Genetic Cancer Susceptibility in Adolescents and Adults 25 Years or Younger With Colorectal Cancer

Colorectal cancer (CRC) diagnosed at age 25 years or younger is rare[1] and suggests a genetic predisposition. However, a high percentage of patients with CRC at age 25 years or younger present without polyposis or microsatellite instability, making a pathogenic variant in *APC* or mismatch repair (MMR) genes unlikely.[1] Other syndromes in which CRC occasionally develops at a young age, such as Li-Fraumeni syndrome, are not always investigated, likely due to their low prevalence. Most case series describing CRC at a young age report low positive familial history rates for CRC, possibly explained by recessively inherited predispositions or by reduced reproduction rates.[2] Here, we performed germline whole-exome sequencing (WES) in individuals with CRC at age 25 years or younger to explore the occurrence of inherited and/or de novo predisposing variants.

We performed WES on genomic DNA from 23 patients and 14 patient-parent trios in which no germline pathogenic variants in MMR genes or *APC* were previously identified. All patients were diagnosed with CRC between 12 and 25 years of age (mean age, 20 years; Supplementary Table 1). This study was approved by the local research ethics committee (CMO Arnhem-Nijmegen; 2014-1183) and the Institutional Review Board of the General Hospital of Beijing Military Region. Patients and/or parents provided written informed consent. Details on sequencing, sequence read mapping, variant calling and annotation, and quality filtering are provided in the Supplementary Material.

To identify (likely) pathogenic variants, we collected all: (1) variants in known hereditary cancer-associated genes (HCGs); (2) hemizygous, homozygous, and potential compound heterozygous variants that were infrequent (<0.15%) in control populations; and (3) de novo variants from patient-parent trios. Detailed variant selection, prioritization, and interpretation are described in the Supplementary Material.

We identified (likely) pathogenic variants in HCGs in 6 patients with CRC (16%; Table 1). Two patients carried a *TP53* variant and 1 other patient had a homozygous pathogenic splice site variant in *BLM*. In 3 other patients we identified heterozygous pathogenic variants in genes not known to be associated with CRC predisposition: pathogenic *BRCA2* and *PALB2* frameshift variants in a 13-year old and 24 year-old male, respectively, and a likely pathogenic missense variant in *NF1* in a patient with a Neurofibromatosis type 1 phenotype. In addition, we identified 2 heterozygous pathogenic variants in recessive HCGs, *MUTYH* and *MSH3.* Furthermore, 1 patient carried a *POLD1* frameshift variant in the POLD1-exonuclease domain. WES on tumor tissue of this patient revealed an in trans somatic *POLD1* (p.(Glu318Gly)) mutation, affecting an active site of the exonuclease domain. The tumor was ultra-hypermutated (379 Mut/Mb) and showed the *POLD1*-associated mutational signature SBS10d[3] (Supplementary Figure 1*A–C*).

In search for novel CRC susceptibility genes, the recessive disease scenario revealed 3 male patients with CRC with a variant in a gene associated with an X-linked disorder and 17 genes in 13 patients that harbored 2 likely pathogenic variants. For none of these genes a link with CRC could be made. However, the epidermolysis bullosa phenotype of AYA018 could be explained by variants in *COL7A1* (Table 1; Supplementary Table 1).

Potential pathogenic de novo variants were identified in 7 patients with CRC (Table 1). Two de novo variants were protein truncating, a nonsense variant in *ROCK1* encoding a protein serine/threonine kinase and a frameshift variant in *SMG6* encoding the telomere maintenance protein EST1A. Both genes are reported to be loss-of-function intolerant and have not been associated with disease. Two de novo missense variants were identified in the missense intolerant genes *PRKCB* and *SOS2,* encoding proteins directly interacting with Ras (Supplementary Figure 1*D*). *SOS2* has been associated with Noonan syndrome, in which CRC is not reported.[4] Features of Noonan syndrome were absent in this patient, but in vitro analyses of SOS2 p.(Gln970Glu) revealed increased extracellular signal-regulated kinase (ERK) phosphorylation after epidermal growth factor (EGF) stimulation (Supplementary Figure 1*E–F*). We identified 3 additional de novo missense variants that are predicted to be likely pathogenic, but do not affect genes that have been associated with oncogenic processes (Table 1).

By performing WES on germline DNA of patients diagnosed with CRC at age 25 years or younger, without pathogenic variants in MMR genes or *APC,* we identified pathogenic variants in other HCGs at a high percentage (19%; n = 7). Furthermore, this is the first study to describe that patients with CRC diagnosed at age 25 years or younger carry potentially pathogenic de novo variants in genes previously not associated with cancer that are well-conserved in pathways known to be involved in cancer development. Moreover, most genes in which we identified de novo variants are suggested to be intolerant to loss-of-function variants. To confirm that these genes play a role in early CRC development, these findings need to be confirmed in other patients with CRC.

Most current article

0016-5085/$36.00
https://doi.org/10.1053/j.gastro.2021.11.009

Table 1. Overview of All (Likely) Pathogenic Variants Identified Using Whole-Exome Sequencing in Individuals With Unexplained Colorectal Cancer at Adolescent or Young Adult Age

Gene	Patient	Sex	Age of onset (*y*)	Tumor information	NM number	cDNA level change	Protein level change	Genotype	PhyloP	CADD	LOF intolerant[a]	Pathway	Disease(s) associated with this gene (inheritance pattern)
Hereditary cancer gene variant													
BLM	AYA041	F	22	Colon N/A, MSS	NM_000057.3	c.3558+1G>T	Exon 18 skip	Hom.	5.13	33.00	Z = 0.89; pLI = 0	Double-strand break repair	Bloom syndrome (AR)
TP53	AYA029	F	24	Rectum, MSS	NM_000546.4	c.872del	p.(Lys291Argfs*54)	Het.	NA	18.10	Z = 0.98; pLI = 0.53	DNA repair response	Li-Fraumeni syndrome (AD)
TP53	AYA010	F	15	Rectum, MSI	NM_000546.4	c.845G>A	p.(Arg282Gln)	Het.	5.70	28.80	Z = 0.98; pLI = 0.53	DNA repair response	Li-Fraumeni syndrome (AD)
BRCA2	AYA003[b]	M	13	Rectum, MSS	NM_000059.3	c.2808_2811 del	p.(Ala938Profs*21)	Het.	3.51	40.00	Z = -1.29; pLI = 0	Double-strand break repair/ homologous recombination	Hereditary breast and ovarian cancer (AD); Fanconi anemia (AR)
PALB2	AYA015[c]	M	24	Rectum, MSS	NM_024675.3	c.886insA	p.(Met296*)	Het.	0.18	11.07	Z = 0.28; pLI = 0	Homologous recombination	Breast cancer (AD); prostate cancer (AD); Fanconi anemia (AR)
NF1	AYA014[d]	M	22	Colon N/A, MSS	NM_001042492.2	c.7930G>C	p.(Ala2644Pro)	Het.	9.05	28.50	Z = 6.54; pLI = 0.9	RAS/MAPK signaling	Neurofibromatosis type 1 (AD)
MUTYH	AYA006	F	20	Rectum, MSS	NM_001128425.1	c.1187G>A	p.(Gly396Asp)	Het.	7.16	23.80	Z = -0.21; pLI =0	Oxidative DNA damage repair	MUTYH-associated polyposis (AR)
MSH3	AYA025	F	23	Hepatic flexure, MSS	NM_002439.4	c.593_594del	p.(Phe198*)	Het.	NA	0.31	Z = -0.8; pLI = 0	DNA mismatch repair	Familial adenomatous polyposis 4 (AR)
POLD1	AYA032	M	21	Rectum, MSS	NM_001308632.1	c.927del	p.(Leu310Cysfs*83)	Het.	NA	18.14	Z = 2.46; pLI = 0	Polymerase proofreading	Polymerase-proof-reading associated polyposis (AD)
X-linked variants													
NOX1	AYA012	M	23	Colon N/A, MSS	NM_007052.4	c.120C>G	p.Tyr40*	Hemi.	2.50	12.98	Z = -0.05; pLI = 0	Oxidoreductase activity	Inflammatory bowel disease
COL4A5	AYA020	M	16	Rectum, MSS	NM_033380.2	c.4414C>T	p.Arg1472Cys	Hemi.	4.81	25.00	Z = 2.5; pLI = 1	PI3K-Akt signaling; Focal adhesion	Alport syndrome 1 (XLD)
CASK	AYA038	M	21	Rectum, MSS	NM_001126054.2	c.1880G>A	p.Arg627Gln	Hemi.	4.58	24.40	Z = 4.25; pLI = 1	Calcium/ calmodulin dependent serine protein kinase	Intellectual disability (XLD)

Table 1. Continued

Gene	Patient	Sex	Age of onset (*y*)	Tumor information	NM number	cDNA level change	Protein level change	Genotype	PhyloP	CADD	LOF intolerant[a]	Pathway	Disease(s) associated with this gene (inheritance pattern)
Recessive variants													
FN1	AYA022	M	17	Rectum, MMR unknown	NM_212482.4	c.4096T>C c.3307A>C	p.Phe1366Leu p.Ile1103Leu	Het. Het.	5.67 6.77	20.60 25.00	Z = 1.65; pLI = 0	PI3K-Akt signaling; focal adhesion	Glomerulopathy with fibronectin (AD)
KCNH6	AYA035	M	22	Rectum, MSS	NM_030779.3	c.2144G>T c.297C>A	p.Arg715Leu p.Tyr99*	Het. Het.	7.97 4.02	23.50 21.00	Z = -0.21; pLI = 0	Potassium channel	
SUGCT	AYA032[e]	M	21	Rectum, MSS	NM_001193311.1	c.286del c.499A>C	p.Val96Leufs*28 p.Ile167Leu	Het. Het.	NA 7.59	14.65 20.90	Z = 0.54; pLI = 0	Succinyl-CoA:glutarate-CoA transferase	Glutaric aciduria III (AR)
MYH8	AYA026	F	25	Transversum, MSS	NM_002472.2	c.3874C>T c.2117G>A	p.Arg1292* p.Arg706His	Het. Het.	0.73 7.53	41.00 28.20	Z = 0.35; pLI = 0	Myosin heavy chain 8	Trismus-pseudocamptodactyly syndrome (AD)
USP34	AYA025[e]	F	23	Hepatic flexure, MSS	NM_014709.3	c.4603G>T c.5851G>A	p.Asp1535Tyr p.Glu1951Lys	Het. Het.	7.82 7.80	25.30 35.00	Z = 2.34; pLI = 1	Ubiquitin modification removal	
SERPINA6	AYA027	M	23	Sigmoid, MSS	NM_001756.3	c.218C>T	p.Ala73Val	Hom.	7.70	34.00	Z = -0.66; pLI = 0	Steroid hormone biosynthesis	Corticosteroid-binding globulin deficiency (AD, AR)
UBAP2	AYA038	M	21	Rectum, MSS	NM_018449.3	c.1322G>A c.2416C>T	p.Arg441Gln p.His806Tyr	Het. Het.	4.98 9.36	29.70 22.20	Z = -0.11; pLI = 0	Likely involved in the ubiquitination pathway	
DUOX2	AYA038	M	21	Rectum, MSS	NM_001363711.2	c.2048G>T c.4027C>T	p.Arg683Leu p.Leu1343Phe	Het. Het.	5.32 6.75	24.60 35.00	Z = -1.17; pLI = 0	Thyroid hormone synthesis	Thyroid dyshormonogenesis 6 (AR)
ALDH1L2	AYA024	M	24	Sigmoid, MSS	NM_001034173.3	c.2270T>C c.1729C>G	p.Ile757Thr p.Leu577Val	Het. Het.	9.25 5.65	23.50 20.50	Z = 1.37; pLI = 1	One carbon pool by folate; aldehyde dehydrogenase 1 family member L2	
NUP93	AYA024	M	24	Sigmoid, MSS	NM_014669.4	c.1954C>T c.1457G>A	p.Gln652* p.Arg486His	Het. Het.	7.43 9.87	37.00 36.00	Z = 1.3; pLI = 0	RNA transport	Nephrotic syndrome, type 12 (AR)
SDK2	AYA001	F	12	Rectum, MSS	NM_001144952.1	c.2200G>A c.1163A>G	p.Gly734Arg p.Glu388Gly	Het. Het.	7.79 7.41	27.90 23.90	Z = 2.68; pLI = 0	Cell adhesion	
UNC13B	AYA007	F	24	Rectum, MMR unknown	NM_006377.3	c.4754G>A	p.Arg1585His	Hom.	9.76	34.00	Z = 1.45; pLI = 0	Synaptic vesicle cycle	
FHIP1A	AYA009	M	25	Colon N/A, MSI	NM_001109977.3	c.202G>A c.2722_2725del	p.Val68Ile p.Leu908Ilefs*7	Het. Het.	9.23 7.73	23.60 25.30	Z = 2.3; pLI = 0	Unknown	

Table 1. Continued

Gene	Patient	Sex	Age of onset (*y*)	Tumor information	NM number	cDNA level change	Protein level change	Genotype	PhyloP	CADD	LOF intolerant[a]	Pathway	Disease(s) associated with this gene (inheritance pattern)
DAPK2	AYA013	F	23	Colon N/A, MSS	NM_014326.3	c.170G>A	p.Arg57Gln	Hom.	7.77	36.00	Z = 0.35; pLI = 0	Autophagy; positive regulator of programmed cell death	
SLC6A20	AYA018	M	23	Sigmoid, MSS	NM_020208.3	c.1102G>A	p.Val368Ile	Hom.	7.76	25.30	Z = 0.5; pLI = 0	Neurotransmitter transport	Hyperglycinuria (AD); Iminoglycinuria (AR)
COL7A1	AYA018	M	23	Sigmoid, MSS	NM_000094.4	c.7828C>T	p.Arg2610*	Hom.	1.94	44.00	Z = 1.59; pLI = 1	Epidermis development; collagen type VII alpha 1 chain	Epidermolysis bullosa (AD, AR)
MYRIP	AYA018	M	23	Sigmoid, MSS	NM_015460.3	c.729+1G>A	Exon 7 skip	Hom.	7.50	19.64	Z = 0.73; pLI = 0	Rab effector protein involved in melanosome transport	
De novo variants													
ROCK1	AYA022	M	17	Rectum, MMR unknown	NM_005406.2	c.745C>T	p.(Gln249*)	Het.	9.60	46.00	Z = 4.59; pLI = 1	Rho-associated protein kinase 1	
SMG6	AYA019	F	24	Ascendens, MSS	NM_017575.4	c.678dup	p.(Arg227Glufs*43)	Het.	0.79	0.01	Z = 2.36; pLI = 0.62	mRNA surveillance pathway/ telomere maintenance	
SOS2	AYA017	M	15	Ascendens, MSS	NM_006939.3	c.2908C>G	p.(Gln970Glu)	Het.	9.41	18.53	Z = 2.53; pLI = 1	RAS/MAPK signaling	Noonan syndrome (AD)
PRKCB	AYA011	M	24	Colon N/A, MSI	NM_002738.6	c.1244G>T	p.(Arg415Leu)	Het.	7.97	27.90	Z = 3.99; pLI = 1	RAS/MAPK signaling	
TTI2	AYA002	M	21	Hepatic flexure, MSS	NM_025115.3	c.520G>A	p.(Glu174Lys)	Het.	0.03	12.24	Z = -0.29; pLI = 1	Regulator of the DNA damage response	Intellectual disability (AR)
CLASP1	AYA002	M	21	Hepatic flexure, MSS	NM_015282.2	c.827G>A	p.(Arg276His)	Het.	6.61	33.00	Z = 3.61; pLI = 1	Regulation of microtubule dynamics	
FBN1	AYA003[e]	M	13	Rectum, MSS	NM_000138.4	c.511G>A	p.(Gly171Arg)	Het.	7.25	33.00	Z = 5.06; pLI = 1	Major constituent of extracellular matrix	Marfan syndrome (AD)

Table 1. Continued

Gene	Patient	Sex	Age of onset (y)	Tumor information	NM number	cDNA level change	Protein level change	Genotype	PhyloP	CADD	LOF intolerant[a]	Pathway	Disease(s) associated with this gene (inheritance pattern)
AP1G1	AYA014[e]	M	22	Colon N/A, MSS	NM_001030007.1	c.43C>T	p.(Arg15Trp)	Het.	7.54	31.00	Z = 2.98; pLI = 1	Transport of synthesized lysosomal enzymes	
LNPEP	AYA014[e]	M	22	Colon N/A, MSS	NM_005575.2	c.1085T>C	p.(Ile362Thr)	Het.	5.91	20.80	Z = 1.44; pLI = 1	Renin-angiotensin system	

AD, autosomal dominant; AR, autosomal recessive; CADD, combined annotation dependent depletion score; cDNA, complementary DNA; Colon N/A, location in colon not specified; F, female; Hemi., hemizygous; Het., heterozygous; Hom., homozygous; LOF, loss-of-function; M, male; MMR, mismatch repair; mRNA, messenger RNA; MSI, microsatellite instable; MSS, microsatellite stable; NA, not available; NM, mRNA record in the nucleotide database; pLI, probability of being loss-of-function intolerant; XLD, X-linked disease.

[a]LOF prediction is based on the synonymous and missense Z scores (see Supplementary Material), the higher a Z score, the more intolerant a transcript is to variation. The pLI scores concern protein-truncating variation, with pLI scores $\geq$ 0.9 being extremely intolerant.

[b]The *BRCA2* variant is inherited from an unaffected father. Two sisters of the father also have the *BRCA2* variant and developed breast cancer and ovarian cancer before the age of 50, respectively. In addition, this patient also carried variants in *PARP1* that were not prioritized in this study.[8]

[c]The *PALB2* variant is inherited from an unaffected father. A sister and brother of the patient also have the *PALB2* variant and developed breast cancer at age 57 and a squamous cell carcinoma of the tongue at age 43, respectively.

[d]The *NF1* variant is inherited from the father, who is also affected by neurofibromatosis type 1.

[e]This patient also carries a (likely) pathogenic variant in a hereditary cancer gene.

Strikingly, in 1 patient with CRC we identified a germline frameshift variant in *POLD1* and an in trans somatic missense variant affecting the exonuclease domain. This indicates that loss-of-function *POLE* and *POLD1* variants may, next to missense variants in the exonuclease domain,[5] also contribute to an increased CRC susceptibility, an effect that was previously seen in an endometrial cancer.[6] This finding adds up to our previous observation of an over-representation of somatic *POLE* mutations in young patients with CRC,[1] which is relevant because the hypermutated tumors resulting from these variants might be sensitive for treatments with immune-checkpoint-inhibitors.

Two identified pathogenic variants affect breast cancer susceptibility genes that lack significant association with CRC susceptibility. However, a presumed enrichment for heterozygous germline *PALB2* and *BRCA2* pathogenic variants has been reported in pediatric and adolescent cancer.[7] It needs to be investigated whether such early-onset tumors are rare presentations of *BRCA2* and *PALB2* variants or the result of additional synergistic germline variants.

In conclusion, WES sequencing combined with HCG analyses may provide a genetic diagnosis in at least 19% of patients diagnosed with CRC at age 25 years or younger without pathogenic variants in the MMR genes or *APC*. Broad implementation of patient-parent sequencing in individuals with extremely young onset of an adult type of cancer and international sharing of data, for instance by gene-matching platforms, will help identify (de novo) pathogenic variants that occur repeatedly in these patients with CRC.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at https://doi.org/10.1053/j.gastro.2021.11.009.

MARJOLIJN C.J. JONGMANS
Department of Human Genetics
Radboud University Medical Center
Radboud Institute for Molecular Life Sciences
Nijmegen, the Netherlands *and*
Princess Máxima Center for Pediatric Oncology
Utrecht, the Netherlands *and*
Department of Genetics
University Medical Center Utrecht
Utrecht, the Netherlands

JUNXIAO ZHANG
Department of Human Genetics
Radboud University Medical Center
Radboud Institute for Molecular Life Sciences
Nijmegen, the Netherlands *and*
Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Diseases

Guangdong Institute of Gastroenterology
Sixth Affiliated Hospital
Sun Yat-Sen University
Guangzhou, Guangdong, China

AYA-CRC STUDY GROUP

ROLAND P. KUIPER
Department of Human Genetics
Radboud University Medical Center
Radboud Institute for Molecular Life Sciences
Nijmegen, the Netherlands *and*
Princess Máxima Center for Pediatric Oncology
Utrecht, the Netherlands

NICOLINE HOOGERBRUGGE
Department of Human Genetics
Radboud University Medical Center
Radboud Institute for Molecular Life Sciences
Nijmegen, the Netherlands

MARJOLIJN J.L. LIGTENBERG[*]
Department of Human Genetics
Radboud University Medical Center
Radboud Institute for Molecular Life Sciences
Nijmegen, the Netherlands *and*
Department of Pathology
Radboud University Medical Center
Radboud Institute for Molecular Life Sciences
Nijmegen, the Netherlands

RICHARDA M. DE VOER[*]
Department of Human Genetics
Radboud University Medical Center
Radboud Institute for Molecular Life Sciences
Nijmegen, the Netherlands

References

1. **de Voer RM, Diets IJ, van der Post RA**, et al. Clin Gastroenterol Hepatol 2021;9:1642–1651.e8.
2. Datta RV, et al. N Engl J Med 2000;342:137–138.
3. COSMIC. Mutational Signatures (v3.2). https://cancer.sanger.ac.uk/signatures/sbs/sbs10d/.
4. **Lissewski C, Chune V, Pantaleoni F**, et al. Eur J Hum Genet 2021;29:51–60.
5. **Palles C, Cazier JB**, et al. Nat Genet 2013; 45:136–144.
6. **Rosa RCA, Yurchenko AA**, et al. Genet Mol Biol 2020; 43:e20200100.
7. Sylvester DE, et al. J Med Genet 2018;55:785–792.
8. Rank L, et al. Nucleic Acids Res 2016; 44:10386–10405.

***These authors contributed equally to this study.**

Author names in bold designate shared co-first authorship.

Received October 5, 2021. Accepted November 4, 2021.

Correspondence
Address correspondence to: Richarda M. de Voer, PhD, Department of Human Genetics, Radboud University Medical Center, Geert Grooteplein Zuid 10, 6525GA Nijmegen, the Netherlands. e-mail: richarda.devoer@radboudumc.nl.

Acknowledgments
The authors thank all patients and their parents for participation in this study. In addition, we would like to thank Robbert D.A. Weren (Radboud University Medical Center, Nijmegen, the Netherlands) and Hans K. Schackert (Universitätsklinikum Carl Gustav Carus an der TU Dresden, Dresden, Germany) for expert assistance.

CRediT Authorship Contributions
Marjolijn Jongmans, MD, PhD (Conceptualization: Equal; Formal analysis: Equal; Investigation: Equal; Methodology: Equal; Writing – original draft: Equal; Writing – review & editing: Equal). Junxiao Zhang, MSc (Formal analysis: Equal; Investigation: Equal). Roland Kuiper, PhD (Conceptualization: Equal; Investigation: Equal; Methodology: Equal; Supervision: Supporting; Writing – review & editing: Equal). Nicoline Hoogerbrugge, MD, PhD (Conceptualization: Equal; Supervision: Supporting; Writing – review & editing: Equal). Marjolijn Ligtenberg, PhD (Conceptualization: Equal; Funding acquisition: Lead; Methodology: Equal; Project administration: Lead; Supervision: Supporting; Writing – review & editing: Equal). Richarda Maria de Voer, PhD (Conceptualization: Equal; Data curation: Lead; Formal analysis: Equal; Investigation: Equal; Methodology: Equal; Writing – original draft: Equal; Writing – review & editing: Equal).

AYA-CRC Study Group: Arjen R. Mensenkamp, Radboud University Medical Center; Ad Geurts van Kessel, Radboud University Medical Center; Liesbeth Spruijt, Radboud University Medical Center; C. Marleen Kets, Radboud University Medical Center; Wendy A.G. van Zelst-Stam, Radboud University Medical Center; Meyke I. Schouten, Radboud University Medical Center; Marijke R. Wevers, Radboud University Medical Center; Maran J.W. Olderode-Berends, University of Groningen, University Medical Center Groningen; Jan C. Oosterwijk, University of Groningen, University Medical Center Groningen; Marrit M. Hitzert, University of Groningen, University Medical Center Groningen; Tom G.W. Letteboer, University Medical Center Utrecht; Snežana Stanković, Radboud University Medical Center; Eveline J. Kamping, Radboud University Medical Center; Laurensia Yuniati, Radboud University Medical Center; Frank N. van Leeuwen, Princess Maxima Center for Pediatric Oncology; Jürgen Weitz, Technische Universität Dresden; Rachel S. van der Post, Radboud University Medical Center; Manuel R. Teixeira, Portuguese Oncology Institute of Porto; Huanliang Liu, Sun Yat-Sen University; and Jianping Wang, Sun Yat-Sen University.

Conflicts of interest
The authors disclose no conflicts.

Funding
Supported by the Dutch Digestive Foundation FP13-19 (M.J.L.L.). The funding source had no role in the study design, conduct, data analysis and interpretation, or manuscript writing.

Supplementary Methods

Recruitment of Patients

Patients (n = 37) and available parents were recruited through the Radboud University Medical Center, the University Medical Center Utrecht, the University Medical Center Groningen, the Carl Gustav Carus University Hospital, the Portuguese Oncology Institute of Porto, and the Sun Yat-Sen University. All patients were diagnosed with CRC ≤25 years of age (Supplementary Table 1). Patients, legal guardians, and/or parents provided written informed consent. This study was approved by the local research ethics committee (CMO Arnhem-Nijmegen; 2014-1183) and the Institutional Review Board of the General Hospital of Beijing Military Region.

WES

Genomic DNA was isolated from peripheral blood cells (hereafter referred to as germline). Exome capture was performed using Agilent SureSelect enrichment v4 (n = 5) and v5 (n = 32), and libraries were sequenced to a median coverage of 75× using an Illumina HiSeq platform with 101–base pair (bp) paired-end reads. Sequencing reads were aligned to the reference genome (GRCh37/hg19) using Burrows-Wheeler analysis (version 0.7.12-r1039). For the germline variant calling, single nucleotide variants and indels were called using the Genome Analysis Toolkit (GATK) HaplotypeCaller (version 3.4-46-gbc02625). Subsequently, all variants were annotated using an in-house annotation pipeline that includes in silico pathogenicity scores, population frequencies (in-house variant database, consisting of >22,000 germline exomes), ExAC, gnomAD, and germline variant databases (HGMD, ClinVar) as described previously.[1]

Germline Variant Analysis

Identification and selection of variants in known HCGs. The annotated germline exomes from each patient were first filtered for quality and the removal of common single-nucleotide polymorphisms by the following steps: (1) all variants covered by <10 reads were removed; (2) all variants with a GATK quality score of <200 were removed; and (3) all variants with a frequency >1% in our in-house variant database and >1% in ExAC and gnomAD-G were removed. All nonsynonymous variants in genes associated with hereditary cancer syndromes located in the coding sequence and/or variants that affected canonical splice sites (within intronic positions −2 till +2) were interpreted using American College of Medical Genetics and Genomics and the Association for Molecular Pathology.[2] Furthermore, for each prioritized variant we assessed how tolerant a gene is to a certain class of variation (eg, loss-of-function or missense) based on the constraint analyses provided by gnomAD.

Identification and selection of variants for recessive disease. Next to the aforementioned filtering for quality and removal of common single-nucleotide polymorphisms, all protein-truncating variants (nonsense or frameshift variants), canonical splice site variants, and in-frame indels and missense variants with a PhyloP of ≥3.0 and a combined annotation dependent depletion (CADD) score of ≥20[3] were prioritized. Furthermore, based on the severity and expected disease penetrance of our patient phenotypes, we consider variants that are present above a frequency of 0.15% in our in-house database (n > 22,000), ExAC, or gnomAD to be noncausative in a recessive state.[4] All cases that were explained by the HCG panel analysis were excluded for further analysis. To assess the recessive disease scenario, all genes with hemizygous, homozygous, or potential compound heterozygous variants were selected. All variants were interpreted using AlamutVisual mutation software (Interactive Biosoftware). Variants with a GATK quality score >200–500 were independently validated using Sanger sequencing if no parental exome was available to match the identified germline variant. All variants with a GATK quality score ≥500 were considered to be true positives as previously determined.[5] To interpret function, each mutated gene was annotated for their Gene Ontology, mouse knock-out phenotypes, and associated Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and a literature study was performed using PubMed. Furthermore, for each prioritized variant we assessed how tolerant a gene is to a certain class of variation (eg, loss-of-function or missense) based on the constraint analyses provided by gnomAD.

Identification and selection of de novo mutations. We collected all de novo mutations from the patient-parent trios (n = 14) as described previously with minor modifications.[5,6] Briefly, we selected all nonsynonymous de novo variants located in the coding sequence and/or those that affected canonical splice sites. Next, we applied several quality and prioritization steps, such that each variant: (1) was covered at least 10× in the patient and matching parental exomes; (2) showed a variant read percentage >30%; and (3) was called in a maximum of 2 samples in our in-house variant database (consisting of >22,000 exomes), ExAC, or gnomAD. The impact of each variant on protein level was assessed based on the majority of 3 in silico pathogenicity scores (Sorting Intolerant From Tolerant [SIFT],[7] PolyPhen-2,[8] and CADD[3]) and mutation severity by genomic conservation (PhyloP ≥3.0). All potential de novo mutations were subjected to Sanger sequencing in the subsequent patient-parent trio for validation purposes. In addition, to interpret function, each mutated gene was annotated for their Gene Ontology, mouse knock-out phenotypes, and associated KEGG pathway and a literature study was performed using PubMed. In addition, each gene was submitted to GeneMatcher to query if the gene was associated with an unsolved cancer phenotype identified in an independent study. Furthermore, for each prioritized variant we assessed how tolerant a gene is to a certain class of variation (eg, loss-of-function or missense) based on the constraint analyses provided by gnomAD.

WES of a POLD1 Mutated Tumor

Genomic DNA was isolated from fresh-frozen tissue slides. Exome capture was performed using Agilent SureSelect enrichment v5, and libraries were sequenced to a

median coverage of 259× using an Illumina HiSeq platform with 101-bp paired-end reads. Sequencing reads were aligned to the reference genome (GRCh37/hg19) using Burrows-Wheeler analysis (version 0.7.12-r1039). Somatic mutations (single nucleotide variants and indels) were called using GATK MuTect2 (version 4.0.6.0) with the paired germline sample. Subsequently, all mutations were annotated as described. Annotated somatic mutations were first filtered for quality and removal of germline variants that were not excluded by the MuTect2 algorithm. The following mutations were removed: (1) variants covered by <15 reads in both the tumor and the germline exome; (2) mutations supported by 1 or more reads in the germline; (3) mutations reported in our in-house database; (4) mutations reported >2 in ExAC; (5) mutations reported in the genomes of gnomAD; and (6) all mutations with a variant allele frequency <0.10. The total number of megabases (Mb) covered with at least 15× reads was calculated using bedtools. Total tumor mutational burden was calculated as the number of high-confident somatic variants per 1 Mb of coding region. Somatic mutational signature analyses, based on all 96 trinucleotide substitutions, were performed using the R package DeconstructSigs with the V3.2 of the mutational signatures as reported in Catalogue of Somatic Mutations in Cancer (COSMIC).[9,10]

Analysis of ERK/MAPK Pathway Activation by Mutated SOS2

Human *SOS2* complementary DNA was purchased from Open Biosystems (Huntsville, AL). The *SOS2* c.2908C>G substitution was introduced by site-directed mutagenesis following the manufacturer's protocol (Stratagene, LaJolla, CA). Both wild-type and mutated *SOS2* were cloned into plasmid DNA vector pcDNA3.1 (Invitrogen; Carlsbad, CA). HEK-293FT cells were maintained in DMEM supplemented with 10% fetal bovine serum (Greiner Bio-One) and 100 U/mL penicillin/streptomycin (Invitrogen). Empty plasmid DNA vector pcDNA3.1, as well as wild-type and mutant *SOS2* constructs, and myc-tagged ERK2 (kindly provided by Dr. Wiljan Hendriks) were transiently transfected into HEK-293FT cells. Seven hours after transfection, cells were split and cells were starved after 24 hours in DMEM + 0.1% fetal bovine serum overnight before being stimulated with 20 ng/mL epidermal growth factor (Peprotech, Rocky Hill, NJ). Cells were collected at 0, 5, 15, and 30 minutes after epidermal growth factor stimulation. Equal amounts of proteins were subjected to sodium dodecyl sulfate polyacrylamide gel electrophoresis and transferred to polyvinylidene difluoride membrane using standard protocols. Immunoblotting was performed with antibodies targeting phospho-ERK1/2 (Cell Signaling, Danvers, MA), MYC (hybridoma supernatant from clone 9E10), and SOS2 (Santa Cruz Biotechnology, Santa Cruz, CA), followed by horseradish peroxidase-conjugated anti-rabbit or anti-mouse antibodies (Dako, Glostrup, Denmark). Statistical analysis was done with 2-tailed Student *t* test.

Supplementary References

1. **De Voer RM, Hahn M-M**, et al. PLoS Genet 2016;12:e1005880.
2. Richards S, et al. Genet Med 2015;17:405–424.
3. Rentzsch P, et al. Nucleic Acids Res 2018;8(47):D886–D894.
4. Te Paske IBAW, et al. Int J Mol Sci 2020;21:8757.
5. **Lelieveld SH, Reijnders MRF**, et al. Nat Neurosci 2016;19:1194–1196.
6. **Gilissen C, Hehir-Kwa JY**, et al. Nature 2014;511:344–347.
7. Kumar P, et al. Nat Protoc 2009;4:1073–1081.
8. **Adzhubei IA, Schmidt S, Peshkin L**, et al. Nat Methods 2010;7:248–249.
9. Rosenthal R, et al. Genome Biol 2016;17:31.
10. Tate JG, et al. Nucleic Acids Res 2019;47:D941–D947.
11. **Campbell BB, Light L**, et al. Cell 2017;171:1042–1056.e10.
12. Mur P, et al. Genet Med 2020;22:2089–2100.

Author names in bold designate shared co-first authorship.

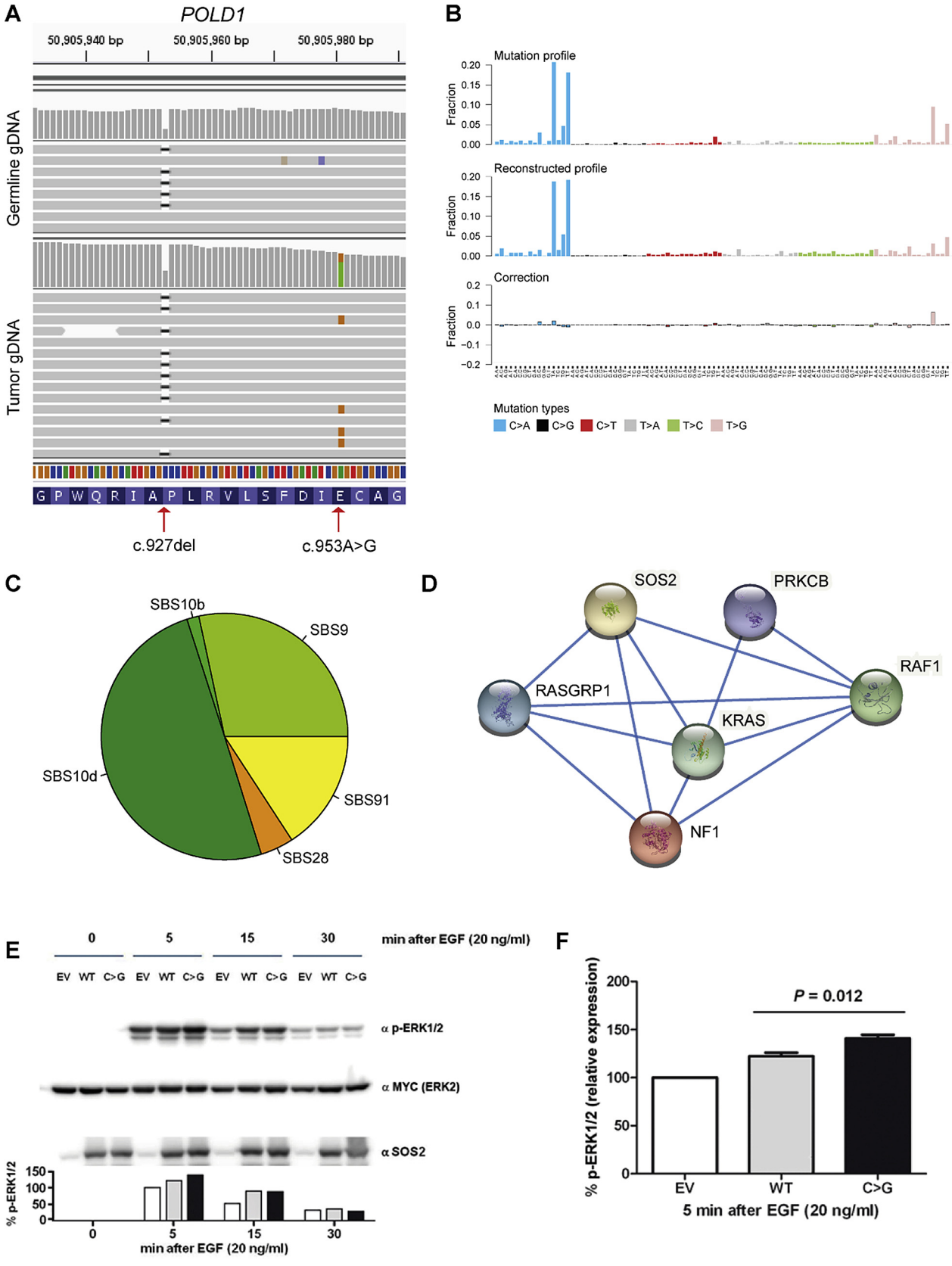
Supplementary Table 1. Clinical Features of Adolescents and Young Adults With CRC

Proband	Gender (M/F)	Age at diagnosis (*y*)	Location	Histology	MMR[a]	Other phenotypes	Trio sequencing
AYA001	F	12	Rectum	N/A	MSS (IHC)	-	Yes
AYA002	M	21	Hepatic flexure	Mucinous adenocarcinoma	N/A	-	Yes
AYA003	M	13	Rectum	Poorly differentiated adenocarcinoma	MSS (IHC)	-	Yes
AYA004	M	15	Colon ascendens	Signet ring cell carcinoma	MSS (IHC)	Pigmentation on abdomen and arm	Yes
AYA005	F	20	Sigmoid	Well-differentiated adenocarcinoma	MSS (IHC)	-	Yes
AYA006	F	20	Rectum	N/A	MSS (IHC)	-	No
AYA007	F	24	Rectum	N/A	N/A	-	No
AYA008	F	21	Colon N/A	N/A	N/A	-	No
AYA009	M	25	Colon N/A	N/A	MSI[b]	-	No
AYA010	F	15	Rectum	N/A	MSI[c]	-	No
AYA011	M	24	Colon N/A	N/A	MSI[d]	-	Yes
AYA012	M	23	Colon N/A	N/A	MSS (IHC)	-	No
AYA013	F	23	Colon N/A	N/A	MSS (IHC)	-	No
AYA014	M	22	Colon N/A	Moderately differentiated adenocarcinoma	MSS (IHC)	Neurofibromatosis (father from proband idem)	Yes
AYA015	M	24	Rectum	Moderately differentiated adenocarcinoma	MSS (IHC)	-	No
AYA016	F	16	Colon N/A	N/A	MSS (IHC)	-	Yes
AYA017	M	15	Colon ascendens	Cribriform carcinoma	N/A	Two primary colon cancers	Yes
AYA018	M	23	Sigmoid	Well-differentiated adenocarcinoma	MSS (IHC)	Epidermolysis bullosa	Yes
AYA019	F	24	Colon ascendens	Well-differentiated adenocarcinoma	MSS (IHC)	-	Yes
AYA020	M	16	Rectum	Signet ring cell carcinoma	MSS (IHC)	-	No
AYA021	M	14	Colon N/A	N/A	N/A	Juvenile polyps, hypogonadism	No
AYA022	M	17	Rectum	N/A	N/A	-	Yes
AYA024	M	24	Sigmoid	N/A	MSS (IHC)	-	No
AYA025	F	23	Hepatic flexure	Moderately differentiated adenocarcinoma	MSS (IHC)	-	No
AYA026	F	25	Colon transversum	Well-differentiated adenocarcinoma	MSS (IHC)	-	No
AYA027	M	23	Sigmoid	Adenocarcinoma	MSS (IHC)	-	No
AYA028	F	18	Colon ascendens	Poorly differentiated adenocarcinoma	N/A	-	Yes
AYA029	F	24	Rectum	Moderately differentiated adenocarcinoma	MSS (IHC)	-	No
AYA032	M	21	Rectum	N/A	MSS (IHC)	-	No
AYA034	M	13	Colon transversum	N/A	MSS (IHC)	-	No
AYA035	M	22	Rectum	N/A	MSS (IHC)	-	No
AYA036	F	25	Rectum	Moderately differentiated adenocarcinoma	N/A	Colon polyps	Yes

Supplementary Table 1. Continued

Proband	Gender (M/F)	Age at diagnosis (*y*)	Location	Histology	MMR[a]	Other phenotypes	Trio sequencing
AYA037	M	21	Colon ascendens	Poorly differentiated adenocarcinoma	MSS (IHC)	-	No
AYA038	M	21	Rectum	N/A	MSS (IHC)	Malignant mesenchymal tumors	No
AYA040	M	21	Rectum	N/A	MSS (IHC)	-	No
AYA041	F	22	Colon N/A	N/A	MSS (IHC)	Psoriasis, microcephaly, unilateral freckling	No
AYA042	F	25	Colon N/A	N/A	N/A	-	No

F, female; IHC, immunohistochemistry; M, male; MMR, mismatch repair; MSI, microsatellite unstable; MSS, microsatellite stable; N/A, not available.
[a]MMR deficiency was in most samples analyzed using IHC. Three samples are MSI, based on the analyses of a pentaplex MSI analyses.
[b]MMR deficiency was the result of somatic inactivation of *MLH1* (NM_000249.3:c.2041G>A (p.(Ala681Thr); 48%) and loss of heterozygosity; 30% neoplastic cells).
[c]The tumor was unavailable to perform somatic mutation analyses of the MMR genes.
[d]MMR deficiency was the result of somatic inactivation of *MLH1* (NM_000249.3: c.1896+1G>T (splice site loss; 28%) and c.1937A>C (p.(Tyr646Ser); 31%); 60% neoplastic cells).


A
POLD1
50,905,940 bp
50,905,960 bp
50,905,980 bp
Germline gDNA
Tumor gDNA
G P W Q R I A P L R V L S F D I E C A G
c.927del
c.953A>G
B
Mutation profile
Reconstructed profile
Correction
Fraction
Mutation types
C>A C>G C>T T>A T>C T>G
C
SBS10b
SBS9
SBS10d
SBS91
SBS28
D
SOS2
PRKCB
RAF1
RASGRP1
KRAS
NF1
E
0 5 15 30 min after EGF (20 ng/ml)
EV WT C>G EV WT C>G EV WT C>G EV WT C>G
α p-ERK1/2
α MYC (ERK2)
α SOS2
% p-ERK1/2
min after EGF (20 ng/ml)
F
P = 0.012
% p-ERK1/2 (relative expression)
EV
WT
C>G
5 min after EGF (20 ng/ml)

Supplementary Figure 1. Analyses of germline potential pathogenic variants in genes identified in adolescents and young adults with CRC. (*A–C*) Analyses of germline and somatic variants in *POLD1*. (*A*) Integrated genome viewer screenshot of coverage and variant detection germline and tumor reads from AYA032. The somatic c.953A>G variant is in trans with the germline c.927del variant. The tumor is ultra-hypermutated with 379 mut/Mb and is among the most highly mutated tumors with a *POLE* or *POLD1* described in literature.[11,12] Usually, only tumors with a pathogenic variant in the *POLE* exonuclease domain are ultra-hypermutated (>100 mut/Mb) and tumors with pathogenic variants in *POLD1* harbor on average 30–60 mut/Mb.[12] This indicates that the germline frameshift variant in *POLD1* in combination with the missense variant in the *POLD1* exonuclease domain has likely resulted in complete lack of exonuclease function by POLD1. (*B*) Mutation profile (*top*) of the high-confident somatic variants in the tumor of AYA032 (*POLD1* mutated). To refit known mutational signatures the mutation profile is reconstructed (*middle*) and minor corrections are made (*lower*). (*C*) Contribution of multiple mutational signatures to the reconstructed mutational profile of AYA032. SBS10d is the major contributor and is previously associated with POLD1 exonuclease deficiency. (*D*) Protein interaction network for de novo mutated protein SOS2 and PRKCB with KRAS (https://string-db.org/). (*E–F*) SOS2 p.Q970E shows enhanced ERK activation. Epidermal growth factor (EGF)–triggered ERK activation in HEK-293FT cells transfected with Empty Vector (EV), wild-type (WT), or mutant (C>G) SOS2 constructs, cotransfected with MYC-ERK2, measured at different time points after EGF stimulation. (*E*) Western blot analysis of 1 representative experiment (from 4 independent determinations). (*Bottom*) A quantification of relative ERK phosphorylation at various time points poststimulation, compared with EV-transfected cells (set as 100%). Exogenous expression of SOS2 and MYC-ERK2 are similar for WT and mutant transfected cells. Mutant SOS2 shows increased phosphorylation of ERK 5 minutes after EGF stimulation compared with the WT. (*F*) Quantification of phosphorylated ERK levels 5 minutes after addition of EGF, compared with EV-transfected cells (set to 100%), from 4 independent experiments. $P = .012$ for c.2908C>G compared with WT SOS2 (2-tailed Student *t* test). Error bars represent standard error of the mean