ORIGINAL ARTICLE

WILEY Molecular Genetics & Genomic Medicine
Open Access

Exome sequencing in 51 early onset non-familial CRC cases

Jessada Thutkawkorapin[1] [ID] | Annika Lindblom[2] | Emma Tham[2] [ID]

[1]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden

[2]Department of Molecular Medicine and Surgery, Karolinska Institutet and Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden

Correspondence
Emma Tham, Department of Molecular Medicine and Surgery, Karolinska Institutet, SE-171 76 Stockholm, Sweden.
Email: emma.tham@ki.se

Abstract

Background: Colorectal cancer (CRC) cases with an age of onset <40 years suggests a germline genetic cause. In total, 51 simplex cases were included to test the hypothesis of CRC as a mendelian trait caused by either heterozygous autosomal dominant or bi-allelic autosomal recessive pathogenic variants.

Methods: The cohort was whole exome sequenced (WES) at 100× coverage. Both a dominant- and recessive model were used for searching predisposing genetic factors. In addition, we assayed recessive variants of potential moderate risk that were enriched in our young-onset CRC cohort. Variants were filtered using a candidate cancer gene list or by selecting variants more likely to be pathogenic based on variant type (e.g., loss-of-function) or allele frequency.

Results: We identified one pathogenic variant in *PTEN* in a patient subsequently confirmed to have a hereditary hamartoma tumor syndrome (Cowden syndrome) and one patient with a pathogenic heterozygous variant in *PMS2* that was originally not identified by WES due to low quality reads resulting from pseudogenes. In addition, we identified three heterozygous candidate missense variants in known cancer susceptibility genes (*BMPR1A*, *BRIP1*, and *SRC*), three truncating variants in possibly novel cancer genes (*CLSPN*, *SEC24B*, *SSH2*) and four candidate missense variants in *ACACA*, *NR2C2*, *INPP4A,* and *DIDO1*. We also identify five possible autosomal recessive candidate genes: *ATP10B*, *PKHD1*, *UGGT2*, *MYH13*, *TFF3*.

Conclusion: Two clear pathogenic variants were identified in patients that had not been identified clinically. Thus, the chance of detecting a hereditary cancer syndrome in patients with CRC at young age but without family history is 2/51 (4%) and therefore the clinical benefit of genetic testing in this patient group is low. Of note, using stringent filtering, we have identified a total of ten candidate heterozygous variants and five possibly biallelic autosomal recessive candidate genes that warrant further study.

KEYWORDS
colorectal cancer, early onset, exome sequencing, non-familial

1 | INTRODUCTION

Colorectal cancer (CRC) is the third most common cancer diagnosed in Sweden ("Cancer fact sheet. World Health Organization", 2010). In Sweden, familial CRC represents approximately 13% of all CRC cases (Frank, Sundquist, Yu, Hemminki, & Hemminki, 2017), while the cases with age of onset <40 years account for 1.4% ("Statistics on Cancer Incidence in Sweden 2016," 2017). Known CRC susceptibility genes are responsible for 3.5% of all cases (Aaltonen, Johns, Jarvinen, Mecklin, & Houlston, 2007).

In a study with over 10,000 CRC cases, Lynch syndrome was responsible for 3.1% overall (Moreira et al., 2012) and in 11/140 (8%) of CRC before 50 years (Giraldez et al., 2010). The same study detected 4/140 (3%) with bi-allelic *MUTYH* mutations. Factors other than monogenic variants that contribute to cancer development are somatic genetic and epigenetic aberrations, combinations of low- and moderate-risk genetic variation, individual life style (Lichtenstein et al., 2000), chronic bowel inflammation (Beaugerie & Itzkowitz, 2015), and environmental factors (Bodmer & Bonilla, 2008). However, CRC development before the age of 40 years and without a family history of CRC is unusual, suggesting either a de novo monogenic factor in a high-risk autosomal dominant cancer gene, an autosomal dominant syndrome with decreased penetrance or a high-risk autosomal recessive gene, or complex disease.

Here, we used exome sequencing of 51 simplex CRC cases with age of onset <40 years in order to search for novel monogenic cancer genes that cause a rare autosomal dominant or autosomal recessive colorectal cancer syndrome. In addition, we looked for moderate recessive risk variants that were enriched in our young-onset CRC cohort.

2 | MATERIALS AND METHODS

2.1 | Ethical compliance

All patients gave written informed consent in accordance with Swedish legislation (2003:460) and the study was approved by the Regional Research Ethics Committee, Dnr: 2002-20489, 2008/125-2031/2, and 2014/1326-32.

2.2 | Early-onset CRC cohort

CRC patients with age of onset before 40 years and without a family history of CRC were recruited through the Department of Clinical Genetics, Karolinska University Hospital Solna (Sweden) ($n = 17$) or recruited in a nationwide study, the Swedish Low-risk Colorectal Cancer Study ($n = 34$). The latter consisted of more than 3,300 consecutive patients operated on for CRC in 14 hospitals in and around Stockholm and Uppsala between 2004 and 2009. All patients gave written informed consent to participate in the study and DNA was isolated from peripheral blood. The cancer diagnosis and histopathology were verified in all cases through medical records (Ghazi, 2010). For the present study, all simplex CRC cases diagnosed before 40 years were selected.

2.3 | Familial Breast cancer (BRC) cohort used as a comparison group in autosomal recessive gene analysis

The familial BRC cohort was defined as individuals from families where at least two first or second-degree relatives were affected with BRC. The families were recruited through the Department of Clinical Genetics, Karolinska University Hospital Solna (Sweden).

2.4 | Whole exome sequencing of early onset CRC and familial BRC samples

DNA was quantified using a Qubit Fluorometer (Life Technologies, USA). Sequencing libraries were prepared according to the TruSeq DNA Sample Preparation Kit EUC 15005180 or EUC 15026489 (Illumina, USA). Briefly, 1–1.5 µg of genomic DNA was fragmented using the Covaris 400 bp protocol (Covaris, Inc., USA). After fragmentation, all samples were subjected to end-repair, A-tailing, and adaptor ligation of Illumina Multiplexing PE adaptors. An additional gel-based size selection step was performed and the adapter-ligated fragments were subsequently enriched by PCR followed by purification using Agencourt AMPure Beads (Beckman Coulter, Sweden). Exome capture was performed by pre-pooling equimolar amounts and performing enrichment in 5- or 6-plex reactions according to the TruSeq Exome Enrichment Kit Protocol (EUC 15013230). Library size was checked on a Bioanalyzer High Sensitivity DNA chip (Agilent Technologies, Sweden) while concentration was calculated by quantitative PCR. The pooled DNA libraries were clustered on a cBot instrument (Illumina) using the TruSeq PE Cluster Kit v3. Paired-end sequencing was performed for 100 cycles using a HiSeq 2000 instrument (Illumina) with TruSeq SBS Chemistry v3, according to the manufacturer's protocol. Base calling was performed with RTA (1.12.4.2 or 1.13.48) and the resulting BCL files were filtered, de-multiplexed, and converted to FASTQ format using CASAVA 1.7 or 1.8 (Illumina). The sequencing was performed at an average coverage of 100x.

2.5 | Bioinformatics workflow

Sequencing reads were aligned to the reference genome GRCh37 using BWA (H. Li & Durbin, 2010). Aligned reads were sorted and PCR-duplicated reads were removed using Picard (http://broadinstitute.github.io/picard/). The calculation of mapping and enrichment statistics were done with Picard and GATK (McKenna et al., 2010). Variants were called using GATK by following the best practice procedure implemented at the Broad Institute (DePristo et al., 2011). Variant Quality Score Recalibration from GATK was used for quality control of the variants. Variant annotation was done by ANNOVAR (Wang, Li, & Hakonarson, 2010). The annotated information includes RefSeq gene annotation (O'Leary et al., 2016), dbSNP rs number (Sherry et al., 2001), COSMIC (Forbes et al., 2017), ClinVar (Landrum et al., 2018), SPIDEX (Xiong

et al., 2015), ExAC conservative constraint (Lek et al., 2016), UniProt (Chen, Huang, & Wu, 2017). Background allele frequencies are from SweGen (Ameur et al., 2017), ExAC (Lek et al., 2016), gnomAD (Lek et al., 2016), and 1000 Genomes Project allele frequencies (1000 Genomes Project Consortium et al., 2012). In silico predictors used for predicting pathogenic effects include SIFT (Kumar, Henikoff, & Ng, 2009), PolyPhen2 (Adzhubei et al., 2010), Phylop (Cooper et al., 2005), LRT (Chun & Fay, 2009), Mutation Taster (Schwarz, Rodelsperger, Schuelke, & Seelow, 2010), Mutation Assessor (Reva, Antipin, & Sander, 2011), FATHMM (Shihab et al., 2015), GERP++ (Davydov et al., 2010), and CADD (Kircher et al., 2014).

2.6 | Allele frequency database

In this study, we have used maximum minor allele frequency (MMAF) from 21 populations (SweGen (Ameur et al., 2017), 1000G (The 1000 Genomes Project Consortium, 2012), ExAC (Lek et al., 2016), gnomAD (Lek et al., 2016), 200Danes (Y. Li et al., 2010), and 249Swedes (http://neotek.scilifelab.se/hbvdb/)) to identify common variants.

2.7 | Analysis workflows

2.7.1 | Rare autosomal dominant and autosomal recessive analysis in cancer susceptibility gene list

To search for causative mutations in autosomal dominant- or autosomal recessive cancer susceptibility genes, we first used an in silico cancer gene list modified from (Vogelstein et al., 2013) in the analysis of 51 early-onset CRC samples. The gene list contains 244 known

cancer-related genes (known somatic cancer driver genes as well as hereditary cancer genes). The candidate variants were primarily selected by filtering splicing and non-silent variants with MMAF <1% in autosomal-recessive cancer genes and <0.1% in autosomal-dominant cancer genes. The variants with MMAF higher than the prevalence of cancer syndrome suggested by the gene were manually excluded. The resulting candidate variants were classified as of uncertain significance, likely pathogenic, or pathogenic using American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG-AMP) guidelines criteria (Richards et al., 2015) (Figure 1). The candidate variants were then confirmed by Sanger sequencing. Manual analysis of poor quality *PMS2* variants using the sequencing file (BAM) was performed.

2.7.2 | Frameshift-, nonsense-, and splice variants analysis

We selected all frameshift-, nonsense-, and splice variants in the exome which have MMAF < 0.1% (for heterozygous variants) and <1% (for recessive variants) (Figure 2). The candidate variants were then confirmed by Sanger sequencing.

2.7.3 | Analysis of missense variants

In this analysis, we selected all missense variants in the exomes of the 51 CRC patients with MMAF < 0.1% in all public databases. Then, we grouped the variants using CADD score (Kircher et al., 2014) ("more than 20", "more than 25", and "more than 30") and ExAC missense Z-score (Lek et al., 2016) ("less than or equal to 3" and "more than 3") (Figure 3).
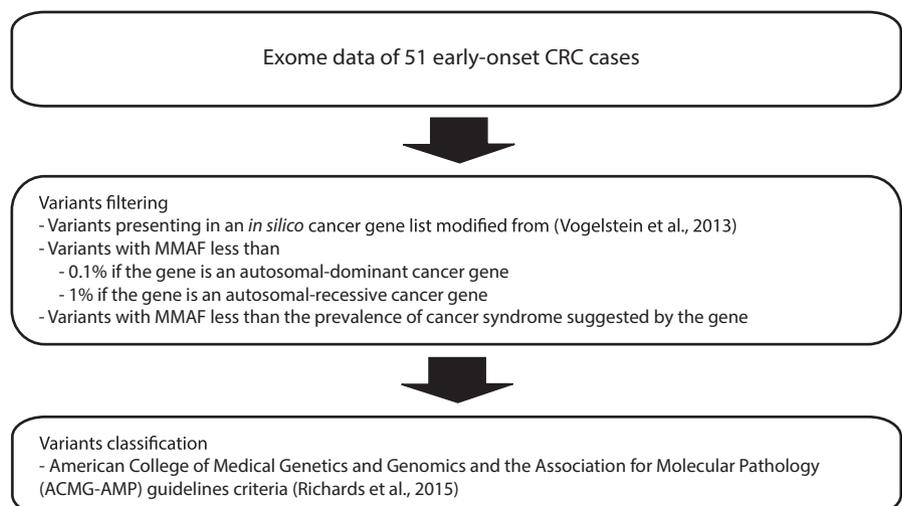

FIGURE 1 Autosomal dominant and autosomal recessive analysis in cancer susceptibility gene list

```
Exome data of 51 early-onset CRC cases
```

```
Variants filtering
- Frameshift variants, nonsense variants, or splicing variants
- Variants with MMAF less than
   - 0.1% for autosomal-dominant variants
   - 1% for autosomal-recessive variants
```
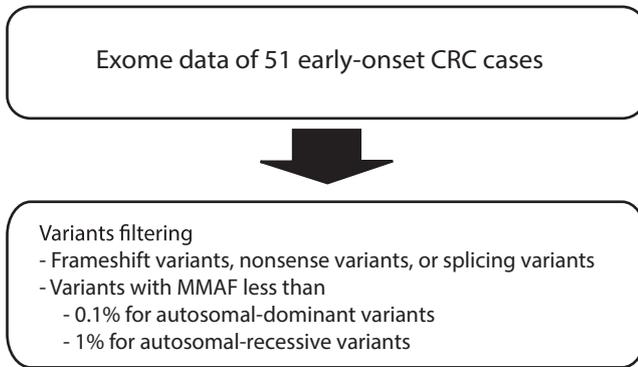
FIGURE 2 Frameshift-, nonsense-, and splice variants analysis

2.7.4 | Autosomal recessive genes analysis (rare monogenic and less common risk genes)

To pinpoint autosomal recessive genes predisposing to early-onset colorectal cancer, we hypothesized that the patients inherited biallelic risk variants, one from each parent in a homozygous or compound heterozygous state. A gene was assumed to be associated with an increased risk of CRC if the number of cases with the possible biallelic variants was more prevalent than in a control population. Due to an absence of individual genotyping data of a normal population, we used a cohort of familial breast cancer processed in the same way as the CRC samples as a comparison group, assuming that the possible biallelic variants predisposing to early-onset colorectal cancer are not associated with familial breast cancer. With such a hypothesis, we searched for possible biallelic variants in the exome in our 51 early-onset CRC cases with 56 familial BRC as a comparison group. During the first step of filtering, genes with splicing and non-silent variants, with a MMAF < 20% and predicted to be pathogenic by more than 4 out of 9 in silico effect predictors (SIFT (Kumar et al., 2009), PolyPhen2 HDIV (Adzhubei et al., 2010), PolyPhen2 HVAR (Adzhubei et al., 2010), Phylop (Cooper et al., 2005), LRT (Chun & Fay, 2009), Mutation Taster (Schwarz et al., 2010), Mutation
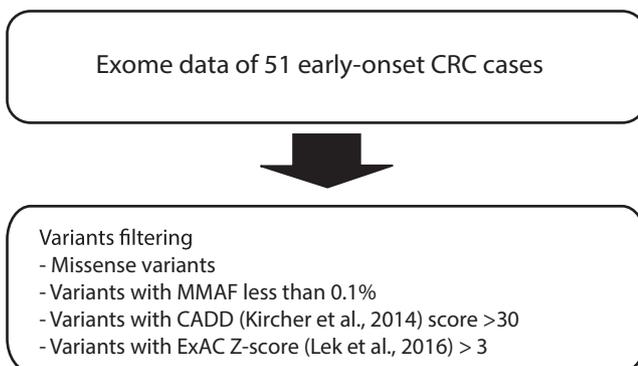


```
Exome data of 51 early-onset CRC cases
```

```
Variants filtering
- Missense variants
- Variants with MMAF less than 0.1%
- Variants with CADD (Kircher et al., 2014) score >30
- Variants with ExAC Z-score (Lek et al., 2016) > 3
```

FIGURE 3 Analysis of missense variants

Assessor (Reva et al., 2011), FATHMM (Shihab et al., 2015), GERP++ (Davydov et al., 2010)), were selected. The next filtering step was to select genes with possible biallelic variants (possible compound heterozygous or homozygous) in at least two CRC cases and where no possible compound or homozygous case was found among the familial breast cancer cohort. If any two variants always had a similar MAF among population allele frequency databases or the variants showed up together in multiple samples, the variants were considered likely to be on the same allele and were excluded (Figure 4). The variants were prioritized based on MMAF and the observed frequency in CRC compared to the statistical likelihood of occurring together (computed by multiplying the MMAF of the two alleles).

2.8 | Sanger sequencing

Primer pairs were designed using Primer3Plus (http://www.bioinformatics.nl/primer3plus) to amplify the candidate pathogenic variants suggested by exome sequencing data. PCR was performed in 25 µl reaction volumes containing AmpliTaq Gold (Life Technologies, Carlsbad, CA), 1.25 µg DNA and 10 µmol l^{-1} of primers. Excess single-stranded DNA, primers, and dNTPs were removed using ExoSAP. The sequences were visualized and analyzed using CodonCode Aligner (http://www.codoncode.com).

2.9 | Database submission of novel variants

Relevant variant information has been uploaded to the Leiden Open Variation Database (LOVD), http://www.lovd.nl/3.0

3 | RESULTS

We first searched for variants in known cancer susceptibility genes (Figure 1), resulting in eight heterozygous variants in eight genes (Table 1, Figure 1). Among them, the variant in *PTEN*, p.Ser59Ter, has been reported to be pathogenic in ClinVar (Landrum et al., 2018) and had been reported as a somatic mutation 15 times in the TCGA dataset in various tumor types in cBioportal (Cerami et al., 2012; Gao et al., 2013). The patient was subsequently examined at the Department of Clinical Genetics at an age of 50 years and found to have macrocephaly (OFC: 65 cm), papillomatous lesions on hands and oral mucosa, intellectual disability (IQ = 84), and hamartomatous intestinal polyps. He was sent to surveillance of the thyroid according to clinical guidelines and multinodular goiter was detected. In all, these findings confirmed the clinical diagnosis of Cowden syndrome/ Hamartoma tumor syndrome (MIM #601728). His parents were both 23 years of age when he was born and neither had had any cancer. Genetic counseling was offered to his family
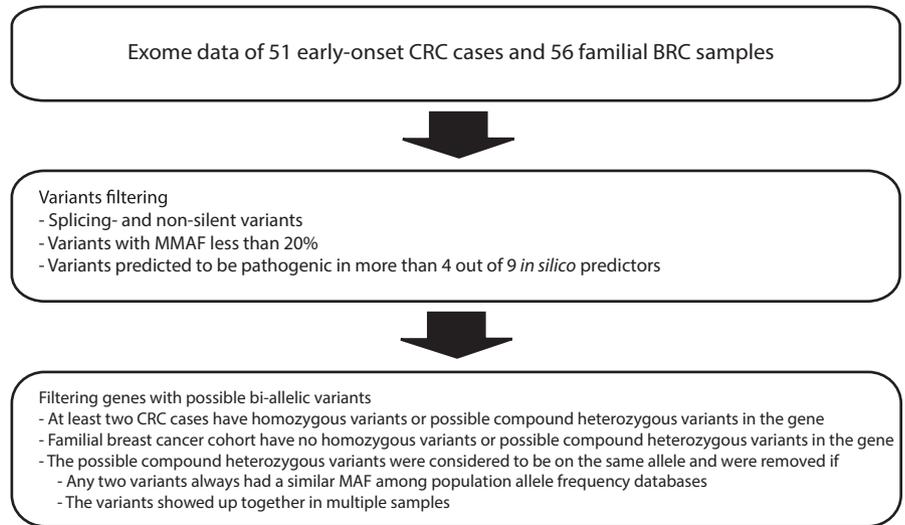
FIGURE 4 Autosomal recessive genes analysis

members. Unfortunately, no family members have contacted us and we have therefore not been able to perform genetic testing of the parents/siblings in order to confirm a de novo or familial variant. There were six candidate missense variants in the following cancer genes: *MSH2*, *APC*, *PTPN12*, *BMPR1A*, *POLE,* and *SRC,* and one inframe duplication in *BRIP1*. All were classified as variants of unknown significance according to ACMG criteria (Table 1). All of these genes are expressed in colon tissue (www.genecards.org), but none of these specific variants have been reported as somatic variants in 3,473 colorectal cancer samples or in any other tumor types in cBioPortal. After our study, immunohistochemistry of mismatch repair genes was performed as part of another research project on one patient with CRC at 37 years. His tumor showed loss of PMS2 protein. Targeted sequencing of *PMS2* (based on a clinical nested PCR approach) detected a NM_000535.5:c.2113G>A, p.Glu705Lys variant that is known to be pathogenic and has been reported as causative in many families, even though the INSIGHT expert panel have interpreted it as a variant of unknown significance (ClinVar). Thereafter, we manually checked the sequencing files of the other patients for poor quality *PMS2* variants and no other possible pathogenic variant in *PMS2* was found. No homozygous/compound heterozygous variants were found among the 244 cancer genes.

In addition to the cancer gene list, we also performed an analysis on frameshift-, nonsense-, and splice variants in the exome (Figure 2). After Sanger verification, apart from the variant in *PTEN*, there were 10 possible candidate variants from 10 genes (*CLSPN, CELSR2, ADAM17, BIRC6, SEC24B, RBM27, PPARGC1B, NCOA7, SSH2,* and *MYO9B*) (1 frameshift deletion, 7 nonsense-, and 2 splice variants) (Table 2). All these genes are expressed in the normal colon tissue (genecards). None of the variants were found in cBioPortal.

In missense variants analysis (Figure 3), there were in total 3,800 missense variants with MMAF < 0.1% in all

public databases. In order to select candidate variants, we filtered the variants with CADD score (Kircher et al., 2014) more than 20 ($n = 2301$) (Supporting Information, Table S1). Among them, there were 248 variants in genes in which the normal population has fewer missense variants than expected, ExAC missense Z-score > 3 (Lek et al., 2016), (Supporting Information, Table S2), suggesting that missense variants might be detrimental to the normal gene function. If we instead categorize the 2,301 variants based on higher CADD score, 1,140 variants had a CADD score >25 and 367 variants had a CADD score of more than 30 (Supporting Information, Table S2). 669/2301 variants had never been reported in a normal population and, of those with CADD >30, 82 had never been reported and eight of these were in a gene where missense variants were expected to be deleterious based on the conservative constraint, ExAC missense Z-score (Lek et al., 2016) resulting in the following candidate gene list: *ACACA, KIAA1109, GPHN, NR2C2, GPR25, ZNF462, INPP4A,* and *DIDO1* (Table 3).

In the autosomal recessive gene analysis (Figure 4), there were 16 genes where at least two CRC cases had homozygous or possible compound heterozygous variants and none in the breast cancer cohort that we used as a comparison group (Table 4). Six genes (*ATP10B*, *PKHD1*, *PTPRQ*, *UGGT2*, *MYH13*, *TFF3*) had variants with MMAF <5%, which were observed together in CRC cases twenty times more than expected by the computed likelihood. Therefore, these six genes are more likely to be candidate autosomal recessive risk factors.

4 | DISCUSSION

Since onset of colorectal cancer before the age of 40 years is infrequent, genetic predisposition may be suspected. Our cohort of 51 simplex cases with early-onset CRC tested two

TABLE 1 Candidate variants from cancer-gene-list study

Gene name	Protein change	cDNA change	Protein domain[b]	MMAF[c]	ClinVar	CADD[d]	Samples
PTEN	**NP_000305.3:p.Ser59Ter**	**NM_000314.3:c.176C>A**	Phosphatase tensin-type	**NA**	Pathogenic	50	**48-07LF**
APC	NP_000029.4:p.Asn944Asp	NM_000038.4:c.2830A>G	No special domain	0.05%	VUS	10.93	329-09F
MSH2	NP_000242.2:p.Val722Phe	NM_000251.2:c.2164G>T	No special domain	0.02%	VUS	34	1068-05o
PTPN12	NP_002826.6:p.Asp26Glu	NM_002835.6:c.78C>G	No special domain	0.01%	NA	11.9	323-09F
POLE	NP_006222.5:p.Pro13Ser	NM_006231.5:c.37C>T	No special domain	0.01%	NA	10.64	Co-121
BMPR1A	NP_004320.1:p.Ala293Val	NM_004329.3:c.878C>T	Cytoplasmic and protein kinase domain	NA	NA	27.2	Co-286
BRIP1	NP_114432.2:p.Asn1161_Asp1163dup	NM_032043.2:c.3481_3489dup	No special domain	NA	NA	NA	4283-13D
SRC	NP_938033.4:p.Gln112His	NM_198291.4:c.336G>C	SH3 domain	NA	NA	12.22	59-04o

VUS, variant of unknown significance.

Bold font indicates a known pathogenic variant based on it being a nonsense mutation, not previously reported in healthy individuals, but reported in patients with Cowden syndrome.

[a]UniProt protein domain. [b]Maximum minor allele frequency from ExAC ALL, ExAC NFE, ExAC AFR, ExAC AMR, ExAC EAS, ExAC FIN, ExAC SAS, ExAC OTH, gnomAD ALL, gnomAD AFR, gnomAD AMR, gnomAD ASJ, gnomAD EAS, gnomAD FIN, gnomAD NFE, gnomAD OTH, gnomAD NFE, 1000G ALL, 1000G EUR, SweGen 249DANES, 249SWEDES. [c]CADD score.

hypotheses: first a pathogenic variant in an autosomal dominant cancer gene; and second an autosomal recessive cancer syndrome.

In order to address the first hypothesis, we searched for heterozygous variants in genes known to be important in hereditary cancer or known somatic driver genes according to the classification by the Vogelstein group (Vogelstein et al., 2013) (Table 1). We found one pathogenic variant in *PTEN*, p.Ser59Ter, known to cause Cowden/hamartoma tumor syndrome and this syndrome could later be confirmed clinically in the patient and genetic counseling could be offered to his relatives. While the other variants (Table 1) suggested by the cancer gene list approach have an unknown clinical significance (Landrum et al., 2018), the genes are known to be cancer-related. As we did not have access to tumor material or DNA from relatives, unfortunately further analyses to test the gene expression in tumors and/or segregation in the family could not be performed. One patient demonstrated loss of PMS2 protein in his tumor and harbored a pathogenic variant in exon 12 in *PMS2* that was detected by a clinical lab. This variant was initially missed by our WES due to poor quality in the mapping step as a result of the pseudogene *PMS2CL* that has a high homology to exon 9 and 11–15 (Takeda et al., 2014). Loss of function in *APC* gene leads to the development of adenomas, a precursor lesion to CRC, in familial adenomatous polyposis (MIM #175100). This condition occurs in <1/10,000, thus the p.Asn944Asp variant with a MMAF of 0.05% is also probably too common to be disease causing. *MSH2* is known to cause Lynch syndrome (MIM #120435). The incidence of Lynch syndrome has been estimated to be 1/660–1/2,000 (de la Chapelle, 2005) and approximately 40% of cases are caused by *MSH2* mutation (i.e., incidence of MSH2-Lynch syndrome: 1/1,650–1/5,000) (Moller et al., 2017). In the ClinVar database, there are 850 pathogenic variants in *MSH2* and thus the p.Val722Phe *MSH2* variant with an incidence of 1/5,000 (0.02%) is probably too common to be disease-causing. *PTPN12* is critical for the regulation of cell proliferation, differentiation, and neoplastic transformation (Takekawa et al., 1992) but is not associated with a known clinical cancer syndrome. *POLE* encodes the catalytic subunit of DNA polymerase epsilon and mutations cause susceptibility to CRC and other tumor types (MIM # 615083). Both of these genes are somatically mutated in CRC, but these variants were also present in 0.01% of a normal population and are likely too common to be cancer predisposition variants. The *BMPR1A, BRIP1* and *SRC* variants have never been reported in healthy individuals which makes them highly interesting. *BMPR1A* variants can cause juvenile intestinal polyposis (MIM #174900). However, the only clinical information we had available from the patient was microsatellite-stable CRC at 35 years with no information

TABLE 2 Candidate truncating variants

Gene name	Protein change	cDNA change	Protein domain[a]	MMAF[c]	Samples
PTEN	**NP_000305.3:p.Ser59Ter**	**NM_000314.3:c.176C>A**	Phosphatase tensin-type	NA	48-07LF
CLSPN	NP_071394.5:p.Arg1040Ter	NM_022111.5:c.3118C>T	No special domain	NA	2750-13D
CELSR2	NP_001399.3:p.Asp843IlefsTer65	NM_001408.3:c.2527del	Cadherin 9	NA	4270-11D
ADAM17	NP_003174.3:p.Gln471Ter	NM_003183.3:c.1411C>T	Peptidase M12B	NA	760-04o
BIRC6	NA	NM_016252.2:c.12811-1G>A	No special domain	0.05%	1307-09D
SEC24B	NP_006314.1:p.Arg307Ter	NM_006323.2:c.919C>T	No special domain	NA	Co-286
RBM27	NA	NM_018989.1:c.590-2A>C	Arg-rich domain	0.05%	4735-11D
PPARGC1B	NP_573570.5:p.Trp328Ter	NM_133263.5:c.983G>A	No special domain	NA	294-07LF
NCOA7	NP_001186548.3:p.Arg913Ter	NM_001199619.3:c.2737C>T	TLD domain	0.02%	664-08F
SSH2	NP_203747.1:p.Tyr1292Ter	NM_033389.1:c.3876C>G	No special domain	NA	49-06o
MYO9B	NP_004136.1:p.Cys1646Ter	NM_004145.1:c.4938C>A	Phorbol-ester/DAG-type	NA	Co-489

All the variants in this table are heterozygous and were not reported in ClinVar. FIN, ExAC SAS, ExAC OTH, gnomAD ALL, gnomAD AFR, gnomAD AMR, gnomAD ASJ, gnomAD EAS, gnomAD FIN, gnomAD NFE, gnomAD OTH, 1000G ALL, 1000G EUR, SweGen 249DANES, 249SWEDES.
Bold font indicates a known pathogenic variant based on it being a nonsense mutation, not previously reported in healthy individuals, but reported in patients with Cowden syndrome.
[a]UniProt protein domain. [b]Maximum minor allele frequency from ExAC ALL, ExAC NFE, ExAC AFR, ExAC AMR, ExAC EAS, ExAC.

TABLE 3 Candidate missense variants from whole exome sequencing

Gene name	Protein change	Protein domain[a]	MMAF[b]	Mis Z[c]	CADD[d]	Samples
KIAA1109	NM_015312:p.S3578F	No special domain	0.0001	1.210177	28.1	711-14D
GPR25	NM_005298:p.Arg228Cys	Cytoplasmic domain	0	3.4183516	35	4270-11D
INPP4A	NM_001134224:p.Arg901His	No special domain	0	3.1907657	34	295-08F
NR2C2	NM_003298:p.Arg411His	Nuclear receptor (NR) ligand-binding (LBD) domain	0	3.481554	35	329-09F
ZNF462	NM_021224:p.Arg2068Gln	In zinc finger: C2H2-type 20	0	3.2241783	35	679-05o
GPHN	NM_020806:p.Gly472Asp	No special domain	0	3.571934	32	376-05o
ACACA	NM_198834:p.Arg2208Gln	No special domain	0.0000899	7.5240764	35	1307-09D
DIDO1	NM_033081:p.Arg1145His	No special domain	0	3.0966303	35	1068-05o

[a]UniProt protein domain. [b]Maximum minor allele frequency from ExAC ALL, ExAC NFE, ExAC AFR, ExAC AMR, ExAC EAS, ExAC FIN, ExAC SAS, ExAC OTH, gnomAD ALL, gnomAD AFR, gnomAD AMR, gnomAD ASJ, gnomAD EAS, gnomAD FIN, gnomAD NFE, gnomAD OTH, 1000G ALL, 1000G EUR, SWEGEN 249DANES, 249SWEDES. [c]ExAC missense Z Score. [d]Combined Annotation Dependent Depletion (CADD) score.

regarding possible polyps. We cannot exclude or verify this potential diagnosis clinically. *BRIP1* has been shown to be a moderate-risk ovarian cancer gene (Ramus et al., 2015; Weber-Lassalle et al., 2018), most often caused by truncating variants (ClinVar). Two CRC patients with truncating *BRIP1* variants and either familial CRC or polyposis have recently been reported (Martin-Morales et al., 2018; Rosenthal et al., 2018). However, the pathogenicity of our in-frame *BRIP1* duplication is not known. *SRC*, encoding a tyrosine kinase receptor, is frequently implicated in cancer (Turro et al., 2016). It was suggested as a putative CRC gene based on proteomic analysis (Zhang et al., 2014) and may have a role in colon cancer progression (Irby et al., 1999).

In an attempt to search for novel cancer genes, we identified 10 truncating variants (in addition to the pathogenic

PTEN variant) (Table 2) in genes which are not known to be susceptibility genes for cancer: *CLSPN, CELSR2, ADAM17, BIRC6, SEC24B, RBM27, PPARGC1B, NCOA7, SSH2,* and *MYO9B*. All these genes are expressed in colon tissue (genecards.org). Seven of these have never been reported in normal databases, supporting their role as rare risk variants. Three have a function highly related to cancer. *CLSPN* is involved in the DNA damage checkpoint (Chini & Chen, 2003) and is required for the activation of *CHK1* during a DNA replication checkpoint response (Kumagai & Dunphy, 2000). It is also involved in DNA replication and repair of DNA damage repair and may function as a tumor suppressor gene (Azenha, Lopes, & Martins, 2017). Nine somatic truncating mutations in *CLSPN* have been reported in CRC in cBioPortal including one at position 1139, that is, more distal to our variant. Therefore, this gene is the most likely candidate

TABLE 4 Sixteen candidate autosomal recessive genes where at least two CRC cases had homozygous or possible compound heterozygous variants and none in the breast cancer cohort that we used as a comparison group

Gene name	Protein change	cDNA change	Coordinate[a]	Likelihood[b]
ATP10B	NP_079429.2:p.Asn865Lys	NM_025153.2:c.2595C>A	5:g.160042903G>T	
	NP_079429.2:p.Gly393Trp	NM_025153.2:c.1177G>T	5:g.160061565C>A	0.0008
PKHD1	NP_619639.3:p.Thr2869Lys	NM_138694.3:c.8606C>A	6:g.51637536G>T	
	NP_619639.3:p.Ser1833Leu	NM_138694.3:c.5498C>T	6:g.51882310G>A	0.0000
	NP_619639.3:p.Leu1709Phe	NM_138694.3:c.5125C>T	6:g.51889483G>A	
	NP_619639.3:p.Cys843Ser	NM_138694.3:c.2527T>A	6:g.51910867A>T	0.0000
PTPRQ	NP_001138498.1:p.Pro98Leu	NM_001145026.1:c.293C>T	12:g.80839400C>T	
	NP_001138498.1:p.Gln425Glu	NM_001145026.1:c.831A>G	12:g.80878310A>G	
	NP_001138498.1:p.Thr1967Met	NM_001145026.1:c.5458C>T	12:g.81028849C>T	0.0001
	NP_001138498.1:p.Ile2207Val	NM_001145026.1:c.6177A>G	12:g.81066976A>G	0.0001
UGGT2	NP_064506.3:p.Asn1268Tyr	NM_020121.3:c.3802A>T	13:g.96511868T>A	0.0002
	NP_064506.3:p.Arg480His	NM_020121.3:c.1439G>A	13:g.96601605C>T	0.0006
MYH13	NP_003793.2:p.Arg1438Cys	NM_003802.2:c.4312C>T	17:g.10215944G>A	
	NP_003793.2:p.Ala1128Val	NM_003802.2:c.3383C>T	17:g.10222462G>A	
	NP_003793.2:p.Lys807del	NM_003802.2:c.2420_2422del	17:g.10233717_10233719del	0.0001
	NP_003793.2:p.Gly701Arg	NM_003802.2:c.2101G>A	17:g.10236464C>T	0.0006
TFF3	NP_003217.2:p.Val116Met	NM_003226.2:c.346G>A	21:g.43733628C>T	0.0001
PCDHGB4	NP_115269.1:p.Pro15_Val16del	NM_032098.1:c.44_49del	5:g.140767492_140767497del	0.0360
BCLAF1	NP_001070908.1:p.Glu401Ter	NM_001077440.1:c.1201G>T	6:g.136597456C>A	
	NP_001070908.1:p.Arg45Met	NM_001077441.1:c.134G>T	6:g.136599885C>A	0.0030
ZNRF2	NP_667339.3:p.Asp123dup	NM_147128.3:c.368_370dup	7:g.30325341_30325343dup	0.0027
AGPAT2	NP_006403.3:p.Val18dup	NM_006412.3:c.50_52dup	9:g.139581758_139581760dup	0.0110
MRGPRX3	NP_473372.3:p.Trp307Ter	NM_054031.3:c.920G>A	11:g.18159669G>A	0.0070
TPSG1	NP_036599.3:p.Arg194Trp	NM_012467.3:c.580C>T	16:g.1272273G>A	0.0016
	NP_036599.3:p.Ser138Phe	NM_012467.3:c.413C>T	16:g.1272750G>A	0.0150
	NP_036599.3:p.Thr75Lys	NM_012467.3:c.224C>A	16:g.1273444G>T	0.0240
FBXW10	NP_001254514.1:p.Arg275Leu	NM_001267585.1:c.824G>T	17:g.18653188G>T	0.0120
	NP_001254514.1:p.Arg607Cys	NM_001267585.1:c.1819C>T	17:g.18671961C>T	0.0150
ZNF730	NP_001264332.1:p.Thr196Ala	NM_001277403.1:c.586A>G	19:g.23328432A>G	
	NP_001264332.1:p.Ser296Phe	NM_001277403.1:c.887C>T	19:g.23328733C>T	0.0060
	NP_001264332.1:p.Ile479AsnfsTer	NM_001277403.1:c.1429_1430insA	19:g.23329275_23329276insA	0.0004
TARM1	NP_001129158.1:p.Arg111His	NM_001135686.1:c.332G>A	19:g.54578105C>T	0.0100
RFPL1	NP_066306.2:p.Gln243Ter	NM_021026.2:c.727C>T	22:g.29837884C>T	0.0370

Six genes (ATP10B, PKHD1, PTPRQ, UGGT2, MYH13 and TFF3) had variants, which were observed together in CRC cases twenty times more than expected by the computed likelihood.

Bold font indicates variants that are rare in the normal population and have a low lieklihoood of occuring together and are therefore possible candidate variants.

[a]Genomics coordinate based on GRCh37. [b]Likelihood to observe homozygous- or possible compound heterozygous variants based on MMAF of the two alleles. [c]Frequency of observed homozygous- or possible compound heterozygous variants in 51 early onset colorectal cancer samples. [d]Maximum obseeved frequency of homozygous samples in gnomAD ALL, gnomAD AFR, gnomAD AMR, gnomAD ASJ, gnomAD EAS, gnomAD FIN, gnomAD NFE, gnomAD OTH. [e]Maximum minor allele frequency from ExAC ALL, ExAC NFE, ExAC AFR, ExAC AMR, ExAC EAS, ExAC FIN, ExAC SAS, ExAC OTH, gnomAD ALL, gnomAD AFR, gnomAD AMR, gnomAD ASJ, gnomAD EAS, gnomAD FIN, gnomAD NFE, gnomAD OTH, 1000G ALL, 1000G EUR, SweGen 249DANES, 249SWEDES. [f]Allele frequency in 51 early onset colorectal cancer samples. [g]Allele frequency in 56 breast cancer samples. [h]Patients with homozygous- and possible compound heterozygous variants.

for loss-of-function mutations. The *CELSR2* protein was identified to contain epidermal growth factor-like domains (Nakayama et al., 1998). It is important for the development of normal cilia and neuronal migration (Goffinet & Tissir, 2017). The expression of *CELSR2* protein was increased in the cytoplasm of breast cancer cells (Jiang et al., 2018), but

obs gnomAD[c]	obs CRC[d]	MMAF[e]	CRC[f]	BRC[g]	Patient1[h]	Patient2[i]
		0.0430	0.0294	NA	48-07LF (het)	569-04o (het)
NA	0.0400	0.0199	0.0196	0.0179	48-07LF (het)	569-04o (het)
		0.0237	0.0490	0.0179	566-04o (het)	
NA	0.0200	0.0005	0.0098	NA	566-04o (het)	
		0.0053	0.0098	NA		1199-05o (het)
NA	0.0200	0.0000	0.0098	NA		1199-05o (het)
		0.0015	0.0098	NA	4773-11D (het)	
		0.0625	0.0294	NA		386-04o (het)
NA	0.0200	0.0012	0.0098	NA		386-04o (het)
NA	0.0200	0.0485	0.0098	NA	4773-11D (het)	
NA	0.0200	0.0310	0.0392	0.0089	Co-1524 (het)	
0.0002	0.0200	0.0078	0.0294	NA	Co-1524 (het)	61-09F (hom)
		0.0169	0.0196	0.0179	784-05o (het)	
		0.0149	0.0098	0.0089		Co-286 (het)
NA	0.0200	0.0065	0.0098	0.0268		Co-286 (het)
NA	0.0200	0.0331	0.0392	NA	784-05o (het)	
0.0001	0.0400	0.0089	0.0568	NA	1205-05o (hom)	760-04o (hom)
0.0180	0.0400	0.1905	0.0784	0.0909	294-07LF (hom)	Co-286 (hom)
		0.0393	0.0196	NA	2014-04018 (het)	91-04o (het)
NA	0.0400	0.1049	0.0588	0.0893	2014-04018 (het)	91-04o (het)
0.0006	0.0400	0.0526	0.0588	0.0577	813-06o (hom)	91-04o (hom)
0.0086	0.0400	0.1054	0.1765	0.0545	749-08F (hom)	760-04o (hom)
0.0060	0.0400	0.0855	0.1275	0.0893	1307-09D (hom)	784-05o (hom)
NA	0.0200	0.0132	0.0098	NA		4270-11D (het)
0.0090	0.0200	0.1225	0.0784	0.0446	1094-08F (hom)	4270-11D (het)
0.0070	0.0200	0.1546	0.0196	0.0364		711-14D (hom)
0.0070	0.0200	0.1134	0.0882	0.0804	Co-1190 (hom)	1199-05o (het)
NA	0.0200	0.1320	0.0686	0.0268		1199-05o (het)
		0.0921	0.0196	0.0446		771-08F (het)
NA	0.0200	0.0649	0.0392	0.0446	711-14D (het)	771-08F (het)
NA	0.0200	0.0065	0.0098	NA	711-14D (het)	
0.0200	0.0400	0.0979	0.1176	0.0638	5053-10D (hom)	Co-489 (hom)
0.0300	0.0400	0.1931	0.1863	0.125	258-06o (hom)	4283-13D (hom)

its possible function in cancer is unclear. The *ADAM17* protein is a metalloprotease first identified as an enzyme that cleaves tumor necrosis factor-alpha (Adrain, Zettl, Christova, Taylor, & Freeman, 2012; Black et al., 1997). Later it was found to have more than 80 substrates and is important for releasing ligands of *EGFR* from the membrane and thus

activating *EGFR* (Zunke & Rose-John, 2017). *ADAM17* is upregulated in colorectal cancer cells (Blanchot-Jossic et al., 2005) and blocking *ADAM17* in mouse colorectal cancer xenografts inhibited tumor growth (Rios-Doria et al., 2015). In humans, loss of both alleles of *ADAM17* can cause a rare neonatal autosomal recessive inflammatory bowel and skin disease (MIM #603639). As both *CELSR2* and *ADAM17* seem to have growth-stimulating functions, the loss-of-function mutations are less likely candidates for cancer predisposition. The variants in *BIRC6, RBM27,* and *NCOA7* were found in 0.02%–0.05% of the reference population which reduces the likelihood of them being pathogenic. Of these, *BIRC6* exhibited a role in resistance to apoptosis (Chen et al., 1999; Hao et al., 2004). The gene has been suggested to be a prognosis predictor in colorectal cancer (Hu et al., 2015) and *BIRC6* mutations have been found in colorectal adenomas (Zhou et al., 2013) and are common in carcinomas (Wolff et al., 2018). *NCOA7/ERAP140* is a nuclear receptor coactivator that binds to the estrogen receptor among others. It has been implicated as a risk factor for breast cancer (Higginbotham et al., 2011) but has as yet no known association to CRC. It has been shown to be over-expressed in oral squamous cell carcinoma where it promotes cell proliferation (Xie et al., 2016). Two genes have some association to cancer but not to CRC: *PPARGC1B* is a co-activator for *ESR1* and variants in the gene have been associated with increased risk for estrogen-positive breast cancer (Li et al., 2011). One study demonstrated that *MYO9B* was involved in the migration of prostate cancer cell lines and might be important for metastasis (Makowska, Hughes, White, Wells, & Peckham, 2015). To date, no studies have implicated *RBM27, SEC24B,* or *SSH2* in CRC or cancer development. As the variants in *SEC24B* and *SSH2* have never been reported in the normal population databases, they are also possible candidate genes.

In total, 3,800 rare missense variants were found in the exomes of the 51 young CRC cases. In order to try to reduce the number of candidates to consider, we filtered using CADD score > 20 which left 2301 missense variants (Supporting Information, Table S2). In order to pinpoint the candidate variants, we used even more stringent criteria (higher CADD, variants that were never previously reported and in genes that normally do not have many missense variants) giving a short-list of eight variants. CADD represents multiple lines of in silico evidence of pathogenicity and can predict pathogenicity reasonably well, however, in silico tools (such as CADD and the ExAC missense *Z*-score) will also miss clinically relevant pathogenic variants (van der Velde et al., 2015) and therefore this way of filtering is not optimal. In this way, we have probably excluded several variants that may have a pathogenic effect and in the future when many hundred thousand individuals have been sequenced in population databases, we may be able to filter solely on MMAF. Among the eight top candidate variants, one has been reported to be a

possible tumor suppressor gene: *INPP4A*. *INPP4A* has been shown to be downregulated in pancreatic cancer and to inhibit cell proliferation and promote apoptosis in bladder and pancreatic cancer cells (Wang, Feng, Jiang, & Zuo, 2017; Wang, Wu, Huang, & Chen, 2018). Three have been described in cancer, but their function remains unclear. *ACACA* (acetyl-CoA carboxylase or *ACC1*) catalyses the rate-limiting reaction in the biogenesis of long-chain fatty acids and is thus vital for cancer cell survival during hypoxia (Gao et al., 2016). Inhibition of *ACACA* can lead to either decreased cell proliferation (Jones et al., 2017; Singh, Yadav, Kumar, & Saini, 2015) or decreased apoptosis (Keenan et al., 2015) and increased risk of metastasis/tumor recurrence (in mice) (Rios Garcia et al., 2017). *NR2C2* can inhibit cancer initiation, but promote cancer progression (Lin et al., 2017). In colon cancer cell lines, *NR2C2* is required for cell survival and its inhibition induces cell death (McNew et al., 2016; Singh et al., 2012). *DIDO1* regulates embryonic stem cell renewal (Liu et al., 2014) and is upregulated in melanoma tissues and cell lines as well as colorectal tumors (Braig & Bosserhoff, 2013; Sillars-Hardebol et al., 2012), although in the latter, there was no correlation to gene copy number and *DIDO1* was suggested to be a passenger on the common 20q duplication seen in CRC. The remaining four have no known cancer-related function (*KIAA1109*, *GPHN*, *GPR25*, *ZNF462*). Thus, whole exome sequencing has demonstrated its capability in identifying a large amount of possible candidate variants in this study. By using very stringent criteria, we could identify four possible candidate genes with a putative role in cancer, however we have probably excluded several candidate variants in the filtering. Thus, the challenge is how to rank the variants in the candidate list to obtain an optimal balance between sensitivity and specificity.

In the search for a recessive syndrome, we found no cases of rare biallelic variants in any known cancer genes in our patients. We also searched the entire exome for more common autosomal recessive alleles and listed 16 possible candidate genes (Table 4) with six of them (*ATP10B*, *PKHD1*, *PTPRQ*, *UGGT2*, *MYH13*, *TFF3*) more likely than the others based on their MMAF <5% and observed frequency 20 times higher than the expected likelihood of occurring together. *ATP10B* is the catalytic part of a complex which catalyzes the hydrolysis of ATP coupled to the transport of aminophospholipids from the outer to the inner leaflet of the late endosomes (UniProt). It has been reported to be mutated somatically in 1.2% of all CRC in cBioportal (accessed 20181218), but its possible function in cancer is not yet known. *PKHD1* is known to cause autosomal recessive Polycystic Kidney Disease (MIM #606702). It encodes a protein which is important for correct mitotic spindle formation and function and inhibition leads to mitotic defects (Zhang et al., 2010). *PKHD1* is mutated in 5% of all CRC in cBioportal and in an analysis of 13,023 genes in 11 colorectal cases, *PKHD1* was ranked as the

seventh most common somatically mutated gene (Sjoblom et al., 2006). However, there was a debate whether the finding reached statistical significance (Forrest & Cavet, 2007; Ward et al., 2011). *PTPRQ,* Protein tyrosine phosphatase receptor-like type Q, was reported to be involved in phosphorylation/dephosphorylation signaling pathways and metastasis (Laczmanska et al., 2014, 2016; Sato et al., 2017). 1.2% of all CRC have somatic mutations in this gene (cBioportal, accessed 181218), however, as a suspected oncogene, this gene is a less likely candidate for an autosomal recessive predisposition gene. *UGGT2* is a uridine diphosphate-glucose:glycoprotein glucosyltransferase (Takeda et al., 2014). It has no known function in cancer although 2.4% of all CRC carry somatic mutations in *UGGT2* (cBioportal). *MYH13* also has no known role in cancer, but somatic mutations in CRC are found in 2.3% (cBioportal). *TFF3*, Trefoil factor 3, is a secreted protein which stimulates cell migration and prevention of apoptosis, enabling repair of the intestinal mucosa. It was suggested to be a risk factor for early recurrence of CRC and to be involved in promoting lymph node metastasis (Huang, Li, Wang, & Zhang, 2013; Morito et al., 2015). although TFF3 is very rarely mutated in CRC tissue (0.1%, cBioportal). In all, the relevance of our detected possibly biallelic variants is unknown. In summary, we have detected six putative candidate genes, of which five (*ATP10B, PKHD1, UGGT2, MYH13, TFF3)* may be candidate risk factor recessive genes for CRC.

Using the breast cancer cohort as a comparison to search for autosomal recessive genes may not be a good approach since there are genes that are predispose to both breast- and colorectal cancer. However, to look for possible compound heterozygous variants, we need a cohort where we have all the genotyping information. We have considered using genotyping data from public databases but there are a few issues. First, the number of samples with available genotyping information is far fewer than those that provide only allele frequency, due to data privacy. The second issue is about platform errors. Not all the variants we see in the analysis are from the samples. The errors can happen in any steps after DNA isolation: errors caused by contamination, by library preparation, by sequencing machines, or errors caused by computational tools. An advantage of our inhouse BRC cohort is that they are Swedish, the same population as the study group. Also, the DNA from the BRC cohort has been collected and processed in the same way as the study group. These advantages have eliminated the artifacts caused by the differences in populations and the difference in platforms.

The number of our proposed candidate truncating variants (10 variants in 51 cases) are slightly fewer than expected compared to those with similar study (26 of the 102 cases) (Adam et al., 2016), considering that our frequency cut-off is less stringent (0.1% vs. 0.01%). This discrepancy can likely be explained by number of databases and number of sub-populations used during filtering of rare variants. Adam et al. filtered their variants using 12 population databases including the seven ExAC cohorts. Our proposed candidate variants were required to have a MMAF <0.1% in 21 population databases. Among them, SweGen (Ameur et al., 2017) and our in-house database of 249 Swedish samples played a major role in removing normal variants belonging to the Swedish population. The ALL population in ExAC and gnomAD (Lek et al., 2016), together with their other seven sub populations, even though they overlapped, contributed significantly with variants belonging to other populations. Databases of 200 Danish samples (Li et al., 2010) and 1000Genomes (ALL and NFE populations) (1000 Genomes Project Consortium et al., 2012) were also included to increase the sensitivity. This shows the importance of using population-specific databases of normal variants in the filtering stage.

Our results (2/51 = 4% with a hereditary cancer syndrome: Cowden and *PMS2*-Lynch syndrome) are comparable to others. For example Pearlman et al. showed that 2% of all CRC cases with onset before 50 years had a hereditary cancer syndrome other than Lynch syndrome (*APC, MUTYH, SMAD4*). They detected Lynch syndrome in 8%, especially in those with a family history (Pearlman et al., 2017). In our cohort, we selected patients with no family history, which largely excludes Lynch syndrome, although CRC caused by *PMS2*-mutation with its low penetrance can occur in the absence of family history. As *PMS2* has multiple pseudogenes, pathogenic variants are often missed by WES and need to be manually analysed.

Simplex CRC cases are often caused by other mechanisms such as somatic mutations (Haraldsdottir et al., 2014; Mensenkamp et al., 2014) or epigenetic aberrations (Moreira et al., 2015). In the near future, patients with cancer will likely undergo paired tumor and germline testing in order to detect somatic aberrations that can guide therapy as part of genomic medicine initiatives across the Western world. As a bonus, both hereditary variants and somatic driver mutations can be identified in the same analysis. Until such pipelines are established, our results and those of others suggest that, in the absence of family history, or other suggestive features suggestive of hereditary syndromes such as polyposis or multiple tumors, genetic testing for gemline mutations in young CRC patients is of limited clinical benefit.

5 | CONCLUSION

Whole exome sequencing in early onset non-familial CRC patients only identified a causative germline mutation in two of 51 (4%). These patients had Cowden syndrome, that had not been diagnosed clinically or *PMS2*-Lynch syndrome that was difficult to detect using WES. Our results suggest

germline analysis using WES or broad gene panels in simplex cases with CRC at young age is of limited value in the clinic. In addition, we propose three candidate variants in known cancer susceptibility genes (*BMPR1A*, *BRIP1,* and *SRC*), up to three truncating variants in possibly novel cancer genes (*CLSPN* possibly *SEC24B*, *SSH2*), four missense variants in genes involved in cancer initiation or progression (*ACACA*, *NR2C2*, *INPP4A,* and *DIDO1*), and five candidate risk factor recessive genes (*ATP10B*, *PKHD1*, *UGGT2*, *MYH13*, *TFF3*). Further studies are needed to find support for the pathogenicity of these variants in novel Mendelian or complex disease in early onset non-familial CRC.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Jessada Thutkawkorapin https://orcid.org/0000-0001-9306-844X

Emma Tham https://orcid.org/0000-0001-6079-164X

REFERENCES

1000 Genomes Project Consortium, Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., … McVean, G. A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, *491*(7422), 56–65. https://doi.org/10.1038/nature11632

Aaltonen, L., Johns, L., Jarvinen, H., Mecklin, J. P., & Houlston, R. (2007). Explaining the familial colorectal cancer risk associated with mismatch repair (MMR)-deficient and MMR-stable tumors. *Clinical Cancer Research*, *13*(1), 356–361. https://doi.org/10.1158/1078-0432.CCR-06-1256

Adam, R., Spier, I., Zhao, B., Kloth, M., Marquez, J., Hinrichsen, I., … Aretz, S. (2016). Exome Sequencing Identifies Biallelic MSH3 Germline Mutations as a Recessive Subtype of Colorectal Adenomatous Polyposis. *The American Journal of Human Genetics*, *99*(2), 337–351. https//doi.org/ 10.1016/j.ajhg.2016.06.015

Adrain, C., Zettl, M., Christova, Y., Taylor, N., & Freeman, M. (2012). Tumor necrosis factor signaling requires iRhom2 to promote trafficking and activation of TACE. *Science*, *335*(6065), 225–228. https://doi.org/10.1126/science.1214400

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., … Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nature Methods*, *7*(4), 248–249. https://doi.org/10.1038/nmeth0410-248

Ameur, A., Dahlberg, J., Olason, P., Vezzi, F., Karlsson, R., Martin, M., … Gyllensten, U. (2017). SweGen: A whole-genome data resource of genetic variability in a cross-section of the Swedish population. *European Journal of Human Genetics*, *25*, 1253–1260. https://doi.org/10.1038/ejhg.2017.130

Azenha, D., Lopes, M. C., & Martins, T. C. (2017). Claspin functions in cell homeostasis-A link to cancer? *DNA Repair (Amst)*, *59*, 27–33. https://doi.org/10.1016/j.dnarep.2017.09.002

Beaugerie, L., & Itzkowitz, S. H. (2015). Cancers complicating inflammatory bowel disease. *New England Journal of Medicine*, *373*(2), 195. https://doi.org/10.1056/NEJMc1505689

Black, R. A., Rauch, C. T., Kozlosky, C. J., Peschon, J. J., Slack, J. L., Wolfson, M. F., … Cerretti, D. P. (1997). A metalloproteinase disintegrin that releases tumour-necrosis factor-alpha from cells. *Nature*, *385*(6618), 729–733. https://doi.org/10.1038/385729a0

Blanchot-Jossic, F., Jarry, A., Masson, D., Bach-Ngohou, K., Paineau, J., Denis, M. G., … Mosnier, J. F. (2005). Up-regulated expression of ADAM17 in human colon carcinoma: Co-expression with EGFR in neoplastic and endothelial cells. *The Journal of Pathology*, *207*(2), 156–163. https://doi.org/10.1002/path.1814

Bodmer, W., & Bonilla, C. (2008). Common and rare variants in multifactorial susceptibility to common diseases. *Nature Genetics*, *40*(6), 695–701. https://doi.org/10.1038/ng.f.136

Braig, S., & Bosserhoff, A. K. (2013). Death inducer-obliterator 1 (Dido1) is a BMP target gene and promotes BMP-induced melanoma progression. *Oncogene*, *32*(7), 837–848. https://doi.org/10.1038/onc.2012.115

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., … Schultz, N. (2012). The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discovery*, *2*(5), 401–404. https://doi.org/10.1158/2159-8290.CD-12-0095

de la Chapelle, A. (2005). The incidence of Lynch syndrome. *Familial Cancer*, *4*(3), 233–237. https://doi.org/10.1007/s10689-004-5811-3

Chen, C., Huang, H., & Wu, C. H. (2017). Protein bioinformatics databases and resources. *Methods in Molecular Biology*, *1558*, 3–39. https://doi.org/10.1007/978-1-4939-6783-4_1

Chen, Z., Naito, M., Hori, S., Mashima, T., Yamori, T., & Tsuruo, T. (1999). A human IAP-family gene, apollon, expressed in human brain cancer cells. *Biochemical and Biophysical Research Communications*, *264*(3), 847–854. https://doi.org/10.1006/bbrc.1999.1585

Chini, C. C., & Chen, J. (2003). Human claspin is required for replication checkpoint control. *Journal of Biological Chemistry*, *278*(32), 30057–30062. https://doi.org/10.1074/jbc.M301136200

Chun, S., & Fay, J. C. (2009). Identification of deleterious mutations within three human genomes. *Genome Research*, *19*(9), 1553–1561. https://doi.org/10.1101/gr.092619.109

Cooper, G. M., Stone, E. A., Asimenos, G., Program, N. C. S., Green, E. D., Batzoglou, S., & Sidow, A. (2005). Distribution and intensity of constraint in mammalian genomic sequence. *Genome Research*, *15*(7), 901–913. https://doi.org/10.1101/gr.3577405

Davydov, E. V., Goode, D. L., Sirota, M., Cooper, G. M., Sidow, A., & Batzoglou, S. (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS*

Computational Biology, *6*(12), e1001025. https://doi.org/10.1371/journal.pcbi.1001025

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., … Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498. https://doi.org/10.1038/ng.806

Forbes, S. A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., … Campbell, P. J. (2017). COSMIC: Somatic cancer genetics at high-resolution. *Nucleic Acids Research*, *45*(D1), D777–D783. https://doi.org/10.1093/nar/gkw1121

Forrest, W. F., & Cavet, G. (2007). Comment on "The consensus coding sequences of human breast and colorectal cancers". *Science*, *317*(5844), 1500; author reply 1500. https://doi.org/10.1126/science.1138179

Frank, C., Sundquist, J., Yu, H., Hemminki, A., & Hemminki, K. (2017). Concordant and discordant familial cancer: Familial risks, proportions and population impact. *International Journal of Cancer*, *140*(7), 1510–1516. https://doi.org/10.1002/ijc.30583

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., … Schultz, N. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science Signalling*, *6*(269), pl1. https://doi.org/10.1126/scisignal.2004088

Gao, X., Lin, S. H., Ren, F., Li, J. T., Chen, J. J., Yao, C. B., … Lei, Q. Y. (2016). Acetate functions as an epigenetic metabolite to promote lipid synthesis under hypoxia. *Nature Communications*, *7*, 11960. https://doi.org/10.1038/ncomms11960

Ghazi, N. G. (2010). Bevacizumab for neovascular age-related macular degeneration (ABC trial): Multicenter randomized double-masked study. *Expert Review of Clinical Pharmacology*, *3*(6), 747–752. https://doi.org/10.1586/ecp.10.58

Giraldez, M. D., Balaguer, F., Bujanda, L., Cuatrecasas, M., Munoz, J., Alonso-Espinaco, V., … Castellvi-Bel, S. (2010). MSH6 and MUTYH deficiency is a frequent event in early-onset colorectal cancer. *Clinical Cancer Research*, *16*(22), 5402–5413. https://doi.org/10.1158/1078-0432.CCR-10-1491

Goffinet, A. M., & Tissir, F. (2017). Seven pass Cadherins CELSR1-3. *Seminars in Cell & Developmental Biology*, *69*, 102–110. https://doi.org/10.1016/j.semcdb.2017.07.014

Hao, Y., Sekine, K., Kawabata, A., Nakamura, H., Ishioka, T., Ohata, H., … Naito, M. (2004). Apollon ubiquitinates SMAC and caspase-9, and has an essential cytoprotection function. *Nature Cell Biology*, *6*(9), 849–860. https://doi.org/10.1038/ncb1159

Haraldsdottir, S., Hampel, H., Tomsic, J., Frankel, W. L., Pearlman, R., de la Chapelle, A., & Pritchard, C. C. (2014). Colon and endometrial cancers with mismatch repair deficiency can arise from somatic, rather than germline, mutations. *Gastroenterology*, *147*(6), 1308–1316 e1301. https://doi.org/10.1053/j.gastro.2014.08.041

Higginbotham, K. S., Breyer, J. P., Bradley, K. M., Schuyler, P. A., Plummer, W. D. Jr, Freudenthal, M. E., … Smith, J. R. (2011). A multistage association study identifies a breast cancer genetic locus at NCOA7. *Cancer Research*, *71*(11), 3881–3888. https://doi.org/10.1158/0008-5472.CAN-10-2653

Hu, T., Weng, S., Tang, W., Xue, R., Chen, S., Cai, G., … Dong, L. (2015). Overexpression of BIRC6 is a predictor of prognosis for colorectal cancer. *PLoS ONE*, *10*(5), e0125281. https://doi.org/10.1371/journal.pone.0125281

Huang, Y. G., Li, Y. F., Wang, L. P., & Zhang, Y. (2013). Aberrant expression of trefoil factor 3 is associated with colorectal carcinoma metastasis. *Journal of Cancer Research & Therapy*, *9*(3), 376–380. https://doi.org/10.4103/0973-1482.119308

Irby, R. B., Mao, W., Coppola, D., Kang, J., Loubeau, J. M., Trudeau, W., … Yeatman, T. J. (1999). Activating SRC mutation in a subset of advanced human colon cancers. *Nature Genetics*, *21*(2), 187–190. https://doi.org/10.1038/5971

Jiang, L., Zhang, X., Xiang, C., Geradts, J., Wei, Q., Liang, Y., … Xu, J. F. (2018). Differential cellular localization of CELSR2 and ING4 and correlations with hormone receptor status in breast cancer. *Histology and Histopathology*, *33*(8), 835–842. https://doi.org/10.14670/HH-11-979

Jones, J. E., Esler, W. P., Patel, R., Lanba, A., Vera, N. B., Pfefferkorn, J. A., & Vernochet, C. (2017). Inhibition of acetyl-CoA carboxylase 1 (ACC1) and 2 (ACC2) reduces proliferation and de novo lipogenesis of EGFRvIII human glioblastoma cells. *PLoS ONE*, *12*(1), e0169566. https://doi.org/10.1371/journal.pone.0169566

Keenan, M. M., Liu, B., Tang, X., Wu, J., Cyr, D., Stevens, R. D., … Chi, J. T. (2015). ACLY and ACC1 regulate hypoxia-induced apoptosis by modulating ETV4 via alpha-ketoglutarate. *PLoS Genetics*, *11*(10), e1005599. https://doi.org/10.1371/journal.pgen.1005599

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*, *46*(3), 310–315. https://doi.org/10.1038/ng.2892

Kumagai, A., & Dunphy, W. G. (2000). Claspin, a novel protein required for the activation of Chk1 during a DNA replication checkpoint response in Xenopus egg extracts. *Molecular Cell*, *6*(4), 839–849.

Kumar, P., Henikoff, S., & Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, *4*(7), 1073–1081. https://doi.org/10.1038/nprot.2009.86

Laczmanska, I., Karpinski, P., Gil, J., Laczmanski, L., Bebenek, M., & Sasiadek, M. M. (2016). High PTPRQ expression and its relationship to expression of PTPRZ1 and the presence of KRAS mutations in colorectal cancer tissues. *Anticancer Research*, *36*(2), 677–681.

Laczmanska, I., Karpinski, P., Kozlowska, J., Bebenek, M., Ramsey, D., Sedziak, T., … Sasiadek, M. M. (2014). Copy number alterations of chromosomal regions enclosing protein tyrosine phosphatase receptor-like genes in colorectal cancer. *Pathology, Research and Practice*, *210*(12), 893–896. https://doi.org/10.1016/j.prp.2014.07.010

Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., … Maglott, D. R. (2018). ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Research*, *46*(D1), D1062–D1067. https://doi.org/10.1093/nar/gkx1153

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., … Exome Aggregation, C. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, *536*(7616), 285–291. https://doi.org/10.1038/nature19057

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26*(5), 589–595. https://doi.org/10.1093/bioinformatics/btp698

Li, Y., Li, Y., Wedren, S., Li, G., Charn, T. H., Desai, K. V., … Liu, J. (2011). Genetic variation of ESR1 and its co-activator PPARGC1B is synergistic in augmenting the risk of estrogen receptor-positive breast cancer. *Breast Cancer Research*, *13*(1), R10. https://doi.org/10.1186/bcr2817

Li, Y., Vinckenbosch, N., Tian, G., Huerta-Sanchez, E., Jiang, T., Jiang, H., … Wang, J. (2010). Resequencing of 200 human exomes

identifies an excess of low-frequency non-synonymous coding variants. *Nature Genetics*, *42*(11), 969–972. https://doi.org/10.1038/ng.680

Lichtenstein, P., Holm, N. V., Verkasalo, P. K., Iliadou, A., Kaprio, J., Koskenvuo, M., … Hemminki, K. (2000). Environmental and heritable factors in the causation of cancer–analyses of cohorts of twins from Sweden, Denmark, and Finland. *The New England Journal of Medicine.*, *343*(2), 78–85. https://doi.org/10.1056/NEJM200007133430201

Lin, S. J., Yang, D. R., Yang, G., Lin, C. Y., Chang, H. C., Li, G., & Chang, C. (2017). TR2 and TR4 orphan nuclear receptors: An overview. *Current Topics in Developmental Biology*, *125*, 357–373. https://doi.org/10.1016/bs.ctdb.2017.02.002

Liu, Y., Kim, H., Liang, J., Lu, W., Ouyang, B., Liu, D., & Songyang, Z. (2014). The death-inducer obliterator 1 (Dido1) gene regulates embryonic stem cell self-renewal. *Journal of Biological Chemistry*, *289*(8), 4778–4786. https://doi.org/10.1074/jbc.M113.486290

Makowska, K. A., Hughes, R. E., White, K. J., Wells, C. M., & Peckham, M. (2015). Specific myosins control actin organization, cell morphology, and migration in prostate cancer cells. *Cell Reports*, *13*(10), 2118–2125. https://doi.org/10.1016/j.celrep.2015.11.012

Martin-Morales, L., Rofes, P., Diaz-Rubio, E., Llovet, P., Lorca, V., Bando, I., … Caldes, T. (2018). Novel genetic mutations detected by multigene panel are associated with hereditary colorectal cancer predisposition. *PLoS ONE*, *13*(9), e0203885. https://doi.org/10.1371/journal.pone.0203885

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

McNew, K. L., Whipple, W. J., Mehta, A. K., Grant, T. J., Ray, L., Kenny, C., & Singh, A. (2016). MEK and TAK1 regulate apoptosis in colon cancer cells with KRAS-dependent activation of proinflammatory signaling. *Molecular Cancer Research*, *14*(12), 1204–1216. https://doi.org/10.1158/1541-7786.MCR-16-0173

Mensenkamp, A. R., Vogelaar, I. P., van Zelst-Stams, W. A., Goossens, M., Ouchene, H., Hendriks-Cornelissen, S. J., … Ligtenberg, M. J. (2014). Somatic mutations in MLH1 and MSH2 are a frequent cause of mismatch-repair deficiency in Lynch syndrome-like tumors. *Gastroenterology*, *146*(3), 643–646 e648. https://doi.org/10.1053/j.gastro.2013.12.002

Moller, P., Seppala, T., Bernstein, I., Holinski-Feder, E., Sala, P., Evans, D. G., … Mallorca, G. (2017). Incidence of and survival after subsequent cancers in carriers of pathogenic MMR variants with previous cancer: A report from the prospective Lynch syndrome database. *Gut*, *66*(9), 1657–1664. https://doi.org/10.1136/gutjnl-2016-311403

Moreira, L., Balaguer, F., Lindor, N., de la Chapelle, A., Hampel, H., Aaltonen, L. A., … Consortium, E. (2012). Identification of Lynch syndrome among patients with colorectal cancer. *JAMA*, *308*(15), 1555–1565. https://doi.org/10.1001/jama.2012.13088

Moreira, L., Munoz, J., Cuatrecasas, M., Quintanilla, I., Leoz, M. L., Carballal, S., … Gastrointestinal Oncology Group of the Spanish Gastroenterological Association (2015). Prevalence of somatic mutl homolog 1 promoter hypermethylation in Lynch syndrome colorectal cancer. *Cancer*, *121*(9), 1395–1404. https://doi.org/10.1002/cncr.29190

Morito, K., Nakamura, J., Kitajima, Y., Kai, K., Tanaka, T., Kubo, H., … Noshiro, H. (2015). The value of trefoil factor 3 expression in predicting the longterm outcome and early recurrence of colorectal cancer. *International Journal of Oncology*, *46*(2), 563–568. https://doi.org/10.3892/ijo.2014.2755

Nakayama, M., Nakajima, D., Nagase, T., Nomura, N., Seki, N., & Ohara, O. (1998). Identification of high-molecular-weight proteins with multiple EGF-like motifs by motif-trap screening. *Genomics*, *51*(1), 27–34. https://doi.org/10.1006/geno.1998.5341

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., … Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, *44*(D1), D733–745. https://doi.org/10.1093/nar/gkv1189

Pearlman, R., Frankel, W. L., Swanson, B., Zhao, W., Yilmaz, A., Miller, K., … Ohio Colorectal Cancer Prevention Initiative Study Group (2017). Prevalence and spectrum of germline cancer susceptibility gene mutations among patients with early-onset colorectal cancer. *JAMA Oncology*, *3*(4), 464–471. https://doi.org/10.1001/jamaoncol.2016.5194

Ramus, S. J., Song, H., Dicks, E., Tyrer, J. P., Rosenthal, A. N., Intermaggio, M. P., … Gayther, S. A. (2015). Germline mutations in the BRIP1, BARD1, PALB2, and NBN genes in women with ovarian cancer. *Journal of the National Cancer Institute*, *107*(11), pii: djv214. https://doi.org/10.1093/jnci/djv214

Reva, B., Antipin, Y., & Sander, C. (2011). Predicting the functional impact of protein mutations: Application to cancer genomics. *Nucleic Acids Research*, *39*(17), e118. https://doi.org/10.1093/nar/gkr407

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., … Committee, A. L. Q. A. (2015). Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, *17*(5), 405–424. https://doi.org/10.1038/gim.2015.30

Rios Garcia, M., Steinbauer, B., Srivastava, K., Singhal, M., Mattijssen, F., Maida, A., … Berriel Diaz, M. (2017). Acetyl-CoA carboxylase 1-dependent protein acetylation controls breast cancer metastasis and recurrence. *Cell Metabolism*, *26*(6), 842–855 e845. https://doi.org/10.1016/j.cmet.2017.09.018

Rios-Doria, J., Sabol, D., Chesebrough, J., Stewart, D., Xu, L., Tammali, R., … Michelotti, E. F. (2015). A Monoclonal antibody to ADAM17 inhibits tumor growth by inhibiting EGFR and non-EGFR-mediated pathways. *Molecular Cancer Therapeutics*, *14*(7), 1637–1649. https://doi.org/10.1158/1535-7163.MCT-14-1040

Rosenthal, E. A., Shirts, B. H., Amendola, L. M., Horike-Pyne, M., Robertson, P. D., Hisama, F. M., … NHLBI GO Exome Sequencing Project (2018). Rare loss of function variants in candidate genes and risk of colorectal cancer. *Human Genetics*, *137*(10), 795–806. https://doi.org/10.1007/s00439-018-1938-4

Sato, R., Nakano, T., Hosonaga, M., Sampetrean, O., Harigai, R., Sasaki, T., … Arima, Y. (2017). RNA sequencing analysis reveals interactions between breast cancer or melanoma cells and the tissue microenvironment during brain metastasis. *BioMed Research International*, *2017*, 8032910. https://doi.org/10.1155/2017/8032910

Schwarz, J. M., Rodelsperger, C., Schuelke, M., & Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. *Nature Methods*, *7*(8), 575–576. https://doi.org/10.1038/nmeth0810-575

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). dbSNP: The NCBI database of genetic variation. *Nucleic Acids Research*, *29*(1), 308–311. https://doi.org/10.1093/nar/29.1.308

Shihab, H. A., Rogers, M. F., Gough, J., Mort, M., Cooper, D. N., Day, I. N., … Campbell, C. (2015). An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics*, *31*(10), 1536–1543. https://doi.org/10.1093/bioinformatics/btv009

Sillars-Hardebol, A. H., Carvalho, B., Belien, J. A., de Wit, M., Delis-van Diemen, P. M., Tijssen, M., … Fijneman, R. J. (2012). CSE1L, DIDO1 and RBM39 in colorectal adenoma to carcinoma progression. *Cell Oncol (Dordr)*, *35*(4), 293–300. https://doi.org/10.1007/s13402-012-0088-2

Singh, A., Sweeney, M. F., Yu, M., Burger, A., Greninger, P., Benes, C., … Settleman, J. (2012). TAK1 inhibition promotes apoptosis in KRAS-dependent colon cancers. *Cell*, *148*(4), 639–650. https://doi.org/10.1016/j.cell.2011.12.033

Singh, R., Yadav, V., Kumar, S., & Saini, N. (2015). MicroRNA-195 inhibits proliferation, invasion and metastasis in breast cancer cells by targeting FASN, HMGCR, ACACA and CYP27B1. *Scientific Reports*, *5*, 17454. https://doi.org/10.1038/srep17454

Sjoblom, T., Jones, S., Wood, L. D., Parsons, D. W., Lin, J., Barber, T. D., … Velculescu, V. E. (2006). The consensus coding sequences of human breast and colorectal cancers. *Science*, *314*(5797), 268–274. https://doi.org/10.1126/science.1133427

Takeda, Y., Seko, A., Hachisu, M., Daikoku, S., Izumi, M., Koizumi, A., … Ito, Y. (2014). Both isoforms of human UDP-glucose:Glycoprotein glucosyltransferase are enzymatically active. *Glycobiology*, *24*(4), 344–350. https://doi.org/10.1093/glycob/cwt163

Takekawa, M., Itoh, F., Hinoda, Y., Arimura, Y., Toyota, M., Sekiya, M., … Yachi, A. (1992). Cloning and characterization of a human cDNA encoding a novel putative cytoplasmic protein-tyrosine-phosphatase. *Biochemical and Biophysical Research Communications*, *189*(2), 1223–1230.

Turro, E., Greene, D., Wijgaerts, A., Thys, C., Lentaigne, C., Bariana, T. K., … Ouwehand, W. H. (2016). A dominant gain-of-function mutation in universal tyrosine kinase SRC causes thrombocytopenia, myelofibrosis, bleeding, and bone pathologies. *Science Translational Medicine*, *8*(328), 328ra330. https://doi.org/10.1126/scitranslmed.aad7666

van der Velde, K. J., Kuiper, J., Thompson, B. A., Plazzer, J. P., van Valkenhoef, G., de Haan, M., … InSi, G. H. T. G. (2015). Evaluation of CADD scores in curated mismatch repair gene variants yields a model for clinical validation and prioritization. *Human Mutation*, *36*(7), 712–719. https://doi.org/10.1002/humu.22798

Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S. B., Diaz, L. A., & Kinzler, K. W. (2013). Cancer genome landscapes. *Science*, *339*(6127), 1546–1558. https://doi.org/10.1126/science.1235122

Wang, C., Feng, Z., Jiang, K., & Zuo, X. (2017). Upregulation of MicroRNA-935 promotes the malignant behaviors of pancreatic carcinoma PANC-1 cells via targeting inositol polyphosphate 4-phosphatase type i gene (INPP4A). *Oncology Research*, *25*(4), 559–569. https://doi.org/10.3727/096504016X14759554689565

Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, *38*(16), e164. https://doi.org/10.1093/nar/gkq603

Wang, R., Wu, Y., Huang, W., & Chen, W. (2018). MicroRNA-940 targets INPP4A or GSK3beta and activates the Wnt/beta-catenin pathway to regulate the malignant behavior of bladder cancer cells. *Oncology Research*, *26*(1), 145–155. https://doi.org/10.3727/096504017X14902261600566

Ward, C. J., Wu, Y., Johnson, R. A., Woollard, J. R., Bergstralh, E. J., Cicek, M. S., … Boardman, L. A. (2011). Germline PKHD1 mutations are protective against colorectal cancer. *Human Genetics*, *129*(3), 345–349. https://doi.org/10.1007/s00439-011-0950-8

Weber-Lassalle, N., Hauke, J., Ramser, J., Richters, L., Gross, E., Blumcke, B., … Hahnen, E. (2018). BRIP1 loss-of-function mutations confer high risk for familial ovarian cancer, but not familial breast cancer. *Breast Cancer Research*, *20*(1), 7. https://doi.org/10.1186/s13058-018-0935-9

Wolff, R. K., Hoffman, M. D., Wolff, E. C., Herrick, J. S., Sakoda, L. C., Samowitz, W. S., & Slattery, M. L. (2018). Mutation analysis of adenomas and carcinomas of the colon: Early and late drivers. *Genes Chromosomes Cancer*, *57*(7), 366–376. https://doi.org/10.1002/gcc.22539

Xie, X., Jiang, Y., Yuan, Y., Wang, P., Li, X., Chen, F., … Chen, Q. (2016). MALDI imaging reveals NCOA7 as a potential biomarker in oral squamous cell carcinoma arising from oral submucous fibrosis. *Oncotarget*, *7*(37), 59987–60004. https://doi.org/10.18632/oncotarget.11046

Xiong, H. Y., Alipanahi, B., Lee, L. J., Bretschneider, H., Merico, D., Yuen, R. K., … Frey, B. J. (2015). RNA splicing the human splicing code reveals new insights into the genetic determinants of disease. *Science*, *347*(6218), 1254806. https://doi.org/10.1126/science.1254806

Zhang, B., Wang, J., Wang, X., Zhu, J., Liu, Q., Shi, Z., … Nci, C. (2014). Proteogenomic characterization of human colon and rectal cancer. *Nature*, *513*(7518), 382–387. https://doi.org/10.1038/nature13438

Zhang, J., Wu, M., Wang, S., Shah, J. V., Wilson, P. D., & Zhou, J. (2010). Polycystic kidney disease protein fibrocystin localizes to the mitotic spindle and regulates spindle bipolarity. *Human Molecular Genetics*, *19*(17), 3306–3319. https://doi.org/10.1093/hmg/ddq233

Zhou, D., Yang, L., Zheng, L., Ge, W., Li, D., Zhang, Y., … Zheng, S. (2013). Exome capture sequencing of adenoma reveals genetic alterations in multiple cellular pathways at the early stage of colorectal tumorigenesis. *PLoS ONE*, *8*(1), e53310. https://doi.org/10.1371/journal.pone.0053310

Zunke, F., & Rose-John, S. (2017). The shedding protease ADAM17: Physiology and pathophysiology. *Biochimica et Biophysica Acta*, *1864*(11 Pt B), 2059–2070. https://doi.org/10.1016/j.bbamcr.2017.07.001

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Thutkawkorapin J, Lindblom A, Tham E. Exome sequencing in 51 early onset non-familial CRC cases. *Mol Genet Genomic Med*. 2019;7:e605. https://doi.org/10.1002/mgg3.605