

Open

Whole-exome sequencing identifies rare pathogenic variants in new predisposition genes for familial colorectal cancer

Clara Esteban-Jurado, MSc¹, Maria Vila-Casadesús, MSc², Pilar Garre, MD, PhD³, Juan José Lozano, PhD², Anna Pristoupilova, MSc^{4,5}, Sergi Beltran, PhD⁴, Jenifer Muñoz, MSc¹, Teresa Ocaña, MSc¹, Francesc Balaguer, MD, PhD¹, Maria López-Cerón, MD, PhD¹, Miriam Cuatrecasas, MD, PhD⁶, Sebastià Franch-Expósito, MSc¹, Josep M. Piqué, MD, PhD¹, Antoni Castells, MD, PhD¹, Angel Carracedo, MD, PhD⁷, Clara Ruiz-Ponte, PhD⁷, Anna Abulí, PhD⁸, Xavier Bessa, MD, PhD⁸, Montserrat Andreu, MD, PhD⁸, the EPICOLON Consortium, Luis Bujanda, MD, PhD⁹, Trinidad Caldés, PhD³ and Sergi Castellví-Bel, PhD¹

Purpose: Colorectal cancer is an important cause of mortality in the developed world. Hereditary forms are due to germ-line mutations in *APC*, *MUTYH*, and the mismatch repair genes, but many cases present familial aggregation but an unknown inherited cause. The hypothesis of rare high-penetrance mutations in new genes is a likely explanation for the underlying predisposition in some of these familial cases.

Methods: Exome sequencing was performed in 43 patients with colorectal cancer from 29 families with strong disease aggregation without mutations in known hereditary colorectal cancer genes. Data analysis selected only very rare variants (0–0.1%), producing a putative loss of function and located in genes with a role compatible with cancer. Variants in genes previously involved in hereditary colorectal cancer or nearby previous colorectal cancer genome-wide association study hits were also chosen.

Results: Twenty-eight final candidate variants were selected and validated by Sanger sequencing. Correct family segregation and somatic studies were used to categorize the most interesting variants in *CDKN1B*, *XRCCA4*, *EPHX1*, *NFKBIZ*, *SMARCA4*, and *BARD1*.

Conclusion: We identified new potential colorectal cancer predisposition variants in genes that have a role in cancer predisposition and are involved in DNA repair and the cell cycle, which supports their putative involvement in germ-line predisposition to this neoplasm.

Genet Med advance online publication 24 July 2014

Key Words: colorectal neoplasm; genetic variant; genetic predisposition to disease; hereditary disease; next-generation sequencing

INTRODUCTION

Colorectal cancer (CRC) is a very common disease, and its associated mortality rate is quite significant in the developed world. It is estimated that around 5% of the general population will be diagnosed with CRC. Also, as life expectancy increases, the number of CRC cases is also presumed to increase. As an illustrative example, there will be ~473,200 new CRC diagnoses and ~233,900 deaths related to this neoplasm in Europe during 2015.¹

Germ-line predisposition and environmental factors affect CRC susceptibility, as established for many other complex diseases. Importantly, the inherited germ-line contribution is known to influence about 35% of all cases.² Included in this previous group, the Mendelian CRC syndromes are the best characterized CRC cases because an inherited cause corresponds to

5% of total CRC cases. Lynch syndrome and familial adenomatous polyposis are the most frequent forms of Mendelian CRC syndromes. Classic hereditary CRC syndromes are mainly due to germ-line mutations in *APC*, *MUTYH*, and the mismatch repair genes (i.e., *MLH1*, *MSH2*, *MSH6*, *PMS2*).^{3,4} Finding the causative mutation in familial CRC also has implications that apply to genetic counseling practices that are of critical importance for the analyzed family. Once it is established in a particular family which individuals are carriers and which are noncarriers, prevention strategies can be directed more precisely to those individuals carrying the causative mutation and who are therefore at risk of developing CRC and other related malignancies. On the other hand, noncarriers can be spared excessive clinical monitoring.

¹Servei de Gastroenterologia, Hospital Clínic, Institut d'Investigacions Biomèdiques August Pi i Sunyer, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas, University of Barcelona, Barcelona, Spain; ²Plataforma de Bioinformática, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas, Barcelona, Spain; ³Molecular Oncology Laboratory, Hospital Clínico San Carlos, Instituto de Investigación Sanitaria del Hospital Clínico San Carlos, Madrid, Spain; ⁴Centre Nacional d'Anàlisi Genòmica, Parc Científic de Barcelona, Barcelona, Spain; ⁵Institute of Inherited Metabolic Disorders, First Faculty of Medicine, Charles University in Prague, Prague, Czech Republic; ⁶Department of Pathology, Hospital Clínic, Barcelona, Spain; ⁷Galician Public Foundation of Genomic Medicine, Centro de Investigación Biomédica en Red de Enfermedades Raras, Genomics Medicine Group, Hospital Clínic, University of Santiago de Compostela, Galicia, Spain; ⁸Department of Gastroenterology, Hospital del Mar-IMIM (Hospital del Mar Medical Research Centre), Pompeu Fabra University, Barcelona, Spain; ⁹Gastroenterology Department, Hospital Donostia – Instituto Biodonostia, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas, Country University (UPV/EHU), San Sebastián, Spain. Correspondence: Sergi Castellví-Bel (sbel@clinic.ub.es)

Submitted 8 April 2014; accepted 10 June 2014; advance online publication 24 July 2014. doi:10.1038/gim.2014.89

In addition to hereditary forms, around 30% of CRC cases also present familial aggregation but an unknown inherited cause. Among these cases, familial CRC type X can be used as an example in which the clinical criteria of Lynch syndrome are fulfilled but no alteration of the mismatch repair system is found.⁵ Then, the hypothesis of rare high-penetrance mutations in genes yet to be discovered is a very likely explanation for the underlying predisposition in a portion of these familial CRC cases. Therefore, past efforts in this direction included some low-throughput sequencing studies in familial CRC cases of some plausible candidate genes such as *EPHB2*, *GALNT12*, *PTPRJ*, *BMP4*, and *BMPRIA*.^{6–10} Next-generation sequencing technologies added a new unbiased approach to facilitate the identification of new genes responsible for predisposition to

human disease. Palles *et al.*¹¹ recently reported the identification of germ-line mutations in the *POLE* and *POLD1* genes in individuals with multiple colorectal adenomas, carcinoma, or both, or early onset of this disease using whole-genome sequencing. Smith *et al.*¹² recently performed exome sequencing in a cohort of patients with sporadic CRC enriched for early onset, and variants in genes showing biallelic inactivation were selected. In addition, exome sequencing was completed in CRC familial cases and shared variants were selected within families in an additional study.¹³ Finally, a Finnish cohort of familial CRC was also sequenced in order to find rare truncating variants present in two or more cases.¹⁴

Accordingly, the aim of our study was to find rare predisposition variants in new genes by performing exome sequencing in patients with familial CRC compatible with an autosomal dominant inheritance and without an alteration in the previously known hereditary CRC genes. In doing so, our final goal is to facilitate genetic counseling and to be able to correctly address prevention strategies in these families.

MATERIALS AND METHODS

Patients

Forty-three CRC patients from 29 families with strong CRC aggregation compatible with an autosomal dominant pattern of inheritance were selected. Alterations in *APC* or the mismatch repair genes, and homozygous or compound heterozygous mutations in *MUTYH* were previously excluded. Families were chosen based on the following criteria: three or more relatives with CRC, two or more consecutive affected generations, and at least one case of CRC diagnosed before the age of 60. In two families, advanced adenomas (*i.e.*, size ≥ 1 cm, villous architecture, or high-grade dysplasia) were taken into account as early disease presentation. In addition, other extracolonic cancers were considered in six families. Fourteen families were collected in high-risk CRC clinics (Hospital Clínico San Carlos in Madrid, Hospital Clinic in Barcelona, and Hospital Donostia in San Sebastián), and two patients with CRC were selected to be sequenced from among available affected individuals, preferentially those most distantly related. On the other hand, 15 families were chosen from the EPICOLON Consortium¹⁵ and one patient with CRC per family was selected to be sequenced. This study was approved by the institutional ethics committee of each participating hospital. Written informed consent was obtained at CRC diagnosis on a systematic basis.

Germ-line DNA samples used for exome sequencing were obtained from peripheral blood, whereas formalin-fixed, paraffin-embedded tumor DNA was isolated in some cases for loss of heterozygosity (LOH) studies using the QIAamp DNA Blood Kit or QIAamp Tissue Kit (Qiagen, Redwood City, CA) and following the manufacturer's instructions.

Exome sequencing

Quality control was applied to DNA samples (3–5 μg needed per reaction at a concentration of 50–300 ng/ μl measured

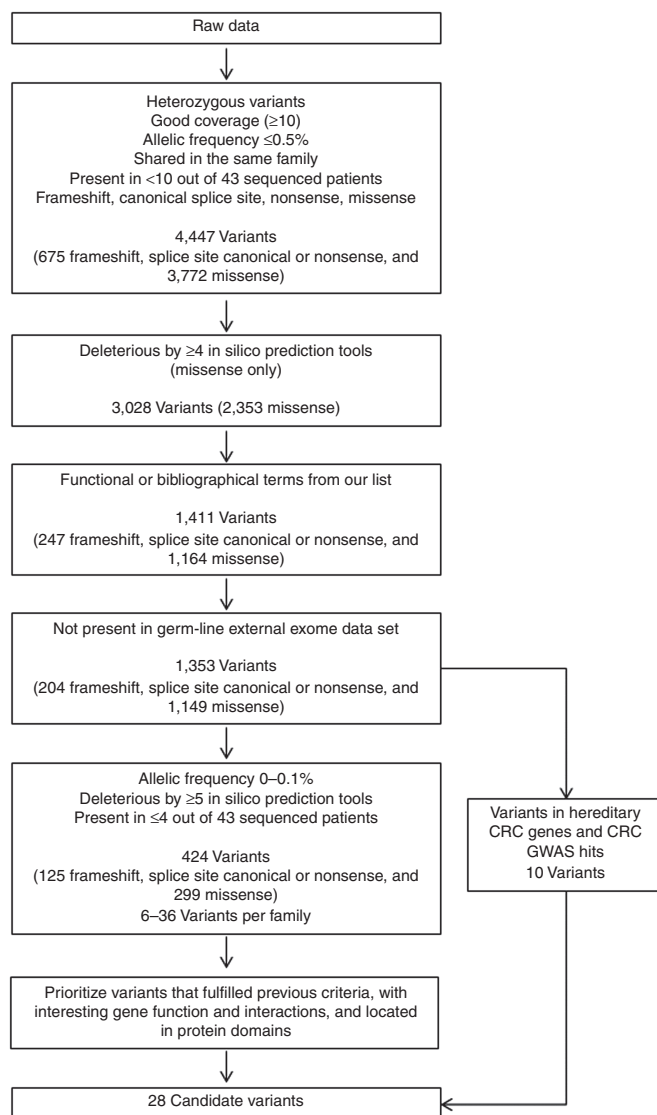


Figure 1 Schematic of the data analysis steps after whole-exome sequencing. Forty-three patients with colorectal cancer (CRC) from 29 families with strong CRC aggregation compatible with an autosomal dominant pattern of inheritance were sequenced. Variants remaining after each filtering step are indicated. GWAS, genome-wide association study.

by PicoGreen, A260/280 = 1.7–2, integrity check by agarose electrophoresis). The whole exome was characterized by using the HiSeq2000 platform (Illumina, San Diego, CA) and SureSelectXT Human All Exon V4 for exon enrichment (Agilent, Santa Clara, CA). Initial DNA shearing was performed using the Covaris S2 equipment, achieving an optimal range in the size distribution of fragments. Library size and concentration were checked by capillary electrophoresis (Bioanalyzer 2100; Agilent). Adapters with different indexes for each sample were incorporated during enrichment, allowing samples to be multiplexed before sequencing. After enrichment, the indexed libraries were pooled and massively parallel sequenced using a paired-end 2 × 75–base pair (bp) read length protocol.

Data analysis

Base calling and quality control were performed using the Real-Time Analysis software sequence pipeline (Illumina). Sequence reads were trimmed to keep only those bases with a quality >10 and then mapped to the human genome build (hg19/GRCh37) using Genome Multitool,¹⁶ allowing up to four mismatches. Reads not mapped by Genome Multitool were submitted to a last round of mapping with BLAT-like Fast Accurate Search Tool.¹⁷ Uniquely mapping nonduplicate read pairs were locally realigned with Genome Analysis Toolkit.¹⁸ The SAMtools suite (<http://samtools.sourceforge.net>) was used to call single-nucleotide variants and short insertions/deletions, taking into account all reads per position.¹⁹ Variants with high strand bias ($P > 0.001$ in at least one sample) or regions with low mappability (identified with the Genome Multitool mappability tool as having 75–bp reads and two mismatches)¹⁶ were filtered out. Variant annotation took into account data available in dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>), the 1000 Genomes Project (<http://www.1000genomes.org>), the Exome Variant Server (<http://evs.gs.washington.edu>), and the Geuvadis European Exome Variants Server (<http://geevs.org.eu>) and from an in-house database (100 whole genomes of Spanish ancestry from Centre Nacional d'Anàlisi Genòmica (<http://www.cnag.cat>)). Functional consequences of variants were also predicted by SnpEff (<http://snpeff.sourceforge.net>) (stop codon, frameshift, splicing, missense, synonymous), as well as by position (coding, intronic, exon–intron junction, untranslated regions). Regarding missense changes, six bioinformatic predictions for pathogenicity were available (PhyloP (<http://compugen.bscb.cornell.edu/phast/help-pages/phyloP.txt>), SIFT (Sorting Intolerant From Tolerant; <http://sift.bii.a-star.edu.sg>), PolyPhen (<http://genetics.bwh.harvard.edu/pph2>), MutationTaster (<http://www.mutationtaster.org>), GERP (Genomic Evolutionary Rate Profiling; <http://mendel.stanford.edu/SidowLab/downloads/gerp>), LRT (likelihood ratio test)).

Because a dominant inheritance pattern was expected, homozygous variants were removed, except for chromosome X non-pseudoautosomal regions in male samples. When analyzing two affected individuals from the same family, only shared variants were selected. Variants with low sequencing coverage (<10) and those with an allelic frequency $\geq 0.5\%$ in the 1000 Genomes

Project, Exome Variant Server, Geuvadis European Exome Variants Server, or the Centre Nacional d'Anàlisi Genòmica in-house database were filtered out. Variants present in >10 of the 43 individuals in our data set were discarded because they most likely corresponded to polymorphisms. Also, only variants predicted to have a strong effect on gene function (frameshift, splice-site canonical, nonsense, and missense) were chosen. Regarding missense variants, we used six bioinformatics tools to select for a deleterious amino acid change, namely, PhyloP (score >0.85), SIFT (score <0.05), PolyPhen (score >0.85), GERP (score >2), Mutation Taster (score >0.5), and LRT (score >0.9), and only those with four or more deleterious predictions were further considered.

Biological functions and pathways of the genes containing variants were annotated with terms and previous bibliography according to NCBI Gene (<http://www.ncbi.nlm.nih.gov/gene>), Gene Ontology (<http://www.geneontology.org/GO>), KEGG (<http://www.genome.jp/kegg/>), and Reactome (<http://www.reactome.org/PathwayBrowser/>). A list of cancer terms was created from these previous databases (**Supplementary Table S1** online) and used to select variants from among genes that had those terms annotated. All previous filters were performed using an automated pipeline encoded with R software (<http://CRAN.R-project.org>). CRC specificity of this pipeline regarding function and bibliography was tested by comparing our data with an external germ-line exome sequencing data set with equivalent coverage, which included the same number of patients with chronic lymphocytic leukemia from the International Cancer Genome Consortium (<https://www.icgc.org/>). Also, variants present in both data sets were filtered out.

Once a variant list per sequenced CRC patient was generated, a thorough manual annotation using NCBI Gene corroborated variant genome position and annotated protein interactions. The amino acid position of missense variants in functional domains, disulfide bonds, or posttranslational modifications was verified, as well as their effect on protein tridimensional structure, when available, using NCBI Protein (<http://www.ncbi.nlm.nih.gov/protein>) and UniProtKB (<http://www.uniprot.org/>). Also, their conservation in 46 vertebrates was checked (comparative alignment UCSC (<https://genome.ucsc.edu/>)).

Variant prioritization

Once all previous information was available, variant prioritization selected those variants more plausible to be causative of CRC genetic predisposition when they fulfilled more stringent criteria (0–0.1% allelic frequency; present in ≤ 4 individuals in our data set; ≥ 5 missense pathogenicity predictions; gene terms and bibliography compatible with cancer; interesting interactions and protein information; and amino acid species conservation). It is noteworthy that variants in genes previously involved in hereditary CRC were carefully checked, as were those genes near previous CRC genome-wide association studies (GWAS) hits (**Supplementary Table S2** online) with less strict criteria (missense considered deleterious by four or more bioinformatics tools). As previously specified, thresholds to select variants were

Table 1 Description of the final 28 prioritized variants, including gene category, frequency, and functional information

Gene	Mutation	Category	Genotype frequency (EVS)	In silico	LD region	Domain/region	Interactions	Function	OMIM
<i>CDKN1B</i>	c.195G>T(p.Q65H)	New	0/6503	5	N	Cyclin-dependent kinase inhibitor region	AKT1, CCND2, CDKN1A, TSC2, CDK2, CCNA1, CCNA2, CCNE1, CDK4, CCND1	DNA damage response, induction of apoptosis, negative regulation of cell proliferation	600778; Multiple endocrine neoplasia, hereditary prostate cancer
<i>XRCC4</i>	c.497_498delTG (p.V166Efs*3)	New	0/6503	FS	N	—	LIG4, APFL, APTX, BIN1, CAAP1, CHD3, DDX5, ERG, JUN, PRKDC	DNA ligation involved in DNA repair, double-strand break repair	194363; Peripheral B-cell lymphoma
<i>EPHX1</i>	c.293G>A(p.R98Q)	New	0/6503	6	N	Epoxide hydrolase N terminus region	CDKN2AIP, ATP4A, NRF1, SUMO1	Metabolism of xenobiotics by cytochrome P450	132810; Lymphoproliferative disorder
<i>NFKB1</i>	c.2153_2154dupAT (p.*719fs*10)	New	2/6258	FS	N	—	NFKB1, NFKB2, STAT3	TNF- α /NF- κ B signaling pathway	608004
<i>SMARCA4</i>	c.295C>T(p.R99W)	New	0/6503	5	N	Interaction with SS18L1/CREST	PMS2, CDKN1A, STK11, MYC, BRCA1, BCL6, TP53, BRCA1, IRF1, CREBBP	Regulation of cell growth, regulation of mitotic cell cycle, chromatin remodeling	603254; Rhabdoid tumor predisposition syndrome, SCCOHT (ref. 31)
<i>BARD1</i>	c.1811-2A>G	New	0/6503	SP	N	BRCT1 domain (exon 9 skipping)	MSH2, BRIP1, MLH1, MSH6, TP53, BRCA2, BRCA1	DNA repair, apoptosis, cell cycle arrest	601593; Breast cancer susceptibility, ovarian cancer susceptibility
<i>BRIP1</i>	c.1702_1703delAA (p.N568Wfs*9)	New	0/6503	FS	N	—	BARD1, PMS2, PMS1, MLH1, BRCA1, LMNA, BLM	Double-strand break repair, DNA damage checkpoint	605882; Breast cancer susceptibility, ovarian cancer susceptibility
<i>RB1</i>	c.491A>G(p.K164R)	New	0/6503	4	N	DUF3452 domain	CDKN1A, RUNX2, PIK3R1, BRCA1, EP300, MYC, SMARCA4, TP53	Negative regulation of cell growth and epithelial cell proliferation, cell cycle checkpoint	614041; Retinoblastoma, pinealoma, sarcoma, melanoma (ref. 43)
<i>AKR1C4</i>	c.667C>T(p.R223*)	New ^a	0/6503	NS	10p15.3-p15.1 (ref. 39)	—	BAG3	Metabolism of xenobiotics by cytochrome P450	600451; CRC
<i>CARD9</i>	c.1228G>T(p.E410*)	New	0/6503	NS	9q33.3-q34.3 (ref. 39)	—	BCL10, CSNK2A1	Regulation of apoptotic process	607212; Innate immunity
<i>NSMCE2</i>	c.346delIT(p.S116Lfs*18)	New	2/6249	FS	N	—	NSMCE1, RAD21, SMC1A, SMC3, SMC5, SMC6, STAG2, SUMO1	Double-strand break repair via homologous recombination, mitosis, telomere maintenance	—

CRC, colorectal cancer; GWAS, genome-wide association study; FS, frameshift; N, no; NS, nonsense; SP, canonical splice site; SCCOHT, ovarian small-cell carcinoma of hypercalcemic type.

^aCategory: new genes corresponded mainly to those not previously involved in CRC predisposition, except for *AKR1C4* (Smith et al.¹², Gylfe et al.¹⁴) and *ATM* (Gala et al.³⁸), with seminal evidence. ^bThis same variant was previously found in two juvenile polyposis patients (Sayed et al.⁴⁵ and Greenman et al.⁴⁶).

EVS: Exome Variant Server (version ESP6500SI-V2). 0/6503 indicates absence of the variant in the database. This data set is listed as representative of variant frequency. The 1000 Genomes Project and Geuvadis European Exome Variants Server were also checked but showed similar results.

In silico: number of deleterious predictions by bioinformatics tools used.

LD region: previously reported linkage disequilibrium regions for familial CRC are listed if the variant was located within.

Domain/region: the protein domain or region where the variant is located. If the position affects a disulfide bond or corresponds to a modified residue, the information is also listed.

Interactions: those reported with proteins with a role in cancer and other relevant partners.

OMIM: entry number (if available), hereditary involvement in malignancies and in some cases also somatic contribution.

Table 1 Continued on next page

Table 1 Continued

Gene	Mutation	Category	Genotype frequency (EVS)	In silico	LD region	Domain/region	Interactions	Function	OMIM
<i>BMPR1A</i>	c.1327C>T(p.R443C)	Hereditary	7/6496	6	N	Protein kinase domain	BMP4, SMAD1, RUNX2, SMURF1, MYC, BMP2, BMP7, BMP6, BMPR1B, BAMBI	Epithelial cell proliferation, cell differentiation, BMP signaling pathway	601299; Juvenile polyposis syndrome ^b , hereditary mixed polyposis syndrome, CRC
<i>CCDC18</i>	c.3268G>A(p.R1036H)	New ^a	1/5839	4	N	Coiled coil/SMC prok B region	UBC	Antigen found differentially expressed in sarcoma	—
<i>MYC</i>	c.906C>A(p.H302Q)	GWAS	4/6499	4	N	Myc amino-terminal region	BRCA1, CDKN1A, MLH1, BMPR1A, MSH2, POLE, POLD1, SMARCA4	Response to DNA damage stimulus, cell cycle arrest, Wnt signaling, TGF-β signaling	190080; Burkitt lymphoma, CRC, prostate cancer
<i>POLE</i>	c.5473A>T(p.H1810L)	Hereditary	0/6503	4	12q24.32 (ref. 40)	DUF1744 domain	MDM2, MYC, RAD17, PCNA, XRCC5	DNA repair, mitotic cell cycle, nucleotide excision repair, telomere maintenance	174762; CRC
<i>TSC2</i>	c.5116C>T(p.R1706C)	New	3/6495	6	N	Rap-GAP domain	AKT1, CDKN1B, GSK3B, MAPK1, CCND1, CCND2, CCND3	Negative regulation of cell proliferation, negative regulation of Wnt receptor signaling pathway	191092; Tuberous sclerosis, renal-cell cancer, angiomyolipoma, subependymal giant-cell astrocytoma, rhabdomyoma (ref. 43)
<i>RAD52</i>	c.590_593dupAACC (p.S199Tfs*88)	New	0/6503	FS	N	—	MCM2, ERCC4, RAD51	DNA repair, double-strand break repair via homologous recombination	600392
<i>BMP4</i>	c.1001C>T(p.A334V)	GWAS	0/6503	6	14q23.1–14q24.1 (ref. 41)	TGF-β region	BMPR1A, BMPR1B, BMPR2	Apoptosis, negative regulation of cell proliferation, BMP signaling pathway	CRC (ref. 9)
<i>DUSP4</i>	c.82G>A(p.A28T)	GWAS	0/6503	4	N	Regulatory rhodanese domain of dual specificity phosphatases	MAPK1, MAPK14, MAPK3, MAPK8, MAPK9	Inactivation of MAPK activity, Toll signaling pathway	602747

CRC, colorectal cancer; GWAS, genome-wide association study; FS, frameshift; N, no; NS, nonsense; SP, canonical splice site; SCCOHT, ovarian small-cell carcinoma of hypercalcemic type.

^aCategory: new genes corresponded mainly to those not previously involved in CRC predisposition, except for *AKR1C4* (Smith et al.¹², Gyffe et al.¹⁴), *CCDC18* (Gyffe et al.¹⁴) and *ATM* (Gala et al.³⁸), with seminal evidence. ^bThis same variant was previously found in two juvenile polyposis patients (Sayed et al.⁴⁵ and Greenman et al.⁴⁶).

EVS: Exome Variant Server (version ESP6500SI-V2). 0/6503 indicates absence of the variant in the database. This data set is listed as representative of variant frequency. The 1000 Genomes Project and Geuvadis European Exome Variants Server were also checked but showed similar results.

In silico: number of deleterious predictions by bioinformatics tools used.

LD region: previously reported linkage disequilibrium regions for familial CRC are listed if the variant was located within.

Domain/region: the protein domain or region where the variant is located. If the position affects a disulfide bond or corresponds to a modified residue, the information is also listed.

Interactions: those reported with proteins with a role in cancer and other relevant partners.

OMIM: entry number (if available), hereditary involvement in malignancies and in some cases also somatic contribution.

Table 1 Continued on next page

Table 1 Continued

Gene	Mutation	Category	Genotype frequency (EVS)	In silico	LD region	Domain/region	Interactions	Function	OMIM
LAMA5	c.4930C>T(p.R1644C)	GWAS	0/6503	6	20q13.11-13.33 (ref. 42)	Laminin IV type A domain	LAMC1, SMAD2	Regulation of cell proliferation, cell differentiation, regulation of cell adhesion	601033
CETN2	c3+2T>C	New	0/6503	SP	N	—	PARP1, POU5F1, PRMT6, RAD23B, SF11, USP39	Mitosis, centriole replication, regulation of cytokinesis	300006
MAP9	c.681dupA (p.A2285fs*4)	New	0/6503	FS	N	—	EP300, TP53, SH3KBP1	Cell cycle	610070
ENG	c.934C>T(p.A312V)	Hereditary	0/6503	4	9q33.3-q34.3 (ref. 39)	Interaction with EGL	TGFBR1, BMP7, BMP2	Regulation of cell proliferation, regulation of cell adhesion, BMP signaling pathway	Hamartomatous or hyperplastic/mixed polyposis (ref. 44)
RUNX2	c.622T>C(p.P208S)	New	0/6503	6	N	Runt domain	BMPR1A, EP300, RB1, SMURF2, SMURF1, JUN, SMAD2, SMAD3, SMAD6, XRCC6	BMP signaling pathway	600211; Osteosarcoma
PIK3R1	c.220G>A(p.V74I)	New	0/6503	6	N	SH3 domain	AKT1, PIK3CA, EGFR, BRCA1, EP300, GSK3B, TGFBR1	Insulin receptor signaling pathway, EGF receptor signaling pathway, focal adhesion	171833; Glioblastoma
SMURF2	c.838C>T(p.H280Y)	New	0/6503	5	N	WW2 domain	SMAD7, GSK3B, RUNX2, TGFBR1, SMAD1, SMAD2, SMAD3, SMAD5, SMAD6, NEDD4	BMP signaling pathway, TGF-β-signaling pathway	605532; Different tumor types
ATM	c.8327T>C(p.I2776T)	New ^a	0/6503	6	11q22.3 (ref. 39)	PI3K catalytic domain	MLH1, MSH2, MSH6, STK11, SMAD7, BRCA2, BRCA1, TP53	DNA repair, cell cycle arrest, mitotic cell cycle spindle assembly checkpoint	607585; Ataxia-telangiectasia lymphoma, familial breast cancer, kidney cancer, multiple sessile serrated adenomas
DHX9	c.286T>G(p.V40G)	New	0/6503	6	N	DSRM1 domain	POLD1, AKT1, BRCA1, JUN, PRKDC	CRD-mediated mRNA stabilization, RNA splicing, innate immune response	603115

CRC, colorectal cancer; GWAS, genome-wide association study; FS, frameshift; N, no; NS, nonsense; SP, canonical splice site; SSCOHT, ovarian small-cell carcinoma of hypercalcemic type.

^aCategory: new genes corresponded mainly to those not previously involved in CRC predisposition, except for *AKR1C4* (Smith et al.¹², Gylfe et al.¹⁴), *CCDC18* (Gylfe et al.¹⁴) and *ATM* (Gala et al.³⁸), with seminal evidence. ^bThis same variant was previously found in two juvenile polyposis patients (Sayed et al.⁴⁵ and Greenman et al.⁴⁶).

EVS: Exome Variant Server (version ESP6500SI-V2). 0/6503 indicates absence of the variant in the database. This data set is listed as representative of variant frequency. The 1000 Genomes Project and Geuvadis European Exome Variants Server were also checked but showed similar results.

In silico: number of deleterious predictions by bioinformatics tools used.

LD region: previously reported linkage disequilibrium regions for familial CRC are listed if the variant was located within.

Domain/region: the protein domain or region where the variant is located. If the position affects a disulfide bond or corresponds to a modified residue, the information is also listed.

Interactions: those reported with proteins with a role in cancer and other relevant partners.

OMIM: entry number (if available), hereditary involvement in malignancies and in some cases also somatic contribution.

applied for sequencing coverage, allelic frequency, presence in our data set, predictions by bioinformatics tools, presence in the functional and bibliography term list, and absence in the external exome set. On the other hand, there were no thresholds for some other additional variant/gene information that was used if available to further select for variants present within each family. This information included protein function and interactions; amino acid position in functional domains, disulfide bonds, or posttranslational modification sites; effect on protein tridimensional structure; and amino acid species conservation. Therefore, variants also complying with these last criteria were considered more interesting functionally and were further selected as final candidates. Some studied CRC families had up to four variants prioritized, whereas other families had none.

Variant validation, segregation analysis, and tumor loss of heterozygosis

Exome sequencing results for prioritized variants were validated using specific primers for polymerase chain reaction amplification designed using Primer3Plus (<http://primer3plus.com/cgi-bin/dev/primer3plus.cgi>) and Sanger sequencing (GATC Biotech, Cologne, Germany).

Segregation analysis of the prioritized variants was performed in additional family members (those with CRC and advanced adenoma) when germ-line DNA was available. When possible, somatic LOH was studied in tumor DNA of patients carrying the selected variants. LOH was tested by comparing Sanger sequencing results for germ-line and tumor DNA of the same individual. In addition, microsatellite markers within and around the gene of interest were used when LOH of the wild-type allele was suspected. Sanger and microsatellite markers results were always concordant. Primer details are listed in **Supplementary Table S3** online.

Network analysis

Ingenuity Pathway Analysis (IPA; Qiagen; <http://www.qiagen.com/ingenuity>) was used to perform a core analysis to check the putative enrichment for canonical pathways, disease and biological functions, and molecular networks among the 18 final candidate genes carrying variants that either fulfilled CRC family segregation or could not be tested (variants without correct family segregation were not included). IPA was run with an experimentally observed filter, aiming to obtain information based on confirmed data. The IPA networks generation algorithm transformed the gene list into a network set using Global Molecular Network connections and Ingenuity Pathways Knowledge Base.

RESULTS

Whole-exome sequencing was performed in 43 patients with CRC from 29 families (2 affected relatives from 14 families and 15 unrelated patients with CRC) with strong disease aggregation compatible with an autosomal dominant pattern of inheritance but without mutations in known hereditary CRC.

After sequencing, mean coverage was $>95\times$ in all samples. Raw data were analyzed using an automatic pipeline that selected only very rare variants (0–0.1%) producing a putative loss of function and located in genes with a role compatible with cancer. Also, variants in genes previously involved in hereditary CRC or nearby previous CRC GWAS hits were prioritized (**Figure 1**). Initial filtering removed variants in homozygosis, those with low coverage, those not shared in the same family, those with a frequency $\geq 0.5\%$ and those present in ≥ 10 of the 43 individuals in our data set. On the other hand, frameshift, nonsense, canonical splice-site, and missense variants were selected (4,447 variants: 675 frameshift, splice-site canonical, or nonsense and 3,772 missense). When missense variants complying with most pathogenicity prediction tools (designated deleterious by at least 4 of 6 tools) were selected, 2,353 remained. Of these, 1,411 variants annotated with functional or bibliographical terms from our cancer list were selected.

CRC specificity of this pipeline regarding function and bibliography was tested by comparing our set with an external germ-line exome sequencing data set for a different disease. After applying frequency, heterozygosity, function, and bibliography filters, a *t* test was used to compare the mean number of frameshift, splice-site canonical, or nonsense variants per individual in the two exome data sets. In doing so, our pipeline selected more variants in our exome data set (mean_{CRC} = 41.87; mean_{external} = 34.05; $P = 3.75 \times 10^{-10}$), supporting the CRC specificity of our pipeline.

After checking the aforementioned pipeline specificity, we continued with variant filtering, and 1,353 variants that were not present in the external data set were further considered. At this stage, 10 variants in genes previously implicated in CRC predisposition and CRC GWAS hits that fulfilled previous criteria had been selected as final candidates. Stricter filtering was applied to prioritize variants in new genes, including allelic frequency 0–0.1%, presence in ≤ 4 of the 43 individuals in our data set, and compliance with most pathogenicity prediction tools for missense classification (designated deleterious by ≥ 5 tools), leaving 424 selected variants (125 frameshift, splice-site canonical, or nonsense and 299 missense), ranging from 6 to 36 variants per family. Filtering for the 10 variants in genes previously implicated in CRC predisposition and CRC GWAS hits was the same except for pathogenicity prediction tools for missense classification (designated deleterious by ≥ 4 tools). Finally, as previously specified, thresholds regarding sequencing coverage, allelic frequency, presence in our data set, prediction by bioinformatics tools, presence in functional and bibliography term lists, and absence from the external exome set were applied in order to select variants. On the other hand, there were no thresholds for some other additional variant/gene information that was used, if available, to further select for variants present within each family. This information included protein function and interactions; amino acid position in functional domains, disulfide bonds, or posttranslational modification sites; effect on protein tridimensional structure; and amino acid species conservation. Therefore, variants also complying with these last

Table 2 Results for the final 28 prioritized variants regarding Sanger validation, family segregation, and somatic status

Gene	Category	Mutation	Samples	Sanger	CRC seg	AA seg	LOH	COSMIC
<i>CDKN1B</i>	New	c.195G>T(p.Q65H)	FAM14	Y	Y 2/2	Y 2/2	Y	7/636 (1.1%)
<i>XRCC4</i>	New	c.497_498delTG(p.V166Efs*3)	FAM10	Y	Y 2/2		Y ^b	9/577 (1.56%)
<i>EPHX1</i>	New	c.293G>A(p.R98Q)	FAM11	Y	Y 2/2		Y ^b	18/576 (3.13%)
<i>NFKBIZ</i>	New	c.2153_2154dupAT (p.*719Ifs*10)	FAM8	Y	Y 4/4		N	11/576 (1.91%)
<i>SMARCA4</i>	New	c.295C>T(p.R99W)	FAM3	Y	Y 2/2	Y 1/1	N	60/646 (9.29%)
<i>BARD1</i>	New	c.1811-2A>G	H458	Y	Y 2/2		N	18/577 (3.12%)
<i>BRIP1</i>	New	c.1702_1703delAA (p.N568Wfs*9)	H463	Y	NA		Y ^b	28/636 (4.4%)
<i>RB1</i>	New	c.491A>G(p.K164R)	H466	Y	NA		Y	126/713 (17.67%)
<i>AKR1C4</i>	New ^a	c.667C>T(p.R223*)	H467	Y	NA		N	13/576 (2.26%)
<i>CARD9</i>	New	c.1228G>T(p.E410*)	H456	Y	NA		N	13/578 (2.25%)
<i>NSMCE2</i>	New	c.346delT(p.S116Lfs*18)	H465	Y	NA		N	3/602 (0.5%)
<i>BMPR1A</i>	Hereditary	c.1327C>T(p.R443C)	H460	Y	NA		N	23/757 (3.04%)
<i>CCDC18</i>	New ^a	c.3268G>A(p.R1036H)	H460	Y	NA		N	22/597 (3.69%)
<i>MYC</i>	GWAS	c.906C>A(p.H302Q)	H460	Y	NA		N	6/636 (0.94%)
<i>POLE</i>	Hereditary	c.5473A>T(p.H1810L)	H462	Y	NA		NA	40/577 (6.93%)
<i>TSC2</i>	New	c.5116C>T(p.R1706C)	I139	Y	NA		NA	28/635 (4.41%)
<i>RAD52</i>	New	c.590_593dupAACC (p.S199Tfs*88)	FAM13	Y	Y 3/3	N 0/1	N	9/577 (1.56%)
<i>BMP4</i>	GWAS	c.1001C>T(p.A334V)	FAM3	Y	Y 2/2	N 0/1	N	8/577 (1.39%)
<i>DUSP4</i>	GWAS	c.82G>A(p.A28T)	H468	Y	N 1/2			5/577 (0.87%)
<i>DUSP4</i>	GWAS	c.82G>A(p.A28T)	H466	Y	NA			5/577 (0.87%)
<i>LAMA5</i>	GWAS	c.4930C>T(p.R1644C)	FAM8	Y	N 3/4			51/576 (8.85%)
<i>CETN2</i>	New	c.3 + 2T>C	FAM8	Y	N 2/4			4/577 (0.69%)
<i>MAP9</i>	New	c.681dupA(p.A228Sfs*4)	FAM8	Y	N 2/4			25/576 (4.34%)
<i>ENG</i>	Hereditary	c.934C>T(p.A312V)	H469	Y	N 1/2			9/635 (1.42%)
<i>RUNX2</i>	New	c.622T>C(p.P208S)	FAM1	Y	N 2/3			13/577 (2.25%)
<i>PIK3R1</i>	New	c.220G>A(p.V74I)	FAM4	Y	N 3/4			83/961 (8.64%)
<i>SMURF2</i>	New	c.838C>T(p.H280Y)	FAM12	Y	N 3/4			16/577 (2.77%)
<i>ATM</i>	New ^a	c.8327T>C(p.I2776T)	H470	Y	N 1/2			210/707 (29.7%)
<i>DHX9</i>	New	c.286T>G(p.V40G)	FAM2	N				21/576 (3.65%)

AA seg, advanced adenoma segregation; CRC seg, colorectal cancer segregation; GWAS, genome-wide association study; LOH, loss of heterozygosity; N, no; NA, not available; Y, yes.

^aCategory: new genes corresponded mainly to those not previously found to be involved in CRC predisposition, except for *AKR1C4* (Smith et al.¹²; Gylfe et al.¹⁴), *CCDC18* (Gylfe et al.¹⁴), and *ATM* (Gala et al.³⁸), with seminal evidence.

Samples: identifier for individuals in which the variant was found in exome sequencing. Samples with the header "FAM" belong to families with two sequenced individuals who share the same variant.

Sanger: refers to Sanger sequencing confirmation of the variant found by exome sequencing.

CRC seg: proportion of CRC cases within the family that carry the variant, including additional CRC cases when available.

AA seg: proportion of advanced adenomas cases within the family that carry the variant.

LOH: depletion of the wild-type allele in tumor DNA in comparison with the germ line.

COSMIC: proportion of somatic mutations found in large-intestine carcinoma for the gene carrying the variant (<http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/>).

^bIndicates results for both Sanger sequencing and microsatellite markers.

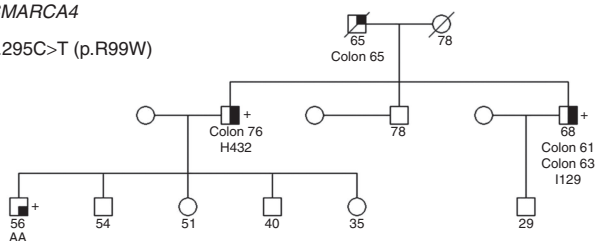
criteria were considered more interesting functionally and were further selected as final candidates. Some CRC families had up to four selected variants, whereas other families had none. The final 28 prioritized variants are shown in **Table 1**.

Candidate variants were subsequently validated by Sanger sequencing, and, if confirmed, segregation was studied in additional affected family members when available (**Table 2**). A variant in *DHX9* was not confirmed (1 of 28 prioritized variants).

FAM3 (I129, H432)

SMARCA4

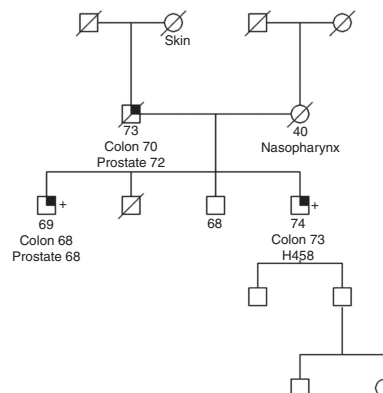
c.295C>T (p.R99W)



H458

BARD1

c.1811-2A>G

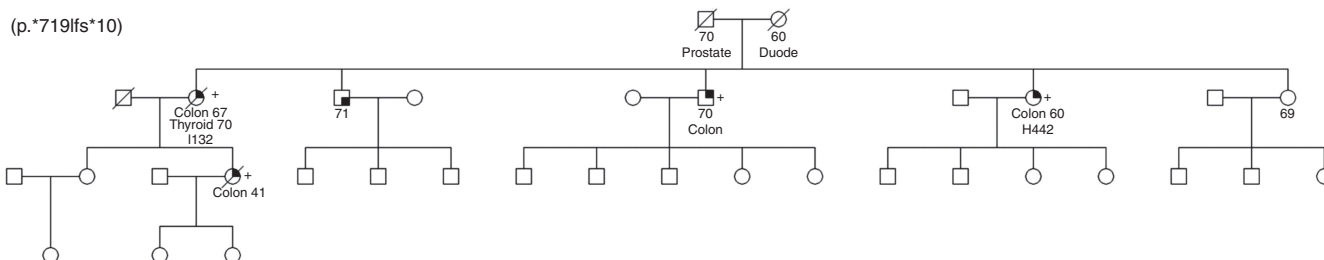


FAM3 (I132, H442)

NFKB1Z

c.2153_2154dupAT

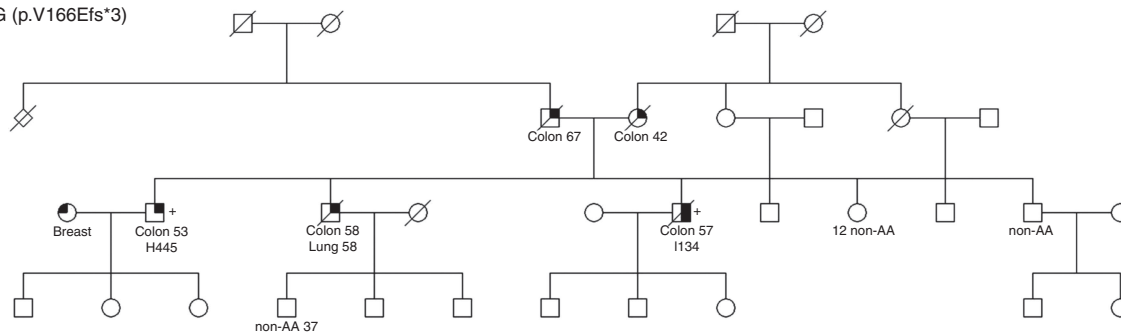
(p.*719lfs*10)



FAM10 (H445, I134)

XRCC4

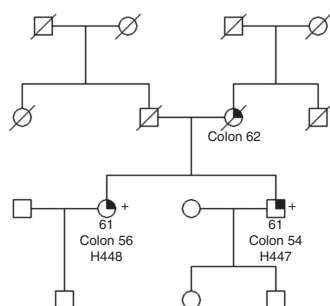
c.497_498delTG (p.V166Efs*3)



FAM11 (H447, H448)

EPHX1

c.293G>A (p.R98Q)



FAM14

CDKN1B

c.195G>T (p.Q65H)

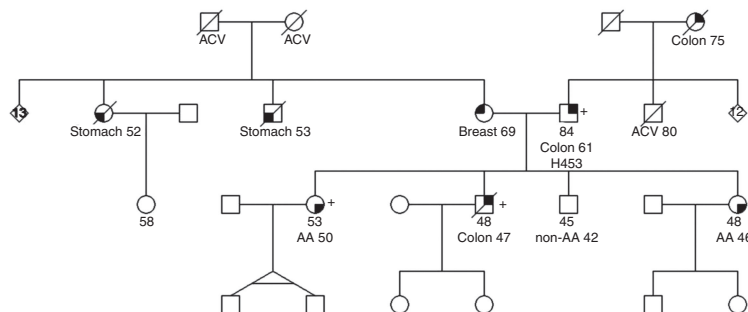


Figure 2 Pedigrees from families FAM3, FAM8, FAM10, FAM11, FAM14, and H458 are shown. Filled symbols indicate those affected by colorectal cancer (upper right quarter), adenoma(s) (lower right quarter), stomach cancer (lower left quarter), or breast cancer (upper left quarter). Colon, breast, stomach, thyroid, lung, prostate, and nasopharynx refer to the type of cancer. (+), mutation carrier; (-), wild type. AA, advanced adenoma; ACV, cerebrovascular accident; Duode, duodenum carcinoma; non-AA, nonadvanced adenoma.

LOH in tumor DNA was analyzed in variants with correct disease segregation when possible (**Table 2**; **Supplementary Figure S1** online). Among the 28 prioritized variants, the best candidates for being involved in CRC genetic predisposition included those located in genes such as *CDKN1B*, *XRCC4*, *EPHX1*, *NFKBIZ*, *SMARCA4*, and *BARD1* because they segregated correctly with disease presentation (**Figure 2**; the rest of families are shown in **Supplementary Figure S2** online). Regarding variants in these genes, it is expected that three of them abolish protein function and the other three are missense changes with strongly deleterious *in silico* predictions. Family segregation and tumor LOH of the wild-type allele was positive for variants in *CDKN1B*, *XRCC4*, and *EPHX1*. Other interesting variants were found in *BRIP1*, *RB1*, *AKR1C4*, *CARD9*, *NSMCE2*, *BMPR1A*, *CCDC18*, *MYC*, *POLE*, and *TSC2*, although segregation analysis was not feasible. Nevertheless, tumor LOH of the wild-type allele was present for the *BRIP1* and *RB1* variants. It is noteworthy that variants in *BMP4* and *RAD52* showed correct family segregation for CRC, but they did not correlate with advanced adenoma presentation, although they can still be considered interesting candidates. As reported in the COSMIC database, somatic mutations in sporadic CRC were more common for the *RB1*, *SMARCA4*, and *POLE* genes (**Table 2**). Candidate variants within genes previously implicated in CRC predisposition and CRC GWAS hits included those located in *AKR1C4*, *BMPR1A*, *CCDC18*, *MYC*, *POLE*, *BMP4*, *DUSP4* (present in two independent families), *LAMA5*, *ENG*, and *ATM*. The variant in the *BMP4* gene segregated with CRC but not with advanced adenoma. *DUSP4*, *LAMA5*, *ENG*, and *ATM* variants did not segregate with disease, whereas disease segregation could not be tested for variants in *AKR1C4*, *BMPR1A*, *CCDC18*, *MYC*, and *POLE*. The *POLE* variant did not correspond to those previously reported and did not fall within the exonuclease or polymerase domains.¹¹

In addition, we performed IPA to test for a putative enrichment for canonical pathways, disease and biological functions, and molecular networks among the 18 final candidate genes carrying variants that either fulfilled CRC family segregation or for which segregation analysis was not possible. A relevant network that contains 9 of the 18 genes was obtained with an overrepresentation of the DNA Replication, Recombination and Repair, Cell Cycle, Connective Tissue Development and Function terms (**Supplementary Figure S3a** online). On the other hand, when testing for canonical pathways in our set, the “Role of BRCA1 in DNA damage response” network included some of our more interesting candidates, such as *SMARCA4*, *BARD1*, *BRIP1*, and *RB1* (**Supplementary Figure S3b** online).

DISCUSSION

Exome sequencing in 43 patients with CRC from 29 families with strong disease aggregation identified new potential CRC predisposition variants in *CDKN1B*, *XRCC4*, *EPHX1*, *NFKBIZ*, *SMARCA4*, and *BARD1*.

CDKN1B (p27, Kip1) binds to cyclin E/A-CDK2 and cyclin D-CDK4 complexes and hinders their activation. By doing

so it exerts control on cell cycle progression.²⁰ The c.195G>T (p.Q65H) mutation is located inside the cyclin-dependent kinase inhibitor region, particularly in the β -hairpin (residues 61–71), which interacts with CDK2.²¹ Therefore, this variant most likely affects the normal interaction between *CDKN1B* and CDK2, causing a deregulation in cell cycle progression. Interestingly, germ-line mutations in this gene have been previously implicated in multiple endocrine neoplasia.²² Moreover, a polymorphism in this gene has been significantly associated with hereditary prostate cancer.²³

XRCC4 is involved in the repair of DNA double-strand breaks by nonhomologous end joining and the completion of V(D)J recombination events, along with DNA ligase IV and the DNA-dependent protein kinase.²⁴ The c.497_498delTG (p.V166Efs*3) mutation is predicted to abolish protein function, and it is likely to contribute to genomic instability and tumorigenesis.

The *EPHX1* enzyme converts epoxides produced by the degradation of aromatic compounds to trans-dihydrodiols, which afterward are conjugated and excreted from the body. Thus, *EPHX1* can be considered an important biotransformation protein.²⁵ The affected residue of the c.293G>A (p.R98Q) mutation is located in the epoxide hydrolase *N*-terminus region. Because *EPHX1* alleles can have a differential efficiency in procarcinogen detoxification, it can be postulated that they may affect cancer risk in a specific manner.²⁶

NFKBIZ is involved in inflammatory response through regulation of nuclear factor- κ B transcription factor complexes.^{27,28} The c.2153_2154dupAT (p.*719Ifs*10) mutation disrupts a stop codon, producing an abnormally long C-terminal region. This could affect the interactions with nuclear factor- κ B complexes that bind to that region, altering the transcriptional regulation of its target genes and leading to cancer predisposition.

The *SMARCA4* protein is a component in the large SNF/SWI complex involved in chromatin remodeling. This complex is necessary to activate the transcription of genes that are usually repressed by chromatin.²⁹ The c.295C>T (p.R99W) mutation is located in the region necessary for the interaction with SS18L1, which inhibits transcription of *c-FOS* and is required for dendritic growth and branching in cortical neurons. It can be hypothesized that this variant may cause predisposition to CRC by impairing this network and causing abnormal cell proliferation. Germ-line mutations in this gene can cause rhabdoid tumor predisposition syndrome type 2³⁰ and small-cell carcinoma of the ovary, hypercalcemic type.³¹

BARD1 interacts with the well-known *BRCA1* protein. Both proteins, along with others, participate in several cellular pathways involved in DNA damage repair, ubiquitination, and transcriptional regulation to preserve genomic stability.³² The c.1811-2A>G mutation is predicted to cause exon 9 skipping, disrupting the BRCT1 domain, which is postulated to participate in ligand binding according its structure.³³ This domain is highly homologous to the *BRCA1* BRCT1 domain, which is considered to bind substrates of DNA damage response kinases such as *ATM*. Moreover, tumor-associated mutations in the

BRCT domains of BRCA1 abolish binding to phosphorylated substrates.³⁴ Thus, disruption of this BARD1 domain likely affects its capacity to interact with other proteins, abolishing its tumor suppressor function. Germ-line mutations in this gene predispose to breast and ovarian cancer,³⁵ and its expression has been involved in differential CRC prognosis.³⁶

Focusing on the best candidates to be involved in CRC genetic predisposition (*CDKN1B*, *XRCC4*, *EPHX1*, *NFKB1Z*, *SMARCA4*, and *BARD1*), it is remarkable that, as highlighted by the IPA analysis and previous studies, most of them have been formerly involved in DNA repair, cell cycle, and predisposition to germ-line cancer, which supports their putative involvement in genetic predisposition to CRC as well. Among them, mutated *BARD1* and *BRIP1* have been found in the germ-line DNA of breast cancer patients described in several reports.^{35,37} In addition, DNA repair constitutes a cellular mechanism with proven importance in the genetic predisposition for CRC.³

Among those variants within genes previously involved in CRC predisposition or located in CRC GWAS hits, it is remarkable that so far three independent studies, including ours, have identified interesting variants in the *AKR1C4* gene.^{12,14}

Taken together, we could conclude that our results highlight some interesting candidates for CRC germ-line predisposition, with an overrepresentation of genes involved in DNA repair and the cell cycle. We identified several putative new genes predisposing to CRC and some with previous involvement in cancer predisposition, including *CDKN1B*, *XRCC4*, *EPHX1*, *NFKB1Z*, *SMARCA4*, and *BARD1*, that deserve to be considered in additional familial CRC cohorts with an unknown hereditary cause. Furthermore, once their role in hereditary CRC is confirmed, more complex functional studies would be warranted to help understand the molecular mechanism of disease predisposition.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at <http://www.nature.com/gim>

ACKNOWLEDGMENTS

We are sincerely grateful to the Centre Nacional d'Anàlisi Genòmica and the Biobank of Hospital Clínic-IDIBAPS, Barcelona, for technical help, and the International Cancer Genome Consortium for access to exome data set. The work was carried out (in part) at the Esther Koplowitz Centre, Barcelona. CEJ and JM are supported by a contract from CIBERehd. MVC is supported by Ministerio de Educación, Cultura y Deporte (FPU12/05138). PG and SCB are supported by a contract from the Fondo de Investigación Sanitaria (JR13/00013 and CP 03-0070, respectively). CIBERehd and CIBERER are funded by the Instituto de Salud Carlos III. This work was supported by grants from the Fondo de Investigación Sanitaria/FEDER (10/00641, 11/00219, 11/00681, RD12/0036/006, 13/02588), the Ministerio de Economía y Competitividad (SAF2010-19273), Fundación Científica de la Asociación Española contra el Cáncer (GCB13131592CAST), COST Action BM1206 (SCB and CRP), Beca Grupo de Trabajo "Oncología" AEG (Asociación Española de Gastroenterología),

and Agència de Gestió d'Ajuts Universitaris i de Recerca (Generalitat de Catalunya, 2014SGR255).

DISCLOSURE

The authors declare no conflict of interest.

REFERENCES

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010;127:2893–2917.
2. 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.
3. Jasperson KW, Tuohy TM, Neklason DW, Burt RW. Hereditary and familial colon cancer. *Gastroenterology* 2010;138:2044–2058.
4. Castells A, Castellví-Bel S, Balaguer F. Concepts in familial colorectal cancer: where do we stand and what is the future? *Gastroenterology* 2009;137:404–409.
5. Lindor NM, Rabe K, Petersen GM, et al. Lower cancer incidence in Amsterdam-I criteria families without mismatch repair deficiency: familial colorectal cancer type X. *JAMA* 2005;293:1979–1985.
6. Zogopoulos G, Jorgensen C, Bacani J, et al. Germline EPHB2 receptor variants in familial colorectal cancer. *PLoS One* 2008;3:e2885.
7. Guda K, Moinova H, He J, et al. Inactivating germ-line and somatic mutations in polypeptide N-acetylgalactosaminyltransferase 12 in human colon cancers. *Proc Natl Acad Sci USA* 2009;106:12921–12925.
8. Venkatchalam R, Ligtenberg MJ, Hoogerbrugge N, et al. Germline epigenetic silencing of the tumor suppressor gene PTPRJ in early-onset familial colorectal cancer. *Gastroenterology* 2010;139:2221–2224.
9. Lubbe SJ, Pittman AM, Matijssen C, et al. Evaluation of germline BMP4 mutation as a cause of colorectal cancer. *Hum Mutat* 2011;32:E1928–E1938.
10. Nieminen TT, Abdel-Rahman WM, Ristimäki A, et al. BMPR1A mutations in hereditary nonpolyposis colorectal cancer without mismatch repair deficiency. *Gastroenterology* 2011;141:e23–e26.
11. Palles C, Cazier JB, Howarth KM, et al.; CORGI Consortium; WGS500 Consortium. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat Genet* 2013;45:136–144.
12. Smith CG, Naven M, Harris R, et al. Exome resequencing identifies potential tumor-suppressor genes that predispose to colorectal cancer. *Hum Mutat* 2013;34:1026–1034.
13. DeRycke MS, Gunawardena SR, Middha S, et al. Identification of novel variants in colorectal cancer families by high-throughput exome sequencing. *Cancer Epidemiol Biomarkers Prev* 2013;22:1239–1251.
14. Gylfe AE, Katainen R, Kondelin J, et al. Eleven candidate susceptibility genes for common familial colorectal cancer. *PLoS Genet* 2013;9:e1003876.
15. Piñol V, Castells A, Andreu M, et al.; Gastrointestinal Oncology Group of the Spanish Gastroenterological Association. Accuracy of revised Bethesda guidelines, microsatellite instability, and immunohistochemistry for the identification of patients with hereditary nonpolyposis colorectal cancer. *JAMA* 2005;293:1986–1994.
16. Marco-Sola S, Sammeth M, Guigó R, Ribeca P. The GEM mapper: fast, accurate and versatile alignment by filtration. *Nat Methods* 2012;9:1185–1188.
17. Homer N, Merriman B, Nelson SF. BFAST: an alignment tool for large scale genome resequencing. *PLoS One* 2009;4:e7767.
18. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–1303.
19. 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.
20. Lee J, Kim SS. The function of p27 KIP1 during tumor development. *Exp Mol Med* 2009;41:765–771.
21. Russo AA, Jeffrey PD, Patten AK, Massagué J, Pavletich NP. Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex. *Nature* 1996;382:325–331.
22. Pellegata NS, Quintanilla-Martinez L, Siggelkow H, et al. Germ-line mutations in p27Kip1 cause a multiple endocrine neoplasia syndrome in rats and humans. *Proc Natl Acad Sci USA* 2006;103:15558–15563.

23. Chang BL, Zheng SL, Isaacs SD, et al. A polymorphism in the CDKN1B gene is associated with increased risk of hereditary prostate cancer. *Cancer Res* 2004;64:1997–1999.
24. Li Z, Otevrel T, Gao Y, et al. The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination. *Cell* 1995;83:1079–1089.
25. Fretland AJ, Omiecinski CJ. Epoxide hydrolases: biochemistry and molecular biology. *Chem Biol Interact* 2000;129:41–59.
26. Liu F, Yuan D, Wei Y, et al. Systematic review and meta-analysis of the relationship between EPHX1 polymorphisms and colorectal cancer risk. *PLoS One* 2012;7:e43821.
27. Totzke G, Essmann F, Pohlmann S, Lindenblatt C, Jänicke RU, Schulze-Osthoff K. A novel member of the I κ B family, human I κ B-zeta, inhibits transactivation of p65 and its DNA binding. *J Biol Chem* 2006;281:12645–12654.
28. Cowland JB, Muta T, Borregaard N. IL-1 β -specific up-regulation of neutrophil gelatinase-associated lipocalin is controlled by I κ B-zeta. *J Immunol* 2006;176:5559–5566.
29. Wilson BG, Roberts CW. SWI/SNF nucleosome remodellers and cancer. *Nat Rev Cancer* 2011;11:481–492.
30. Schneppenheim R, Frühwald MC, Gesk S, et al. Germline nonsense mutation and somatic inactivation of SMARCA4/BRG1 in a family with rhabdoid tumor predisposition syndrome. *Am J Hum Genet* 2010;86:279–284.
31. Witkowski L, Carrot-Zhang J, Albrecht S, et al. Germline and somatic SMARCA4 mutations characterize small cell carcinoma of the ovary, hypercalcemic type. *Nat Genet* 2014;46:438–443.
32. Irminger-Finger I, Jefford CE. Is there more to BARD1 than BRCA1? *Nat Rev Cancer* 2006;6:382–391.
33. Birrane G, Varma AK, Soni A, Ladias JA. Crystal structure of the BARD1 BRCT domains. *Biochemistry* 2007;46:7706–7712.
34. Manke IA, Lowery DM, Nguyen A, Yaffe MB. BRCT repeats as phosphopeptide-binding modules involved in protein targeting. *Science* 2003;302:636–639.
35. Ratajska M, Antoszewska E, Piskorz A, et al. Cancer predisposing BARD1 mutations in breast-ovarian cancer families. *Breast Cancer Res Treat* 2012;131:89–97.
36. Sporn JC, Hothorn T, Jung B. BARD1 expression predicts outcome in colon cancer. *Clin Cancer Res* 2011;17:5451–5462.
37. Cantor SB, Bell DV, Ganesan S, et al. BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. *Cell* 2001;105:149–160.
38. Gala MK, Mizukami Y, Le LP, et al. Germline mutations in oncogene-induced senescence pathways are associated with multiple sessile serrated adenomas. *Gastroenterology* 2014;146:520–529.
39. Saunders IW, Ross J, Macrae F, et al. Evidence of linkage to chromosomes 10p15.3-p15.1, 14q24.3-q31.1 and 9q33.3-q34.3 in non-syndromic colorectal cancer families. *Eur J Hum Genet* 2012;20:91–96.
40. Cicek MS, Cunningham JM, Fridley BL, et al. Colorectal cancer linkage on chromosomes 4q21, 8q13, 12q24, and 15q22. *PLoS One* 2012;7:e38175.
41. Djureinovic T, Skoglund J, Vandrovцова J, et al. A genome wide linkage analysis in Swedish families with hereditary non-familial adenomatous polyposis/non-hereditary non-polyposis colorectal cancer. *Gut* 2006;55:362–366.
42. Laiho P, Hienonen T, Karhu A, et al. Genome-wide allelotyping of 104 Finnish colorectal cancers reveals an excess of allelic imbalance in chromosome 20q in familial cases. *Oncogene* 2003;22:2206–2214.
43. Rahman N. Realizing the promise of cancer predisposition genes. *Nature* 2014;505:302–308.
44. Ngeow J, Heald B, Rybicki LA, et al. Prevalence of germline PTEN, BMPR1A, SMAD4, STK11, and ENG mutations in patients with moderate-load colorectal polyps. *Gastroenterol* 2013;144:1402–1409.
45. Sayed MG, Ahmed AF, Anderson ME, et al. Germline SMAD4 or BMPR1A mutations and phenotype of juvenile polyposis. *Ann Surg Oncol* 2002;9:901–906.
46. Greenman C, Stephens P, Smith R, et al. Patterns of somatic mutation in human cancer genomes. *Nature* 2007;446:153–158.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>