



**Universiteit
Leiden**
The Netherlands

Characterization of candidate genes in unexplained polyposis and colorectal cancer

Abayzeed Elsayed Osman, F.

Citation

Abayzeed Elsayed Osman, F. (2023, November 28). *Characterization of candidate genes in unexplained polyposis and colorectal cancer*. Retrieved from <https://hdl.handle.net/1887/3665175>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3665175>

Note: To cite this publication please use the final published version (if applicable).

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 Kenwick, 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

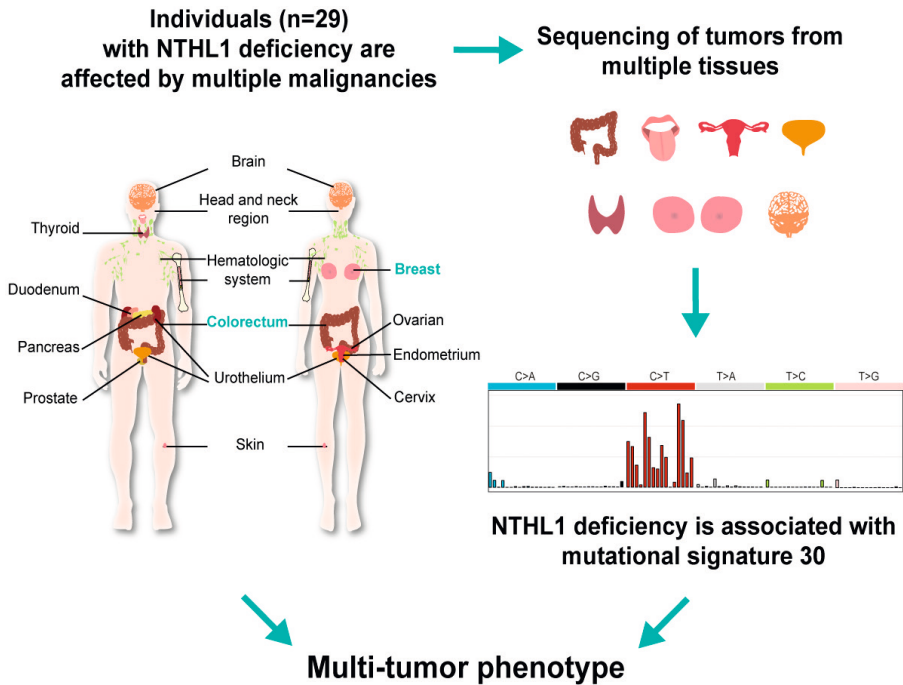
Summary

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 genes³. 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 state². 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.

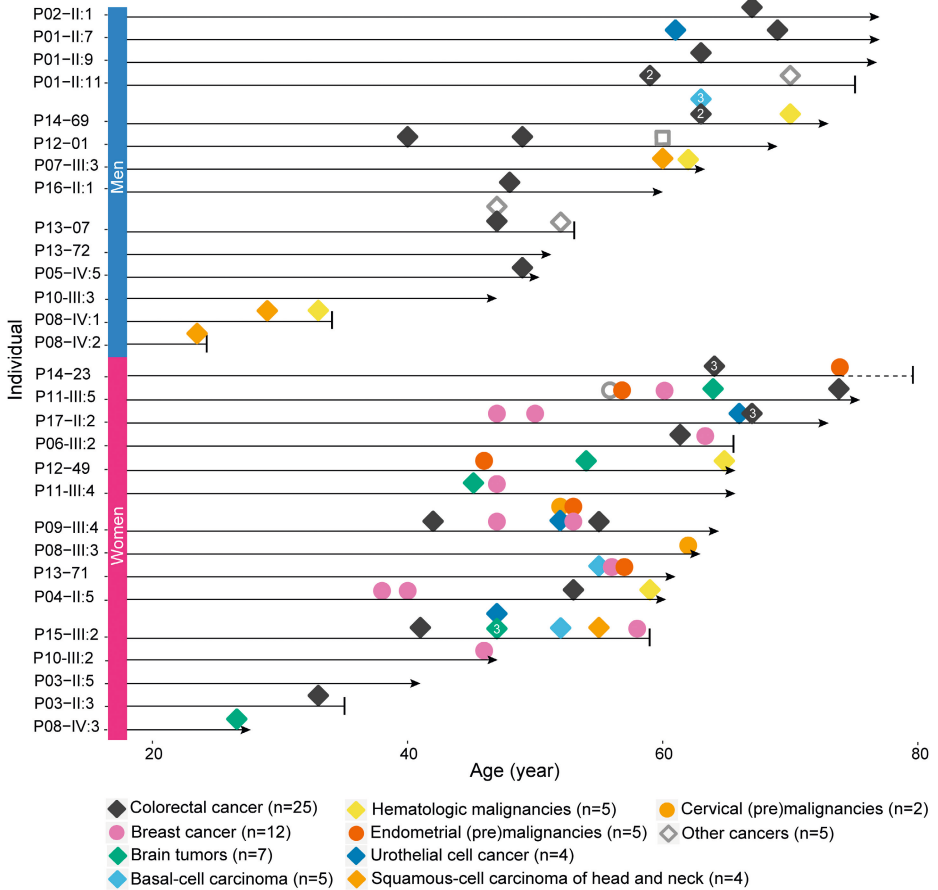


Figure 1. Age of diagnosis of benign meningiomas and (pre)malignant tumors per classification of all 29 individuals with biallelic germline *NTHL1* mutations.

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

Table 1. (continued)

	Family cDNA change (NIM002528.6) change	Amino Acid	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 ^e (29)	No colonoscopy performed	MDS ^e (33)	This study				
			P08-IV:2	M	SCC of the tongue tip ^e (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	11A >4H		This study				
					BC (left, 47), lobular							
					BIC ^g (52), papillary							
					Endocervical adenocarcinoma in situ ^h (52)							
					BC (right, 53), ductal							
EC (53), serous												
CRC (55), transversum												
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
					P11-III:4				F	BC (right, 47), mixed ductal/papillary	13A 2H	Meningioma ^f (45) Breast papilloma (left, 49) Uterine polyps
11	c..268C>T/ c..390>A p.Gln90*/ p.Tyr130*		P11-III:5	F	OC (57), mixed endometrioid/mucinous	No colonoscopy performed	Meningioma ^f (64), right parasellar meninges	This study				
					EC (57), mixed endometrioid/mucinous							
					BC (left, 60), papillary and triple negative							
					CRC (73), ascendens							

Table 1. (continued)

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

Table 1. (continued)

Family	cDNA change (NM002528.6)	Amino Acid change	Patient ID ^a	M/F	Malignancies and pre-malignancies ^b	Polypos ^c	Benign lesions ^b	Publication
15	c.268C>T/	p.Gln90*/	P15-III:2	F	CRC (41) BIC ^d (47) BCC (52) SCC of head and neck ^e (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
	c.709+1G>A	Abnormal splicing						
16	c.268C>T	p.Gln90*	P16-II:1	M	CRC, ascendens (48) BC (left, 47) BC (right, 50), lobular BIC ^d (66), papillary CRC (67), ascendens CRC (67), ascendens CRC (67), ascendens	30A 1H		(Belhadjet al., 2017) ⁵
17	c.268C>T	p.Gln90*	P17-II:2	F		>15A 5H		(Belhadjet al., 2017) ⁵

^aThe index patient is shown in bold. ^bNumbers 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. ^cNumbers 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). ^dClassified as cervical urothelial cell cancer. ^eClassified as hematologic malignancies. ^fClassified as brain tumors. ^gClassified as head and neck squamous cell carcinoma. ^hClassified as cervical (pre)malignancies. ⁱClassified as endometrial (pre)malignancies. ^jPathology reports suggest two individual primary tumors. ^kp12-01 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 spectrum¹². 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* mutations². Next, we jointly extracted the mutational signatures from six colon tumors, of which three were previously sequenced and yielded sufficient mutations² (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 COSMIC¹¹ 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.

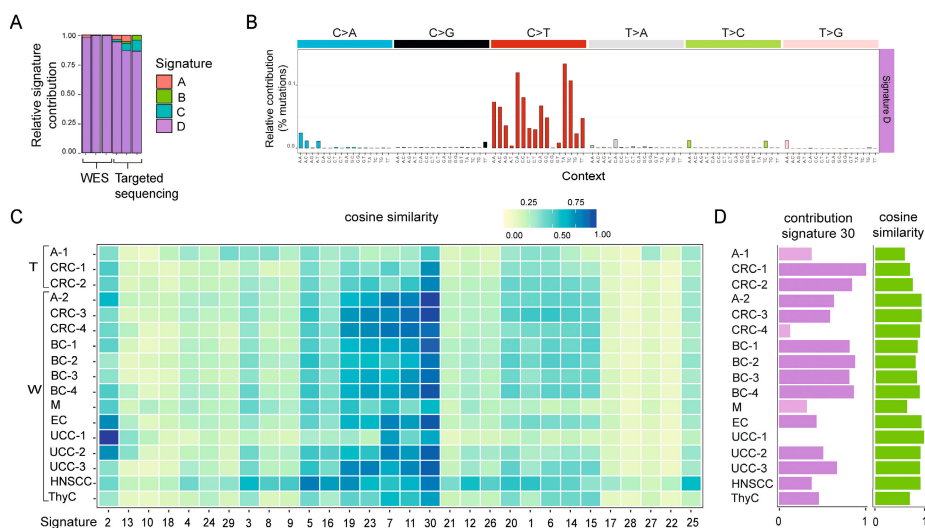


Figure 2. Mutational signature analysis of colonic and extracolonic *NTHL1* deficient tumors.

(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 COSMIC¹¹ 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 <i>NTHL1</i> patients (range)			Median age of diagnosis in the population
	M	F	M	F	M+F	
	(n=14)	(n=15)	(n=14)	(n=15)	(n=29)	
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

^aSEER data, period 2010-2014. ^bDutch cancer registry data, period 2010-2016, data from the South of the Netherlands. ^cDutch 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 out¹². Furthermore, it was found that the single breast cancer sample in which signature 30 originally was identified¹⁴ 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 tumor¹². 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 MAP¹⁷. 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* carriers²³. 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		
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'-ACCAGCTGTGCTGCCAGTCCT-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 used³⁰. 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 $\leq 10\%$ 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 Netherlands², 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 previously³². High confident somatic variant calls, i.e. ≥ 15 fold coverage, with $\geq 20\%$ or $\leq 80\%$ variant reads, of the corresponding genomic position in both the tumor and corresponding germline sample, were selected with the same approach as described previously³³. Subsequently, variant calls observed in our in-house database of germline variants³², 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 $\leq 10\%$ or $\geq 80\%$ 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 substitutions³⁴ was performed using nonnegative matrix factorization²⁷. 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 tool²⁸. 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*, *POLD1* 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 previously³⁷. 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 *RAD51C*. 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*, *POLD1* 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 0.003¹⁷. 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 Mendel³⁸, and other analyses were performed in R.

Data and software availability

Data resources

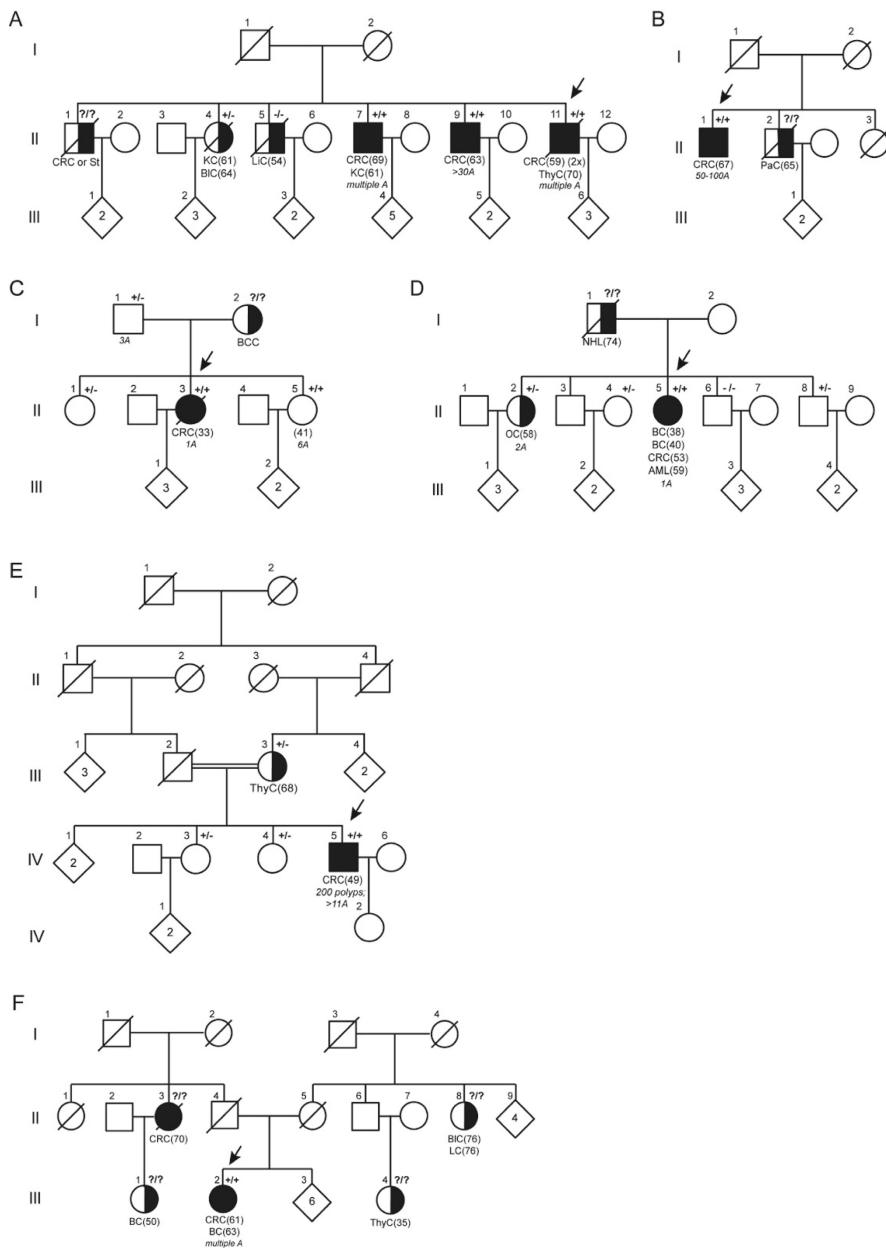
The analyzed whole-exome sequencing data are available in Table S3. The accession number of the raw whole-exome sequencing data reported in this paper is: EGAD00001004534.

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



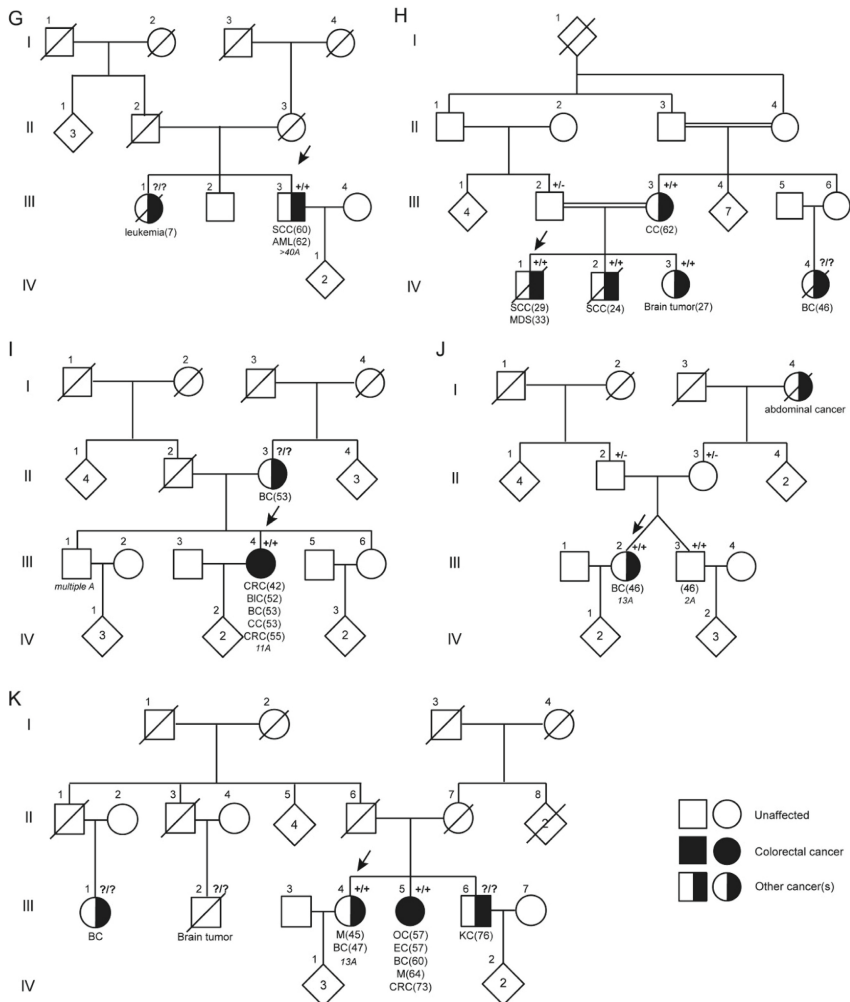


Figure S1, related to Figure 1 and Table 1. Pedigrees of 11 novel families with biallelic germline *NTHL1* mutations.

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, BIC^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: myelodysplastic 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. ^aClassified as hematologic malignancies, ^bClassified as urothelial cell cancer. ^cClassified as cervical (pre)malignancies. ^dClassified as endometrial (pre)malignancies. ^eClassified 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 <i>NTHL1</i> 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
		BIC (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
<u>P14-23</u> (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
<u>P15-III:2</u> (Rivera et al., 2015) ²	F	CRC (41)	NA	-
		BIC (47)	NA	Very unlikely
		BCC (52)	NA	Unlikely
		SCC of head and neck (55)	NA	Unlikely
		BC (58)	NA	Unlikely
<u>P17-II:2</u> (Belhadj et al., 2017) ³	F	BC (47)	NA	-
		BC (50)	Resection (bilateral)	Unlikely
		BIC (66)	NA	Unlikely
		CRC (67, 67, 67)	NA	Unlikely

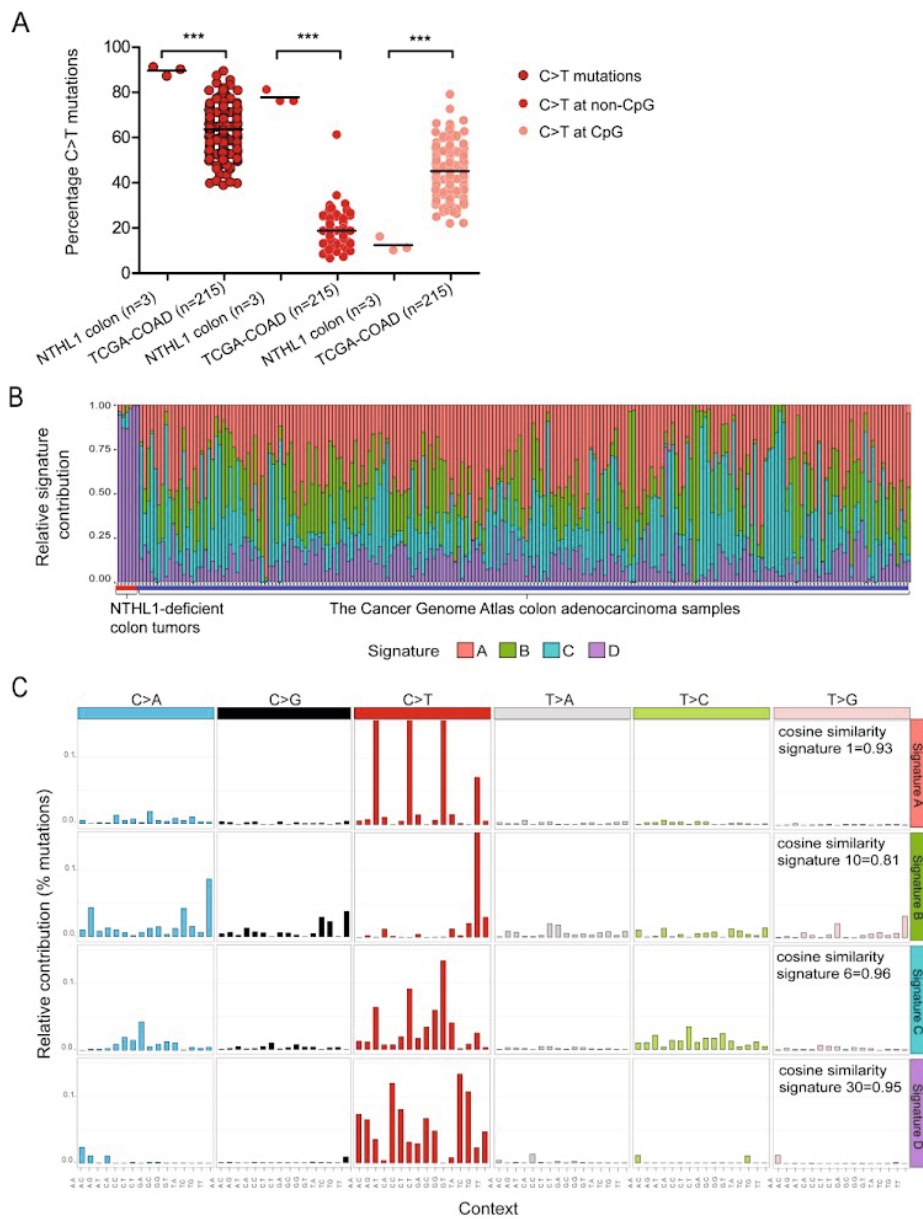
^aThe index patient is underlined. ^bAML: acute myeloid leukemia, BC: breast cancer, BCC: basal-cell carcinoma, BIC: bladder cancer, CC: cervical cancer, CRC: colorectal cancer, DC: duodenal cancer, EC: endometrium cancer, MDS: myelodysplastic syndrome, OC: ovarian cancer, PaC: pancreas cancer, PC: prostate cancer, SCC: squamous cell carcinoma, ThyC: thyroid cancer. Numbers represent age of diagnosis. ^cInformation may be incomplete. NA: treatment information not available. ^dNo; 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; ^aSamples previously sequenced by Weren et al., Nature Genetics, 2015. ^bFrom these tumors we were unable to retrieve at least 10 somatic mutations to perform mutational signature analyses. ^cThere 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.



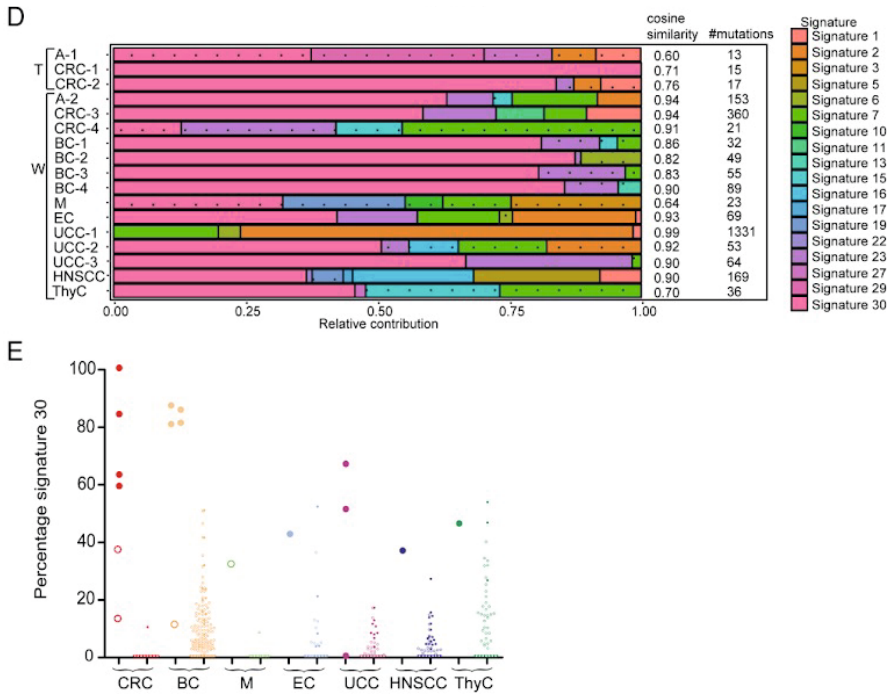


Figure S2, related to Figure 2. Mutation spectrum and mutational signature analysis of NTHL1-deficient colon tumors and sporadic TCGA colon adenocarcinoma samples.

(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

Age	Cumulative risk extracolonic lifetime	
	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 <i>NTHL1</i>
Skopje, Macedonia	12	Polyposis, recessive inheritance	MIP-based NGS	MMR genes, <i>APC</i> , <i>TP53</i> , <i>MUTYH</i> , <i>POLE</i> , <i>POLD1</i>	p.Gln90*/p.Trp269* (CH)
Nijmegen, the Netherlands	169	Polyposis or familial CRC	MIP-based NGS	<i>APC</i> , <i>MUTYH</i>	p.Gln90* (hom)
Nijmegen, the Netherlands	348	Polyposis or familial CRC	KASPAR assay p.Gln90*	<i>APC</i> , <i>MUTYH</i> , <i>POLE</i> , <i>POLD1</i> , MMR genes	none
Dresden, Germany	105	Polyposis or familial CRC	MIP-based NGS	<i>APC</i> , <i>MUTYH</i>	none
Oxford, United Kingdom	273	Polyposis	MIP-based NGS	<i>APC</i> , <i>MUTYH</i>	none
Szczecin, Poland	145	Familial CRC	MIP-based NGS	<i>POLE</i> , <i>POLD1</i> , MMR genes*	none
Santiago de Compostela, Spain	36	Polyposis or familial CRC	MIP-based NGS	<i>APC</i> , <i>MUTYH</i> (partly), <i>POLE</i> , <i>POLD1</i> , <i>BMPR1A</i> , <i>SMAD4</i> , <i>PTEN</i>	none
Trondheim, Norway	61	Polyposis or familial CRC	Sanger Sequencing	<i>APC</i> , <i>MUTYH</i> , MMR genes	none
Bergen, Norway	27	Polyposis or familial CRC	Sanger Sequencing	<i>APC</i> , <i>MUTYH</i> , MMR genes*	none
Leiden, Netherlands	1,494 ^a	Polyposis, or familial/young CRC	KASPAR assay p.Gln90* Sanger Sequencing	<i>APC</i> , <i>MUTYH</i> , <i>POLE</i> , <i>POLD1</i> , 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*. ^aDNA 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. ^bPolyposis 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.

