BCC (63), ear BCC (63), ear Non-Hodgkin lymphoma[e] (70)	8A		(Weren et al., 2015)[2]

Table 1. (continued)

Family	cDNA change (NM002528.6)	Amino Acid change	Patient ID[a]	M/F	Malignancies and pre-malignancies[b]	Polyps[c]	Benign lesions[b]	Publication
15	c.268C>T/ c.709+1G>A	p.Gln90*/ Abnormal splicing	**P15-III:2**	F	CRC (41) BIC[d] (47) BCC (52) SCC of head and neck[g] (55) BC (58)	Multiple A	Ovary cystadenoma (41) Intradermal nevi (42, 55) Meningioma[f] (47, 47, 47) Seborrheic keratosis (47)	(Rivera et al., 2015)[10]
16	c.268C>T	p.Gln90*	**P16-II:1**	M	CRC, ascendens (48)	30A 1H		(Belhadj et al., 2017)[5]
17	c.268C>T	p.Gln90*	**P17-II:2**	F	BC (left, 47) BC (right, 50), lobular BIC[d] (66), pappilary CRC (67), ascendens CRC (67), ascendens CRC (67), ascendens	>15A 5H		(Belhadj et al., 2017)[5]

[a]The index patient is shown in bold. [b]Numbers between brackets represent age of diagnosis. AML: acute myeloid leukemia, BC: breast cancer, BCC: basal cell carcinoma, BIC: bladder cancer, CC: cervical cancer, CRC: colorectal cancer, DC: duodenal cancer, EC: endometrium cancer, MDS: myelodysplastic syndrome, OC: ovarian cancer, PaC: pancreatic cancer, PC: prostate cancer, SCC: squamous cell carcinoma, ThyC: thyroid cancer. [c]Numbers represent the number of polyps present at time of diagnosis. A: adenomatous polyps, H: hyperplastic polyps, S: serrated polyps. Unspecified numbers of polyps is indicated as "multiple" (see also STAR Methods). [d]Classified as urothelial cell cancer. [e]Classified as hematologic malignancies. [f]Classified as brain tumors. [g]Classified as head and neck squamous cell carcinoma. [h]Classified as cervical (pre)malignancies. [i]Classified as endometrial (pre)malignancies. [j]Pathology reports suggest two individual primary tumors. [#]P12-01 developed colon, esophagus, and duodenal adenomas. See also Tables S1 and S2.

NTHL1 deficiency underlies the main mutational process in tumors from individuals with a biallelic germline *NTHL1* mutation

The clinical phenotypes of the aforementioned individuals with a biallelic germline *NTHL1* mutation suggest a predisposition to a multi-tumor phenotype, not limited to polyposis and CRC. However, the prevalence of this syndrome is infrequent and thus it remains a challenge to delineate which tumor appearances are truly the result of a deficiency of NTHL1. Very recently, it was described that *NTHL1* knockout (KO) cells generated from intestinal organoids harbor a distinct mutational signature (signature 30 of the Catalogue of Somatic Mutations in Cancer [COSMIC] [11] database), which is characterized by C>T transitions at non-CpG sites, as the main contributor to the mutation spectrum [12]. However, whether signature 30 is also the main contributor to the mutation spectrum in colon tumors of individuals with biallelic germline *NTHL1* mutations is still unknown. Therefore, we performed whole-exome sequencing (WES) on one colonic adenoma (P01-II:7; A-2) and two CRCs (P01-II:7; CRC-3 and P03-II:3; CRC-4) from two individuals with biallelic germline *NTHL1* mutations. We detected 153 (A-2), 360 (CRC-3), and 21 (CRC-4) somatic mutations in these tumors, including several known CRC driver mutations in *APC*, *KRAS* and *SMAD4* (Tables S3 and S4). Most somatic mutations were C>T transitions (87-91%; Figure S2A), predominantly located at non-CpG sites, confirming our previous observations in adenomas and CRCs from individuals with biallelic germline *NTHL1* mutations [2]. Next, we jointly extracted the mutational signatures from six colon tumors, of which three were previously sequenced and yielded sufficient mutations [2] (Table S4), together with a cohort of 215 publicly available CRC samples. Four distinct mutational signatures were identified, of which three comprised the majority of mutations in the sporadic CRC cases, as reported previously (Figure S2B-C) [13]. However, all six tumors with biallelic germline *NTHL1* mutations predominantly exhibited the fourth signature that strongly resembles signature 30 reported in COSMIC [11] and in *NTHL1*-KO organoids (both cosine similarities 0.95; Figure 2A-B) [12]. These data confirm that the absence of NTHL1-driven DNA repair gives rise to signature 30 resulting from the main mutational process in these colonic tumors from individuals with biallelic germline *NTHL1* mutations.

To determine whether NTHL1 deficiency elicits the same mutational process in extracolonic tumors, we performed WES on 17 extracolonic tumors from 11 individuals. As in the CRC tumors, multiple driver mutations were identified in the extracolonic tumors, including *PIK3CA* hotspot mutations in multiple breast cancers (Table S3). For 14 tumors, originating from seven different tissue types, we were able to retrieve sufficient somatic mutations to perform mutational

signature analyses (Table S4). The mutation spectrum of most tumors highly resembled that of signature 30 (Figure 2C). Furthermore, after refitting of the somatic mutation spectrum of all sequenced tumors to the known mutational signatures we found that signature 30 emerged as the main mutational process in 13 tumors (93%; Figure 2D and Figure S2D). We also assessed the contribution of signature 30 to the mutation spectrum in sporadic cancers of these tissues and this contribution turned out to be substantially lower compared with the tumors with biallelic germline *NTHL1* mutations (Figure S2E). Together, these data reveal a correlation between mutation spectrum and defective base excision repair caused by biallelic germline *NTHL1* mutations, both in colonic and extracolonic malignancies.

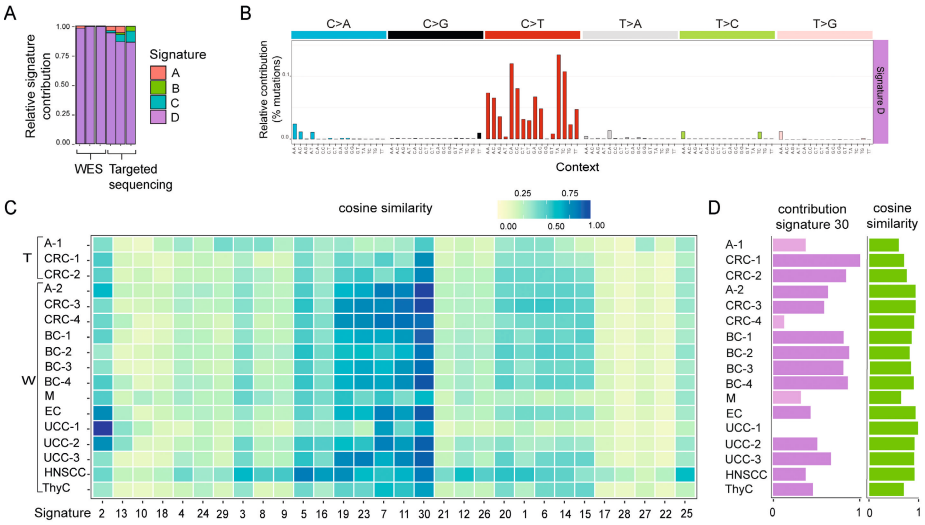


Figure 2. Mutational signature analysis of colonic and extracolonic NTHL1 deficient tumors.

(A) The relative contribution of six NTHL1 deficient colon tumors, three WES (left) and three targeted sequencing (right), to the four de novo extracted signatures from a joint analyses with the somatic mutation spectrums identified in CRCs from the TCGA. (B) Extracted de novo signature D that predominantly represents mutations in NTHL1 deficient colon tumors. This signature has a cosine similarity to the COSMIC [11] and *NTHL1*-KO organoid signature 30 of 0.95. (C) Heatmap showing the cosine similarity scores for each indicated tumor sample from biallelic germline *NTHL1* mutation carriers and the 30 COSMIC signatures. Signatures have been ordered according to their similarity, such that very similar signatures cluster together. T: targeted sequenced tumors, W: whole-exome sequenced tumors (this study). (D) The estimated relative contribution of COSMIC signature 30 to the mutation spectrum of each indicated tumor sample after refitting to 30 COSMIC signatures. Cosine similarity scores on the right indicate the closeness of the reconstruction with the mutation spectrum of each tumor. Light-colored bars represent tumors with less than 10 mutations contributing to signature 30. A: adenomatous polyp, CRC: colorectal cancer, BC: breast cancer, EC: endometrial cancer, HNSCC: head and neck squamous-cell carcinoma, M: meningioma, ThyC: thyroid cancer, UCC: urothelial cell cancer. See also Tables S3, S4 and Figure S2.

Substantial extracolonic cancer risk in individuals with biallelic germline *NTHL1* mutations

The incidence of extracolonic tumors in individuals with biallelic germline *NTHL1* mutations and the prominent presence of signature 30 in these tumors strongly suggest a high tumor risk that clearly extends beyond the gastrointestinal tract (Figure 1 and Table 2). Particularly, the high incidence of breast cancer among women with biallelic germline *NTHL1* mutations was unexpected and is potentially of high clinical relevance. The median age at diagnosis for breast cancer in these women was also found to be lower than expected in the general population (48.5 years [SD 8.2, range: 38-63] compared with 62 years, respectively; Table 2). In addition, three women were diagnosed with bilateral breast cancer, and the four breast tumors that were sequenced showed the highest contributions of signature 30 (Figure 2D and Figure S2D). These data suggest that the risk for breast cancer in women with biallelic germline *NTHL1* mutations is substantial. These findings are highly relevant for the counseling and surveillance of these patients. So far, however, no clear recommendations for clinical management have been reported. Therefore, we performed first risk analyses for all extracolonic cancers combined. We found that the median age at diagnosis for any extracolonic malignancy in the group of patients in this study was 53 (range: 24-74) years (Table 2; see STAR Methods for details). These extracolonic cancers were evenly distributed between probands (13 out of 17 individuals) and non-probands (8 out of 12 individuals). The cumulative risk for an extracolonic cancer was estimated to be between 35% and 78% (95% confidence interval [CI]) by the age of 60 years and, when accounting for ascertainment bias, between 6% and 56% (95% CI) (Table S5). Together, these data further illustrate that the cancer risk in individuals with biallelic germline *NTHL1* mutations involves a wide range of tissues including breast in women.

Table 2. Summary of clinical features of tumor types reported in individuals with biallelic germline *NTHL1* mutations

	Reported frequency		**Median age of diagnosis in *NTHL1* patients (range)**			**Median age of diagnosis**
	M (n=14)	**F (n=15)**	**M (n=14)**	**F (n=15)**	**M+F (n=29)**	**in the population**
Colorectal cancer	9	7	59 (40-69)	64 (33-73)	61 (33-73)	67[a]
Extracolonic cancer	12	29	60.5 (24-70)	53 (27-74)	53 (24-74)	
Breast cancer	0	9	NA	48.5 (38-63)	48.5 (38-63)	62[a]
Endometrial (pre) malignancies	NA	5	NA	57 (46-74)	57 (46-74)	62[a]
Urothelial cell cancer	1	3	61	52 (47-66)	56.5 (47-66)	73[a]
Brain tumors	0	4	NA	47 (27-64)	47 (27-64)	58[a]
Basal-cell carcinoma	1	2	63	53.5 (52-55)	63 (52-63)	67[b]
Head and neck squamous cell carcinoma	3	1	29 (24-60)	55	42 (24-60)	66[c]
Hematologic malignancies	3	2	62 (33-70)	62 (59-65)	62 (33-70)	67.5[a]
Cervical (pre) malignancies	NA	2	NA	57 (52-62)	57 (52-62)	47[c]
Duodenal cancer	1	NA	52	NA	52	66[a]
Prostate cancer	1	NA	60	NA	60	66[a]
Thyroid cancer	1	NA	70	NA	70	51[a]
Pancreatic cancer	1	NA	47	NA	47	70[a]
Ovarian cancer	NA	1	NA	57	57	63[a]

[a]SEER data, period 2010-2014. [b]Dutch cancer registry data, period 2010-2016, data from the South of the Netherlands. [c]Dutch cancer registry data, period 2010-2016, data from whole of the Netherlands. NA: not applicable for gender-specific malignancies. See also Table S5.

Discussion

Following the initial discovery that biallelic germline *NTHL1* mutations predispose to the development of polyposis and CRC [2], we here present a molecular and clinical characterization of the tumor spectrum of 29 individuals with biallelic germline *NTHL1* mutations from 17 unrelated families, including 11 previously unreported families. Next to adenomatous polyposis and CRC, we show that many patients develop multiple primary tumors at various sites, of which the majority is extracolonic (66%). Nine tissues were recurrently affected, with a remarkably high incidence of breast cancer. Initial cancer risk estimates for extracolonic tissues strongly suggest that clinical management for individuals with biallelic germline *NTHL1* mutations should be extended beyond the colon.

In this study, we have obtained additional evidence for causality of NTHL1 deficiency for specific malignancies by analyzing the somatic mutational patterns in tumors from seven different tissues. This analysis revealed mutational signature 30 to be prominent in most of these tumors, suggesting that deficiency of NTHL1 elicits the same mutational process in multiple tissues. A causal link between NTHL1 deficiency and mutational signature 30 has recently been suggested by a study using colonic organoids in which *NTHL1* was knocked out [12]. Furthermore, it was found that the single breast cancer sample in which signature 30 originally was identified [14] was NTHL1 deficient upon retrospective analysis of the sequencing data, due to a germline p.Gln287* mutation and loss of the wild type allele in the tumor [12]. We now show that in four breast cancer samples from four individuals with biallelic germline *NTHL1* mutations, more than 80% of the mutations can be assigned to signature 30, suggesting that this base excision repair defect has driven breast cancer formation in these patients. Importantly, this cross cancer *NTHL1*-associated signature may be used to determine whether a (rare) tumor encountered in an individual with biallelic germline *NTHL1* mutations is likely to be initiated by the absence of functional NTHL1. Similarly, in CRCs from patients with *MUTYH*-associated polyposis (MAP), where biallelic germline mutations in the base excision repair gene *MUTYH* cause a distinct somatic mutational signature characterized by an accumulation of C>A transversions [1, 15, 16]. Together, these findings suggests that the somatic mutation spectra and mutational signatures identified in patients with an unexplained cancer phenotype could facilitate the identification of an underlying constitutional DNA repair defect.

The size and variability of our polyposis cohorts and the differences in mutation detection methodology used prevent us from making accurate estimates of the incidence of this NTHL1-associated tumor syndrome in polyposis patients. However, based on the prevalence of pathogenic base excision repair gene mutations in the population, we have previously estimated that NTHL1-associated tumor syndrome is approximately five times less frequent than MAP [17]. Eight different pathogenic germline *NTHL1* mutations have now been described, all resulting in truncation of the gene (Table S1). The p.Gln90* mutation has been encountered in 18 families, and is predominantly observed in a homozygous state (n=12). Interestingly, two of the families with homozygous p.Gln90* mutations originated from Qatar and Kazakhstan, confirming earlier reports that this mutation exhibits a wide global distribution [5, 10]. It can be anticipated that the relative frequency of *NTHL1* mutations will show variation between populations, and additional pathogenic mutations may turn out to

play an important role in the prevalence of this syndrome in relatively isolated populations, as illustrated by our finding of a truncating mutation (p.Trp182*) in a consanguineous Turkish family (Family 7). Therefore, if *NTHL1* is considered for testing in new families, we recommend sequencing of the entire open reading frame.

Next to breast cancer (60% of the women), we encountered endometrial (pre)malignancies, UCCs, brain tumors, hematologic malignancies, basal cell carcinomas, HNSCCs, and cervical cancers in multiple individuals, and at least five other cancers in single individuals, including duodenal cancer. While not all observed malignancies may be the result of the NTHL1 deficiency, as for example shown by the mutation spectrum in one of the three UCCs, the range of malignancies in individuals with an NTHL1 deficiency is striking. Extracolonic malignancies appear to occur more frequently than what is described for other Mendelian CRC syndromes, such as Lynch syndrome, polymerase proofreading-associated polyposis, and MAP [1, 3, 18-22]. Particularly, breast cancer seems to occur much less in these syndromes compared with what we observe in females with NTHL1 deficiency.

We are aware that a selection bias in our study partially explains the high frequency of CRCs in our cohort, particularly in the index patients. Nevertheless, many individuals developed other malignancies at first diagnosis or no CRC at all. Due to ascertainment bias, caused by the selection of patients with cancer or polyposis, the risk calculations for extracolonic malignancies should be treated with caution. We applied stringent ascertainment bias correction considering all cancer estimates. Therefore, the lower limit of the risk range might be an underestimation, as the clinic-based population that is offered genetic counseling is most likely a selected higher risk population out of all *NTHL1* mutation carriers present in the general population. Even though this is the largest cohort of individuals with biallelic germline *NTHL1* mutations reported to date, the sample size and follow-up time is still too limited to present precise, site-specific, cancer risk estimates. Hence, once more families will be identified, updates of these calculations may be required. Eventually, this may also allow us to determine cancer risk estimates for heterozygous *NTHL1* carriers, as a subtle increased cancer risk has been reported for monoallelic *MUTYH* carriers [23]. Nevertheless, our data indicate that constitutional NTHL1 deficiency underlies a high-risk hereditary multi-tumor syndrome. Therefore, we recommend germline testing of *NTHL1* for patients with multiple primary

malignancies, independent of tissue type and, especially, in the case of recessive inheritance.

Considering the spectrum of malignancies observed in the 17 families described thus far, additional surveillance of these patients might be considered beyond that offered to patients with polyposis. Both NTHL1- and MUTYH-deficiency syndromes are characterized by a high risk of CRC with an attenuated polyposis phenotype. However, whereas for MAP patients only a significant higher risk for bladder and ovarian cancer has been reported, the risks in other tissues, such as breast, endometrium and bone marrow are less clear or absent [7, 21, 24]. For colon surveillance, we propose that the established surveillance guidelines for MAP should be extended to individuals with biallelic germline *NTHL1* mutations [5, 7], which includes colonoscopy surveillance beginning at age 18-20 years. Based on the median age and age range of breast cancer diagnosis in our study, we suggest breast cancer screening depending on local guidelines, at least based on moderate risk. There may be an increased risk of endometrial cancer in these patients, potentially comparable with Lynch syndrome. Yearly ultrasound and endometrial biopsy may be considered, albeit that its efficacy remains to be determined (NCCN) [25]. For the other cancers no advice for surveillance schedules can be provided due to uncertainty about exact cumulative cancer risks and/or lack of evidence for the efficacy of screening methods for these cancers. Revision of the surveillance recommendations may be needed once more families with biallelic germline *NTHL1* mutations have been identified.

We conclude that individuals with biallelic germline *NTHL1* mutations present with adenomatous polyposis and multiple primary tumors, including colon cancer and breast cancer. We found tumor mutational signature analysis to be very suitable for obtaining additional support for a causative link between NTHL1 deficiency and tumor development. We recommend *NTHL1* mutation testing for individuals with multiple primary malignancies, either with or without adenomatous polyposis and/or a family history of cancer. The suggested high lifetime risk of (multiple) malignancies associated with this *NTHL1*-associated tumor syndrome requires awareness and surveillance for colonic and extracolonic cancers, including breast cancer.

STAR methods

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
p.Gln90* genotyping: FFPE and blood-derived DNA (see Table S6)	LUMC	N/A
NTHL1 targeted Sanger sequencing and Molecular Inversion Probe: blood-derived DNA (see Table S6)	Participating institutes	N/A
WES: tumor material from NTHL1 patients	Participating institutes	N/A
Chemicals, Peptides, and Recombinant proteins		
KASP V4.0 2X Master mix	LGC	Cat# KBS-1016-002
Critical Commercial Assays		
WES: SureSelectXT Human All Exon V5 enrichment kit	Agilent Technologies	https://www.agilent.com/
WES: SureSelectXT^{HS} Target enrichment system for Illumina paired end multiplexed sequencing library	Agilent Technologies	https://www.agilent.com/
WES: SureSelectXT Human All Exon V6 enrichment kit	Agilent Technologies	https://www.agilent.com/
DNA isolation: QIAamp DNA mini kit	QIAGEN	Cat# 51304
Identification family 5 adn 10: TruSightTMCancer Sequencing Panel	Illumina	https://www.illumina.com
Identification family 6: HiPlex	Hiplex	www.HiPLEX.org
Identification family 7: TruSight One sequencing panel	Illumina	https://www.illumina.com
Identification family 8: Agilent SureSelect Human Exon V4 enrichment kit	Agilent Technologies	https://www.agilent.com
Identification family 9: custom designed HaloPlex Targeted Enrichment Assays	Agilent Technologies	N/A
Identification family 10: custom Agilent capture array enrichment	Agilent Technologies	N/A
Deposited Data		
Analyzed WES data	This paper	Table S3
Raw WES data	This paper	EGAD00001004534
Human Reference Genome (NCBI build 37, CRch37)	Genome Reference Consortium	http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/
MIP analysis and WES filtering: Exome Aggregation Consortium (ExAC) database (version 0.3)	Exome Aggregation Consortium	http://exac.broadinstitute.org
WES filtering: gnomAD database (version 2.0)	The Genome Aggregation Database	http://gnomad.broadinstitute.org/
Control data somatic mutations: The Cancer Genome Atlas (TCGA) database (see Figure S2E)	The Cancer Genome Atlas	https://gdc-portal.nci.nih.gov/legacy-archive/files/

REAGENT or RESOURCE	SOURCE	IDENTIFIER
30 COSMIC signatures	Catalogue of Somatic Mutations in Cancer	http://cancer.sanger.ac.uk/cancergenome/assets/signatures_probabilities.txt
Risk assessment: Comprehensive Cancer Center the Netherlands [26]: Dutch cancer incidence	The Netherlands Cancer Registry	http://www.cijfersoverkanker.nl.
Oligonucleotides		
KASPar assay: NTHL1_p.Gln90*_A1: 5'–AAGGTGACCAAGTTCATGCTGTGCCAGTCTGGGAGCCCT–3')	This paper	N/A
KASPar assay: NTHL1_p.Gln90*_A2: 5'–GAAGGTCGGAGTCAACGGATTGCCAGTCTGGGAGCCCC–3'	This paper	N/A
KASPar assay: common reverse primer: 5'–ACCAGCTGTTGCTGCCAGTCCT–3'	This paper	N/A
Software algorithms		
De novo signature analysis: Non negative matrix	Gaujoux and Seoighe, 2010 [27]	https://doi.org/10.1186/1471-2105-11-367
Signature reconstruction: R package DeconstructSigs	Rosenthal et al., 2016 [28]	https://cran.r-project.org/web/packages/deconstructSigs/index.html
GraphPad PRISM (version 5)	GraphPad Software	www.graphpad.com
Mendel	OMICtools	https://omictools.com/mendel-tool
R (version 3.4)	R [29]	https://www.r-project.org/
KASPar primers design: PrimerPicker Lite Beta (version 0.1)	KBioscience	www.kbiosciences.co.uk
KASPAr data analysis: Bio-Rad CFX manager software (version 3.0)	Bio-Rad	www.bio-rad.com
MIP analysis: SeqNext (version 4.2.2, build 502)	JSI medical systems	https://jsi-medisys.de/
Variant calling WES: UnifiedGenotyper	Broad Institute, Genome Analysis Toolkit (GATK)	https://software.broadinstitute.org/gatk
WES filtering: integrative genome viewer (IGV)	Broad Institute	http://software.broadinstitute.org/software/igv
Identification family 8: NextGENe Software (v.2.3.4.4)	Softgenetics	https://softgenetics.com/NextGENe.php

Contact for reagent and resource sharing

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact Richarda M. de Voer (richarda.devoer@radboudumc.nl).

Experimental model and subject details

Patient cohorts

We have ascertained patients with unexplained polyposis (cumulative occurrence of at least 10 polyps but no germline mutations in known CRC/polyposis-predisposing genes), young CRC (diagnosis ≤40) and/or familial CRC (CRC ≤50 + first degree relative with CRC ≤60). Blood-derived DNA from 828 unrelated patients from the United Kingdom (n=273), the Netherlands (n=169), Poland (n=145), Germany (n=105), Norway (n=88), Spain (n=36), and Macedonia (n=12) was used for targeted sequencing of *NTHL1* (Table S6). Furthermore, a total of 1,842 Dutch index patients with unexplained colorectal polyposis or familial CRC were genotyped for the p.Gln90* mutation in *NTHL1* (Table S6). These approaches revealed four previously unreported families with truncating biallelic germline *NTHL1* mutations. Seven additional families with confirmed biallelic *NTHL1* mutations were referred by different centers, as described in more detail in the Method details. This study was approved by local medical ethics committees (CMO; study numbers 2014/032 and 2015/1748 of the Radboudumc Nijmegen, and P01-019 of the LUMC Leiden). All participants provided written informed consent.

Method details

NTHL1 targeted sequencing

Targeted sequencing of 88 of 828 patients was performed by Sanger sequencing, and in the remaining 740 patients Molecular Inversion Probe-based sequencing on a NextSeq500 platform was used [30]. Twenty-three Molecular Inversion Probes were designed according to a previously published methodology [30, 31] with minor modifications, covering all coding regions and intron-exon boundaries of *NTHL1* (NM_002528.6, sequences available upon request). For MIP-based sequencing, fastq files containing all reads split per barcode, were analyzed using SeqNext (JSI medical systems; version 4.2.2, build 502). The average fold coverage in the open reading frame of *NTHL1* was variable, but on average above 100x. Reads fulfilling predetermined quality settings (max. 5% mismatches; min. 95% matching bases) were mapped to the regions of interest (NM_002528). At least

40-fold absolute coverage, 30% variant reads and 30 variant reads were required for variant calling. All variants called in ≤10% of all samples and resulting in missense mutations, nonsense mutations, frame-shift mutations (insertions/deletions), or those affecting canonical splice sites were included for further analyses. The control dataset used consists of whole-exome sequencing data derived from 60,706 individuals listed in the Exome Aggregation Consortium (ExAC) database (http://exac.broadinstitute.org, version 0.3). Subsequently, in line with a recessive inheritance pattern, it was determined if two pathogenic *NTHL1* alleles were present. Validation of germline *NTHL1* mutations was performed by Sanger sequencing on probands and available family members. Primer sequences used for validation of variant calls using Sanger sequencing are available upon request.

NTHL1 p.Gln90* genotyping

Considering the high frequency of the p.Gln90* mutation in the Netherlands [2], the p.Gln90* mutation was genotyped in 1,842 Dutch index patients with unexplained colorectal polyposis or familial CRC. A KBioscience Competitive Allele-Specific Polymerase chain reaction (KASPar) assay was performed using DNA extracted from leukocytes or formalin-fixed paraffin embedded (FFPE) surgical specimens according to standard procedures. Two allele-specific forward primers were designed using Primerpicker (see Key Resources Table) (KBioscience, Hoddesdon, UK). Subsequently, the genotyping was carried out using the manufacturer's protocol (KBioscience, Hoddesdon, UK), the PCR was performed in a total reaction volume of 8.11 µl containing 4 µL of 2.5-10 ng/µl of genomic DNA, 0.11 µl of assay mixture (12 µM each allele-specific forward primer and 30 µM reverse primer) and 4 µL of KASP 2X reaction mix. Finally, a thermal cycling program was performed on these samples (available upon request) and data were analyzed using Bio-Rad CFX manager software version 3.0 under the allelic discrimination mode (Bio-Rad, Veenendaal, the Netherlands). If the p.Gln90* mutation was detected in a sample, the entire open reading frame of *NTHL1* was sequenced using Sanger sequencing on tumor DNA as well as DNA isolated from peripheral blood or histologically normal, macrodissected FFPE tissue.

Whole-exome sequencing and bioinformatic analysis

DNA was isolated from 17 primary tumor samples from nine different tissues (Table S4). Exome capturing was performed on genomic DNA derived from peripheral blood cells and (fresh frozen or FFPE) tumor samples using the Agilent SureSelectXT Human All Exon V5 (50Mb) enrichment kit (Agilent Technologies).

Whole-exome sequencing of these libraries was performed using the Illumina HiSeq 4000 sequencing platform (2×100 bp, paired end; BGI, Copenhagen, Denmark and BGI, Hong Kong, China). Since we only had a limited amount of FFPE DNA from P17-II:2, P11-III:4, P11-III:5, and P15:III:2, sample preparation was done using the SureSelectXT^{HS} Target Enrichment System for Illumina Paired-End Multiplexed Sequencing Library (Agilent Technologies). Subsequent exome capture was performed using the Agilent SureSelectXT Human All Exon V6 (50Mb) enrichment kit (Agilent Technologies). Whole-exome sequencing of these libraries was performed using the NextSeq 500 sequencing platform (2×150 bp, paired end). At least a 50-fold coverage was obtained for the libraries generated using DNA derived from peripheral blood cells and a fresh frozen tumor sample, whereas at least a 100-fold read depth was achieved for the libraries obtained from DNA derived from FFPE tumor samples. We only sequenced tumor samples with high tumor purity (>50%) to guarantee the identification of high-quality variants, without tumor admixture correction in the variant calling process.

Sequencing reads with a quality score cutoff of 60 were mapped to the reference genome (UCSC build hg19). Variant calling was performed using UnifiedGenotyper, a robust SNP caller that outperforms in low quality samples. Annotation was performed as described previously [32]. High confident somatic variant calls, i.e. ≥15 fold coverage, with ≥20% or ≤80% variant reads, of the corresponding genomic position in both the tumor and corresponding germline sample, were selected with the same approach as described previously [33]. Subsequently, variant calls observed in our in-house database of germline variants [32], or present with >0.01% in the general population (the ExAC database, version 0.3; the gnomAD database version 2.0) were excluded. Reliability of variant calls was further improved by excluding variants with a quality score below 200 and variants that were shared between tumors of different tissue types of different indexes. Variants were manually checked using the integrative genome viewer (IGV) when subsequent Sanger sequencing revealed that >20% of the randomly selected somatic variants were not validated.

For patient P03-II:3, variants with ≤10% or ≥80% variant reads were excluded. For the patient P08-IV:2, for which we sequenced the squamous cell carcinoma of the tongue tip, matching normal DNA was not available. We identified somatic variants in this sample by using the whole-exome sequence of the normal DNA from the brother (P08-IV:1).

For each tumor, the somatic mutation status of a representative selection of variant calls, of both tumor and germline DNA, was confirmed by Sanger sequencing (Table S4). Somatic mutational signature extraction based on all 96 trinucleotide substitutions [34] was performed using nonnegative matrix factorization [27]. To infer the contribution of the 30 previously identified mutational signatures available at the Catalogue of Somatic Mutations in Cancer (COSMIC)[11], we used the R package DeconstructSigs tool [28]. Control data of somatic mutations from The Cancer Genome Atlas (TCGA) database were used to support signature analyses (Figure S2E).

Molecular and clinical analysis of novel families

Targeted sequencing (n=828) or p.Gln90* genotyping (n=1,842) of individuals with adenomatous polyposis and/or familial CRC revealed four novel unrelated families with biallelic germline *NTHL1* mutations (families 1-4; Table S1).

Family 1: Three brothers with a homozygous p.Gln90* *NTHL1* mutation developed adenomatous polyposis and CRC (Figure S1A). The index patient (P01-II:11) developed CRC twice at age 59, and was subsequently diagnosed with a thyroid cancer. One brother (P01-II:7) also developed urothelial cell cancer (UCC). Notably, a sister carrying a heterozygous p.Gln90* *NTHL1* mutation was also diagnosed with two different tumors.

Family 2: the index patient (P02-II:1; p.Gln90*/Trp269*) developed adenomatous polyposis and CRC (Figure S1B). Both his siblings are deceased and their germline *NTHL1* mutation status is unknown.

Family 3: two sisters, both with a homozygous p.Gln90* *NTHL1* mutation, were diagnosed with adenomatous and hyperplastic polyps. One sister (P03-II:3) developed CRC at age 33, whereas the other sister (P03-II:5, age 41) had no malignancies (Figure S1C).

Family 4: the index patient of family 4 (P04-II:5; p.Gln90*/p.Ile245fs) developed bilateral breast cancer at age 38 and 40, CRC at age 53, and an acute myeloid leukemia at age 59 (Figure S1D).

Seven additional families (numbered 5-11 in this study) were identified independently in different diagnostic or research-based settings, for which a detailed description is given below:

Family 5: The index patient of family 5 (P05-IV:5; p.Gln90*/p.Gln90*) was diagnosed with adenomatous polyps and CRC (Figure S1E), and referred for routine diagnostic testing of relevant polyposis genes (*APC, MUTYH, MSH3, NTHL1, POLD,1* and *POLE*) using a customized add on version of the TruSightTMCancer Sequencing Panel (Illumina, San Diego), including 145 genes for hereditary tumor syndromes on blood-derived DNA from these patients. A homozygous c.268C>T (p.Gln90*) mutation in *NTHL1* was identified and subsequently confirmed by Sanger sequencing.

Family 6: The index patient from family 6 (P06-III:2) was recruited to the Genetics of Colonic Polyposis Study through the Ohio State Medical Centre based on fulfilling WHO criteria 3 for Serrated Polyposis Syndrome. In addition to multiple adenomas, hyperplastic polyps, and CRC, P06-III:2 also developed breast cancer at age 63 (Figure S1F). Blood lymphocyte-derived DNA was tested in a research setting for germline mutations in colonic polyposis-associated genes, including *NTHL1*, using HiPlex (www.HiPLEX.org), a highly multiplexed PCR-based targeted sequencing approach [35, 36]. Compound heterozygous mutations in *NTHL1* (c.235_236insG; p.Ala79Glyfs*2 and c.268C>T; p.Gln90*) were identified and subsequently confirmed by Sanger sequencing. Due to their proximity to each other, both mutations were captured by the same HiPLEX amplicon, and their biallelic nature was confirmed as each read only contained one of the two mutations.

Family 7: The index patient from family 7 (P07-III:3) was a 62-year-old man of Jewish origin, who presented with a positive fecal occult blood test and was found to have multiple adenomatous polyps. Therefore, this patient was referred to the East Anglian Medical Genetics Service, after which blood-derived DNA was sequenced using the TruSight One sequencing panel (Illumina). Two nonsense mutations in *NTHL1* (c.806G>A; p.Trp269* and c.859C>T; p.Gln287*) were identified *in trans* and subsequently validated by Sanger sequencing. The patient also developed a head and neck squamous cell carcinoma (HNSCC) and, later, he was diagnosed with acute myeloid leukemia (Figure S1G).

Family 8: Two brothers of Turkish origin were diagnosed with a HNSCC at the ages of 29 and 24, respectively. Fanconi anemia was suspected based on cisplatin hypersensitivity in one of these brothers, but no mutations affecting any of the Fanconi anemia genes was identified. To identify a causative mutation for the phenotype in the two brothers, whole-exome sequencing on fibroblast-derived DNA from patient P08-IV:1 was performed on a HiSeq2000 platform

(BGI, Copenhagen). Exome capturing was performed with the Agilent SureSelect Human Exon V4 enrichment kit. For sequence alignment and mutation detection, NextGENe Software v.2.3.4.4 (Softgenetics) was used. Following data analysis, a homozygous nonsense mutation in *NTHL1* (c.545G>A; p.Trp182*) was identified. Sanger sequencing confirmed the homozygous mutation in the proband, and demonstrated that his brother and sister, as well as his mother were homozygous for this mutation, illustrating the high degree of consanguinity in this family (Figure S1H). His father carried the mutation in a heterozygous state.

Family 9: The index case from family 9 (P09-III:4) was a breast cancer affected patient, and also developed multiple primary cancers, including CRC, UCC, cervical cancer, and an endocervical premalignancy (Figure S1I). She was ascertained from the Variants in Practice (ViP) Study which is a familial breast cancer cohort of the combined Familial Cancer Centres, Melbourne, Australia. Participants were assessed by a specialist Familial Cancer Clinic before clinical genetic testing for hereditary breast cancer genes. Initially, the patient was tested negative for pathogenic mutations in *BRCA1*, *BRCA2*, and *PALB2*. The coding regions and exon-intron boundaries (10 bp each side) of *NTHL1* were amplified from germline DNA using custom designed HaloPlex Targeted Enrichment Assays (Agilent Technologies, Santa Clara, CA). Subsequently, sequencing was performed on a HiSeq2500 Genome Analyzer (Illumina, San Diego, CA), sequence alignment and variant calling was performed as described previously [37]. To remove likely false positives, called variants were only retained if they had quality score >60 and an overall read depth ≥30, with a minimum of 8 reads and 20% of all reads supporting the alternate allele, as well as no obvious bias in strand of origin. The index case from family 9 was found to be homozygous for the p.Gln90* mutation which was confirmed by Sanger Sequencing.

Family 10: The index patient from family 10 (P10-III:2) was first diagnosed with breast cancer (Figure S1J). She tested negative for pathogenic variants in *BRCA1, BRCA2, CHEK2, PALB2, PTEN,* and *RAD51*C. In a subsequent CT-scan of the abdomen a suspicious finding in the area of the coecum was detected, after which a colonoscopy was performed which revealed adenomatous polyps. Based on this finding, the polyposis genes *APC, MUTYH, MSH3, NTHL1, POLD,1* and *POLE* were tested using a customized add on version of the TruSightTMCancer Sequencing Panel (as for family 5). The nonsense homozygous mutation in *NTHL1* (p.Gln90*) was identified. Sanger sequencing confirmed the homozygous mutation in the index patient as well as in her twin brother.

Family 11: The index patient from family 11 (P11-III:4) was identified as breast cancer patient (Figure S1K). Panel testing of *BRCA1,* and *BRCA2* was performed because of the history of breast cancer. Thereafter, bowel polyps were identified, and a custom Agilent capture array enrichment, including *APC, BMPR1A, CDH1, EPCAM, PALB2, PMS2, POLD1, POLE, PTEN, SMAD4, STK11*, *TP53,* and *NTHL1* was done followed by targeted next generation sequencing. Compound heterozygous nonsense mutations in *NTHL1* (p.Gln90*/p.Tyr130*) were identified. Subsequent Sanger sequencing confirmed the compound heterozygous mutations to be present in the index patient and her affected sister (P11-III:5), who was diagnosed with multiple primary cancers, including colorectal-, breast-, endometrial-, ovarian cancer and a meningioma (Figure S1K).

Collection of clinical and pathological data

For all novel families included in this study, a clinical information sheet was sent to local clinical geneticists and/or pathologists in order to collect detailed information related to the composition of the family including current age or reason of death of all family members, all known diagnoses of malignancies in the family with age of diagnosis, and results from colonoscopies that were performed. When the number and types of polyps identified were reported in the colonoscopy report as 'some', 'several', or 'many', we used the common term 'multiple'.

Quantification and statistical analysis

Statistical parameters including the exact value of n, and statistical significance are reported in the Figure 2A and S2A. Data is judged to be statistically significant when $p < 0.05$ by two-tailed Student's t test. The asterisks denote statistical significance as calculated by Student's t test (***, $p < 0.0001$). Statistical analysis was performed in GraphPad PRISM 5. Cosine similarity scores were calculated using R studio version 3.4.

Calculation of cancer risks

The age-related cumulative lifetime risks (CLTR) for extracolonic malignancy were calculated using Kaplan-Meier analyses. Censoring was applied at age of first extracolonic malignancy, last moment of follow-up information, or death, whichever occurred first. Basal-cell carcinomas were excluded from this analysis, whereas meningiomas were taken into account as they can be lethal. To correct for ascertainment bias, modified segregation analyses (MSA) were performed with maximizing the conditional likelihood of observing the genotypes and phenotypes in each pedigree given the phenotypes of all

relatives in the pedigree, using a population *NTHL1* cumulative mutant allele frequency of 0.003 [17]. CLTRs for extracolonic cancer were calculated based on the estimated age-group specific hazard ratios for biallelic carriers versus non-carriers and heterozygous carriers, for which we assumed no additive risk effect. The cancer risk of non-carriers and heterozygous carriers was assumed to be equal to the cancer incidence in the general population (CCCN). MSA was performed with Mendel [38], and other analyses were performed in R.

Data and software availability

Data resources

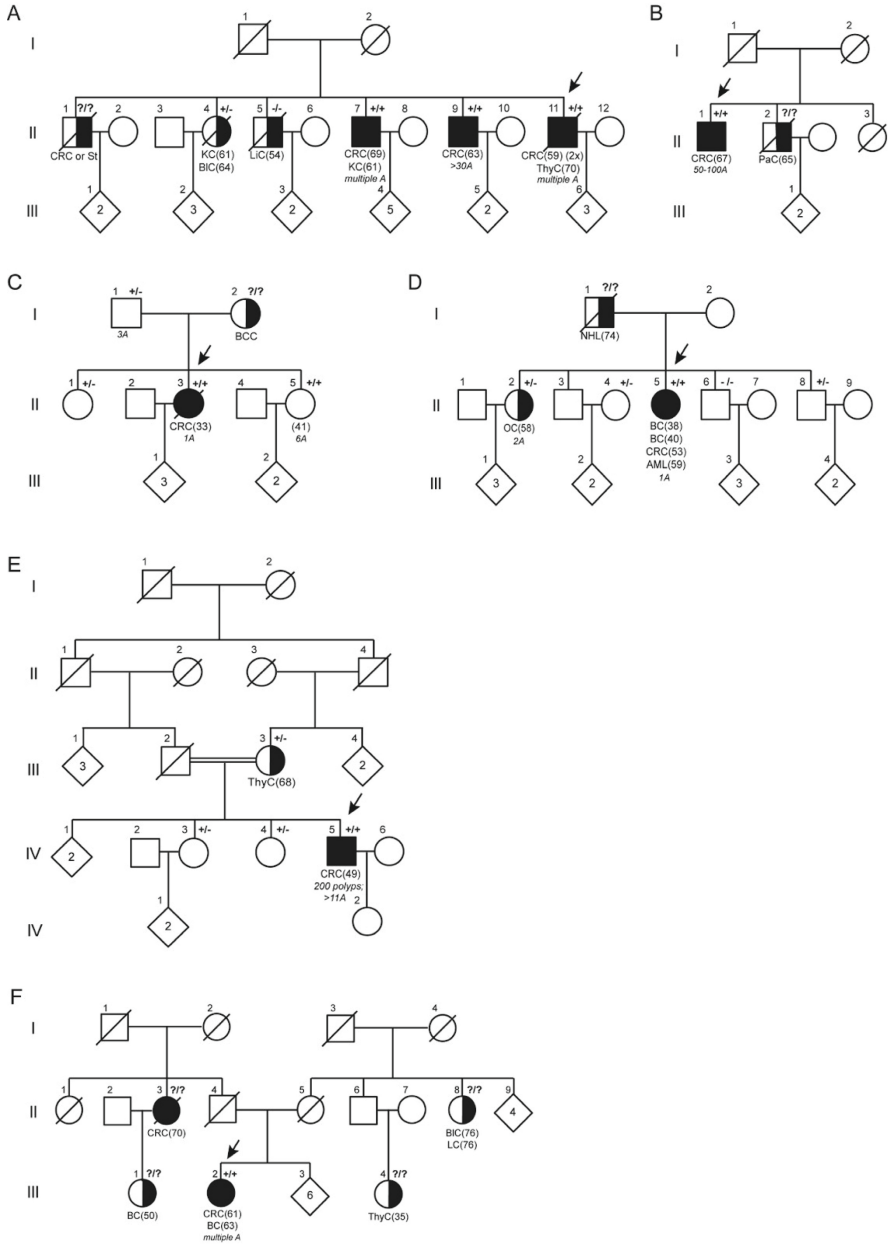
The analyzed whole-exome sequencing data are available in Table S3. The accession number of the raw whole-exome sequencing data reported in this paper is: EGAD00001004534.

References

1. 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.
2. 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.
3. 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.
4. 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.
5. Belhadj S, Mur P, Navarro M, et al. Delineating the Phenotypic Spectrum of the NTHL1-Associated Polyposis. Clin Gastroenterol Hepatol 2017;15:461-462.
6. Briggs S, Tomlinson I. Germline and somatic polymerase epsilon and delta mutations define a new class of hypermutated colorectal and endometrial cancers. J Pathol 2013;230:148-53.
7. Nielsen M, Lynch H, Infante E, et al. MUTYH-Associated Polyposis. In: Pagon RA, Adam MP, Ardinger HH, Wallace SE, Amemiya A, Bean LJH, Bird TD, Ledbetter N, Mefford HC, Smith RJH, Stephens K, eds. GeneReviews(R). Seattle (WA), 1993.
8. Fostira F, Kontopodis E, Apostolou P, et al. Extending the clinical phenotype associated with biallelic NTHL1 germline mutations. Clin Genet 2018;94:588-589.
9. Chubb D, Broderick P, Dobbins SE, et al. Rare disruptive mutations and their contribution to the heritable risk of colorectal cancer. Nat Commun 2016;7:11883.
10. Rivera B, Castellsagué E, Bah I, et al. Biallelic NTHL1 Mutations in a Woman with Multiple Primary Tumors. N Engl J Med 2015;373:1985-6.
11. COSMIC. Signatures probabilities, 2018.
12. 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.
13. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. Nature 2013;500:415-21.
14. Nik-Zainal S, Davies H, Staaf J, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature 2016;534:47-54.
15. Pilati C, Shinde J, Alexandrov LB, et al. Mutational signature analysis identifies MUTYH deficiency in colorectal cancers and adrenocortical carcinomas. J Pathol 2017;242:10-15.
16. 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.
17. Weren RD, Ligtenberg MJ, Geurts van Kessel A, et al. NTHL1 and MUTYH polyposis syndromes: two sides of the same coin? J Pathol 2018;244:135-142.
18. Barrow E, Robinson L, Alduaij W, et al. Cumulative lifetime incidence of extracolonic cancers in Lynch syndrome: a report of 121 families with proven mutations. Clin Genet 2009;75:141-9.
19. Bellido F, Pineda M, Aiza G, et al. POLE and POLD1 mutations in 529 kindred with familial colorectal cancer and/or polyposis: review of reported cases and recommendations for genetic testing and surveillance. Genet Med 2016;18:325-32.
20. Kempers MJ, Kuiper RP, Ockeloen CW, et al. Risk of colorectal and endometrial cancers in EPCAM deletion-positive Lynch syndrome: a cohort study. Lancet Oncol 2011;12:49-55.

21. Vogt S, Jones N, Christian D, et al. Expanded extracolonic tumor spectrum in MUTYH-associated polyposis. Gastroenterology 2009;137:1976-85.e1-10.
22. Watson P, Vasen HF, Mecklin JP, et al. The risk of extra-colonic, extra-endometrial cancer in the Lynch syndrome. Int J Cancer 2008;123:444-9.
23. Win AK, Cleary SP, Dowty JG, et al. Cancer risks for monoallelic MUTYH mutation carriers with a family history of colorectal cancer. Int J Cancer 2011;129:2256-62.
24. Win AK, Dowty JG, Cleary SP, et al. Risk of colorectal cancer for carriers of mutations in MUTYH, with and without a family history of cancer. Gastroenterology 2014;146:1208-11. e1-5.
25. NCCN. Clinical Practice Guidelines in Oncology, Genetic/Familial High-Risk Assessment: Colorectal.
26. Netherlands CCCt. Dutch cancer incidence, 2018.
27. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. BMC Bioinformatics 2010;11:367.
28. Rosenthal R, McGranahan N, Herrero J, et al. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol 2016;17:31.
29. Team RC. R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing). 2016.
30. O'Roak BJ, Vives L, Fu W, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science 2012;338:1619-22.
31. Boyle EA, O'Roak BJ, Martin BK, et al. MIPgen: optimized modeling and design of molecular inversion probes for targeted resequencing. Bioinformatics 2014;30:2670-2.
32. de Voer RM, Hahn MM, Weren RD, et al. Identification of Novel Candidate Genes for Early-Onset Colorectal Cancer Susceptibility. PLoS Genet 2016;12:e1005880.
33. de Ligt J, Willemsen MH, van Bon BW, et al. Diagnostic exome sequencing in persons with severe intellectual disability. N Engl J Med 2012;367:1921-9.
34. Lawrence M, Huber W, Pagès H, et al. Software for computing and annotating genomic ranges. PLoS Comput Biol 2013;9:e1003118.
35. Nguyen-Dumont T, Pope BJ, Hammet F, et al. A high-plex PCR approach for massively parallel sequencing. Biotechniques 2013;55:69-74.
36. Nguyen-Dumont T, Teo ZL, Pope BJ, et al. Hi-Plex for high-throughput mutation screening: application to the breast cancer susceptibility gene PALB2. BMC Med Genomics 2013;6:48.
37. Li N, Thompson ER, Rowley SM, et al. Reevaluation of RINT1 as a breast cancer predisposition gene. Breast Cancer Res Treat 2016;159:385-92.
38. Lange K, Weeks D, Boehnke M. Programs for Pedigree Analysis: MENDEL, FISHER, and dGENE. Genet Epidemiol 1988;5:471-2.

Supplementary information



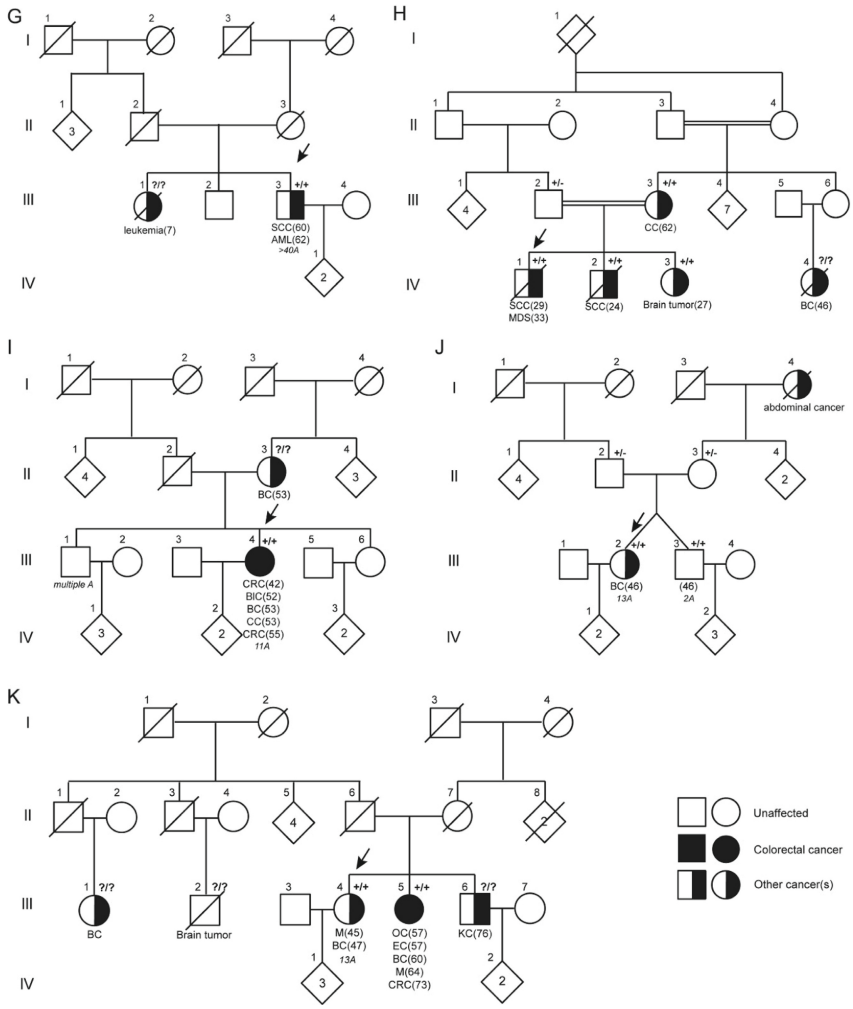


Figure S1, related to Figure 1 and Table 1. Pedigrees of 11 novel families with biallelic germline *NTHL1* mutations.

Depicted pedigrees represent the families of *NTHL1* index patients (A) P01-II:11, (B) P02-II:1, (C) P03-II:3, (D) P4-II:5, (E) P05-IV:5, (F) P06-III:2, (G) P07-III:3, (H) P08-IV:1, (I) P09-III:4, (J) P10-III:2, (K) P11-III:4. Individuals tested positive for a homozygous *NTHL1* mutation are indicated with '$^{+/+}$'. '$^{+/-}$' indicates heterozygous carriers of an *NTHL1* mutation. '$^{-/-}$' are non-carriers, and individuals indicated with '?/?' are not tested affected individuals. Abbreviations used are: AML[a]: acute myeloid leukemia, BC: breast cancer, BlC[b]: bladder cancer, LC: lung cancer, CC[c]: cervical cancer, CRC: colorectal cancer, EC[d]: endometrium cancer, KC[b]: kidney cancer, LiC: liver cancer, M[e]: meningioma, MDS[a]: myelodysplatic syndrome, NHL[a]: non-Hodgkin lymphoma, OC: ovarian cancer, PaC: pancreas cancer, SCC: squamous cell carcinoma, St: stomach cancer, and ThyC: thyroid cancer. Numbers correspond to age of onset. Number of colorectal adenomatous polyps (A) present at time of diagnosis are in italics. Diamonds represent pooled individuals with no report of cancer. [a]Classified as hematologic malignancies, [b]Classified as urothelial cell cancer. [c]Classified as cervical (pre)malignancies. [d]Classified as endometrial (pre)malignancies. [e]Classified as brain tumors.

Table S1, related to Table 1. Novel families with individuals with biallelic germline *NTHL1* mutations

Family	Ethnic origin[a]	cDNA change (NM _002528.6)	Amino Acid change	# individuals with biallelic NTHL1 mutations	Ascertainment	Identification method
Fam1	NL	c.268C>T	p.Gln90*	3	Polyposis and CRC	Cohort screening, this study
Fam2	MK	c.268C>T c.806G>A	p.Gln90* p.Trp269*	1	Polyposis and recessive inheritance	Cohort screening, this study
Fam3	NL	c.268C>T	p.Gln90*	2	Young CRC	Cohort screening, this study
Fam4	NL	c.268C>T c.733dup	p.Gln90* p.Ile245Asnfs*28	1	CRC	Cohort screening, this study
Fam5	QA	c.268C>T	p.Gln90*	1	Polyposis and CRC	Diagnostics
Fam6	US	c.268C>T c.235_236insG	p.Gln90* p.Ala79Glyfs*2	1	Polyposis and CRC	Study cohort
Fam7	UK	c.806G>A c.859C>T	p.Trp269* p.Gln287*	1	Polyposis	Diagnostics
Fam8	TR	c.545G>A	p.Trp182*	4	Suspected Fanconi anemia	Diagnostics
Fam9	AU	c.268C>T	p.Gln90*	1	BC	Study cohort
Fam10	KZ	c.268C>T	p.Gln90*	2	BC (polyposis)*	Diagnostics
Fam11	UK	c.268C>T c.390G>A	p.Gln90* p.Tyr130*	2	BC (polyposis)*	Diagnostics

[a] NL: Netherlands, MK: Macedonia, QA: Qatar, US: United States of America, UK: United Kingdom, KZ: Kazakhstan, TR: Turkey, AU: Australia. *Patients were identified as breast cancer patients and tested for *NTHL1* upon the diagnosis of polyps.

Table S2, related to Table 1. Treatment details of 16 NTHL1-deficient individuals with multiple primary tumors

Patient ID[a]	M/F	Cancer (age)[b]	Treatment information as far as available[c]	Previous treatment related to current malignancy[d]
P01-II:11	M	CRC (59)	Resection	-
		ThyC (70)	Resection and radiotherapy	No
P01-II:7	M	Renal pyelum cancer (61)	Resection	-
		CRC (69)	Resection	No
P04-II:5	F	BC (right, 38)	Resection and radiotherapy	-
		—(39)	Hysterectomy because of hypermenorrhoea	-
		BC (left, 40)	Resection	Very unlikely

Table S2, (continued)

Patient ID[a]	M/F	Cancer (age)[b]	Treatment information as far as available[c]	Previous treatment related to current malignancy[d]
		CRC (53)	Resection and chemotherapy	Very unlikely
		AML (59)	Chemotherapy	Yes, may be related to previous chemotherapy (53)
P06-III:2	F	CRC (61)	Resection	-
		BC (63)	Resection	No
P07-III:3	M	SCC of the parotid gland (60)	Resection and regional radiotherapy	-
		AML (62)	Chemotherapy	Very unlikely
P08-IV:1	M	SCC of the mouth base (29)	Adjuvant regional radiotherapy	-
		MDS (33)		Very unlikely
P09-III:4	F	CRC (42)	Resection and chemoradiotherapy	-
		BC (47)	Resection and endocrine therapy (Tamoxifen)	Very unlikely
		BlC (52)	Resection	Very unlikely
		Endocervical adenocarcinoma in situ (52)	Resection	Very unlikely
		EC (53)	Resection	Possible, but unlikely, after less than 5 years use of Tamoxifen
		CRC(55)	NA	Very unlikely
P11-III:5		OC (57)	Resection	-
		EC (57)	Resection	-
		BC (60)	Resection, radiotherapy and endocrine therapy (Tamoxifen)	Very unlikely
		CRC (73)	Resection	Very unlikely
P12-01 (Weren et al., 2015) [1]	M	CRC (40)	Resection	-
		CRC (49)	Resection	No
		PC (60)	NA	No
P12-49 (Weren et al., 2015) [1]	F	Endometrial complex hyperplasia (46)	NA	-
		Non-Hodgkin lymphoma (65)	NA	No
P13-07 (Weren et al., 2015) [1]	M	CRC (47)	Radiotherapy and resection	-
		PaC (47)	NA	Very unlikely
		DC (52)	NA	Unlikely

Table S2, (continued)

Patient ID[a]	M/F	Cancer (age)[b]	Treatment information as far as available[c]	Previous treatment related to current malignancy[d]
P13-71 (Weren et al., 2015) [1]	F	BCC (55)	NA	-
		BC (56)	NA	No
		EC (57)	NA	Very unlikely
P14-23 (Weren et al., 2015) [1]	F	CRC (64, 64, 64)	Resection	-
		EC (74)	Resection	No
P14-69 (Weren et al., 2015) [1]	M	CRC (63, 63)	Resection	-
		BCC (63, 63, 63)	NA	No
		Non-Hodgkin lymphoma (70)	NA	Very unlikely
P15-III:2 (Rivera et al., 2015) [2]	F	CRC (41)	NA	-
		BlC (47)	NA	Very unlikely
		BCC (52)	NA	Unlikely
		SCC of head and neck (55)	NA	Unlikely
		BC (58)	NA	Unlikely
P17-II:2 (Belhadj et al., 2017) [3]	F	BC (47)	NA	-
		BC (50)	Resection (bilateral)	Unlikely
		BlC (66)	NA	Unlikely
		CRC (67, 67, 67)	NA	Unlikely

[a]The index patient is underlined. [b]AML: acute myeloid leukemia, BC: breast cancer, BCC: basal-cell carcinoma, BlC: bladder cancer, CC: cervical cancer, CRC: colorectal cancer, DC: duodenal cancer, EC: endometrium cancer, MDS: myeloidysplastic syndrome, OC: ovarian cancer, PaC: pancreas cancer, PC: prostate cancer, SCC: squamous cell carcinoma, ThyC: thyroid cancer. Numbers represent age of diagnosis. [c]Information may be incomplete. NA: treatment information not available. [d]No; if no chemo- or radiotherapy was administered before tumor diagnosis. (Very) unlikely; if time to previous chemo- or radiotherapy was very short, previous treatment was only administered locally, or there is no known link between previous treatment(s) and the induction of the specific malignancy.

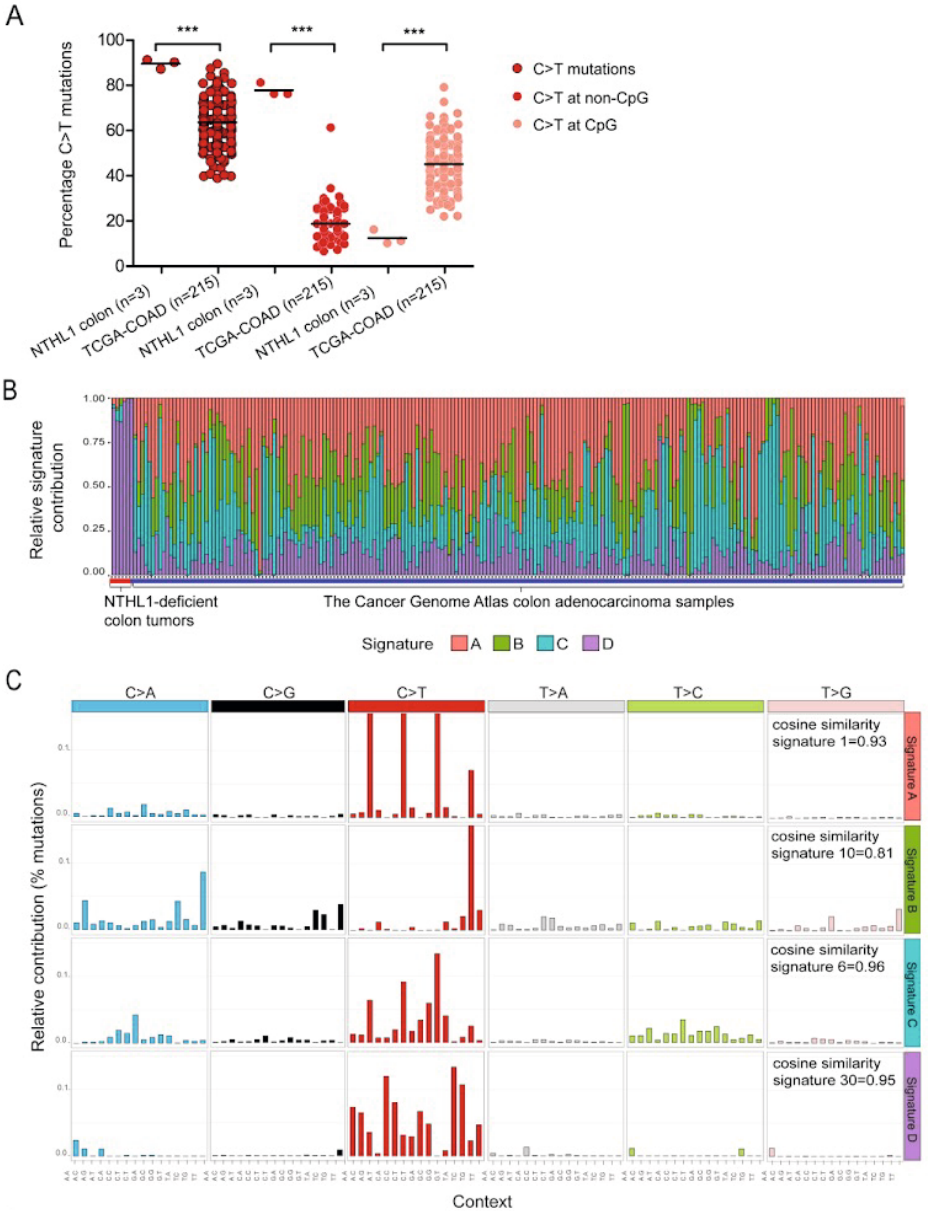
Table S3, related to Figure 2. Analyzed somatic mutations from NTHL1-deficient tumors

Supplementary Table S3 can be found with this article online at doi: 10.1016/j.ccell.2018.12.011.

Table S4, related to Figure 2. NTHL1-deficient tumors used for signature analysis and validated somatic mutations

Sample	Patient	Tumor type	Tumor material	DNA used for whole exome sequencing	Number of somatic mutations*	C>T	C>A	C>G	T>A	T>C	T>G
A-1[a]	P14-69	Adenomatous polyp	FFPE	No, targeted sequencing	13	NA	NA	NA	NA	NA	NA
CRC-1[a]	P14-23	Colorectal cancer	FFPE	No, targeted sequencing	15	NA	NA	NA	NA	NA	NA
CRC-2[a]	P13-07	Colorectal cancer	FFPE	No, targeted sequencing	17	NA	NA	NA	NA	NA	NA
A-2	P01-II:7	Adenomatous polyp	FFPE	Yes	153	133 (9/9)	9 (2/2)	3	3 (2/2)	3	2
CRC-3	P01-II:7	Colorectal cancer	FFPE	Yes	360	19 (1/2)	0	1	0	1 (1/1)	0
CRC-4	P03-II:3	Colorectal cancer	fresh-frozen	Yes	21	334 (15/15)	15 (1/1)	9 (1/1)	1	7	2
BC-1	P13-71	Breast cancer	FFPE	Yes	32	27 (5/7)	2	3	1	12 (0/2)	0
BC-2	P17-II:2	Breast cancer	FFPE	Yes	49	41 (12/16)	2 (1/1)	0	2 (1/1)	7 (0/2)	3
BC-3	P11-III:4	Breast cancer	FFPE	Yes	55	51 (10/10)	1	0	0	3	0
BC-4	P11-III:5	Breast cancer	FFPE	Yes	89	70 (9/9)	4	8	4	5	4 (1/1)
M	P12-49	Meningioma	FFPE	Yes	23	13 (6/6)	2 (1/1)	4 (2/2)	1 (1/1)	2 (1/1)	1 (1/1)
EC-1	P13-71	Endometrial cancer	FFPE	Yes	69	66 (12/15)	5 (1/1)	1	3	4	1
EC-2[b]	P12-49	Endometrial cancer	FFPE	Yes	7	NA	NA	NA	NA	NA	NA
UCC-1	P01-II:7	Urothelial cell cancer	FFPE	Yes	1331	1180 (11/11)	19	97 (1/1)	3	19	13
UCC-2	P17-II:2	Urothelial cell cancer	FFPE	Yes	53	42 (11/13)	1 (0/1)	4 (1/1)	1	7 (2/2)	1
UCC-3	P15-III:2	Urothelial cell cancer	FFPE	Yes	64	55 (3/4)**	1	2	1	5 (1/1)	0
HNSCC[c]	P08-IV:2	Head and neck squamous cell carcinoma	FFPE	Yes	169	79 (6/6)	13	15	13	36	13 (1/1)
ThyC	P01-II:11	Thyroid cancer	FFPE	Yes	36	27 (7/7)	1	3	2	2	1
NF[b]	P01-II:7	Neurofibroma	FFPE	Yes	9	NA	NA	NA	NA	NA	NA
PaC[b]	P13-07	Pancreatic cancer	FFPE	Yes	9	NA	NA	NA	NA	NA	NA

FFPE: formalin-fixed paraffin embedded; [a]Samples previously sequenced by Weren et al., Nature Genetics, 2015. [b]From these tumors we were unable to retrieve at least 10 somatic mutations to perform mutational signature analyses. [c]There was no germline DNA available of patient P08-IV:2, instead WES on germline DNA from the brother was used for somatic variant calling (see also STAR Methods). For each point mutation (C>T; C>A; C>G; T>A; T>C; T>G), the total number of mutations identified in each sample is indicated. Shown between brackets is the number of mutations that were confirmed from the subset selected for validation. *The number of somatic mutations after validation and manual check of variants with IGV. **Variant could not be validated due to the poor quality of the DNA. Abbreviations used for sample types are A: colorectal adenomatous polyps, BC: breast cancer, CRC: colorectal cancer, EC: endometrium cancer, HNSCC: head and neck squamous cell carcinoma, M: meningioma, PaC: pancreas cancer, ThyC: thyroid cancer, UCC: urothelial cell carcinoma.


A
Percentage C>T mutations
100
80
60
40
20
0
***
***
***
C>T mutations
C>T at non-CpG
C>T at CpG
NTHL1 colon (n=3)
TCGA-COAD (n=215)
NTHL1 colon (n=3)
TCGA-COAD (n=215)
NTHL1 colon (n=3)
TCGA-COAD (n=215)
B
Relative signature contribution
1.00
0.75
0.50
0.25
0.00
NTHL1-deficient colon tumors
The Cancer Genome Atlas colon adenocarcinoma samples
Signature A B C D
C
C>A
C>G
C>T
T>A
T>C
T>G
Relative contribution (% mutations)
cosine similarity signature 1=0.93
cosine similarity signature 10=0.81
cosine similarity signature 6=0.96
cosine similarity signature 30=0.95
Signature A
Signature B
Signature C
Signature D
Context

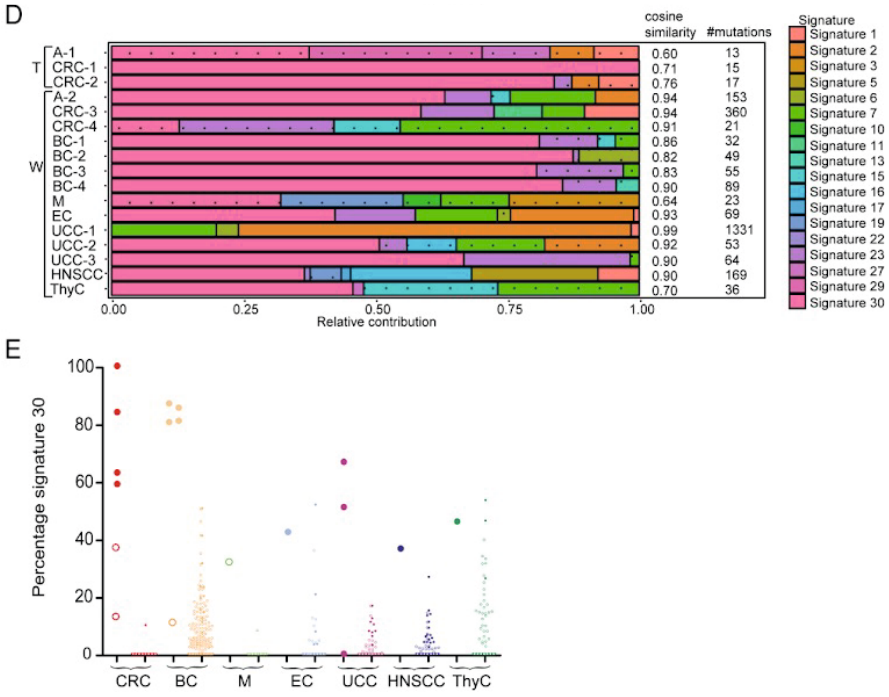


Figure S2, related to Figure 2. Mutation spectrum and mutational signature analysis of NTHL1-deficient colon tumors and sporadic TCGA colon adenocarcinoma samples.

(A) Percentage C>T mutations of colon tumors derived from individuals with a biallelic germline *NTHL1* mutation and from TCGA colon adenocarcinoma samples. *** denotes a p-value < 0.0001. (B) The relative contribution of each sample to the *de novo* extracted signatures. Included are six colon tumors from NTHL1-deficient patients (left) and 215 sporadic CRCs from TCGA (right). (C) The relative contribution of the four *de novo* extracted mutational signatures. The cosine similarity between the extracted signatures and signature 1, 10, 6, and 30 is given in each plot, respectively. (D) The estimated relative contribution of all known COSMIC mutational signatures, including signature 30 (pink), to mutations in 3 targeted and 14 whole-exome sequenced tumors from 11 individuals with biallelic germline *NTHL1* mutations (see also Tables S4 and S5). Signatures with a contribution of less than 10 mutations are indicated by spotted bars. Cosine similarity scores on the right indicate the closeness of the reconstructed mutation profile with the original mutation spectra observed in these tumors. CRC: colorectal cancer, BC: breast cancer, M: meningioma, EC: endometrium cancer, UCC: urothelial cell carcinoma, HNSCC: head and neck squamous cell carcinoma, ThyC: thyroid cancer. T: targeted sequenced tumors, W: whole-exome sequenced tumors. (E) Contribution to signature 30 in NTHL1-deficient tumors and corresponding tumors from TCGA. CRC: colorectal cancer, BC: breast cancer, M: meningioma, EC: endometrial cancer, UCC: urothelial cell cancer, HNSCC: head and neck squamous cell carcinoma, ThyC: thyroid cancer. These respectively correspond to TCGA data COAD (access date 01-09-2017), BRCA (access date 05-29-2017), LGG (access date 05-29-2017), UCEC (access date 05-29-2017), BLCA (access date 01-09-2017), HNSC (access date 01-09-2017), and THCA (access date 01-09-2017). Only samples containing a total of more than 10 mutations are plotted, which was the case in 215, 761, 14, 282, 405, 510, and 417 samples for COAD, BRCA, LGG, UCEC, BLCA, HNSC, and THCA, respectively. NTHL1-deficient tumors are plotted in large dots, whereas TCGA tumors are plotted in smaller dots. Closed dots represent samples with a contribution of more than 10 mutations for signature 30, whereas open dots represent samples with a contribution of less than 10 mutations for signature 30.

Table S5, related to Table 2. The cumulative lifetime risk for extracolonic malignancies for individuals with biallelic germline *NTHL1* mutations

	Cumulative risk extracolonic lifetime	
Age	MSA	KM
20	0 (95%CI 0-0)	0 (95%CI 0-0)
30	9 (95%CI 0-17)	10 (95%CI 3-29)
40	11 (95%CI 0-21)	14 (95%CI 5-33)
50	24 (95%CI 4-36)	32 (95%CI 17-54)
60	44 (95%CI 6-56)	55 (95%CI 35-78)

Shown are the estimated risks to develop extracolonic cancer (excluding basal-cell carcinomas), based on Kaplan-Meier (KM) and modified segregation analyses (MSA) when accounting for ascertainment bias.

Table S6, related to STAR methods. Patient cohort inclusion and results

Cohort	# samples	Selection Criteria[b]	Sequencing technique	Genes tested negative	Biallelic germline mutations in *NTHL1*
Skopje, Macedonia	12	Polyposis, recessive inheritance	MIP-based NGS	MMR genes, *APC*, *TP53*, *MUTYH*, *POLE*, *POLD1*	p.Gln90*/p. Trp269* (CH)
Nijmegen, the Netherlands	169	Polyposis or familial CRC	MIP-based NGS	*APC*, *MUTYH*	p.Gln90* (hom)
Nijmegen, the Netherlands	348	Polyposis or familial CRC	KASPAR assay p.Gln90*	*APC*, *MUTYH*, *POLE*, *POLD1*, MMR genes	none
Dresden, Germany	105	Polyposis or familial CRC	MIP-based NGS	*APC*, *MUTYH*	none
Oxford, United Kingdom	273	Polyposis	MIP-based NGS	*APC*, *MUTYH*	none
Szczecin, Poland	145	Familial CRC	MIP-based NGS	*POLE*, *POLD1*, MMR genes*	none
Santiago de Compostela, Spain	36	Polyposis or familial CRC	MIP-based NGS	*APC*, *MUTYH* (partly), *POLE*, *POLD1, BMPR1A, SMAD4, PTEN*	none
Trondheim, Norway	61	Polyposis or familial CRC	Sanger Sequencing	*APC*, *MUTYH*, MMR genes	none
Bergen, Norway	27	Polyposis or familial CRC	Sanger Sequencing	*APC*, *MUTYH*, MMR genes*	none
Leiden, Netherlands	1,494[a]	Polyposis, or familial/young CRC	KASPAR assay p.Gln90* Sanger Sequencing	*APC*, *MUTYH*, *POLE*, *POLD1*, MMR genes	p.Gln90* (hom) p.Gln90*/p. Ile245Asnfs*28 (CH)
Total	2,670				

CH: compound heterozygous, hom: homozygous, MMR genes: *MLH1*, *MSH2*, *MSH6*, and *PMS2*. [a]DNA from either leukocytes or formalin-fixed paraffin embedded (FFPE) surgical specimen. Validations were performed on tumor DNA as well as DNA isolated from peripheral blood or histologically normal, macrodissected FFPE tissue. [b]Polyposis is defined as the cumulative occurrence of at least 10 polyps. Familial CRC is defined as the proband having a CRC ≤50 years of age and at least one first degree relative with CRC ≤60 years of age. Young CRC is defined as CRC at an age ≤40 years of age. *Most patients were tested for these genes.

References

1. 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.
2. Rivera B, Castellsagué E, Bah I, et al. Biallelic NTHL1 Mutations in a Woman with Multiple Primary Tumors. N Engl J Med 2015;373:1985-6.
3. Belhadj S, Mur P, Navarro M, et al. Delineating the Phenotypic Spectrum of the NTHL1-Associated Polyposis. Clin Gastroenterol Hepatol 2017;15:461-462.

Chapter 6

Monoallelic *NTHL1* loss-of-function variants and risk of polyposis and colorectal cancer

Fadwa A. Elsayed*, Judith E. Grolleman*, Abiramy Ragunathan*, *NTHL1* study group, Daniel D. Buchanan#, Tom van Wezel#, Richarda M. de Voer#

*Authors share co-first authorship
#Authors share co-senior authorship

Gastroenterology, 2020; 159(6): 2241-2243.e6

Introduction

The endonuclease III-like protein 1, encoded by *NTHL1*, is a bifunctional glycosylase involved in base-excision repair (BER) that recognizes and removes oxidized pyrimidines [1]. Similar to biallelic loss-of-function (LoF) variants in *MUTYH* [2], biallelic LoF variants in *NTHL1* predispose to colorectal polyps and colorectal cancer (CRC) [3]. Recently, a multitumor phenotype was observed in individuals diagnosed with NTHL1 deficiency [4]. Carriers of monoallelic pathogenic variants in *MUTYH* have an increased, albeit small, risk of CRC [5]. Thus far, it is unknown if monoallelic *NTHL1* LoF variants also increase the risk of polyposis and/or CRC. This information is especially important for carriers of the most common LoF variant in *NTHL1* (p.(Gln90*); NM_002528.5), which is heterozygous in approximately 0.28% of the general population [6]. Identification of monoallelic *NTHL1* LoF variants currently presents a clinical conundrum regarding how best to counsel carriers with respect to their cancer risk because of the lack of published evidence. Here, we show that monoallelic LoF variants in *NTHL1* are not enriched in individuals with polyposis and/or CRC compared to the general population. Furthermore, 13 colorectal tumors from *NTHL1* LoF carriers did not show a somatic second hit, and we did not find evidence of a main contribution of mutational signature SBS30, the signature associated with NTHL1 deficiency, suggesting that monoallelic loss of *NTHL1* does not substantially contribute to colorectal tumor development.

Methods

A total of 5,942 individuals with unexplained polyposis, familial CRC, or sporadic CRC at young age or suspected of having Lynch syndrome with CRC or multiple adenomas were included in this study and defined as case patients (individual studies and their ascertainment are described in Supplementary Methods and Supplementary Table 1). Three independent data sets were used as controls, including (1) the non-Finnish European subpopulation of the genome aggregation database (gnomAD: n = 64,328) [6], (2) a Dutch cohort of individuals without a suspicion of hereditary cancer who underwent whole-exome sequencing (WES) (Dutch WES; n = 2,329) [7], and (3) a population-based and cancer-unaffected cohort from the Colon Cancer Family Registry Cohort (CCFRC; n = 1,207) (Supplementary Methods and Supplementary Table 1).

Pathogenic *NTHL1* LoF variants were identified in case patients by sequencing the exonic regions of *NTHL1* (n = 3,439) or by genotyping of 2 LoF variants in

NTHL1 (c.268C>T, p.(Gln90*); n = 2503 and c.806G>A, p.(Trp269*); n = 261) (Supplementary Table 1). For control individuals, all pathogenic LoF variants were retrieved from gnomAD and the Dutch WES-cohort [6,7], and for the CCFRC control individuals, the exonic regions of *NTHL1* were sequenced (Supplementary Table 1). Odds ratios between case patients and control groups were calculated and a Fisher exact test was performed to assess the significance of difference in carrier rates. Cosegregation analysis was performed by using Sanger sequencing. Two adenomas and 11 primary CRCs from *NTHL1* LoF variant carriers were subjected to WES, and subsequently, mutational signature analysis was performed (Supplementary Methods and Supplementary Table 2). For signature analysis comparison, we included 3 CRCs from individuals with a biallelic *NTHL1* LoF variant.

Results

Monoallelic *NTHL1* LoF variants were identified in 11 of 3,439 case patients (0.32%) and in 5 of 1,207 (0.41%) of CCFRC control individuals, indicating no significant difference (*P* = .784) (Figure 1A, Supplementary Table 1). Genotyping of the *NTHL1* p.(Gln90*) variant in another 2,503 case patients identified 7 additional carriers (0.28%). The overall frequency of *NTHL1* p.(Gln90*) in case patients was not different from the frequency in the gnomAD (17/5,942 vs 250/64,328; *P* = .914), CCFRC (17/5,942 vs 3/1,207; *P* = .556) or Dutch WES control individuals (17/5,942; vs 17/2,329; *P* = .998) (Figure 1A and Supplementary Table 1).

Via cosegregation analysis, we identified 3 additional *NTHL1* p.(Gln90*) carriers. The phenotype of all carriers identified in this study is described in Supplementary Table 2. Thirteen colorectal tumors from *NTHL1* LoF carriers underwent WES (details in Supplementary Table 2). The *NTHL1* wild-type allele was unaffected by somatic mutations or loss of heterozygosity in all tumors tested. In contrast to *NTHL1*-deficient tumors, in none of the tumors of the carriers was mutational signature SBS30 the main signature, because it was only present in 1 tumor, where it had a minor contribution (Figure 1B and Supplementary Table 2) [4]. These observations indicate that biallelic inactivation of *NTHL1* through a somatic second hit was not evident and that monoallelic inactivation of *NTHL1* was insufficient to result in the accumulation of somatic mutations that are characteristic of an *NTHL1*-deficiency phenotype.

A

	Monoallelic *NTHL1* LoF variant carriers (*n* = 11/3,439)		
	OR	95% CI	P-value
gnomAD non-Finnish European (*n* = 311/64,328)	0.66	0.36-1.21	0.939
Colon Cancer Family Registry Cohort controls (*n* = 5/1,207)	0.77	0.27-2.22	0.784
Dutch WES controls (*n* = 17/2,329)	0.44	0.20-0.93	0.991

	Monoallelic *NTHL1* p.(Gln90*) carriers (*n* = 17/5,942)		
	OR	95% CI	P-value
gnomAD non-Finnish European (*n* = 250/64,328)	0.74	0.40-1.20	0.914
Colon Cancer Family Registry Cohort controls (*n* = 3/1,207)	1.15	0.34-3.94	0.556
Dutch WES controls (*n* = 17/2,329)	0.39	0.20-0.77	0.998

B

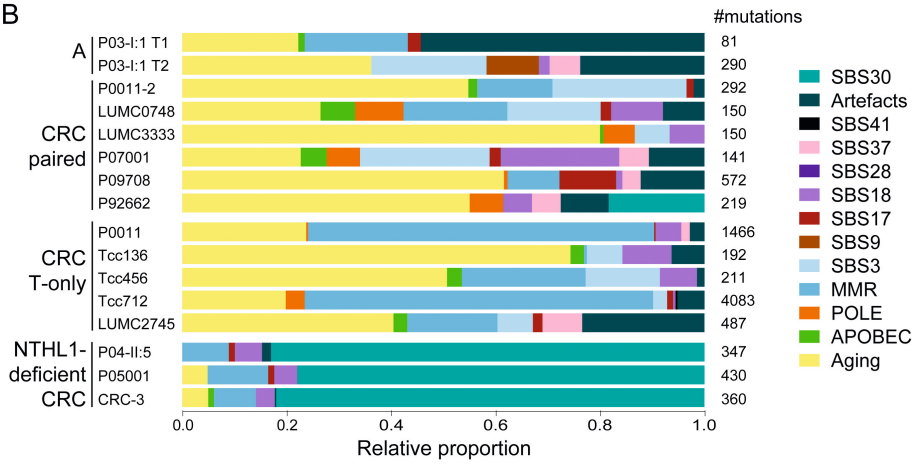


Figure 1. Enrichment and mutational signature analysis of *NTHL1* LoF variants in individuals with polyposis and/or CRC (case patients).

(A) Frequencies of germline monoallelic *NTHL1* LoF variants and monoallelic *NTHL1* p.(Gln90*) variants in individuals with polyposis and/or CRC (case patients) compared with control populations. (B) Mutational signature analysis of tumors from carriers with a monoallelic *NTHL1* LoF variant. Mutational signatures with shared etiologies were grouped for display purposes, which are the signatures associated with aging (SBS1, SBS5 and SBS40), DNA mismatch repair deficiency (SBS6, SBS15, SBS20, SBS21, SBS26 and SBS44), Polymerase Epsilon (POLE) exonuclease domain deficiency (SBS10a and SBS10b), Apolipoprotein B mRNA editing enzyme (APOBEC) activity (SBS2 and SBS13), and artifact signatures (SBS45, SBS51, SBS52, SBS54 and SBS58). Data availability: paired: tumor and normal or tumor data were available; T-only: only data from 1 tumor tissue were available. A, adenomatous polyp; CI, confidence interval; OR, odds ratio.

Discussion

In this study, the largest investigating monoallelic LoF variants in *NTHL1* to date to our knowledge, we observed no evidence of an association between carriers and the risk of polyposis and/or CRC. In our case patients, the prevalence of

pathogenic *NTHL1* LoF variant alleles is comparable to that of the general population. However, we cannot rule out that a small risk for CRC, similar to what is observed for *MUTYH* carriers, still exists.

Colorectal tumors from monoallelic *NTHL1* LoF variant carriers did not show evidence of a somatic second hit in *NTHL1* nor of defective base-excision repair, which is typically associated with biallelic *NTHL1* inactivation. Only 1 tumor showed a minor SBS30 contribution to the mutation profile, but this contribution was far less significant compared to NTHL1-deficient CRC and is likely the result of multiple testing correction. Our data suggest that inactivation of the *NTHL1* wild-type allele is a rare event in colorectal tumors, which is in agreement with the observation that loss of heterozygosity of chromosome arm 16p is not frequently observed in CRC [8]. We were unable to discriminate between individuals with polyposis or CRC due to the historical nature of the case collections. Therefore, differences in frequencies of monoallelic *NTHL1* LoF variants between control individuals and these 2 phenotypes were not made separately. However, because we identified *NTHL1* LoF variants in individuals with polyposis or CRC, we do not consider a major difference between these 2 phenotypes. Because NTHL1 deficiency may also predispose to extracolonic tumors, the risk for these tumor types in monoallelic *NTHL1* carriers still needs further assessment.

In conclusion, the evidence to date does not support an increased risk of polyposis and/or CRC for carriers of monoallelic *NTHL1* LoF variants, and consequently, no additional surveillance is currently warranted beyond population screening for CRC, unless family history characteristics point to a reason for colonoscopy.

Acknowledgements:

The authors thank all study participants, the CCFRC and staff, and the Dutch Parelsnoer Institute Biobank Hereditary Colorectal Cancer for their contributions to this project. Furthermore, we would like to thank Robbert Weren, Eveline Kamping, M. Elisa Vink-Börger, Riki Willems, Christian Gillissen, Peggy Manders, Dina Ruano, Ruud van der Breggen, Marina Ventayol, Sanne ten Broeke, Allyson Templeton, Maggie Angelakos, members of the Colorectal Oncogenomics Group, Sharelle Joseland, Susan Preston, Julia Como, Thomas Green, Magda Kloc and Chris Cotsopoulos for their contributions to this project. The author(s) would further like to acknowledge networking support by the Cooperation in

Science and Technology Action CA17118, supported the European Cooperation in Science and Technology.

NTHL1 study group: Arnoud Boot, Marija Staninova Stojovska, Khalid Mahmood, Mark Clendenning, Noel de Miranda, Dagmara Dymerska, Demi van Egmond, Steven Gallinger, Peter Georgeson, Nicoline Hoogerbrugge, John L. Hopper, Erik A.M. Jansen, Mark A. Jenkins, Jihoon E. Joo, Roland P. Kuiper, Marjolijn J.L. Ligtenberg, Jan Lubinski, Finlay A. Macrae, Hans Morreau, Polly Newcomb, Maartje Nielsen, Claire Palles, Daniel J. Park, Bernard J. Pope, Christophe Rosty, Clara Ruiz Ponte, Hans K. Schackert, Rolf H. Sijmons, Ian P. Tomlinson, Carli M. J. Tops, Lilian Vreede, Romy Walker, Aung K. Win, Colon Cancer Family Registry Cohort Investigators, Aleksandar J. Dimovski, Ingrid M. Winship.

References

1. Krokan HE, Bjørås M. Base excision repair. Cold Spring Harb Perspect Biol 2013;5:a012583.
2. 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.
3. 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.
4. 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.
5. Win AK, Cleary SP, Dowty JG, et al. Cancer risks for monoallelic MUTYH mutation carriers with a family history of colorectal cancer. Int J Cancer 2011;129:2256-62.
6. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020;581:434-443.
7. de Voer RM, Hahn MM, Mensenkamp AR, et al. Deleterious Germline BLM Mutations and the Risk for Early-onset Colorectal Cancer. Sci Rep 2015;5:14060.
8. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2012;2:401-4.

Supplementary methods

Study cohorts

We included 5,942 patients with unexplained polyposis, familial CRC, or sporadic CRC at a young age or suspected of having Lynch syndrome with CRC or multiple adenomas (Supplementary Table 1) from the Netherlands (n = 3,158); United Kingdom (n = 275); Poland (n = 144); Germany (n = 104); Spain (n = 35); North Macedonia (n = 273); and North America, Canada, and Australia (CCFRC; n = 1,953) [1-3]. All participants provided written informed consent. Local medical ethical committees approved this study (Radboudumc [Commissie mensgebonden onderzoek (CMO)-light, 2015/2172 and 2015/1748], Leiden University Medical Center (LUMC) [P01-019], and Ontario Cancer Research Ethics Board, University of Melbourne Human Research Ethics Committee, and Fred Hutchinson Cancer Research Center Institutional review board).

A total of 1,207 cancer-unaffected control individuals were available from the population-based recruitment arms of the CCFRC [2, 3]. From the Netherlands, 2,329 WES control individuals with a >90-fold median coverage without a suspicion of hereditary cancer were available [4]. The European non-Finnish population of gnomAD was used to determine overall frequencies of LoF variants [5].

Targeted resequencing

Hi-Plex

Leukocyte DNA from 1,953 CRC-affected case patients and 1,207 control individuals was used to screen the coding regions of *NTHL1* by using multiplex polymerase chain reaction (PCR)-based targeted sequencing and variant calling approach (HiPlex2 and Hiplexpipe, hiplex.org, github.com/khalidm/hiplexpipe) [6]. Germline variants in *NTHL1* (NM_002528.5) were prioritized according to quality–the sequence depth of >30 reads and variant frequency of >30%.

Molecular Inversion Probe-Based sequencing

Leukocyte DNA from 1,486 polyposis and/or CRC cases was screened for all coding regions and intron-exon boundaries of *NTHL1* (NM_002528.5) by using molecular inversion probe MIPsequencing, combined with a panel of base excision repair genes, as described previously [1]. Reads were mapped with Burrows-Wheeler Aligner (BWA), and variant calling was performed with UnifiedGenotyper [7]. Somatic variants in *NTHL1* were prioritized according to

quality: sequence depth of > 40 reads, > 20 variant reads, variant frequency of > 25% and quality by depth scores > 8,000.

Variants from HiPlex and MIP screenings were further selected based on predicted LoF of *NTHL1*. We selected all nonsense, frameshift canonical splice sites and included only coding and noncoding splice site region variants with a predicted change of > 20%, based on Alamut (Interactive Biosoftware, Rouen, France) (MaxEnt, NNSplice, and Human Splicesite Finder [HSF]).

KASPar assay

Leukocyte DNA (n = 1,260) or germline DNA extracted from formalin-fixed, paraffin embedded (FFPE) surgical specimens (n = 982) was genotyped for *NTHL1* p.(Gln90*) by using KBioscience Competitive Allele-Specific PCR (KASPar) assay [1].

Allele-Specific Polymerase Chain Reaction

Leukocyte DNA from 261 individuals with sporadic or familial CRC was subjected to an allele- specific PCR (AS-PCR) specific for *NTHL1* p.(Gln90*) and p.(Trp269*); primers are available upon request.

Sanger sequencing

Sanger sequencing was used for variant validation and to sequence the entire open reading frame of *NTHL1* in confirmed heterozygous cases. In addition, when available, family members were sequenced by using Sanger sequencing for cosegregation purposes.

Statistical analysis

A one-sided Fisher exact test was performed to determine differences in the frequency of monoallelic *NTHL1* germline LoF variants in carriers with polyposis and/or CRC compared to control individuals. We calculated the *P* value, odds ratio, and the 95% confidence interval using R (R Foundation for Statistical Computing, Vienna, Austria; http://www.R-project.org). Three control data sets were used in this comparison.

First, we retrieved all LoF variants (nonsense, frameshift canonical splice sites, and coding or noncoding splice site regions with > 20% splice site change) in canonical transcripts of *NTHL1* listed in the non-Finnish European subpopulation of the genome aggregation database (gnomAD) [5]. All variants were checked manually in gnomAD for their quality. Second, LoF variants in *NTHL1* identified in the Dutch WES cohort (n = 2,329 individuals without a suspicion of hereditary

cancer) were extracted in a similar way as described earlier [4]. Third, LoF variants in *NTHL1* identified in the CCFRC control group of 1,207 individuals, sequenced in this study, were used.

Whole-Exome sequencing

Exome captures (Supplementary Table 2) were performed according to the manufacturer by using either Agilent Clinical Research Exome (CRE) V2 (Agilent, Santa Clara, CA) in combination with sequencing on a NovaSeq 6000 (Illumina, San Diego, CA), Agilent SureSelect XT^{HS} Human All Exon V6 enrichment Kit in combination with sequencing on a NextSeq 500, or xGEN Exome Research Panel (Integrated DNA Technology [IDT]), Coralville, IA) in combination with sequencing on a NovaSeq 6000.

Novaseq 6000 sequencing reads were trimmed by using Trimmomaticv0.36 and aligned to hs37d5 by using BWA-MEM, followed by merging and PCR duplicate removal with Sambamba (version 0.5.8) [8, 9]. Variant calling was performed by using Strelka (version 2.017) and Freebayes for paired samples; only variants called by both callers were reported [10, 11]. For LUMC2745, no paired sample was available, and variant calling was performed with Mutect2 (GATK version 4.1.0.0; GATK, Broadinstitute, Cambridge, MA). Trimmed NextSeq 500 sequencing reads were aligned to GRCh37 by using BWA-MEM, and duplicates were flagged by using Picard Tools, version 1.90. Variants were called with Mutect2 (GATK version 4.1.0.0), with or without matched germline samples; variant filtering was performed as described [1], with minor modifications. Variants in dbSNPv132 (minus catalogue of somatic mutations in cancer [COSMIC]), microsatellites, homopolymers, simple repeats and variants called outside of the respective exome capture target were removed. Somatic variants with a variant allele frequency of < 10%, < 20x coverage in both normal and tumor, and fewer than 4 reads supporting the variant were removed. For tumor-only analysis, variants shared by more than 1 individual and variants with a variant allele frequency of > 80% were removed to reduce germline leakage.

Mutational signature analysis

Mutation spectra were generated by using In-depth characterization and analysis of mutational signatures (ICAMS), version 2.1.2 (github.com/steverozen/ICAMS), and mutational signature analysis was performed by using mSigAct v2.0.0.9018 [12]. Tissue-specific CRC signature universes were inferred from the Pan-cancer analysis of whole genomes (PCAWG) signature assignments [13]. The signature universe was extended with SBS30 and potential artefact

signatures SBS45, SBS51, SBS52, SBS54, and SBS58, which were present in a subset of the samples of this cohort. Signatures were normalized to the trinucleotide abundance of the respective exome capture panel used. Per mutation spectrum, mutational signature assignment was performed by using mSigAct::SparseAssignActivity, with $p = .5$ to reduce sparsity. The presence of SBS30 was then determined using mSigAct::SignaturePresenceTest using the signatures determined by mSigAct::SparseAssignActivity plus SBS30 as well as the aging-associated signatures SBS1, SBS5, and SBS40 (Supplementary Table 2). Multiple testing correction was done according to Benjamini-Hochberg.

Supplementary Table 1. Characteristics of case and control cohorts and identified case patients and control individuals with monoallelic *NTHL1* LoF variants in this study

Approach	Sequencing method and cohorts	Samples, n	Selection[a] criteria	Genes tested	Monoallelic *NTHL1* p.(Gln90*), n	Other monoallelic *NTHL1* LoF variants, n	Total monoallelic *NTHL1* LoF variants, n
NTHL1- targeted resequencing (n = 3,439 cases)	**Hi-Plex multiplex PCR-based sequence screening of *NTHL1* exons (control individuals)**						
	Colon Cancer Family Registry	1,207	Population-based healthy individuals with no history of polyposis and/or CRC	NA	3	2	5
	Hi-Plex multiplex PCR based sequence screening of *NTHL1* exons (case patients)						
	Colon Cancer Family Registry	1,953	Population-based CRC	*APC*, *MUTYH*, *POLE*, *POLD1*, MMR*[b]	4	1	5
	MIP-based sequence screening of *NTHL1* (case patients)						
	ParelBED (the Netherlands)	600	Polyposis, CRC, or CRC and additional tumor	No disease-causing mutation found after routine diagnostics	0	0	0
	Oxford (United Kingdom)	275	Polyposis	*APC*, *MUTYH*	4	0	4
	Leiden (the Netherlands)	150	Polyposis or familial CRC	*APC*, *MUTYH*	0	0	0
	Nijmegen (the Netherlands)	147	Polyposis or familial CRC	*APC*, *MUTYH*	0	0	0
	Szczecin (Poland)	144	Familial CRC	*POLE*, *POLD1*, MMR*[b]	1	0	1
	Dresden (Germany)	104	Polyposis or familial CRC	*APC*, *MUTYH*	0	0	0
	Santiago de Compostela (Spain)	35	Polyposis or familial CRC	*APC*, *MUTYH* (in part), *POLE*, *POLD1*, *BMPR1A*, *SMAD4*, *PTEN*	0	0	0
	Groningen (the Netherlands)	19	Polyposis or familial CRC	*APC*, *MUTYH*	0	0	0
	Skopje (North Macedonia)	12	Polyposis, recessive inheritance	MMR*[b], *APC*, *TP53*, *MUTYH*, *POLE*, *POLD1*	1	0	1

Supplementary Table 1. (continued)

Approach	Sequencing method and cohorts	Samples, n	Selection[a] criteria	Genes tested	Monoallelic *NTHL1* p.(Gln90*), n	Other monoallelic *NTHL1* LoF variants, n	Total monoallelic *NTHL1* LoF variants, n
NTHL1 genotyping (n = 2,503 cases)	***NTHL1* p.(Gln90*) genotyping by KASPar assay (case patients)**						
	Leiden (the Netherlands)	1,894	Polyposis or familial CRC, with or without suspected Lynch syndrome	*APC*, *MUTYH*, *POLE*, *POLD1*, MMR*[b]	3	NA	3
	Nijmegen (the Netherlands)	348	Polyposis or familial CRC	*APC*, *MUTYH*, *POLE*, *POLD1*, MMR*[b]	1	NA	1
	***NTHL1* p.(Gln90*) and p.(Trp269*) genotyping by allele specific-PCR (case patients)**						
	Skopje (North Macedonia)	200	Sporadic CRC	None	2	0	2
	Skopje (North Macedonia)	61	Polyposis or familial CRC	TruSight Hereditary Cancer Panel (Illumina)	1	0	1

NA, not applicable; ParelBED, The Dutch Parelsnoer Institute Biobank Hereditary Colorectal Cancer [14].

[a] Polyposis is defined as the cumulative occurrence of at least 10 polyps. Familial CRC is defined as the proband having a CRC ≤50 years of age and at least one first degree relative with CRC ≤60 years of age. Sporadic CRC is defined as patients with CRC without a family history, irrespective of age. [b]MMR* genes: *MLH1*, *MSH2*, *MSH6* and *PMS2*.

Supplementary Table 2. Phenotypic description and details on the tumors subjected to WES of identified carriers of a monoallelic *NTHL1* LoF variant

Number	Patient ID	Identification method	Amino acid change	Sex	Polyps	Malignancies[e]	Tumor type for WGS	Matched normal available	Exome enrichment kit	Sequening platform	Median coverage tumor(s)[i]	Number of somatic variant calls	*P*-value SBS30[j]
1	P09708	Hi-Plex	p.(Gln287*)	M		Cecum (73), CRC (73)	CRC	Yes, blood	Agilent CRE V2	Novaseq 6000	221	572	0.976
2	P92662	Hi-Plex	p.(Gln90*)	M		CRC (53)	CRC	Yes, blood	Agilent CRE V2	Novaseq 6000	189	219	1.61×10^{3}
3	P07001	Hi-Plex	p.(Gln90*)	M		CRC (43)	CRC	Yes, blood	Agilent CRE V2	Novaseq 6000	116	141	0.331
4	P58832	Hi-Plex	p.(Gln90*)	F		CRC (46), UC (29)	--	--	--	--	--	--	--
5	P00387	Hi-Plex	p.(Gln90*)	F		Cecum (42), UC (23), LC (53)	--	--	--	--	--	--	--
6	P0011[a]	MIP screen	p.(Gln90*)	M		CRC (56), LiC (unk)	CRC	No[h]	Agilent V6	NextSeq500	133	1466	0.976
7	P0011-2[a]	Co-segregation	p.(Gln90*)	F		CRC (55)	CRC	Yes, FFPE	Agilent V6	NextSeq500	86	292	0.953
8	P0804	MIP screen	p.(Gln90*)	F		CRC (50)	CRC[f]	Yes, FFPE	Agilent V6	NextSeq500	--	--	--
9	P0468[b]	MIP screen	p.(Gln90*)	M	A (43)		--	--	--	--	--	--	--
10	P0567[b]	Co-segregation	p.(Gln90*)	F	A (55)		--	--	--	--	--	--	--

Supplementary Table 2. (continued)

Number	Patient ID	Identification method	Amino acid change	Sex	Polyps	Malignancies[e]	Tumor type for WGS	Matched normal available	Exome enrichment kit	Sequening platform	Median coverage tumor(s)[i]	Number of somatic variant calls	*P*-value SBS30[j]
11	P0567-2[b]	Co-segregation	p.(Gln90*)	F	A (61)		--	--	--	--	--	--	--
12	P0523	MIP screen	p.(Gln90*)	M	A (59)	CRC (58)	--	--	--	--	--	--	--
13	P0568	MIP screen	p.(Gln90*)	M	A (unk)		--	--	--	--	--	--	--
14	P0602	MIP screen	p.(Gln90*)	F	A (unk)		--	--	--	--	--	--	--
15	K134	KASPar assay	p.(Gln90*)	F	A (48-56)	CRC (49)	--	--	--	--	--	--	--
16	LUMC3333	KASPar assay	p.(Gln90*)	M		CRC (<69), Cecum (69)	CRC	Yes, FFPE	IDT xGEN	Novaseq 6000	131	150	0.888
17	LUMC2745	KASPar assay	p.(Gln90*)	M		CRC (72); CRC, SCC (61)	CRC	No	IDT xGEN	Novaseq 6000	99	487	0.053
18	LUMC0748	KASPar assay	p.(Gln90*)	F		CRC (56), OvC (56), CRC (56), CRC (68)	CRC	Yes, FFPE	IDT xGEN	Novaseq 6000	84	150	1
19	Tcc136	AS-PCR	p.(Gln90*)	M		CRC (75)	CRC[g]	No	Agilent V6	NextSeq500	195	192	0.331
20	Tcc456	AS-PCR	p.(Gln90*)	M		PC, CRC (72)	CRC	No	Agilent V6	NextSeq500	140	211	0.052
21	Tcc712	AS-PCR	p.(Gln90*)	F	7A (71)	EC (66), CRC (71)	CRC[g]	No	Agilent V6	NextSeq500	180	4083	1
22	P03-I:1	c	p.(Gln90*)	M	A, HP		A	No	IDT xGEN	Novaseq 6000	T1=64 T2=39	T1=81 T2=290	T1=1 T2=0.088

Supplementary Table 2. (continued)

Number	Patient ID	Identification method	Amino acid change	Sex	Polyps	Malignancies[e]	Tumor type for WGS	Matched normal available	Exome enrichment kit	Sequening platform	Median coverage tumor(s)[i]	Number of somatic variant calls	*P*-value SBS30[j]
--	P04-II:5	c	p.Gln90*/ p.Ile245Asnfs*28	F	--	--	NTHL1-deficient CRC	Yes, FFPE	IDT xGEN	Novaseq 6000	162	347	$3.11x10^{-45}$
--	P05001	Hi-Plex	p.(Gln90*)/ p.(Ala79fs)	F	A, HP (61)	CRC (61), BC (63)	NTHL1-deficient CRC	Yes, blood	Agilent CRE V2	Novaseq 6000	108	430	$1.82x10^{-39}$
--	CRC-3	d	p.(Gln90*)/ p.(Gln90*)	M	--	--	NTHL1-deficient CRC	Chapter 5 [1]	Chapter 5 [1]	Chapter 5 [1]	Chapter 5 [1]	360	$3.08x10^{-38}$

A: colorectal adenomatous polyps; BCC: basal cell carcinoma; CRC: colorectal cancer; EC: endometrial cancer; HP: hyperplastic polyps; LC: lung cancer; LiC: liver cancer; OvC: ovarian cancer; PC: prostate cancer; SCC: squamous cell carcinoma; UC: uterus cancer; unk: age unknown; --: not applicable.

[a]Sibling.

[b]Sibling.

[c]Identified by Grolleman et al., 2019 [1].

[d]Tumor data from Grolleman et al., 2019 [1].

[e]Numbers in parenthesis indicate age at diagnosis.

[f]Tumor P0804 was excluded from further analysis due to insufficient data quality.

[g]Fresh-frozen tumor material.

[h]The normal sample of the sibling was used for somatic variant extraction.

[i]Median read coverage (units=reads).

[j]Multiple testing correction was done according to Benjamini-Hochberg.

References

1. 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.
2. Jenkins MA, Win AK, Templeton AS, et al. Cohort Profile: The Colon Cancer Family Registry Cohort (CCFRC). Int J Epidemiol 2018;47:387-388i.
3. Newcomb PA, Baron J, Cotterchio M, et al. Colon Cancer Family Registry: an international resource for studies of the genetic epidemiology of colon cancer. Cancer Epidemiol Biomarkers Prev 2007;16:2331-43.
4. de Voer RM, Hahn MM, Mensenkamp AR, et al. Deleterious Germline BLM Mutations and the Risk for Early-onset Colorectal Cancer. Sci Rep 2015;5:14060.
5. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020;581:434-443.
6. Hammet F, Mahmood K, Green TR, et al. Hi-Plex2: a simple and robust approach to targeted sequencing-based genetic screening. Biotechniques 2019;67:118-122.
7. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011;43:491-8.
8. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114-20.
9. Tarasov A, Vilella AJ, Cuppen E, et al. Sambamba: fast processing of NGS alignment formats. Bioinformatics 2015;31:2032-4.
10. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907 2012.
11. Saunders CT, Wong WS, Swamy S, et al. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics 2012;28:1811-7.
12. Ng AWT, Poon SL, Huang MN, et al. Aristolochic acids and their derivatives are widely implicated in liver cancers in Taiwan and throughout Asia. Sci Transl Med 2017;9.
13. Alexandrov LB, Kim J, Haradhvala NJ, et al. The repertoire of mutational signatures in human cancer. Nature 2020;578:94-101.
14. Manders P, Vos J, de Voer RM, et al. Parelsnoer institute biobank hereditary colorectal cancer: A joint infrastructure for patient data and biomaterial on hereditary colorectal cancer in the Netherlands. Open Journal of Bioresources 2019;6.

Chapter 7

Discussion and future perspectives

Discussion

The work described in this thesis aims to determine the underlying genetic causes of polyposis and colorectal cancer (CRC) in unexplained cases by screening known high-risk genes such as *POLE*, *POLD1*, *APC* and *NTHL1*.

POLE and *POLD1*

Palles et al. identified that germline variants affecting the proofreading domains of *POLE* and *POLD1* predispose to colorectal adenomas and carcinomas [1]. *POLE* p.(L424Val) and *POLD1* p.(Ser478Asn) were established as new high-penetrance causes of germline CRC predisposition with an autosomal dominant pattern of inheritance [1]. In **Chapter 2**, we show that germline variants in *POLE* are also associated with early-onset mismatch repair (MMR)-deficient colorectal cancer [2]. In a cohort of 1188 unexplained index patients enriched for inherited CRC and polyposis, we identified three *POLE* p.(Leu424Val) carriers at a frequency (0.25%), comparable to reported frequencies [1, 3]. Interestingly, *POLE* carriers from two families displayed a Lynch syndrome-like phenotype with MMR-deficient tumors. MMR deficiency in these tumors resulted from secondary somatic MMR variants due to the proofreading defect. In a study by Jansen et al. [4], a similar Lynch syndrome-like phenotype in *POLE* variant carriers was described.

DNA proofreading defects result in ultramutated tumor phenotypes with an increase in C:T>A:G mutations [5]. Recently, genomic sequencing of tumors with concurrent activity loss of one of the MMR genes and *POLE* or *POLD1* revealed the distinct mutational signatures SBS14 and SBS20, respectively, different from the signatures SBS10 or SBS6 for *POLE* or MMR deficiency, respectively [6-9]. Previously, MMR-deficient tumors with somatic MMR variants or *MLH1* promoter hypermethylation have been reported for patients with biallelic variants in the base excision repair (BER) gene *MUTYH* [10, 11]. The somatic MMR variants were MAP-specific G>T variants, indicating that impaired BER was the primary defect followed by MMR deficiency [10]. *POLE* DNA analysis now seems warranted in microsatellite-unstable CRC, especially in the absence of a causative DNA mismatch repair germline variant.

In **Chapter 3**, in search for additional *POLE*/*POLD1* pathogenic variants other than Leu424Val and Ser478Asn, we sequenced the exonuclease domains of *POLE* and *POLD1* in unexplained patients with multiple colorectal polyps. We describe two variants of unknown significance (VUS) in *POLD1* [12]. However, the available evidence is insufficient to evaluate the pathogenicity of these variants due to a

lack of cosegregation information and functional analysis. Sequencing of *POLE* and *POLD1* results in VUS variants rather than pathogenic variants, suggesting that pathogenic variants in *POLE* and *POLD1* probably occur at low frequencies. The assessment of the pathogenicity of variants of unknown significance remains a significant challenge in the investigation of hereditary CRC (and any other cancer syndrome). Interestingly, we found that one patient in addition to the *POLD1* VUS variant also carried a monoallelic *MUTYH* pathogenic variant, possibly suggesting that both genes could act cooperatively and together to confer an increased CRC risk. Hamzaoui et al. reported the cooccurrence of a *POLE* VUS variant and a pathogenic *MSH2* variant in CRC patients [13].

APC

In addition to classic *APC* germline variants, a few deep intronic variants contribute substantially to the *APC* mutation spectrum [14, 15]. In a study by Spier et al., the first systematic analysis of intronic variants that may affect RNA splicing in *APC* was performed. They investigated the frequency and type of deep intronic splice variants of *APC* in polyposis patients and highlighted the relevance of studying deep intronic *APC* splice variants in FAP, which cannot be identified by conventional routine screening methods [14]. In a study by Nieminen et al., pseudoexons in *APC* were successfully identified using next-generation sequencing, and this was the second study to reveal *APC*-related pseudoexons in FAP [15]. In **Chapter 4**, we attempt to investigate the roles of these deep intronic germline *APC* variants described by Spier et al. and Nieminen et al. [14, 15] as possible genetic causes of colorectal polyposis in a Dutch cohort of unexplained patients with more than 50 polyps. We did not detect any one of these variants in our cohort as a cause of colorectal polyposis. It is possible that either the frequency of intronic variants is lower in the Dutch population and the sample size of our cohort is not large enough or these intronic *APC* variants are local founder variants [16].

In 10-25% of the index patients with FAP, a de novo *APC* variant is identified [17-19]. Among those, there is a substantial but still underestimated proportion of mosaic carriers [20, 21]. Recent reports using methods that are able to detect germline variants with low allele frequencies, as well as variants only present in tumor material, indicate that many mosaic patients are undiagnosed [22, 23]. With the advantage of NGS technology, which allows for deep sequencing of selected regions, mosaic variants in *APC* are detected more frequently [22, 23]. In **Chapter 4**, we investigate the role of mosaic *APC* variants as possible genetic causes of colorectal polyposis in the same cohort that we screened

for deep intronic germline *APC* variants. We performed deep NGS of *APC* to identify possible undetected pathogenic mosaic variants in leukocyte DNA of unexplained index patients with colorectal polyposis. We did not detect mosaic *APC* variants. A limitation of this study is that we screened only the available leukocyte DNA for mosaicism due to the scarcity of tumor tissue for our study cohort [16]. The strategy of sequencing multiple adenomas of the same patients has been proven to be more sensitive and specific than sequencing leukocyte DNA for variants with low variant allele frequencies and can detect mosaicism confined to the colon [22-24].

Biallelic *NTHL1* LoF variants

In 2015, it was shown that germline biallelic loss-of-function (LoF) variants in *NTHL1* predispose to adenomatous polyposis and CRC, but the phenotypic spectrum remained to be elucidated, as patient numbers for this rare syndrome were low [25, 26]. Hence, large-scale studies are needed to further delineate this recently identified syndrome. In **Chapter 5**, using a large cohort of patients, we aimed to define the molecular and clinical characteristics of individuals with germline *NTHL1* LoF variants, and we found that NTHL1 deficiency predisposes them to multiple tumor types, including colon and breast cancer.

We screened our cohort for the most common LoF variant in *NTHL1* (p.Q90*) and studied the genotype-phenotype relationship in *NTHL1* biallelic LoF variant carriers. For a comprehensive analysis with sufficient cases, our data were combined with the data from an international consortium. In this chapter, we present a molecular and clinical characterization of the tumor spectrum of a total of 29 individuals with biallelic LoF variants in *NTHL1* from 17 unrelated families, including 11 previously unreported families, of which 26 developed one (n=10) or multiple (n=16) malignancies in 14 different tissues. We found that the majority of individuals developed one or more CRCs (59%). We show that 55% of the individuals with biallelic LoF variants in *NTHL1* developed multiple primary tumors at various sites, of which the majority were extracolonic (66%), while for MUTYH-associated polyposis, no more than 13% of the individuals developed an extracolonic malignancy [27]. An unexpectedly high breast cancer incidence was observed in female carriers (60%). In addition to breast cancer, we encountered endometrial (pre)malignancies, urothelial cell cancers, brain tumors, hematologic malignancies, basal cell carcinomas, head and neck squamous cell carcinomas, cervical cancers in multiple individuals and five other cancers in single individuals, including duodenal cancer.

We obtained additional evidence for causality of NTHL1 deficiency for specific malignancies by analyzing somatic mutational patterns using whole-exome sequencing from 14 tumors from seven different tissues (adenomatous/colorectal cancer, breast cancer, endometrial cancer, head and neck squamous cell carcinoma, meningioma, thyroid cancer, and urothelial cell cancer). We identified signature SBS30 in 13 out of the 14 tumors (93%). This signature is associated with NTHL1 deficiency and is characterized by C:G>T:A transitions at non-CpG sites. This suggests that deficiency of NTHL1 elicits the same mutational process in multiple tissues. The tumor without signature SBS30 was a urinary cell carcinoma in which signature 2 was the most prominent signature. This signature is commonly observed in sporadic urothelial cell cancers and suggests that this tumor developed sporadically [28]. A study in which *NTHL1* was knocked out in human intestinal organoids revealed that NTHL1 deficiency is the mutational process underlying signature SBS30 [29]. Signature SBS30 was previously identified in a single breast cancer case [30]. Retrospective analysis of that single breast cancer sample revealed an *NTHL1* germline LoF variant with loss of heterozygosity in tumors [29]. We show that in four breast cancer samples from four individuals with biallelic LoF variants in *NTHL1* that were sequenced, more than 80% of the mutations can be assigned to signature SBS30, suggesting that this base excision repair defect has driven breast cancer formation in these patients. We found a high incidence of breast cancer among women with biallelic *NTHL1* LoF variants (60%), and the median age at diagnosis for breast cancer in these women was found to be lower than in the general population (48.5 years [range: 38-63] compared with 62 years, respectively). This observation suggests a high penetrance for breast cancer for individuals with biallelic *NTHL1* LoF variants compared to, for example, the risks of breast cancer for *BRCA1* and *BRCA2* carriers of 57% and 49% by the age of 70 years, respectively [31]. We estimated the cumulative risk for extracolonic cancer to be between 35% and 78% by the age of 60 years, which highlights the importance of surveillance for extracolonic malignancies in patients with NTHL1 deficiency.

The tumor spectrum of individuals with biallelic *NTHL1* LoF variants was shown to be broader than polyposis and colorectal carcinomas, as has also been observed for other CRC syndromes associated with DNA repair defects. For example, MUTYH-associated polyposis patients have an increased lifetime risk of developing duodenal, ovarian, bladder, skin and possibly breast cancer [27]. Lynch syndrome patients have an increased lifetime risk of developing cancer of the endometrium, small bowel, urinary tract, stomach and ovaries [32, 33]. It has been postulated that polymerase proofreading-associated polyposis patients

may, next to endometrial cancer, be at an increased lifetime risk of developing brain tumors and cutaneous tumors [1, 34].

We conclude that biallelic germline *NTHL1* LoF variants predispose patients to multiple primary tumors, including colon cancer and breast cancer (**Chapter 5**)[28], and recent studies confirmed our findings [35-37]. Consequently, germline testing of *NTHL1* for individuals with multiple primary malignancies, either with or without adenomatous polyposis and/or a family history of cancer, might be considered.

Additionally, in **Chapter 5**, we demonstrate that mutational signatures in tumors can be used as a tool to corroborate a genetic predisposition. We found tumor mutational signature analysis to be suitable for obtaining additional support for a causative link between NTHL1 deficiency and tumor development. We showed that the presence of a unique mutational signature that is associated with a germline defect can distinguish these tumors from those that developed sporadically, as somatic inactivation of *NTHL1* is not a frequent event.

Monoallelic *NTHL1* LoF variants in polyposis and CRC

The list of genes associated with adenomatous polyposis and colorectal cancer now includes two recessive cancer-predisposing base-excision repair genes, i.e., *MUTYH* and *NTHL1*. For *MUTYH*, it is suggested that individuals with monoallelic LoF variants may have an increased, albeit small, risk of developing CRC compared to the general population [38-40]. Thus far, it is unknown whether monoallelic *NTHL1* LoF variants increase the risk of polyposis and/or CRC and whether carriers of monoallelic *NTHL1* LoF variants and their family members need additional counseling. While the prevalence of biallelic *NTHL1* LoF variants is low, the identification of monoallelic *NTHL1* LoF variant carriers from multigene panel testing is more common. The most common LoF variant in *NTHL1* is p.(Gln90*), which is heterozygous in approximately 0.28% of the general population [41]. The analysis of a breast cancer from an individual with a monoallelic *NTHL1* LoF variant suggests that these alleles may play a potential role in tumor development [29]. Therefore, it is of clinical importance to know whether carriers of monoallelic LoF variants in *NTHL1* are at increased risk of developing polyposis and/or CRC.

In **Chapter 6**, we investigated whether individuals with polyposis and/or CRC more frequently carry monoallelic LoF variants in *NTHL1* than the general population and whether monoallelic *NTHL1* LoF variants increase the risk of

polyposis and/or CRC in carriers. To address this question, an international collaboration between various research groups (the Netherlands, the United Kingdom, Poland, Germany, North Macedonia, North America, Canada and Australia) established a large cohort of 5,942 cases. The cohort consisted of individuals with unexplained polyposis, familial CRC, or sporadic CRC at a young age or those suspected of having Lynch syndrome with CRC or multiple adenomas. The cohort was investigated for monoallelic LoF variants in *NTHL1.* We did not find significant enrichment of monoallelic *NTHL1* LoF variant carriers in our cohort compared to control datasets. Furthermore, mutational signature analysis of 13 colorectal tumors from monoallelic *NTHL1* LoF variant carriers did not show a somatic second hit, and we did not find evidence of a main contribution of the mutational signature SBS30, the signature associated with NTHL1 deficiency, suggesting that monoallelic loss of *NTHL1* does not substantially contribute to colorectal tumor development [42]. Thus, we found no evidence that monoallelic *NTHL1* LoF variant carriers are at increased risk of developing polyposis and/or CRC; consequently, no additional surveillance is currently warranted. However, we cannot rule out that a small risk for CRC, similar to what is observed for *MUTYH* carriers, still exists. To date, screening cohorts of patients and tumors with a monoallelic pathogenic variant in *MUTYH* have been larger than those for *NTHL1*. Therefore, screening more patients for *NTHL1* is needed. From our data, we suggest that inactivation of the *NTHL1* wild-type allele (via LOH) is a rare event in colorectal tumors, which is in agreement with the observation that loss of 16p, the chromosome arm on which *NTHL1* is located, does not frequently occur in CRC [43]. Monoallelic LoF variants in *MUTYH* with LOH (on chromosome arm 1p) and high levels of signature SBS18 or combined SBS18/SBS36 have been reported in colorectal tumors [44, 45]. Loss of 1p is reported to occur in only approximately 10% of CRCs [46], which may explain the only slightly increased CRC risk reported for *MUTYH* [40]. In a recent study, molecular analysis of breast cancers from carriers indicated that *NTHL1* may be included in the growing list of low-penetrance breast cancer genes that appear to function via haploinsufficiency rather than the somatic biallelic inactivation mechanism almost universally observed for high-risk breast cancer predisposition genes [47]. The absence of a second hit in *NTHL1* may be a generic feature of low- to moderate-penetrance alleles, and these alleles are less prone to obtain second hits leading to a complete loss of function, always retaining some activity in the tumor [47]. To conclude, there is no evidence that monoallelic germline *NTHL1* LoF variant carriers are at increased risk of developing polyposis and/or CRC. To date, there is no evidence supporting specific surveillance for monoallelic carriers.

Monoallelic *NTHL1* LoF variants in the risk of extracolonic cancer

The biallelic *NTHL1* LoF variants predispose to a multitumor phenotype, but whether monoallelic carriers are at increased risk of developing other extracolonic malignancies remains to be elucidated. We investigated the role of the monoallelic *NTHL1 c.268C*>T, p.(Gln90*) variant in the risk of extracolonic cancers, but we found that the monoallelic *NTHL1* p.(Gln90*) variant does not seem to predispose patients to extracolonic cancer (unpublished data). In a cohort of cases with extracolonic cancer and suspected Lynch syndrome (N= 327), two monoallelic *NTHL1* p.(Gln90*) carriers were detected (2/327, 61%). One patient developed urothelial cell cancer (UCC), and the second patient developed adenosquamous carcinoma (ASC) of the mouth. We found no significant enrichment of monoallelic *NTHL1* p.(Gln90*) carriers in our cohort compared to a genome aggregation database (gnomAD) non-Finnish European control population (2/327; 0.61% versus 250/64,328; 0.39%; *P* = 0.36). Further exome sequencing for the available tumor (ASC) did not detect the *NTHL1* deficiency-related mutational signature SBS30 and LOH of the wild-type *NTHL1* allele, which indicates that monoallelic *NTHL1* did not play a role in tumor development in this patient. Following the initial discovery that biallelic LoF variants in *NTHL1* predispose to breast cancer, we genotyped *NTHL1* p.(Gln90*) in a cohort of 692 individuals with ductal carcinoma in situ (DCIS) and detected one biallelic (1/692; 0.14%) and three monoallelic carriers (3/692; 0.4%). The frequency of monoallelic *NTHL1* p.(Gln90*) was not significantly enriched in our DCIS cohort compared to gnomAD non-Finnish European controls (3/692; 0.4% versus 250/64,328; 0.39%; *P* = 0.75). We found no evidence that monoallelic *NTHL1* p.(Gln90*) carriers are at increased risk of developing DCIS. A recent study suggested that carriers of monoallelic *NTHL1* p.(Gln90*) do not have an increased risk for breast cancer [48]. An even more recent study suggested that monoallelic LoF variants in *NTHL1* may be associated with a low to moderate increased risk of breast cancer [47]. Salo-Mullen et al. identified a woman with high-grade serous ovarian carcinoma harboring monoallelic *NTHL1* p.(Gln90*) with corresponding LOH of the wild-type allele in the tumor resulting in signature 30 [49]. Based on data from cBioPortal, loss of 16p, the chromosome arm on which *NTHL1* is located, mainly occurs in ovarian serous cystadenocarcinoma and uterine carcinosarcoma, while in colorectal adenocarcinoma and breast invasive ductal carcinoma, this loss is only 6%. It is possible that monoallelic *NTHL1* carriers are at risk of developing ovarian cancer when loss of 16p occurs as an early event in tumorigenesis. Salo-Mullen et al. identified a prostate cancer patient with monoallelic *NTHL1* p.(Gln90*) and signature 30 but without LoF of the wild-type allele [49]. The contradictory results from these studies may be

explained by differences in tumorigenesis, including that different mechanisms can drive tumor development in monoallelic carriers, such as the timing of a potential second hit. In conclusion, our results indicate that monoallelic *NTHL1* p.(Gln90*) is unlikely to be a significant contributor to extracolonic cancer, which is in line with results obtained for CRC cancer in **Chapter 6**.

Future perspectives

In this thesis, we illustrate the power of mutational signature analysis in defining tumor phenotypes in rare cancer predisposition syndromes and provide proof of principle for recognizing new patients with cancer syndromes based on tumor sequencing data. In the future, mutational signature analysis will assist in the identification of novel cancer syndromes, including adenomatous polyposis and/or CRC syndromes caused by DNA repair deficiency.

Studying the mutation signatures in tumors could confirm the pathogenicity of VUS variants and mark them as causal variants in the predisposition for multiple colorectal polyps.

Recent reports using methods that are able to detect germline variants with low allele frequencies, as well as variants only present in tumor material, indicate that many mosaic patients are undiagnosed. Testing tumor DNA, rather than leukocyte DNA, will provide greater knowledge about the true incidence of mosaicism in *APC*. In-depth analysis of adenomas of patients could lead to the detection of more mosaic *APC* carriers. Recently, the recurrent *APC* splice variant c.835-8A>G in a patient with unexplained colorectal polyposis fulfilling the colibactin mutational signature was reported [50]. The presence of pks + E coli, causing a specific mutational signature, might be an additional explanation for unexplained polyposis patients.

The use of novel sequencing techniques will possibly enable the detection of rare variants and germline aberrations in noncoding regions in the near future. Well-defined patient cohorts and families with multiple affected members will help in the identification of novel polyposis- and CRC-predisposing germline aberrations. Joint efforts screening for variants in larger cohorts and data sharing are essential to find underlying genetic causes of colorectal polyposis and CRC. Hopefully, the results and knowledge gathered will ultimately contribute to the significant clinical management and prevention of CRC.

References

1. 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.
2. Elsayed FA, Kets CM, Ruano D, et al. Germline variants in POLE are associated with early onset mismatch repair deficient colorectal cancer. Eur J Hum Genet 2015;23:1080-4.
3. Chubb D, Broderick P, Frampton M, et al. Genetic diagnosis of high-penetrance susceptibility for colorectal cancer (CRC) is achievable for a high proportion of familial CRC by exome sequencing. J Clin Oncol 2015;33:426-32.
4. Jansen AM, van Wezel T, van den Akker BE, et al. Combined mismatch repair and POLE/POLD1 defects explain unresolved suspected Lynch syndrome cancers. Eur J Hum Genet 2016;24:1089-92.
5. Shinbrot E, Henninger EE, Weinhold N, et al. Exonuclease mutations in DNA polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication. Genome Res 2014;24:1740-50.
6. Campbell BB, Light N, Fabrizio D, et al. Comprehensive Analysis of Hypermutation in Human Cancer. Cell 2017;171:1042-1056.e10.
7. Haradhvala NJ, Kim J, Maruvka YE, et al. Distinct mutational signatures characterize concurrent loss of polymerase proofreading and mismatch repair. Nat Commun 2018;9:1746.
8. Castellsagué E, Li R, Aligue R, et al. Novel POLE pathogenic germline variant in a family with multiple primary tumors results in distinct mutational signatures. Hum Mutat 2019;40:36-41.
9. Alexandrov LB, Kim J, Haradhvala NJ, et al. The repertoire of mutational signatures in human cancer. Nature 2020;578:94-101.
10. Morak M, Heidenreich B, Keller G, et al. Biallelic MUTYH mutations can mimic Lynch syndrome. Eur J Hum Genet 2014;22:1334-7.
11. Colebatch A, Hitchins M, Williams R, et al. The role of MYH and microsatellite instability in the development of sporadic colorectal cancer. Br J Cancer 2006;95:1239-43.
12. Elsayed FA, Tops CMJ, Nielsen M, et al. Low frequency of POLD1 and POLE exonuclease domain variants in patients with multiple colorectal polyps. Mol Genet Genomic Med 2019:e603.
13. Hamzaoui N, Alarcon F, Leulliot N, et al. Genetic, structural, and functional characterization of POLE polymerase proofreading variants allows cancer risk prediction. Genet Med 2020;22:1533-1541.
14. 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.
15. Nieminen TT, Pavicic W, Porkka N, et al. Pseudoexons provide a mechanism for allele-specific expression of APC in familial adenomatous polyposis. Oncotarget 2016;7:70685-70698.
16. Elsayed FA, Tops CMJ, Nielsen M, et al. Use of sanger and next-generation sequencing to screen for mosaic and intronic APC variants in unexplained colorectal polyposis patients. Fam Cancer 2021.
17. Bisgaard ML, Fenger K, Bülow S, et al. Familial adenomatous polyposis (FAP): frequency, penetrance, and mutation rate. Hum Mutat 1994;3:121-5.
18. Ripa R, Bisgaard ML, Bülow S, et al. De novo mutations in familial adenomatous polyposis (FAP). Eur J Hum Genet 2002;10:631-7.

19. Aretz S, Uhlhaas S, Caspari R, et al. Frequency and parental origin of de novo APC mutations in familial adenomatous polyposis. Eur J Hum Genet 2004;12:52-8.
20. Hes FJ, Nielsen M, Bik EC, et al. Somatic APC mosaicism: an underestimated cause of polyposis coli. Gut 2008;57:71-6.
21. Jansen AML, Goel A. Mosaicism in Patients With Colorectal Cancer or Polyposis Syndromes: A Systematic Review. Clin Gastroenterol Hepatol 2020;18:1949-1960.
22. 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.
23. 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.
24. Aretz S, Stienen D, Friedrichs N, et al. Somatic APC mosaicism: a frequent cause of familial adenomatous polyposis (FAP). Hum Mutat 2007;28:985-92.
25. 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.
26. Rivera B, Castellsagué E, Bah I, et al. Biallelic NTHL1 Mutations in a Woman with Multiple Primary Tumors. N Engl J Med 2015;373:1985-6.
27. Vogt S, Jones N, Christian D, et al. Expanded extracolonic tumor spectrum in MUTYH-associated polyposis. Gastroenterology 2009;137:1976-85.e1-10.
28. 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.
29. 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.
30. Nik-Zainal S, Davies H, Staaf J, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature 2016;534:47-54.
31. Chen S, Parmigiani G. Meta-analysis of BRCA1 and BRCA2 penetrance. J Clin Oncol 2007;25:1329-33.
32. Vasen HF, Wijnen JT, Menko FH, et al. Cancer risk in families with hereditary nonpolyposis colorectal cancer diagnosed by mutation analysis. Gastroenterology 1996;110:1020-7.
33. Watson P, Vasen HF, Mecklin JP, et al. The risk of endometrial cancer in hereditary nonpolyposis colorectal cancer. Am J Med 1994;96:516-20.
34. Aoude LG, Heitzer E, Johansson P, et al. POLE mutations in families predisposed to cutaneous melanoma. Fam Cancer 2015;14:621-8.
35. Boulouard F, Kasper E, Buisine MP, et al. Further delineation of the NTHL1 associated syndrome: A report from the French Oncogenetic Consortium. Clin Genet 2021;99:662-672.
36. Beck SH, Jelsig AM, Yassin HM, et al. Intestinal and extraintestinal neoplasms in patients with NTHL1 tumor syndrome: a systematic review. Fam Cancer 2022.
37. Altaraihi M, Gerdes AM, Wadt K. A new family with a homozygous nonsense variant in NTHL1 further delineated the clinical phenotype of NTHL1-associated polyposis. Hum Genome Var 2019;6:46.
38. Win AK, Hopper JL, Jenkins MA. Association between monoallelic MUTYH mutation and colorectal cancer risk: a meta-regression analysis. Fam Cancer 2011;10:1-9.
39. Win AK, Cleary SP, Dowty JG, et al. Cancer risks for monoallelic MUTYH mutation carriers with a family history of colorectal cancer. Int J Cancer 2011;129:2256-62.
40. Win AK, Dowty JG, Cleary SP, et al. Risk of colorectal cancer for carriers of mutations in MUTYH, with and without a family history of cancer. Gastroenterology 2014;146:1208-11. e1-5.

41. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020;581:434-443.
42. 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.
43. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2012;2:401-4.
44. Pilati C, Shinde J, Alexandrov LB, et al. Mutational signature analysis identifies MUTYH deficiency in colorectal cancers and adrenocortical carcinomas. J Pathol 2017;242:10-15.
45. Georgeson P, Pope BJ, Rosty C, et al. Evaluating the utility of tumour mutational signatures for identifying hereditary colorectal cancer and polyposis syndrome carriers. Gut 2021;70:2138-2149.
46. De Angelis PM, Clausen OP, Schjølberg A, et al. Chromosomal gains and losses in primary colorectal carcinomas detected by CGH and their associations with tumour DNA ploidy, genotypes and phenotypes. Br J Cancer 1999;80:526-35.
47. Li N, Zethoven M, McInerny S, et al. Evaluation of the association of heterozygous germline variants in NTHL1 with breast cancer predisposition: an international multi-center study of 47,180 subjects. NPJ Breast Cancer 2021;7:52.
48. Kumpula T, Tervasmäki A, Mantere T, et al. Evaluating the role of NTHL1 p.Q90* allele in inherited breast cancer predisposition. Mol Genet Genomic Med 2020:e1493.
49. Salo-Mullen EE, Maio A, Mukherjee S, et al. Prevalence and Characterization of Biallelic and Monoallelic NTHL1 and MSH3 Variant Carriers From a Pan-Cancer Patient Population. JCO Precis Oncol 2021;5.
50. 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.

Chapter 8

Summary
Nederlandse samenvatting
List of publications
Curriculum vitae
Acknowledgements

Summary

Heritable factors account for approximately 35% of colorectal cancer (CRC) risk. Around 5 to 10% of CRC cases are associated with highly penetrant dominant or recessive inherited syndromes, caused by germline variants in known high-penetrance CRC genes. The etiology of the remaining 20%-30% of inherited CRC risk is not completely understood. In recent years, advances were made in discovering the genetic causes for CRC and polyposis. Germline variants in *POLE*, *POLD1* and biallelic variants in *NTHL1* were discovered underlying polymerase proofreading associated polyposis syndrome and NTHL1-assciated tumor syndrome respectively, and new genes are still being described. A precise understanding of the genetics of inherited CRCs is important for identifying at risk individuals, improving cancer surveillance and prevention strategies, and developing better diagnostic and therapeutic approaches. The studies described in this thesis focus on characterizing variants in known high risk genes such as *POLE*, *POLD1*, *APC* and *NTHL1 as* genetic causes of polyposis and CRC in unexplained cases.

Germline pathogenic variants in DNA polymerase ε (*POLE*) and ẟ (*POLD1*) have been identified in families with multiple colorectal adenomas and CRC, in **Chapter 2** we screened the pathogenic germline variants in *POLE* and *POLD1* that were identified by Palles et al. in our cohort of unexplained familial, early onset CRC and polyposis cases. The frequency of the variants we report is comparable to those previously reported, despite an enrichment in our cohort for inherited CRC and polyposis. Interestingly we showed that the tumors associated with *POLE* germline variants can show a Lynch syndrome-like phenotype with mismatch repair (MMR) deficiency due to somatic mutation in MMR genes which results from the proofreading deficiency caused by *POLE* inactivation.

In **Chapter 3**, with the aim to find additional pathogenic variants in *POLE* and *POLD1* using next-generation sequencing (NGS) we sequenced the exonuclease domains of *POLE* and *POLD1* on a cohort of unexplained index patients diagnosed with multiple colorectal polyps. Germline variants of uncertain significance were found in *POLD1*, but no further testing was possible to assess the functional relevance of these variants as tumors were not available for further studies. This study confirms the low frequency of causal variants in these genes in the predisposition for multiple colorectal polyps, and established that these genes are a rare cause of colorectal polyps or CRC.

In **Chapter 4** we screened for previously reported pathogenic deep intronic germline *APC* variants in a cohort of unexplained colorectal polyposis patients. Using deep NGS we furthermore screened for *APC* mosaic variants. We did not detect mosaic or intronic *APC* variants in the screened unexplained colorectal polyposis patients. The limitation of this study was that we screened only leukocyte DNA for mosaic variants. Consequently, *APC* mosaic variants solely confined to the colon could have been missed with this approach because we could not screen the DNA from adenomas of the patients.

In 2015, biallelic germline loss-of-function (LoF) variants in *NTHL1* were shown to predispose to adenomatous polyposis and CRC, but the exact clinical phenotype was unclear as the patient numbers for this syndrome were low. In **Chapter 5** we characterized *NTHL1* tumor syndrome with the use of mutational signature analysis. To define the molecular and clinical characterization of tumor spectrum of the individuals with biallelic germline LoF variants in *NTHL1*, a large collaborative study involving research groups from Netherlands, United Kingdom, Poland, Germany, Norway, Spain and Macedonia was established. We collected clinical and molecular data of 29 individuals with biallelic germline *NTHL1* LoF variants from 17 families. We found that 55% of the individuals developed multiple primary tumors at various sites, of which the majority was extracolonic (66%). In addition to colorectal tumors we found tumors in 13 tissue types. Most individuals developed one or more CRCs (59%) and high breast cancer incidence was observed in female carriers (60%). We identified a unique mutational signature (SBS30) that was associated with NTHL1-deficiency in 13 tumors from seven organs. Our study demonstrates that *NTHL1* is a multi-tumor predisposition gene with a high lifetime risk for extracolonic cancers.

While biallelic germline *NTHL1* LoF variants are causal to adenomatous polyposis and CRC, the adenomatous polyposis and CRC risk for carriers of monoallelic germline *NTHL1* LoF variants remained to be established. As carriers of monoallelic germline LoF variants in *MUTYH* were previously found to have a small increased risk for CRC as well, we investigated the role of monoallelic germline LoF variants in *NTHL1* on the risk of adenomatous polyposis and CRC in **Chapter 6.** To establish a large cohort to investigate the monoallelic *NTHL1* LoF variants role we established the collaborative *NTHL1* study group. In total 5,942 individuals with unexplained polyposis and/or CRC were screened. We demonstrated that monoallelic LoF variants in *NTHL1* are not enriched in individuals with polyposis and/or CRC compared to the general populations. Furthermore, mutational signature analysis on 13 colorectal tumors of

individuals with a monoallelic *NTHL1* LoF variant did not show a somatic second hit, nor did we find evidence of a main contribution of mutational signature SBS30, the signature associated with NTHL1 deficiency, indicating that monoallelic loss of *NTHL1* does not substantially contribute to colorectal tumor development. Thus, we found no evidence that monoallelic *NTHL1* LoF variant carriers are at increased risk to develop polyposis and/or CRC.

Nederlandse samenvatting

Erfelijke factoren zijn verantwoordelijk voor ongeveer 35% van het risico op dikkedarmkanker. In 5 tot 10% van de darmkankers is er sprake van een kiembaanvariant in een bekend dikkedarmkanker gen, bij de overige is het erfelijke risico nog niet volledig begrepen. In de afgelopen jaren is vooruitgang geboekt bij het ontdekken van nieuwe genetische oorzaken van dikkedarmkanker en het hebben van veel darmpoliepen (polyposis), zoals kiembaanvarianten in de genen *POLE*, *POLD1* en "biallelische" varianten in *NTHL1*, en nieuwe genen worden nog steeds beschreven. Kennis van de genetica van erfelijk dikkedarmkanker is essentieel voor het identificeren van personen met een verhoogd risico, het verbeteren van kankersurveillance- en preventiestrategieën, en bepalen van effectievere diagnostische en therapeutische benaderingen. De studies in dit proefschrift zijn gericht op het karakteriseren van varianten in bekende hoogrisicogenen zoals *POLE*, *POLD1*, *APC* en *NTHL1* als genetische oorzaken van polyposis en dikkedarmkanker

In **Hoofdstuk 2** hebben we een onderzoek uitgevoerd naar een groep patiënten met dikkedarmkanker en polyposis zonder bekende erfelijke oorzaak. We hebben gekeken naar de aanwezigheid van door de auteurs Palles en medewerkers ontdekte pathogene kiembaanvarianten in *POLE* en *POLD1*. Deze studie laat zien dat tumoren met *POLE*-kiembaanvarianten een fenotype kunnen vertonen dat lijkt op het Lynch-syndroom. Dit komt doordat de POLE DNA verandering (mutatie) leidt tot opvolgende somatische mutaties in de zogenaamde DNA-schadeherstel (MMR)-genen en dus tot DNA mismatch-repair-deficiëntie. Hoewel de patiënten die werden onderzocht geselecteerd waren op leeftijd en aangedane familie anamnese, bleek de frequentie van POLE-varianten laag te zijn en vergelijkbaar met eerder gerapporteerde frequenties in dikke darmkanker en/of polyposis studies.

In **Hoofdstuk 3** hebben we de DNA "exonuclease-domeinen" van *POLE* en *POLD1* onderzocht bij een groep indexpatiënten met meerdere colorectale poliepen. Het doel was om aanvullende pathogene varianten in deze genen te identificeren met behulp van nieuw generatie DNA sequentieanalyse (NGS). We vonden geen nieuwe ziekte veroorzakende (pathogene) DNA varianten. Wel vonden we DNA varianten in *POLD1* die van onzekere betekenis waren. Helaas konden we de functionele relevantie van deze varianten niet verder onderzoeken, omdat er geen tumormateriaal beschikbaar was. De resultaten van deze studie bevestigen de lage frequentie van pathogene varianten in *POLE*

en *POLD1*. Varianten in deze genen zijn een zeldzame oorzaak van colorectale poliepen en dikkedarmkanker.

In **Hoofdstuk 4** hebben we eerder gerapporteerde pathogene "diep-intronische" DNA kiembaanvarianten in *APC* gescreend in een groep onverklaarde patiënten met colorectale polyposis. Met behulp van diepe NGS hebben we bovendien gezocht naar kiembaan *"APC*-mozaïek" varianten, varianten die slechts in een deel van de lichaamscellen aanwezig is. We hebben geen *APC*-varianten gedetecteerd bij de gescreende patiënten met onverklaarde colorectale polyposis. De beperking van deze studie was dat we alleen bloed (leukocyten) DNA hebben gescreend op mozaïek varianten. Als gevolg hiervan zouden *APC*-mozaïek varianten die beperkt zijn tot de dikke darm gemist kunnen zijn omdat we het DNA van de darmpoliepen van de patiënten niet konden onderzoeken.

In **Hoofdstuk 5** hebben we het *NTHL1*-tumorsyndroom gekarakteriseerd. In 2015 werd aangetoond dat biallelische (twee varianten) "loss-of-function" (LoF) varianten in *NTHL1* verhoogd risico op polyposis en dikkedarmkanker geven, maar omdat het aantal patiënten voor dit zeldzame syndroom klein was, is het exacte klinische fenotype onduidelijk. In een internationale samenwerking met onderzoeksgroepen uit Nederland, het Verenigd Koninkrijk, Polen, Duitsland, Noorwegen, Spanje en Macedonië zijn klinische en moleculaire gegevens verzameld van 29 individuen uit 17 families met biallelische kiembaan *NTHL1* LoF-varianten. De studie toont aan dat 55% van de mutatiedragers meerdere primaire tumoren op verschillende locaties ontwikkeld. Naast colorectale tumoren hebben we tumoren gevonden in 13 verschillende weefseltypes. Veel individuen ontwikkelden dikkedarmkanker (59%), en bij vrouwelijke dragers werd een hoge incidentie van borstkanker waargenomen (60%). Ook werd een unieke DNA schade blauwdruk (SBS30) gevonden die geassocieerd is met NTHL1-deficiëntie.

Hoewel biallelische kiembaan *NTHL1* LoF-varianten de oorzaak zijn van het krijgen van veel darmpoliepen en dikkedarmkanker, is het risico op het krijgen van deze darmpoliepen en dikkedarmkanker bij dragers van enkelvoudige (monoallelische) kiembaan *NTHL1* LoF-varianten nog onbekend. In **Hoofdstuk 6** hebben we daarom onderzocht wat de rol is van monoallelische kiembaan *NTHL1* LoF-varianten op het risico van adenomateuze polyposis en dikkedarmkanker binnen de internationale *NTHL1*-samenwerking. Bijna 6000 individuen met onverklaarde polyposis en/of dikkedarmkanker werden gescreend op monoallelische *NTHL1* LoF-varianten. In deze groep werd aangetoond dat

monoallelische LoF-varianten in *NTHL1* niet verrijkt zijn in vergelijking met de algemene bevolking. Bovendien toonde mutatie analyse van de colorectale tumoren in deze groep ook geen andere *NTHL1* DNA mutaties aan en we vonden ook geen bewijs van een belangrijke bijdrage van de *NTHL1*- mutatiesignatuur SBS30. Dit laat zien dat monoallelisch verlies van *NTHL1* niet in belangrijke mate bijdraagt aan de ontwikkeling van colorectale tumoren en dat dragers van monoallelische *NTHL1* LoF-varianten een verhoogd risico hebben op het ontwikkelen van polyposis en/of dikkedarmkanker.

List of publications

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. Human pathology. 2022 Jul;125:11-17. doi: 10.1016/j.humpath.2022.04.003.

Use of sanger and next-generation sequencing to screen for mosaic and intronic APC variants in unexplained colorectal polyposis patients.
Elsayed FA, Tops CMJ, Nielsen M, Morreau H, Hes FJ, van Wezel T. Familial Cancer. 2022 Jan; 21(1):79-83. doi: 10.1007/s10689-021-00236-2.

Monoallelic NTHL1 loss-of-function variants and risk of polyposis and colorectal cancer.
Elsayed FA*, Grolleman JE*, Ragunathan A*; NTHL1 study group; Buchanan DD, van Wezel T, de Voer RM. Gastroenterology. 2020 Dec;159(6):2241-2243.e6. doi: 10.1053/j.gastro.2020.08.042.

Low frequency of POLD1 and POLE exonuclease domain variants in patients with multiple colorectal polyps.
Elsayed FA, Tops CMJ, Nielsen M, Ruano D, Vasen HFA, Morreau H, Hes FJ, van Wezel T. Molecular Genetics & Genomic Medicine. 2019 Apr;7(4):e00603. doi: 10.1002/mgg3.603.

Mutational signature analysis reveals NTHL1 deficiency to cause a multi-tumor phenotype.
Grolleman JE*, de Voer RM*, **Elsayed FA***, Nielsen M*, Weren RDA*, Palles C, Ligtenberg MJL, Vos JR, Ten Broeke SW, de Miranda NFCC, Kuiper RA, Kamping EJ, Jansen EAM, Vink-Börger ME, Popp I, Lang A, Spier I, Hüneburg R, James PA, Li N, Staninova M, Lindsay H, Cockburn D, Spasic-Boskovic O, Clendenning M, Sweet K, Capellá G, Sjursen W, Høberg-Vetti H, Jongmans MC, Neveling K, Geurts van Kessel A, Morreau H, Hes FJ, Sijmons RH, Schackert HK, Ruiz-Ponte C, Dymerska D, Lubinski J, Rivera B, Foulkes WD, Tomlinson IP, Valle L, Buchanan DD, Kenwrick S, Adlard J, Dimovski AJ, Campbell IG, Aretz S, Schindler D, van Wezel T, Hoogerbrugge N, Kuiper RP. Cancer Cell. 2019 Feb 11;35(2):256-266.e5. doi: 10.1016/j.ccell.2018.12.011. *These authors contributed equally.

SNP association study in PMS2-associated Lynch syndrome.
Ten Broeke SW, **Elsayed FA**, Pagan L, Olderode-Berends MJW, Garcia EG, Gille HJP, van Hest LP, Letteboer TGW, van der Kolk LE, Mensenkamp AR, van Os TA, Spruijt L, Redeker BJW, Suerink M, Vos YJ, Wagner A, Wijnen JT, Steyerberg EW, Tops CMJ, van Wezel T, Nielsen M. Familial Cancer. 2018 Oct;17(4):507-515. doi: 10.1007/s10689-017-0061-3.

Evidence for genetic association between chromosome 1q loci and predisposition to colorectal neoplasia.
Schubert SA, Ruano D, **Elsayed FA**, Boot A, Crobach S, Sarasqueta AF, Wolffenbuttel B, van der Klauw MM, Oosting J, Tops CM, van Eijk R, Vasen HF, Vossen RH, Nielsen M, Castellví-Bel S, Ruiz-Ponte C, Tomlinson I, Dunlop MG, Vodicka P, Wijnen JT, Hes FJ, Morreau H, de Miranda NF, Sijmons RH, van Wezel T. British Journal of Cancer. 2017 Sep 5;117(6):1215-1223. doi: 10.1038/bjc.2017.240.

Germline variants in POLE are associated with early onset mismatch repair deficient colorectal cancer.
Elsayed FA, Kets CM, Ruano D, van den Akker B, Mensenkamp AR, Schrumpf M, Nielsen M, Wijnen JT, Tops CM, Ligtenberg MJ, Vasen HF, Hes FJ, Morreau H, van Wezel T. European Journal of Human Genetics. 2015 Aug;23(8):1080-4. doi: 10.1038/ejhg.2014.242.

Curriculum vitae

Fadwa was born in Khartoum, Sudan. She obtained the degree of bachelor of science (honours) from Faculty of Science, University of Khartoum, Sudan. After which she started working as a teaching assistant in Department of Biology and Biotechnology, Faculty of Science and Technology, Al Neelain University, Sudan. While working there she obtained her master degree from Faculty of Science, University of Khartoum, after which she was promoted to lecturer in the Department of Biology and Biotechnology, Faculty of Science and Technology, Al Neelain University. To further build her career she came to the Leiden University Medical Center, Department of Pathology in 2012 and she worked as researcher under supervision of Dr. Tom van Wezel and Dr. Frederik Hes in a collaborative project between the Departments of Pathology and Clinical Genetics. The project aimed at identification of new genes predisposing for polyposis and colorectal cancer. In 2015 she continued working on this project as a PhD candidate in the Department of Pathology and in collaboration with the Department of Clinical Genetics under supervision of Dr. Tom van Wezel, Dr. Maartje Nielsen and Prof. dr. Hans Morreau. The results of this PhD research are presented in this thesis. During her PhD she also worked on other projects not included in this thesis such as a SNP association studies in PMS2-associated Lynch syndrome and in MAP patients, and the role of digenic inheritance of *NTHL1* and *MUTYH* in predisposition colorectal cancer.

Acknowledgements

I would like to express my deepest appreciation to all the people who contributed to this thesis, “Together everyone achieves more”.