Identification of known and novel familial cancer genes in Swedish colorectal cancer families

Hafdis T. Helgadottir1,2 | Jessada Thutkawkorapin1 | Anna Rohlin3,4 | Margareta Nordling3,4 | Kristina Lagerstedt-Robinson1,2 | Annika Lindblom1,2

1Department of Molecular Medicine and Surgery, Karolinska Institute, Stockholm, Sweden
2Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden
3Department of Laboratory Medicine, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
4Department of Clinical Genetics and Genomics, Sahlgrenska University Hospital, Gothenburg, Sweden

Abstract
Identifying new candidate colorectal cancer (CRC) genes and mutations are important for clinical cancer prevention as well as in cancer care. Genetic counseling is already implemented for known high-risk variants; however, the majority of CRC are of unknown causes. In our study, 110 CRC patients in 55 Swedish families with a strong history of CRC but unknown genetic causes were analyzed with the aim of identifying novel candidate CRC predisposing genes. Exome sequencing was used to identify rare and high-impact variants enriched in the families. No clear pathogenic variants were found in known CRC predisposing genes; however, potential pathogenic variants in novel CRC predisposing genes were identified. Over 3000 variants with minor allele frequency (MAF) <0.01 and Combined Annotation Dependent Depletion (CADD) > 20 were seen aggregating in the CRC families. Of those, 27 variants with MAF < 0.001 and CADD>25 were considered high-risk mutations. Interestingly, more than half of the high-risk variants were detected in three families, suggesting cumulating contribution of several variants to CRC. In summary, our study shows that despite a strong history of CRC within families, identifying pathogenic variants is challenging. In a small number of families, few rare mutations were shared by affected family members. This could indicate that in the absence of known CRC predisposing genes, a cumulating contribution of mutations leads to CRC observed in these families.

Keywords
colorectal cancer, exome-sequencing, germline, hereditary

1 | INTRODUCTION

Colorectal cancer (CRC) is the third most common cancer in the world. Both environmental and genetic risk factors have been proposed to contribute to CRC. The majority of the incidences are sporadic; however, hereditary factors cause about 30% of reported cases.1 The family history of cancer is the major risk factor of CRC. The relative risk of getting CRC is 2-fold if a first-degree relative has CRC and 4-fold if more than one relative has CRC.2

Colorectal cancer syndromes, such as Lynch syndrome, FAP (familial adenomatous polyposis) and MAP (MUTYH-associated polyposis), account for around 5% of the CRC incidents.3 The genetic causes of the Lynch syndromes are mutations in the MLH1, MSH2, MSH6 or PMS2 genes involved in the DNA mismatch repair (MMR), while

Received: 22 November 2020 | Revised: 17 February 2021 | Accepted: 26 February 2021
DOI: 10.1002/ijc.33567

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. International Journal of Cancer published by John Wiley & Sons Ltd on behalf of Union for International Cancer Control.

Int. J. Cancer. 2021;149:627–634. wileyonlinelibrary.com/journal/ijc 627
mutations in the APC and MUTYH genes cause FAP and MAP, respectively.\textsuperscript{3,4}

Traditionally, high-risk disease genes, including cancer genes, have been identified using linkage analysis in families with the disease. On occasion a candidate gene approach has been used to find a high-risk gene in families. Several high-risk colorectal cancer genes are known today and genetic testing in families with known mutations is possible.\textsuperscript{3,4} However, most of the CRC with a typical high-risk pedigree do not show segregated mutations in known genes and traditional linkage analysis has failed to demonstrate causative genes.\textsuperscript{5,6}

As a consequence, cancer and other diseases were considered to be complex diseases and a different strategy using genome-wide association studies (GWAS) have been used to define the genetic risk factors involved. Large GWAS have identified more than 100 loci associated with an increased risk of CRC.\textsuperscript{7-9} However, most of the loci confer low risk and do not explain the seemingly increased risk in cancer families. Therefore, when massive parallel sequencing became possible, families have been used for studies using whole-exome sequencing (WES) and whole genome sequencing (WGS) to find the responsible genes. However, this approach has rarely been successful in identifying a new high-risk gene.\textsuperscript{10-12} Thus, today most families with strong cancer history do not appear to segregate dominant risk genes. This has led to the hypothesis of a complex disease with several genes involved, varying among the family members and resulting in different degrees of risk between carriers.\textsuperscript{13,14}

The aim of our study was to identify high-risk variants that contribute to increased risk of cancer in Swedish CRC families. Exome sequencing was performed on 110 patients from 55 different families to identify variants likely to contribute to CRC risk in the families. In addition to known CRC predisposing genes, rare and high-impact variants in new CRC candidate genes were analyzed. To find support for the most likely predisposing colorectal high-risk genes, the strategy was to select rare variants segregating within family members or genes that were recurrent among families.

2 | MATERIALS AND METHODS

2.1 | Families

The individuals in our study were CRC patients from families that had undergone genetic counseling at the Department of Clinical Genetics, Karolinska University Hospital Solna (Sweden). As a part of the study, additional family members were recruited when possible. All families had a history of at least three close relatives with CRC. For each family, one to four individuals were whole exome sequenced, resulting in 110 CRC patients from a total of 55 families. Four families with four sequenced individuals (WES-4s, mainly consisting of first-degree relatives, average age of onset 61 years), eight families with three sequenced individuals (WES-3s, mainly consisting of first- and second-degree relatives, average age of onset 66 years), 27 families with two sequenced individuals (WES-2s, mainly consisting of first- and second-degree relatives, average age of onset 60 years) and 16 families with one sequenced individual (WES-1s, average age of onset 61 years) were analyzed (Table S1). An additional cohort consisting of 63 CRC patients with a history of at least two close relatives with CRC was whole genome sequenced to search for variants in high-risk genes.

All individuals included in the study had CRC and no specific risk factors for CRC, such as IBD, were known. All individuals gave written informed consent to participate in the study and to donate blood samples.

2.2 | Exome sequencing of blood samples from CRC families

DNA was quantified using a Qubit Fluorometer (Life Technologies). Sequencing libraries were prepared according to the TruSeq DNA Sample Preparation Kit EUC 15005180 or EUC 15026489 (Illumina) at an average coverage of ×100. Briefly, 1 to 1.5 μg of genomic DNA was fragmented (Covaris 400 bp protocol, Covaris, Inc.) and all samples subjected to end-repair, A-tailing and adaptor ligation (Illumina Multiplexing PE adaptors). A gel-based size selection step was performed, and the adapter-ligated fragments 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 analyzed on a Bioanalyzer High Sensitivity DNA chip (Agilent Technologies, Sweden) and concentration 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, demultiplexed and converted to FASTQ format using CASAVA 1.7 or 1.8 (Illumina).
2.3 | Bioinformatics workflow

Sequencing reads were aligned to the reference genome GRCh37 using BWA. Aligned reads were sorted and PCR-duplicated reads were marked using Picard. The sequencing coverage and quality statistics were estimated with Qualimap and are summarized in Supplementary Table S1.

Variants were called using GATK by following the best practice procedure implemented at the Broad Institute for individual samples as well as together in joint genotyping calling with default settings. Variant Quality Score Recalibration from GATK was used for quality control of the variants, and variants with GATK-VQSR “PASS” filter selected for further analysis. Variant annotation was done by ANNOVAR, including RefSeq gene annotation and dbSNP rs numbers. Max minor allele frequency (MMAF) was calculated from 21 genome databases: ExAC, 2100Danes, SweGen, and 1000 Genomes Project allele frequencies. To predict pathogenic effects of the variants, in silico predictors ClinVar and CADD were used. Information about genes expressed in the intestines and gene function was obtained from the Protein Atlas v19.3.

2.4 | Additional CRC cohort

The patient samples were investigated through WGS, performed by Novogene (Oxford, UK) using PCR-free library prep and sequencing on the Illumina platform. Bioinformatic analysis included BWA, Samtools, GATK and ANNOVAR. Vcf-files and CNV-files from the WGS data were analyzed in the software Moon (Diploid, Leuven, Belgium) with the HPO (Human Phenotype Ontology) terms “cancer” and “colon cancer.” Moon suggested four to five causative variants in each patient. Further evaluation and classification of the variants were performed using the ACMG criteria.

2.5 | Known CRC-predisposing genes—variant selection

Variants in 17 CRC genes commonly analyzed at Karolinska University Hospital as a part of genetic screening (APC, BMP1R1A, EPAC, MLH1, MSH2, MSH3, MSH6, MUTYH, NTHL1, PM2, POLD1, POLE, PTEN, RNF43, RPS20, SMAD4 and STK11) were investigated in CRC families. All variants that (a) had MMAF < 0.2; (b) were not considered benign according to ClinVar; and (c) had CADD > 20 were selected for further analysis.

2.6 | Novel CRC-predisposing genes—variant selection

Variants that were (a) detected in all affected family members, (b) with MMAF < 0.01, and (c) with CADD > 20 were selected for further analysis. Additionally, variants that were (a) detected in all family members of WES-3s and WES-4s; (2) with MMAF < 0.001; and (c) with CADD > 25 were defined as high-risk variants.

2.7 | Pathway analysis

Pathway analysis was performed on recurrent genes using the default settings for Wikipathway Cancer on the website WebGestalt (http://www.webgestalt.org/).

3 | RESULTS

3.1 | Variants in the clinical CRC gene panel were not able to explain the increased risk in any of the families

Since only one of the affected individuals in each family had previously been tested using the clinical panel, we first searched for variants in the 17 genes in the clinical panel in all 110 CRC patients (Table S2).

In total, 27 variants were identified in altogether 32 individuals, where 8 individuals had two variants. No clear pathogenic mutation was found that could explain the CRC risk in the families. Two variants in the MUTYH gene, rs36053993 (c.1145G>A;p.G382D) and rs34612342 (c.494A>G;p.Y165C), have previously been described as pathogenic with increased risk of CRC. However, since MAP is inherited as an autosomal recessive disease and the three individuals were heterozygote carriers, they had no increased risk of CRC. Also, in the other genes with an autosomal recessive inheritance (MSH3, NTHL1) all individuals were heterozygous carriers (Table S2).

In the genes known to increase risk in a dominant mode, only two of the variants were found in all tested family members, one in APC (rs748745776 [c.6458G>A;p.G2153E] in both members of the WES-2 Family-35) and one in MSH6 (rs761622304 [c.2342G>A;p.R781Q] in all members of the WES-3 Family-10). Furthermore, two variants were found in families where only one individual was sequenced, rs137854567 in APC and rs752015385 in STK11. Four variants, in the APC, POLE and NTHL1 genes (rs137854567, rs139075637, rs61732929, rs150766139), were recurrent and detected in more than one family. However, the variants did not segregate in all affected members of the same family (Table S2).

3.2 | The search for novel CRC candidate genes reveals 27 candidate high-risk genes in CRC families

To identify new CRC predisposing genes, we looked for risk variants enriched in the families. All variants that (a) were shared by all family members within each family; (b) had MMAF<0.01 and (c) CADD>20 were selected for further analysis.

In total, 3089 variants in 2594 genes were detected in the 55 families (Table 1, S3-S4). Several variants were recurrent (n = 108) and...
detected in two to five families resulting in total 3208 variant calls. Most of the variants were missense (n = 2847) where the majority of them were detected in WES-1s. Additionally, 242 potentially deleterious variants (stop-gain, splicing and frameshift indels) were detected enriched in the families (Table 1, S3-S4).

Variants with MMAF < 0.001 and CADD > 25 were considered the most likely high-risk variants/gene. In total, 27 variants in 27 genes were identified in 9 of the 12 WES-3s and WES-4s families (Table 2). Most variants were detected in Family 11, followed by Families 8 and 12 (n = 7, 5 and 4, respectively) (Table 2). The variants detected in Families 11 and 12 were all missense with an average CADD of 30.4 and 30, respectively, while three variants with average CADD of 30.4 and one frameshift variant were detected in Family 12 (Table 2).

The majority of the possibly pathogenic variants were missense; however, six variants with potentially deleterious consequences were detected. Two stop-gain variants chr15:39910412A/C in the FSIP1 gene, and rs762833274 in the SCN11A gene, were detected in Families 8 and 12, respectively, two frameshift deletions, rs1398727115 in the AKR1B10 gene, and chr19:56515227TA/G in the TSC22D2 gene were detected in Families 2 and 10, respectively, and two frameshift insertions, chr3:150176386−/G in the TTN gene, and rs78602043, in the PLIN2 gene, observed in Families 9 and 12, respectively (Table 2).

The 3089 unique variants total calls WES-4s WES-3s WES-2s WES-1s

| Number of families | Unique variants | Total calls | WES-4s | WES-3s | WES-2s | WES-1s |
|-------------------|----------------|-------------|--------|--------|--------|--------|
| 55                | 55             | 4           | 8      | 27     | 16     |
| 117               | 121            | 0           | 10     | 35     | 76     |
| 52                | 55             | 0           | 2      | 10     | 43     |
| 54                | 55             | 3           | 1      | 17     | 34     |
| 19                | 20             | 0           | 2      | 6      | 12     |
| 2847              | 2957           | 36          | 137    | 783    | 2001   |
| Total             | 3089           | 3208        | 39     | 152    | 851    | 2166   |

Note: WES-4s, families with four sequenced individuals; WES-3s, families with three sequenced individuals; WES-2s, families with two sequenced individuals; WES-1s, families with one sequenced individual.

3.3 | Recurrent genes were observed in the CRC families

Several recurrent genes were observed in the CRC families. First, we looked at the variants located within genes expressed in the intestine (Table S3). In total, we identified one to seven variants in 381 different genes in two to eight families. The most frequent genes were the CDH23 gene detected in eight families, and the ADCYS5, FAM132A, MICAL1, PLXNA2 and SYNE1 detected in five families each (Table S3).

Seven variants in eight families were identified in the CDH23 gene. One variant, rs111033369, was detected in two families (Families 10 and 15), while the remaining six variants were detected in one family each. One of them, a splice donor variant, rs764824311, was detected in Family 42 (Table S3). In the ADCYS5 gene, four missense variants were detected in five families (two WES-2s and three WES-1s) and the variants detected in the FAM132A, MICAL1, PLXNA2 and SYNE1 were all missense variants detected in WES-1s families (Table S3).

Next, we looked at variants located within genes not expressed in the intestine (Table S4). In total, we identified 69 genes with 1 to 19 variants detected in 2 to 15 families. The most frequent genes were the TTN gene detected in 15 families, the OBSCN and SPTB genes detected in five families each and ABCA13 and DNAH3 genes detected in four families each (Table S4). The 19 missense variants detected in the TTN gene were detected in 15 families. The four missense variants observed in the OBSCN gene were detected in four families. Two of the variants were detected in the same family (Family 55) and one variant, rs553216325, was detected in two families (Families 52 and 55). Five missense variants were detected in the SPTB gene in five WES-1s families (Table S4).

Several recurrent variants were observed. The most frequent variants were rs763221717, a missense variant in FAM132A gene, and rs201545668, a missense variant in MICAL1 gene, observed in five and four WES-1s families, respectively (Table S3). Furthermore, six variants were detected in three families each. A splice donor variant, rs78602043, in the LCORL gene, and five missense variants in the NEIL3, PLA2G4F, RTTN, SPTBN2 and TBC1D5 genes were detected in WES-2s and WES-1s families (Table S3). Pathway analysis on recurrent genes using Wikipathway Cancer revealed no significant association with any pathways.

3.4 | Variants in cancer-predisposing genes were observed in the CRC families

In 16 families, 29 variants in 26 known cancer-predisposing genes were observed. In most genes, only one variant was detected; however, two variants were seen in the DOCK8, PTCH1 and RECQL4 genes, and the same variants in the FANCC and TP53 genes were detected in two families (Table S3). These variants do not have known pathogenic consequences according to ClinVar. Interestingly, we observed families that carried more than one sequence variant in the cancer-predisposing genes. Four variants were observed in Family...
**Table 2**  High-risk variants with MMAF < 0.001, CADD > 25 and shared by all family members within families of 3 and 4

| Gene     | Location       | Ref/Alt | SNP-id    | Function       | Change                     | MMAF    | CADD | Family | WES     |
|----------|----------------|---------|-----------|----------------|----------------------------|---------|------|--------|---------|
| FAM150A  | chr8:53452429  | C/T     | rs145116532 | Missense       | NM_207413:exon3:c.G287A:p. R96Q | 0.0004  | 25.6 | 1      | WES-4s  |
| NR1D2    | chr3:24003501  | A/T     |           | Missense       | NM_001145425:exon5:c.A326T:p. K109I | 0       | 31   | 1      | WES-4s  |
| AKR1B10  | chr7:13422968  | AGAG/−  | rs1398727115 | Deletion       | NM_020929:exon8:c.764_766del:p.255_256del | 9.00E-04 | na   | 2      | WES-4s  |
| ERCC6    | chr10:50686495 | C/T     | rs114423177 | Missense       | NM_000124:exon11:c.G2191A:p. A731T | 5.00E-04 | 29   | 6      | WES-3s  |
| GORASP1  | chr3:39144216  | G/A     | rs150534574 | Missense       | NM_031899:exon3:c.C301T:p. R101C | 2.00E-04 | 26.8 | 7      | WES-3s  |
| ITGA3    | chr17:48145642 | G/A     | rs772771598 | Missense       | NM_001145425:exon4:c.G637A:p. G213S | 6.08E-05 | 29.4 | 7      | WES-3s  |
| TSHR     | chr10:50686495 | G/A     | rs114423177 | Missense       | NM_000124:exon11:c.G2191A:p. A731T | 5.00E-04 | 29   | 6      | WES-3s  |
| TCR6     | chr12:118639247 | G/A | rs537291817 | Missense       | NM_016281:exon12:c.C841T:p. R281W | 2.00E-04 | 26.4 | 8      | WES-3s  |
| FSIP1    | chr15:39910412 | A/C     |           | Stopgain       | NM_152597:exon11:c.T1223G:p. L408X | 0       | 44   | 8      | WES-3s  |
| OBL1     | chr2:220422128 | C/T     |           | Missense       | NM_01173431:exon12:c.G4003A:p.D1335N | 0       | 27.4 | 8      | WES-3s  |
| PK3C28   | chr1:204425110 | G/C     |           | Missense       | NM_002646:exon12:c.C301T:p. A101C | 0       | 26.3 | 7      | WES-3s  |
| LMO7     | chr13:76419480 | C/T     | rs763854453 | Missense       | NM_015842:exon23:c.C3416T:p. T1139I | 6.00E-04 | 26.9 | 9      | WES-3s  |
| TP1      | chr15:30092859 | G/A     | rs375824999 | Missense       | NM_0003257:exon12:c.C741T:p.T25M | 1.00E-04 | 28.3 | 9      | WES-3s  |
| TSC22D2  | chr3:150176386 | A/G     |           | Insertion      | NM_011302364:exon3:c.2235dupG:p.T45fs | 0       | na   | 9      | WES-3s  |
| NLRP5    | chr19:56515227 | TA/−    | rs1173056419 | Deletion       | NM_153447:exon2:c.208_209del:p. Y70fs | 0       | na   | 10     | WES-3s  |
| TGM2     | chr20:36789908 | C/T     |           | Missense       | NM_004613:exon2:c.G104A:p. R35Q | 0       | 28.1 | 10     | WES-3s  |
| PDR3GL   | chr1:145460201 | G/A     | rs367697753 | Missense       | NM_032305:exon12:c.C417T:p.R84W | 8.00E-04 | 32   | 11     | WES-3s  |
| PHK1     | chr7:56149726  | C/T     | rs141344297 | Missense       | NM_001258460:exon8:c.G625A:p. V209I | 0.0008  | 26.7 | 11     | WES-3s  |
| TFR2     | chr7:100218569 | G/A     | rs140161160 | Missense       | NM_01206855:exon15:c.C1804T:p.R660W | 5.00E-04 | 29.3 | 11     | WES-3s  |
| IKKBAP   | chr9:111653612 | G/A     | rs763981711 | Missense       | NM_003640:exon28:c.G3031T:p. R101C | 1.00E-04 | 32   | 11     | WES-3s  |
| KCTD18   | chr2:201371607 | G/A     | rs770190157 | Missense       | NM_01258460:exon8:c.G625A:p. V209I | 0.0008  | 26.7 | 11     | WES-3s  |
| FN1      | chr2:216240042 | G/A     | rs372237449 | Missense       | NM_212474:exon36:c.C5509T:p. R1837C | 4.50E-05 | 32   | 11     | WES-3s  |
| FAM131B  | chr7:143056075 | G/T     | rs774485532 | Missense       | NM_001278297:exon4:c.C9A:p. A10D | 1.52E-05 | 27.7 | 11     | WES-3s  |
| SCN11A   | chr3:38913731  | G/A     | rs762833274 | Stopgain       | NM_014139:exon20:c.C3448T:p. R1150X | 5.00E-04 | 37   | 12     | WES-3s  |
| SCN5A    | chr3:38628991  | C/T     | rs771339055 | Missense       | NM_000335:exon15:c.G2416A:p. V806M | 5.00E-04 | 26.4 | 12     | WES-3s  |
| PLIN2    | chr9:19119812  | −/T     | rs550011861 | Insertion      | NM_00122:exon6:c.612dupA:p. V205fs | 6.67E-05 | na   | 12     | WES-3s  |
| BCA1     | chr1:156617309 | T/G     |           | Missense       | NM_021948:exon4:c.T476G:p. F159C | 0       | 27.7 | 12     | WES-3s  |

Note: Location according to hg19, SNPid according to dbSNP150, change shows transcript, exon, amino acid change and protein change. MMAF: Max minor allele frequency indicates the highest minor allele frequency in 20 population (see methods). WES-4s, families with four sequenced individuals; WES-3s, families with three sequenced individuals; WES-2s, families with two sequenced individuals; WES-1s, families with one sequenced individual.
10, three siblings diagnosed with CRC at the ages of 69 to 80 years. The individuals were heterozygous carriers, while only biallelic variants were detected in seven families, the most interesting ones were the BRCA1-rs28897689 and CHEK2-rs28909982 variants detected in Family 44, and the BRCA2-rs748816192 and PALB2-rs515726123 observed in Family 13 (Table S3).

3.5 Variants in high-risk genes were detected in additional Swedish CRC families

Finally, we searched for variants in the high-risk genes in additional set of Swedish CRC patients with an early-onset CRC, or at least one additional affected relative. No pathogenic variants were observed in the clinical CRC predisposing genes, we searched for rare variants enriched in CRC families. No pathogenic variants were observed in the clinical CRC predisposing genes, apart from the MUTYH and MSH3 genes. The individuals were heterozygous carriers, while only biallelic variants increase the risk of CRC.

Two missense variants in the MSH6 and APC genes were enriched in the families. Both variants have uncertain clinical significance, are rare and have high CADD scores (22.8 and 28, respectively). The variant in the MSH6 gene was detected in all sequenced family members of Family 10, three siblings diagnosed with CRC at the ages of 69 to 80 years. The variant is one of many in the MSH6 gene that has been identified in Swedish families with familial bowel cancer; however, its contribution to the disease in this family as well as in the previously reported individuals is unknown. The variant in the APC gene was detected in two family members of Family 35, sisters diagnosed with CRC at the ages of 43 and 49. The contribution of these two variants to the disease in these two families is unknown and further studies are needed to resolve their importance.

To find novel CRC predisposing genes in the CRC families, a selection criteria (MMAF < 0.01 and CADD > 25) was used to identify high-risk variants, even stricter criteria (MMAF < 0.001 and CADD > 25) was applied in the families with at least three affected relatives. That resulted in the identification of 27 very rare and high-impact variants located in 27 genes, many of which previously linked with CRC. Among the high-risk variants, the most interesting candidates were the four frameshift indels and two stopgain variants. These variants have potentially deleterious effects and two of them are located in the genes AKR1B10 and TSC22D2 that previously have been linked to CRC. A 4-bp deletion in the AKR1B10 gene was detected in a family consisting of two brothers diagnosed with CRC in their sixties, and two with severe polyposis. AKR1B10 is highly expressed in the colon and has been suggested to protect the colon cells from DNA damage. Downregulation of AKR1B10 has been observed in CRC and to be correlated with poor prognosis. It is believed to affect the tumor suppressor function of TP53 and suggested to contribute to chronic inflammation and dysplasia in ulcerative colitis. The other gene, TSC22D2, has previously been linked to multicancer families and CRC. A 1-bp insert was observed in the last exon of the gene in Family 9, consisting of two siblings diagnosed with CRC at the ages of 69 and 75, and their niece diagnosed with CRC at the age of 40. TSC22D2 suppresses CRC cell growth and is downregulated in CRC; however, its contribution to tumorigenesis is unknown. Analysis on the high-risk genes in additional CRC cohort did not reveal the same variants as identified here. However, three rare and high-impact missense variants were detected.

Interestingly, more than half of the high-risk variants were observed in three families. Most of the variants were detected in Family 11, among others in the FAN1 gene, a cancer gene that has been linked to CRC, and the POLR3GL gene, a subunit of RNA pol III and important for cell growth. The accumulating effect of risk variants to the contribution of CRC has been suggested in studies on polygenic risk scores. However, the studies have been performed with common variants identified in GWAS while here in our study, the detected variants were rare. Therefore, as with using polygenic risk scores for individual risk prediction, it is hard to estimate the cumulative contribution of the rare high-risk variants to CRC in these families.

Another approach to find the most likely genes/variants contributing to disease in the families is to search for recurrent genes and variants. The most mutated genes were the CDH23 and TTN genes. Both genes are large, consisting of 69 and 363 exons, respectively. The CDH23 gene encodes for cadherin 23 protein, important for cell adhesion. Mutations in the gene have been linked with hearing loss. Moreover, the gene has been linked to familial and sporadic pituitary cancer and to contribute to breast cancer metastasis. Furthermore, it is located within a linkage peak reported in CRC families. The TTN gene is commonly mutated in cancer although its role in cancer is unknown. Mutations in the CDH23 and TTN genes were found in 8 and 15 families, respectively. Several other recurrent genes were found in five or fewer families, including the OBSCN gene that recently was identified as one of the genes frequently mutated in Chinese CRC patients.

Finally, we observed mutations in cancer-predisposing genes that are linked to several hereditary cancer and believed to be involved in...
~3% of cancer incidences. Two of the genes, PALB2 and BRCA2, are important for DNA repair response and are best known as breast cancer-risk genes. Here, they were detected in a CRC family with no history of breast cancer, the two carriers of the variants are second-degree relatives diagnosed with CRC at the ages of 64 and 70. Although the BRCA2 mutation is of uncertain consequence according to ClinVar, the frameshift deletion in the PALB2 gene has previously been linked to breast cancer and CRC.

Our study has several limitations. First of all, the cohort consists of exome sequenced individuals, we cannot rule out that variants outside of the exons contribute to the disease. Second, this cohort consists of a limited number of CRC cases in each family and does not include unaffected individuals; and finally, when searching for rare variants the selection criteria (MMAF<0.01) common variants that might have contributed to the disease were excluded.

Here, we show that CRC families with strong family history do not carry mutations in known CRC genes and that few rare mutations are shared between affected family members. We observed families with several high-risk genes that could contribute to CRC. Our study indicates that when in the absence of mutations in known cancer-predisposing genes, CRC is a complex disease where many mutations with moderate effect could contribute.

ACKNOWLEDGEMENTS
The authors would like to acknowledge the support from Science for Life Laboratory, the National Genomics Infrastructure (NGI) Stockholm (funded by the Swedish Research Council), Swedish National Infrastructure for Computing (SNIC) (partially funded by the Swedish Research Council through grant agreement no. 2018-05973) and the Uppsala Multi-disciplinary Center for Advanced Computational Science (UPPMAX) for assistance with massively parallel sequencing. The computations were performed under Project SNIC sens2018560. This work was mainly funded by grants to M.N. from the Swedish Cancer Society and the Cancer Research Funds of Radiumhemmet. Furthermore, grants to M.N. from the Swedish Cancer Society and The Swedish state under the ALF agreement concerning research and education of doctors.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT
Access to the data is controlled. Variants that fulfilled our selection criteria can be found in the supplementary tables. However, Swedish laws and regulations prohibit the release of individual and personally identifying data. Therefore, the whole data cannot be made publicly available. The data that support the findings of this study are available from the corresponding authors upon a reasonable request.

ETHICAL STATEMENT
All patients gave written informed consent to participate in the study and to donate blood samples. The study was approved by the research ethics committee at Karolinska Institutet (2002-489), the regional ethics committee in Stockholm (2014-928-32) and the regional ethics committee in Gothenburg (227-10).

ORCID
Hafdis T. Helgadottir https://orcid.org/0000-0003-4352-152X

REFERENCES
1. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med. 2000;343:78-85.
2. Johns LE, Houlston RS. A systematic review and meta-analysis of familial colorectal cancer risk. Am J Gastroenterol. 2001;96:2992-3003.
3. Kastrinos F, Syngal S. Inherited colorectal cancer syndromes. Cancer J. 2011;17:405-415.
4. Valle L, Vilar E, Tavtigian SV, Stoffel EM. Genetic predisposition to colorectal cancer: syndromes, genes, classification of genetic variants and implications for precision medicine. J Pathol. 2019;247:574-588.
5. von Holst S, Jiao X, Liu W, et al. Linkage analysis revealed risk loci on 6p21 and 18p11.2-q11.2 in familial colon and rectal cancer, respectively. Eur J Hum Genet. 2019;27:1286-1295.
6. Kontham V, von Holst S, Lindblom A. Linkage analysis in familial non-Lynch syndrome colorectal cancer families from Sweden. PLoS One. 2013;8:e83936.
7. Schmit SL, Edlund CK, Schumacher FR, et al. Novel common genetic susceptibility loci for colorectal cancer. J Natl Cancer Inst. 2019;111:146-157.
8. Huyghe JR, Bien SA, Harrison TA, et al. Discovery of common and rare genetic risk variants for colorectal cancer. Nat Genet. 2019;51:76-87.
9. Li X, Timofeeva M, Spiliopoulou A, et al. Prediction of colorectal cancer risk based on profiling with common genetic variants. Int J Cancer. 2020;147(12):3431-3437.
10. Palles C, Cazier J-B, Howarth KM, et al. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. Nat Genet. 2013;45:136-144.
11. Yan HHN, Lai JCW, Ho SL, et al. NRFN43 germline and somatic mutation in serrated neoplasia pathway and its association with BRAF mutation. Gut. 2017;66:1645-1656.
12. Weren RDA, Litgengen MJL, Kets CM, et al. A germline homozygous mutation in the base-excision repair gene NTHL1 causes adenomatous polyposis and colorectal cancer. Nat Genet. 2015;47:668-671.
13. Frampton MJE, Law P, Litchfield K, et al. Implications of polygenic risk for personalised colorectal cancer screening. Ann Oncol. 2016;27:429-434.
14. Balmain A. Cancer as a complex genetic trait. Cell. 2002;108:145-152.
15. Li H, Durbin R. Fast and accurate long-read alignment with burrows–wheeler transform. Bioinformatics. 2010;26:589-595.
16. Okonechnikov K, Conesa A, Garcia-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics. 2016;32:292-294.
17. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43:491-498.
18. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of variant genotypes from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164-e164.
19. O’Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016;44:D733-D745.
20. Sherry ST. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29:308-311.
21. Exome Aggregation Consortium, Lek M, Karcewski KJ, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016; 536:285-291.
22. Li Y, Vinckenbosch N, Tian G, et al. Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. Nat Genet. 2010;42:969-972.
23. Ameur A, Dahlberg J, Olason P, et al. SweGen: a whole-genome data resource of genetic variability in a cross-section of the Swedish population. Eur J Hum Genet. 2017;25:1253-1260.
24. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015;526:68-74.
25. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018;46:D1062-D1067.
26. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46:310-315.
27. Uhlen M, Fagerberg L, Hallström BM, et al. Tissue-based map of the human proteome. Science. 2015;347:1260419.
28. Li H, Handsaker B, Wysoker A, et al. 1000 genome project data processing subgroup. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078-2079.
29. Lagerstedt-Robinson K, Rohlin A, Nordling M, Lagerstedt-Robinson K, Lindblom A. Mismatch repair defects in colorectal cancer. Mol Carcinog. 2017;56:118-129.
30. Ohashi T, Idogawa M, Sasaki Y, Suzuki H, Tokino T. AKR1B10, a transcriptional target of p53, is downregulated in colorectal cancers associated with poor prognosis. Mol Cancer Res. 2013;11:1554-1563.
31. Xiao L, Wei F, Liang F, et al. TSC22D2 identified as a candidate susceptibility gene of multi-cancer pedigree using genome-wide linkage analysis and whole-exome sequencing. Carcinogenesis. 2019;40:819-827.
32. Liang F, Li Q, Li X, et al. TSC22D2 interacts with PKM2 and inhibits cell growth in colorectal cancer. Int J Oncol. 2016;49:1046-1056.
33. Dobkins SE, Broderick P, Chubb D, Kinnersley B, Sherborne AL, Houlston RS. Undefined familial colorectal cancer and the role of pleiotropy in cancer susceptibility genes. Fam Cancer. 2016;15:593-599.
34. Neklason DW, Tuohy TM, Stevens J, et al. Colorectal adenomas and cancer link to chromosomes 13q22.1-13q31.3 in a large family with excess colorectal cancer. J Med Genet. 2010;47:692-699.
35. Shen Y, Liao D-F, Cao D. AKR1B10 in gastrointestinal diseases. Aging. 2015;7:221-222.
36. Zu X, Yan R, Pan J, et al. Aldo-keto reductase 1B10 protects human colon cells from DNA damage induced by electrophilic carbonyl compounds. Mol Carcinog. 2017;56:118-129.
37. Kropotova ES, Tychko RA, Zinov’eva OL, et al. Downregulation of AKR1B10 expression in colorectal cancer. Mol Biol. 2010;44:216-222.
38. Shen Y, Ma J, Yan R, et al. Impaired self-renewal and increased colitis and dysplastic lesions in colonic mucosa of AKR1B8-deficient mice. Clin Cancer Res. 2015;21:1446-1476.
39. Yi W, Xiao E, Ding R, Luo P, Yang Y. High expression of fibronectin is associated with poor prognosis, cell proliferation and malignancy via the NF-κB/p53-apoptosis signaling pathway in colorectal cancer. Oncol Rep. 2016;36:3145-3153.
40. Haurie V, Durrieu-Gaillard S, Dumay-Odelot H, et al. Two isoforms of human RNA polymerase III with specific functions in cell growth and transformation. Proc Natl Acad Sci U S A. 2010;107:4176-4181.
41. Archambault AN, Su Y-R, Jeon J, et al. Cumulative burden of colorectal-cancer-associated genetic variants is more strongly associated with early-onset vs late-onset cancer. Gastroenterology. 2020;158:1274-1286.e12.
42. Weigl K, Chang-Claude J, Knebel P, Hsu L, Hoffmeister M, Brenner H. Strongly enhanced colorectal cancer risk stratification by combining family history and genetic risk score. Clin Epidemiol. 2018;10:143-152.
43. Dunlop MG, Tenesa A, Farrington SM, et al. Cumulative impact of common genetic variants and other risk factors on colorectal cancer risk in 42,103 individuals. Gut. 2013;62:871-881.
44. Schultz JM, Bhatti R, Madeo AC, et al. Allelic hierarchy of CDH23 mutations causing non-syndromic deafness DFNB12 or usher syndrome USH1D in compound heterozygotes. J Med Genet. 2011;48:767-775.
45. Zhang Q, Peng C, Song J, et al. Germline mutations in CDH23, encoding cadherin-related 23, are associated with both familial and sporadic pituitary adenomas. Am J Hum Genet. 2017;100:817-823.
46. Apostolopoulou M, Ligon L. Cadherin-23 mediates heterotypic cell-cell adhesion between breast cancer epithelial cells and fibroblasts. PLoS One. 2012;7:e33289.
47. Toma C, Díaz-Gay M, Franch-Expósito S, et al. Using linkage studies combined with whole-exome sequencing to identify novel candidate genes for familial colorectal cancer. Int J Cancer. 2020;146:1568-1577.
48. Greenman C, Stephens P, Smith R, et al. Patterns of somatic mutation in human cancer genomes. Nature. 2007;446:153-158.
49. Ma R, Jing C, Zhang Y, et al. The somatic mutation landscape of Chinese colorectal cancer. J Cancer. 2020;11:1038-1046.
50. Rahman N. Realizing the promise of cancer predisposition genes. Nature. 2014;505:302-308.
51. Antoniou AC, Casadel S, Heikkinen T, et al. Breast-cancer risk in families with mutations in PALB2. N Engl J Med. 2014;371:497-506.
52. Phelan CM, Igbal J, Lynch HT, et al. Incidence of colorectal cancer in BRCA1 and BRCA2 mutation carriers: results from a follow-up study. Br J Cancer. 2014;110:530-534.
53. Yang X, Leslie G, Doroszuk A, et al. Cancer risks associated with Germline PALB2 pathogenic variants: an international study of 524 families. J Clin Oncol. 2020;38:674-685.
54. Yurgelun MB, Kulke MH, Fuchs CS, et al. Cancer susceptibility gene mutations in individuals with colorectal cancer. J Clin Oncol. 2017;35:1086-1095.
55. AlDubayan SH, Giannakis M, Moore ND, et al. Inherited DNA-repair defects in colorectal cancer. Am J Hum Genet. 2018;102:401-414.

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

How to cite this article: Helgadottir HT, Thutkawkorapin J, Rohlín A, Nordling M, Lagerstedt-Robinson K, Lindblom A. Identification of known and novel familial cancer genes in Swedish colorectal cancer families. Int. J. Cancer. 2021;149:627–634. https://doi.org/10.1002/ijc.33567