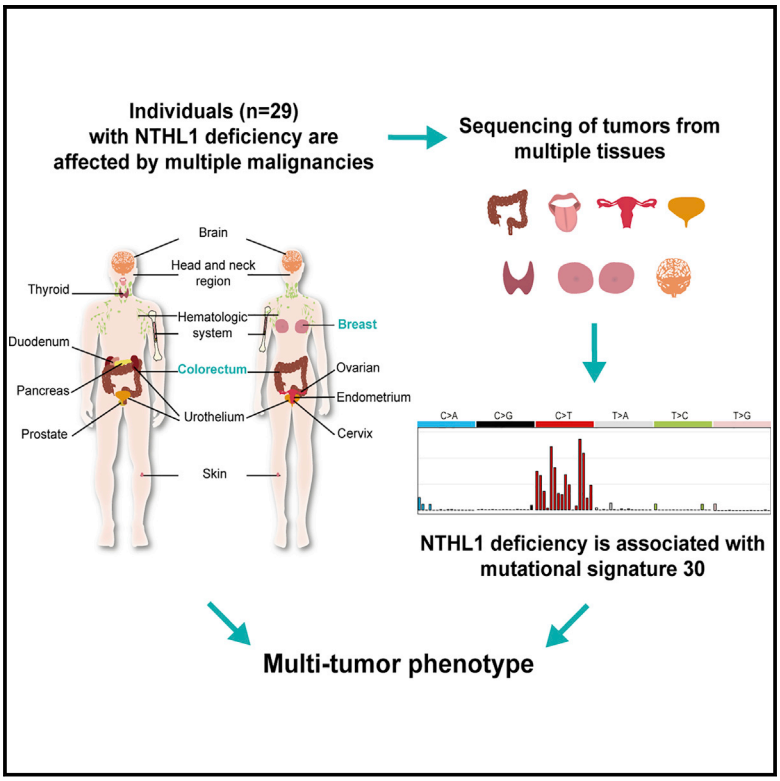
Article

Cancer Cell

Mutational Signature Analysis Reveals NTHL1 Deficiency to Cause a Multi-tumor Phenotype

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

Authors

Judith E. Grolleman,
Richarda M. de Voer,
Fadwa A. Elsayed, ..., Tom van Wezel,
Nicoline Hoogerbrugge,
Roland P. Kuiper

Correspondence

richarda.devoer@radboudumc.nl (R.M.d.V.),
r.kuiper@prinsesmaximacentrum.nl (R.P.K.)

In Brief

In addition to the know colorectal tumors, Grolleman et al. find tumors in 13 tissue types, including a high breast cancer incidence, among 29 carriers of biallelic germline *NTHL1* mutations and identify a mutation signature across tumor types, which may facilitate the identification and management of new cases.

Grolleman et al., 2019, Cancer Cell *35*, 256–266
February 11, 2019
https://doi.org/10.1016/j.ccell.2018.12.011

CellPress

CellPress

Cancer Cell
Article

Mutational Signature Analysis Reveals NTHL1 Deficiency to Cause a Multi-tumor Phenotype

Judith E. Grolleman,[1,36] Richarda M. de Voer,[1,36,38,*] Fadwa A. Elsayed,[2,36] Maartje Nielsen,[3,36] Robbert D.A. Weren,[1,36] Claire Palles,[4] Marjolijn J.L. Ligtenberg,[1,5] Janet R. Vos,[6] Sanne W. ten Broeke,[3] Noel F.C.C. de Miranda,[2] Renske A. Kuiper,[1] Eveline J. Kamping,[1] Erik A.M. Jansen,[1] M. Elisa Vink-Börger,[5] Isabell Popp,[7] Alois Lang,[8] Isabel Spier,[9,10] Robert Hüneburg,[10,11] Paul A. James,[12] Na Li,[13,14] Marija Staninova,[15] Helen Lindsay,[16]

(Author list continued on next page)

[1]Department of Human Genetics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, 6525 GA Nijmegen, the Netherlands
[2]Department of Pathology, Leiden University Medical Center, 2300 RC Leiden, the Netherlands
[3]Department of Clinical Genetics, Leiden University Medical Center, 2300 RC Leiden, the Netherlands
[4]Molecular and Population Genetics Laboratory, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK
[5]Department of Pathology, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, 6525 GA Nijmegen, the Netherlands
[6]Department of Human Genetics, Radboud Institute for Health Sciences, Radboud University Medical Center, 6525 GA Nijmegen, the Netherlands
[7]Department of Human Genetics, University of Würzburg, 97074 Würzburg, Germany
[8]Vorarlberg Cancer Registry, Agency for Preventive and Social Medicine, Bregenz 6900, Austria
[9]Institute of Human Genetics, University of Bonn, 53127 Bonn, Germany
[10]Center for Hereditary Tumor Syndromes, University of Bonn, 53127 Bonn, Germany

(Affiliations continued on next page)

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 7 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.

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* (Al-Tassan et al., 2002; Weren et al., 2015), and dominantly inherited mutations in the polymerase proofreading domains of the *POLE* and *POLD1* polymerase genes (Palles et al., 2013). In addition to adenomatous polyposis and CRC, these syndromes appear to predispose to the development of other types of cancer (Adam et al., 2016; Belhadj et al., 2017; Briggs and Tomlinson, 2013; Nielsen et al., 1993; Weren et al., 2015).

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 seven 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.

David Cockburn,[16] Olivera Spasic-Boskovic,[17] Mark Clendenning,[18,19] Kevin Sweet,[20] Gabriel Capellá,[21] Wenche Sjursen,[22,23] Hildegunn Høberg-Vetti,[24] Marjolijn C. Jongmans,[1] Kornelia Neveling,[1] Ad Geurts van Kessel,[1] Hans Morreau,[2] Frederik J. Hes,[3] Rolf H. Sijmons,[25] Hans K. Schackert,[26] Clara Ruiz-Ponte,[27] Dagmara Dymerska,[28] Jan Lubinski,[28] Barbara Rivera,[29] William D. Foulkes,[30] Ian P. Tomlinson,[4,31] Laura Valle,[21] Daniel. D. Buchanan,[18,19,32,33] Sue Kenwrick,[17] Julian Adlard,[34] Aleksandar J. Dimovski,[15] Ian G. Campbell,[13,14] Stefan Aretz,[9,10] Detlev Schindler,[7] Tom van Wezel,[2] Nicoline Hoogerbrugge,[1,37] and Roland P. Kuiper[1,35,37,*]

[11]Department of Internal Medicine I, University of Bonn, 53127 Bonn, Germany
[12]Familial Cancer Centre, Peter MacCallum Cancer Centre, Melbournem, VIC 3000, Australia
[13]Cancer Genetics Laboratory, Peter MacCallum Cancer Centre, Melbourne, VIC 3000, Australia
[14]Sir Peter MacCallum Department of Oncology, University of Melbourne, Melbourne, VIC 3000, Australia
[15]Center for Biomolecular Pharmaceutical Analyzes, UKIM Faculty of Pharmacy, 1000 Skopje, Republic of Macedonia
[16]Leeds Genetics Laboratory, Leeds Teaching Hospitals NHS Trust, Leeds LS9 7TF, UK
[17]East Anglian Medical Genetics Service, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK
[18]Colorectal Oncogenomics Group, Department of Clinical Pathology, The University of Melbourne, Parkville, VIC 3010, Australia
[19]University of Melbourne Centre for Cancer Research, Victorian Comprehensive Cancer Centre, Parkville, VIC 3010, Australia
[20]Division of Human Genetics, Ohio State University Medical Centre, Columbus, OH 43221, USA
[21]Hereditary Cancer Program, Catalan Institute of Oncology, IDIBELL, CIBERONC, Hospitalet de Llobregat, Barcelona 08908, Spain
[22]Department of Medical Genetics, St Olavs University Hospital, 7030 Trondheim, Norway
[23]Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology (NTNU), 7491, Trondheim, Norway
[24]Western Norway Familial Cancer Center, Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, 5021 Bergen, Norway
[25]Department of Genetics, University Medical Center Groningen, University of Groningen, 9700 RB Groningen, The Netherlands
[26]Department of Surgical Research, Universitätsklinikum Carl Gustav Carus, Technische Universität Dresden, 01307 Dresden, Germany
[27]Fundación Pública Galega de Medicina Xenómica (FPGMX)-SERGAS, Grupo de Medicina Xenómica-USC, Instituto de Investigación Sanitaria de Santiago (IDIS), Santiago de Compostela, Galicia 15706, Spain
[28]Hereditary Cancer Center, Department of Genetics and Pathology, Pomeranian Medical University, 70-111 Szczecin, Poland
[29]Gerald Bronfman Department of Oncology, McGill University, Montreal, QC H3A 0G4, Canada
[30]Department of Human Genetics, McGill University, Montreal, QC H3A 0C7, Canada
[31]Oxford National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centre, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK
[32]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Parkville, VIC 3010, Australia
[33]Genomic Medicine and Family Cancer Clinic, Royal Melbourne Hospital, Parkville, VIC 3010, Australia
[34]Yorkshire Regional Genetics Service and University of Leeds, Leeds LS7 4SA, UK
[35]Princess Máxima Center for Pediatric Oncology, 3584 CT Utrecht, The Netherlands
[36]These authors contributed equally
[37]Co-senior author
[38]Lead Contact
*Correspondence: richarda.devoer@radboudumc.nl (R.M.d.V.), r.kuiper@prinsesmaximacentrum.nl (R.P.K.)
https://doi.org/10.1016/j.ccell.2018.12.011

The first families described with *NTHL1* mutations were of Dutch origin, all having the same truncating germline mutation (p.Gln90*) in a homozygous state (Weren et al., 2015). 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) (Fostira et al., 2018; Belhadj et al., 2017; Chubb et al., 2016; Rivera et al., 2015; Weren et al., 2015). Three of these families have previously been described in detail (Belhadj et al., 2017; Rivera et al., 2015). 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; Table 1). Only 1 of 33

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	multiple a		this study
					CRC (59), transversum			
					ThyC (70), follicular			
			P01-II:7	M	renal pyelum cancer[d] (61), papillary	multiple a	neurofibroma	this study
					CRC (69), ileocecal			
			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		this study
						2H		
			P03-II:5	F	none (41)	6A		this study
						7H		
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
6	c.268C > T/c.235_236insG	p.Gln90*/ p.Ala79Glyfs*2	**P06-III:2**	F	CRC (61), transversum	multiple a >30H		this study
					BC (right, 63), triple-negative			
7	c.806G > A/c.859C > T	p.Trp269*/p.Gln287*	**P07-III:3**	M	SCC of the parotid gland[f] (60),	>40A		this study
					AML[e] (62)			
8	c.545G > A	p.Trp182*	**P08-IV:1**	M	SCC of the mouth base[f] (29)	no colonoscopy performed	MDS[e] (33)	this study
			P08-IV:2	M	SCC of the tongue tip[f] (24)	no colonoscopy performed		this study
			P08-IV:3	F	brain tumor[g] (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[d] (52), papillary			
					endocervical adenocarcinoma *in situ*[h] (52)			
					BC (right, 53), ductal			
					EC[i] (53), serous			
					CRC (55), transversum			

(*Continued on next page*)

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
10	c.268C > T	p.Gln90*	**P10-III:2**	F	BC (right, 46)	13A	skin hemangiomas (3×) 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[g] (45) breast papilloma (left, 49) uterine polyps	this study
			P11-III:5	F	OC (57), mixed endometrioid/mucinous[i] EC[i] (57), mixed endometrioid/mucinous[i] BC (left, 60), papillary and triple-negative CRC (73), ascendens	no colonoscopy performed	meningioma[g] (64), right parasellar meninges	this study
12	c.268C > T	p.Gln90*	**P12-01**	M	CRC (40), rectum CRC (49), cecum PC (60)	15A[k]		Weren et al. (2015)
			P12-49	F	endometrial complex hyperplasia[i] (46) non-Hodgkin's lymphoma[e](65)	40A	psammomatous meningioma[g] (54)	Weren et al. (2015)
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)
			P13-71	F	BCC (55) BC (56) EC[i] (57)	50A		Weren et al. (2015)
			P13-72	M	none	10A		Weren et al. (2015)
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)
			P14-69	M	CRC (63), cecum CRC (63), ascendens BCC (63), nose tip BCC (63), ear BCC (63), ear non-Hodgkin's lymphoma[e] (70)	8A		Weren et al. (2015)

(*Continued on next page*)

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[f] (55) BC (58)	multiple a	ovary cystadenoma (41) intradermal nevi (42, 55) meningioma[g] (47, 47, 47) seborrheic keratosis (47)	Rivera et al. (2015)
16	c.268C > T	p.Gln90*	**P16-II:1**	M	CRC, ascendens (48)	30A 1H		Belhadj et al. (2017)
17	c.268C > T	p.Gln90*	**P17-II:2**	F	BC (left, 47) BC (right, 50), lobular BICd (66), papillary CRC (67), ascendens CRC (67), ascendens CRC (67), ascendens	>15A 5H		Belhadj et al. (2017)

See also Tables S1 and S2.

[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 the STAR Methods).

[d]Classified as urothelial cell cancer.

[e]Classified as hematologic malignancies.

[f]Classified as head and neck squamous cell carcinoma.

[g]Classified as brain tumors.

[h]Classified as cervical (pre)malignancies.

[i]Pathology reports suggest two individual primary tumors.

[j]Classified as endometrial (pre)malignancies.

[k]P12-01 developed colon, esophagus, and duodenal adenomas.

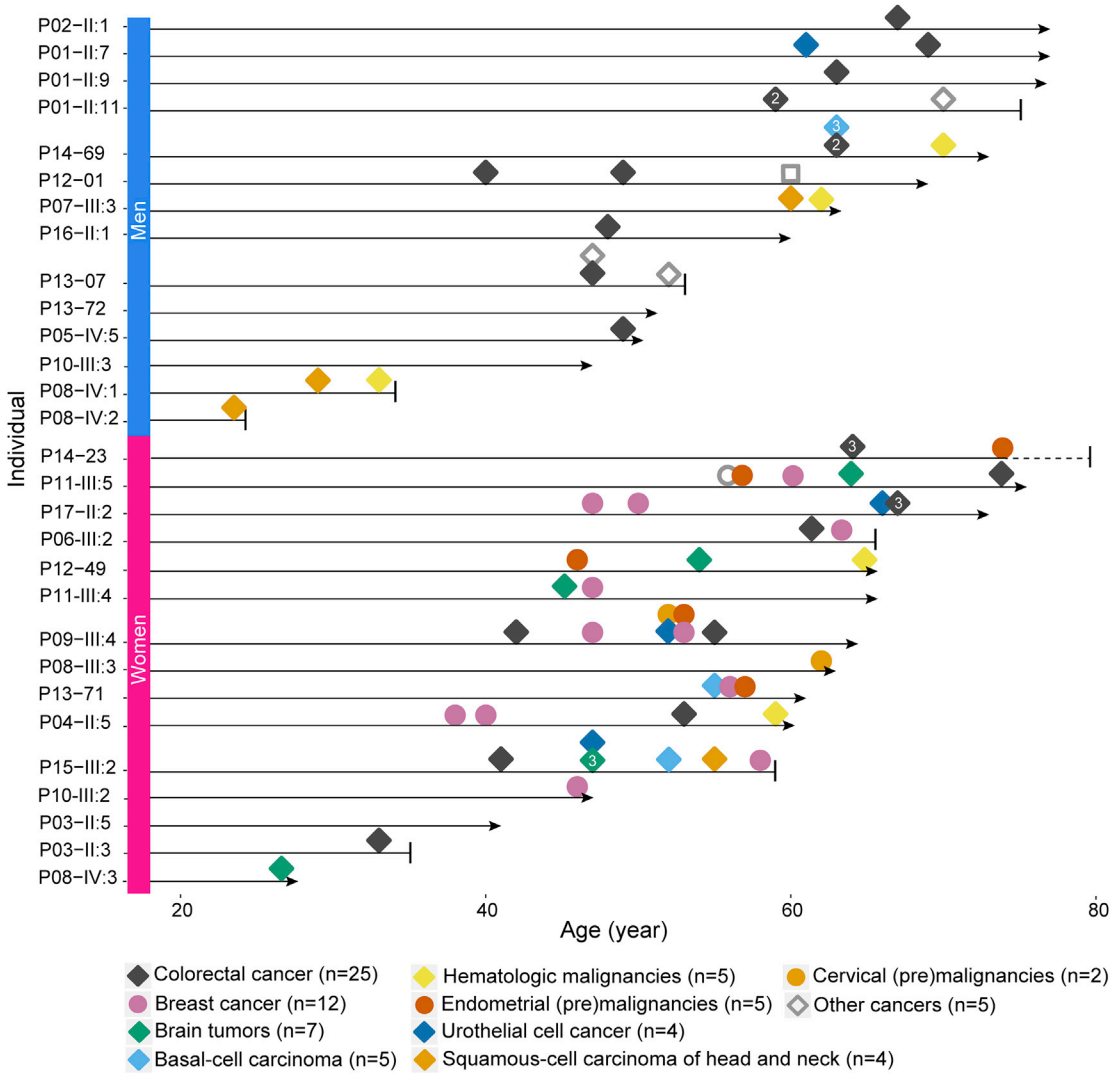


Figure 1. Age of Diagnosis of Benign Meningiomas and (Pre)malignant Tumors per Classification of all 29 Individuals with Biallelic Germline *NTHL1* Mutations

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.

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 9 were recurrently encountered (Figure 1; 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, 9 out of 15 women (60%) were diagnosed with breast cancer.

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, 2018] database), which is characterized by C > T transitions at non-CpG sites, as the main contributor to the mutation spectrum

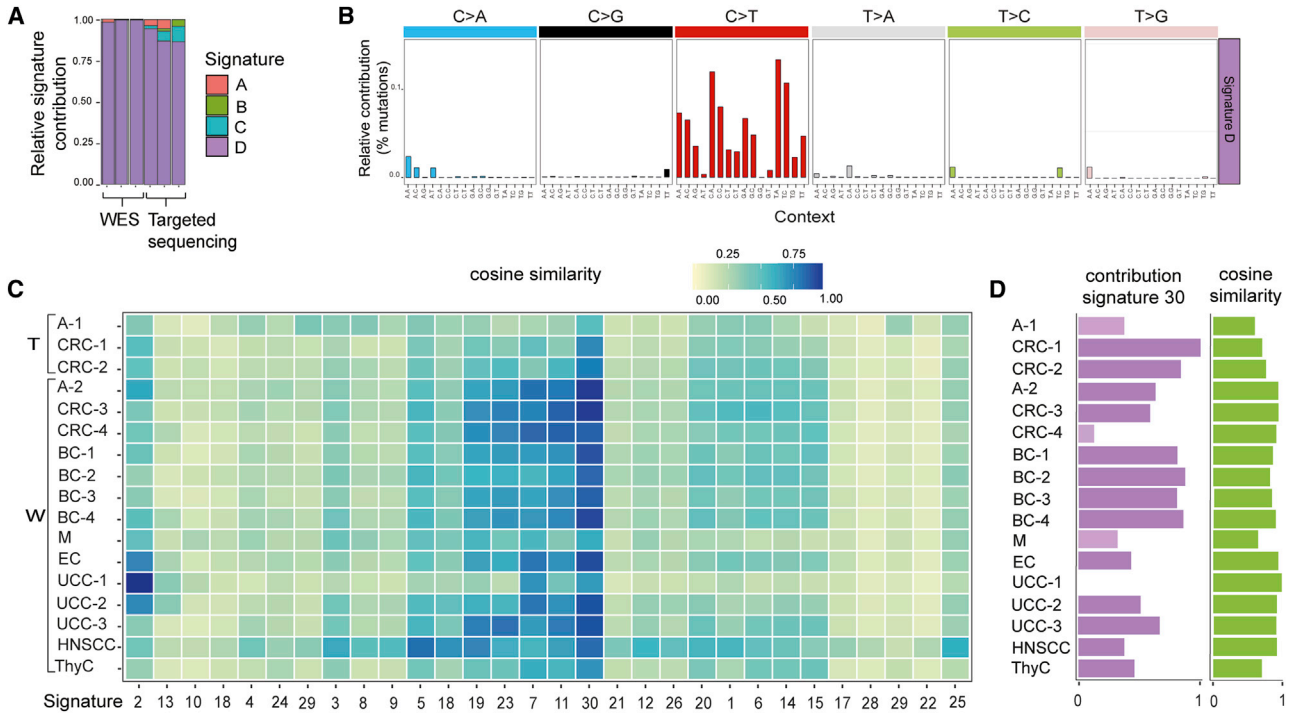


Figure 2. Mutational Signature Analysis of Colonic and Extracolonic NTHL1-Deficient Tumors

(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 spectra identified in CRCs from The Cancer Genome Atlas.

(B) Extracted *de novo* signature D that predominantly represents mutations in NTHL1-deficient colon tumors. This signature has a cosine similarity to the COSMIC, 2018 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 and S4 and Figure S2.

(Drost et al., 2017). 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 (Weren et al., 2015). Next, we jointly extracted the mutational signatures from six colon tumors, of which three were previously sequenced and yielded sufficient mutations (Weren et al., 2015) (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 (Figures S2B and S2C) (Alexandrov et al., 2013). However, all six tumors with biallelic germline *NTHL1* mutations predominantly exhibited the fourth signature that strongly resembles signature 30 reported in COSMIC, 2018 and in *NTHL1*-KO organoids (both cosine similarities 0.95; Figures 2A and 2B) (Drost et al., 2017). 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 7 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

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 in the Population
	M (n = 14)	F (n = 15)	M (n = 14)	F (n = 15)	M + F (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]

NA, not applicable for gender-specific malignancies. See also Table S5.
[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.

known mutational signatures we found that signature 30 emerged as the main mutational process in 13 tumors (93%; Figures 2D and 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.

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; 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 (Figures 2D and 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 the 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.

DISCUSSION

Following the initial discovery that biallelic germline *NTHL1* mutations predispose to the development of polyposis and CRC (Weren et al., 2015), 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 (Drost et al., 2017). Furthermore, it was found that the single breast cancer sample in which signature 30 originally was identified (Nik-Zainal et al., 2016) 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 (Drost et al., 2017). 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 (Al-Tassan et al., 2002; Pilati et al., 2017; Viel et al., 2017). 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 (Weren et al., 2018). 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 (Belhadj et al., 2017; Rivera et al., 2015). 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 (Al-Tassan et al., 2002; Barrow et al., 2009; Bellido et al., 2016; Kempers et al., 2011; Palles et al., 2013; Vogt et al., 2009; Watson et al., 2008). 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 (Win et al., 2011). 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 (Nielsen et al., 1993; Vogt et al., 2009; Win et al., 2014). For colon surveillance, we propose that the established surveillance guidelines for MAP should be extended to individuals with biallelic germline *NTHL1* mutations (Belhadj et al., 2017; Nielsen et al., 1993), 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 (Guidelines, 2018 National Comprehensive Cancer Network, 2018). 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

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Patient Cohorts
- METHOD DETAILS
  - *NTHL1* Targeted Sequencing
  - *NTHL1* p.Gln90* Genotyping
  - Whole-Exome Sequencing and Bioinformatic Analysis
  - Molecular and Clinical Analysis of Novel Families
  - Collection of Clinical and Pathological Data
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Calculation of Cancer Risks
- DATA AND SOFTWARE AVAILABILITY
  - Data Resources

SUPPLEMENTAL INFORMATION

Supplemental Information includes two figures and six tables and can be found with this article online at https://doi.org/10.1016/j.ccell.2018.12.011.

ACKNOWLEDGMENTS

We thank the patients for their cooperation, Drs. Wendy van Zelst-Stams and Marleen Kets for providing and updating clinical information on the Dutch families, Dr. Alexander Hoischen for assisting in the MIP design, the Genome Technology Platform for MIP sequencing support and Dr. Christian Gilissen for the use of the variant calling and annotation pipeline in the Radboudumc, Nijmegen. We acknowledge the networking support of the COST Actions BM1206 and CA17118. This work was supported by grants from the Dutch Cancer Society (KUN2015-7740), and the Sacha Swarttouw-Hijmans Foundation. R.M.d.V. holds a Fellowship from the Dutch Cancer Society (KWF; KUN2014-6666). A.D. and M.S. hold a grant from the International Centre for Genetic Engineering and Biotechnology (ICGEB), Trieste, Italy, on molecular aspects of inherited colorectal carcinoma in Macedonia. C.R.P. holds a grant from Fondo de Investigacion Sanitaria/Plan Estatal de I + D + I 2013–2016/FEDER (PI14/00230). D.D.B. holds a Senior Research Fellowship from the University of Melbourne Research at Melbourne Accelerator Program (R@MAP) and a Fellowship from NHMRC R.D. Wright Career Development. I.P. and D.S. are supported by the Schroeder-Kurth Fund. I.G.C. and P.A.J. are supported by grants from the National Breast Cancer Foundation of Australia and Cancer Australia. L.V. and G.C. hold grants from the Spanish Ministry of Economy and Competitiveness, co-funded by FEDER funds – a way to build Europe (SAF2016-80888-R and SAF2015-68016-R). H.K.S. is supported by the Deutsche Krebshilfe.

AUTHOR CONTRIBUTIONS

Conceptualization, R.M.d.V., N.H., and R.P.K.; Methodology, M.J.L.L., N.F.C.C.d.M., M.S., W.S., H.H.V., R.H.S., H.K.S., C.R.P., D.D., J.L., I.P.T., A.J.D., and T.v.W.; Formal Analysis, J.E.G., F.A.E., R.A.K., E.A.M.J., and J.R.V.; Investigation, J.E.G., R.M.d.V., F.A.E., M.N., R.D.A.W., C.P., M.J.L.L., S.W.t.B., E.J.K., M.E.V.B., I.P., A.L., I.S., R.H., P.A.J., N.L., H.L., D.C., O.S.B., M.C., K.S., G.C., M.C.J., K.N., A.G.v.K., F.J.H., H.M., B.R., W.D.F., I.P.T., L.V., D.D.B., S.K., J.A., I.G.C., S.A., D.S., T.v.W., N.H., and R.P.K.; Writing – Original Draft, J.E.G., R.M.d.V., R.D.A.W., N.H., and R.P.K.; Writing – Review & Editing, all authors.; Supervision, R.M.d.V., T.v.W., N.H., and R.P.K.; Funding Acquisition, M.J.L.L., N.H., and R.P.K.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 1, 2018
Revised: September 17, 2018
Accepted: December 26, 2018
Published: February 11, 2019

REFERENCES

Adam, R., Spier, I., Zhao, B., Kloth, M., Marquez, J., Hinrichsen, I., Kirfel, J., Tafazzoli, A., Horpaopan, S., Uhlhaas, S., et al. (2016). Exome sequencing identifies biallelic MSH3 germline mutations as a recessive subtype of colorectal adenomatous polyposis. Am. J. Hum. Genet. *99*, 337–351.

Al-Tassan, N., Chmiel, N.H., Maynard, J., Fleming, N., Livingston, A.L., Williams, G.T., Hodges, A.K., Davies, D.R., David, S.S., Sampson, J.R., and Cheadle, J.P. (2002). Inherited variants of MYH associated with somatic G:C–>T: a mutations in colorectal tumors. Nat. Genet. *30*, 227–232.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Borresen-Dale, A.L., et al. (2013). Signatures of mutational processes in human cancer. Nature *500*, 415–421.

Barrow, E., Robinson, L., Alduaij, W., Shenton, A., Clancy, T., Lalloo, F., Hill, J., and Evans, D.G. (2009). Cumulative lifetime incidence of extracolonic cancers in Lynch syndrome: a report of 121 families with proven mutations. Clin. Genet. *75*, 141–149.

Belhadj, S., Mur, P., Navarro, M., Gonzalez, S., Moreno, V., Capella, G., and Valle, L. (2017). Delineating the phenotypic spectrum of the NTHL1-associated polyposis. Clin. Gastroenterol. Hepatol. *15*, 461–462.

Bellido, F., Pineda, M., Aiza, G., Valdes-Mas, R., Navarro, M., Puente, D.A., Pons, T., Gonzalez, S., Iglesias, S., Darder, E., et al. (2016). 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. *18*, 325–332.

Boyle, E.A., O'Roak, B.J., Martin, B.K., Kumar, A., and Shendure, J. (2014). MIPgen: optimized modeling and design of molecular inversion probes for targeted resequencing. Bioinformatics *30*, 2670–2672.

Briggs, S., and Tomlinson, I. (2013). Germline and somatic polymerase epsilon and delta mutations define a new class of hypermutated colorectal and endometrial cancers. J. Pathol. *230*, 148–153.

Chubb, D., Broderick, P., Dobbins, S.E., Frampton, M., Kinnersley, B., Penegar, S., Price, A., Ma, Y.P., Sherborne, A.L., Palles, C., et al. (2016). Rare disruptive mutations and their contribution to the heritable risk of colorectal cancer. Nat. Commun. *7*, 11883.

Comprehensive Cancer Center the Netherlands: Dutch cancer incidence. (2018). Available at: http://www.cijfersoverkanker.nl.

COSMIC. Signatures probabilities. (2018). Available at: http://cancer.sanger.ac.uk/cancergenome/assets/signatures_probabilities.txt.

Drost, J., van Boxtel, R., Blokzijl, F., Mizutani, T., Sasaki, N., Sasselli, V., de Ligt, J., Behjati, S., Grolleman, J.E., van Wezel, T., et al. (2017). Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. Science *358*, 234–238.

Fostira, F., Kontopodis, E., Apostolou, P., Fragkaki, M., Androulakis, N., Yannoukakos, D., Konstantopoulou, I., and Saloustros, E. (2018). Extending the clinical phenotype associated with biallelic NTHL1 germline mutations. Clin. Genet. *94*, 588–589.

Gaujoux, R., and Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. BMC Bioinformatics *11*, 367.

Kempers, M.J., Kuiper, R.P., Ockeloen, C.W., Chappuis, P.O., Hutter, P., Rahner, N., Schackert, H.K., Steinke, V., Holinski-Feder, E., Morak, M., et al. (2011). Risk of colorectal and endometrial cancers in EPCAM deletion-positive Lynch syndrome: a cohort study. Lancet Oncol. *12*, 49–55.

Lange, K., Weeks, D., and Boehnke, M. (1988). Programs for pedigree analysis: MENDEL, FISHER, and dGENE. Genet. Epidemiol. *5*, 471–472.

Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. PLoS Comput. Biol. *9*, e1003118.

Li, N., Thompson, E.R., Rowley, S.M., McInerny, S., Devereux, L., Goode, D., Investigators, L., Wong-Brown, M.W., Scott, R.J., Trainer, A.H., et al. (2016). Reevaluation of RINT1 as a breast cancer predisposition gene. Breast Cancer Res. Treat. *159*, 385–392.

de Ligt, J., Willemsen, M.H., van Bon, B.W., Kleefstra, T., Yntema, H.G., Kroes, T., Vulto-van Silfhout, A.T., Koolen, D.A., et al. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. N. Engl. J. Med. *367*, 1921–1929.

National Comprehensive Cancer Network (2018). NCCN Clinical Practice Guidelines in Oncology, Genetic/Familial High-Risk Assessment: Colorectal. Available at: https://www.nccn.org/professionals/physician_gls/pdf/genetics_colon.pdf.

Nguyen-Dumont, T., Pope, B.J., Hammet, F., Southey, M.C., and Park, D.J. (2013a). A high-plex PCR approach for massively parallel sequencing. Biotechniques *55*, 69–74.

Nguyen-Dumont, T., Teo, Z.L., Pope, B.J., Hammet, F., Mahmoodi, M., Tsimiklis, H., Sabbaghian, N., Tischkowitz, M., Foulkes, W.D., Kathleen Cuningham Foundation Consortium for research into Familial Breast (kConFab), et al. (2013b). Hi-Plex for high-throughput mutation screening: application to the breast cancer susceptibility gene PALB2. BMC Med. Genomics *6*, 48.

Nielsen, M., Lynch, H., Infante, E., and Brand, R. (1993). MUTYH-associated polyposis. In GeneReviews(R), R.A. Pagon, M.P. Adam, H.H. Ardinger, S.E. Wallace, A. Amemiya, L.J.H. Bean, T.D. Bird, N. Ledbetter, H.C. Mefford, R.J.H. Smith, and K. Stephens, eds. (University of Washington, Seattle).

Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., Martincorena, I., Alexandrov, L.B., Martin, S., Wedge, D.C., et al. (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature *534*, 47–54.

O'Roak, B.J., Vives, L., Fu, W., Egertson, J.D., Stanaway, I.B., Phelps, I.G., Carvill, G., Kumar, A., Lee, C., Ankenman, K., et al. (2012). Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science *338*, 1619–1622.

Palles, C., Cazier, J.B., Howarth, K.M., Domingo, E., Jones, A.M., Broderick, P., Kemp, Z., Spain, S.L., Guarino, E., Salguero, I., et al. (2013). Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. Nat. Genet. *45*, 136–144.

Pilati, C., Shinde, J., Alexandrov, L.B., Assie, G., Andre, T., Helias-Rodzewicz, Z., Ducoudray, R., Le Corre, D., Zucman-Rossi, J., Emile, J.F., et al. (2017). Mutational signature analysis identifies MUTYH deficiency in colorectal cancers and adrenocortical carcinomas. J. Pathol. *242*, 10–15.

R Core Team. (2016). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing). https://www.R-project.org/.

Rivera, B., Castellsague, E., Bah, I., van Kempen, L.C., and Foulkes, W.D. (2015). Biallelic NTHL1 mutations in a woman with multiple primary tumors. N. Engl. J. Med. *373*, 1985–1986.

Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B.S., and Swanton, C. (2016). DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. *17*, 31.

Viel, A., Bruselles, A., Meccia, E., Fornasarig, M., Quaia, M., Canzonieri, V., Policicchio, E., Urso, E.D., Agostini, M., Genuardi, M., et al. (2017). A specific mutational signature associated with DNA 8-oxoguanine persistence in MUTYH-defective colorectal cancer. EBioMedicine *20*, 39–49.

de Voer, R.M., Hahn, M.M., Weren, R.D., Mensenkamp, A.R., Gilissen, C., van Zelst-Stams, W.A., Spruijt, L., Kets, C.M., et al. (2016). Identification of novel candidate genes for early-onset colorectal cancer susceptibility. PLoS Genet. *12*, e1005880.

Vogt, S., Jones, N., Christian, D., Engel, C., Nielsen, M., Kaufmann, A., Steinke, V., Vasen, H.F., Propping, P., Sampson, J.R., et al. (2009). Expanded extracolonic tumor spectrum in MUTYH-associated polyposis. Gastroenterology *137*, 1976–1985.e1–10.

Watson, P., Vasen, H.F., Mecklin, J.P., Bernstein, I., Aarnio, M., Jarvinen, H.J., Myrhoj, T., Sunde, L., Wijnen, J.T., and Lynch, H.T. (2008). The risk of extracolonic, extra-endometrial cancer in the Lynch syndrome. Int. J. Cancer *123*, 444–449.

Weren, R.D., Ligtenberg, M.J., Kets, C.M., de Voer, R.M., Verwiel, E.T., Spruijt, L., van Zelst-Stams, W.A., Jongmans, M.C., et al. (2015). A germline homozygous mutation in the base-excision repair gene NTHL1 causes adenomatous polyposis and colorectal cancer. Nat. Genet. *47*, 668–671.

Weren, R.D., Ligtenberg, M.J., Geurts van Kessel, A., De Voer, R.M., Hoogerbrugge, N., and Kuiper, R.P. (2018). NTHL1 and MUTYH polyposis syndromes: two sides of the same coin? J. Pathol. *244*, 135–142.

Win, A.K., Cleary, S.P., Dowty, J.G., Baron, J.A., Young, J.P., Buchanan, D.D., Southey, M.C., Burnett, T., Parfrey, P.S., Green, R.C., et al. (2011). Cancer risks for monoallelic MUTYH mutation carriers with a family history of colorectal cancer. Int. J. Cancer *129*, 2256–2262.

Win, A.K., Dowty, J.G., Cleary, S.P., Kim, H., Buchanan, D.D., Young, J.P., Clendenning, M., Rosty, C., MacInnis, R.J., Giles, G.G., et al. (2014). Risk of colorectal cancer for carriers of mutations in MUTYH, with and without a family history of cancer. Gastroenterology *146*, 1208–1211.e1–5.

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/
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 (2018): Dutch cancer incidence	The Netherlands Cancer Registry	http://www.cijfersoverkanker.nl
Oligonucleotides		
KASPar assay: NTHL1_p.Gln90*_A1: 5′–AAGGTGAC CAAGTTCATGCTGTGCCAGTCTGGGAGCCCT–3′)	This paper	N/A
KASPar assay: NTHL1_p.Gln90*_A2: 5′– GAAGGTC GGAGTCAACGGATTGCCAGTCTGGGAGCCCC–3′	This paper	N/A
KASPar assay: common reverse primer: 5′– ACCAG CTGTTGCTGCCAGTCCT-3′	This paper	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software Algorithms		
De novo signature analysis: Non negative matrix	Gaujoux and Seoighe, 2010	https://doi.org/10.1186/1471-2105-11-367
Signature reconstruction: R package DeconstructSigs	Rosenthal et al., 2016	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 Core Team, 2016	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 $\leq$40) and/or familial CRC (CRC $\leq$50 + first degree relative with CRC $\leq$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 (O'Roak et al., 2012). Twenty-three Molecular Inversion Probes were designed according to a previously published methodology (Boyle et al., 2014; O'Roak et al., 2012) 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 Netherlands (Weren et al., 2015), 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 (de Voer et al., 2016). 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 (de Ligt et al., 2012). Subsequently, variant calls observed in our in-house database of germline variants (de Voer et al., 2016), 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 (Lawrence et al., 2013) was performed using nonnegative matrix factorization (Gaujoux and Seoighe, 2010). To infer the contribution of the 30 previously identified mutational signatures available at the Catalogue of Somatic Mutations in Cancer (COSMIC, 2018), we used the R package DeconstructSigs tool (Rosenthal et al., 2016). 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 (Nguyen-Dumont et al., 2013a, 2013b). 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 (Li et al., 2016). To remove likely false positives, called variants were only retained if they had quality score >60 and an overall read depth $\geq$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*, *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 Figures 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 (Weren et al., 2018). 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 (Lange et al., 1988), 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 for the raw whole-exome sequencing data reported in this paper is: EGAD00001004534.