Identification of a Novel Candidate Gene for Serrated Polyposis Syndrome Germline Predisposition by Performing Linkage Analysis Combined With Whole-Exome Sequencing

Claudio Toma, PhD[1,2], Marcos Díaz-Gay, MSc[3], Yasmin Soares de Lima, MSc[3], Coral Arnau-Collell, MSc[3], Sebastià Franch-Expósito, PhD[3], Jenifer Muñoz, BSc[3], Bronwyn Overs, MSc[1,2], Laia Bonjoch, PhD[3], Sabela Carballal, MD, PhD[3], Teresa Ocaña, BSc[3], Miriam Cuatrecasas, MD, PhD[4], Aránzazu Díaz de Bustamante, PhD[5], Antoni Castells, MD, PhD[3], Luis Bujanda, MD, PhD[6], Joaquín Cubiella, MD, PhD[7], Francesc Balaguer, MD, PhD[3], Daniel Rodríguez-Alcalde, MD[8], Janice M. Fullerton, PhD[1,2] and Sergi Castellví-Bel, PhD[3]

OBJECTIVES: **Serrated polyposis syndrome (SPS) is a complex disorder with a high risk of colorectal cancer for which the germline factors remain largely unknown. Here, we combined whole-exome sequencing (WES) and linkage studies in families with multiple members affected by SPS to identify candidate genes harboring rare variants with higher penetrance effects.**

METHODS: **Thirty-nine affected subjects from 16 extended SPS families underwent WES. Genome-wide linkage analysis was performed under linear and exponential models. The contribution of rare coding variants selected to be highly pathogenic was assessed using the gene-based segregation test.**

RESULTS: **A significant linkage peak was identified on chromosome 3p25.2-p22.3 (maxSNP = rs2293787; LOD_{linear} = 2.311, LOD_{exp} = 2.11), which logarithm of the odds (LOD) score increased after fine mapping for the same marker (maxSNP = rs2293787; LOD_{linear} = 2.4, LOD_{exp} = 2.25). This linkage signal was replicated in 10 independent sets of random markers from this locus. To assess the contribution of rare variants predicted to be pathogenic, we performed a family-based segregation test with 11 rare variants predicted to be deleterious from 10 genes under the linkage intervals. This analysis showed significant segregation of rare variants with SPS in *CAPT7*, *TMEM43*, *NGLY1*, and *FBLN2* genes (weighted *P* value > 0.007).**

DISCUSSION: **Protein network analysis suggested *FBLN2* as the most plausible candidate genes for germline SPS predisposition. Etiologic rare variants implicated in disease predisposition may be identified by combining traditional linkage with WES data. This powerful approach was effective for the identification of a new candidate gene for hereditary SPS.**

SUPPLEMENTARY MATERIAL accompanies this paper at http://links.lww.com/CTG/A114

Clinical and Translational Gastroenterology 2019;10:e00100. https://doi.org/10.14309/ctg.0000000000000100

[1]Neuroscience Research Australia, Sydney, Australia; [2]School of Medical Sciences, University of New South Wales, Sydney, Australia; [3]Gastroenterology Department, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Hospital Clínic, University of Barcelona, Barcelona, Spain; [4]Pathology Department, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd) and Tumor Bank-Biobank, Hospital Clínic, Barcelona, Spain; [5]Genetics Unit, Hospital Universitario de Móstoles, Madrid, Spain; [6]Gastroenterology Department, Hospital Donostia-Instituto Biodonostia, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Basque Country University (UPV/EHU), San Sebastián, Spain; [7]Gastroenterology Department, Complexo Hospitalario Universitario de Ourense, Instituto de Investigación Sanitaria Galicia Sur, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Ourense, Spain; [8]Digestive Disease Section, Hospital Universitario de Móstoles, Móstoles, Spain. **Correspondence:** Sergi Castellví-Bel, PhD. E-mail: sbel@clinic.cat.

Received May 2, 2019; accepted September 19, 2019; published online October 28, 2019

INTRODUCTION

Colorectal cancer (CRC) is one of the most frequent neoplasms and an important cause of mortality in the developed world. Approximately 5% of the population develops CRC, and this figure is expected to rise as life expectancy increases. For 2018, over 1.8 million new CRC cases and 881,000 related-deaths were estimated to occur in the world (1).

Most CRCs progress through the adenoma-carcinoma sequence (2). The malignant transformation of conventional adenomas was considered for decades to be the single mechanism underlying the genesis of CRC. Serrated polyps, also previously known as hyperplastic polyps, were first described in the early seventies (3). These polyps exhibit a serrated architecture with infolding of colonocytes in the lumen of the crypts showing a stellate or star-like appearance on cross section and a saw-toothed form in a longitudinal section. Nowadays, they are also considered CRC precursor lesions and an alternative pathway to CRC (4).

Serrated polyposis syndrome (SPS) is a disease with unknown inherited genetic basis characterized by the presence of multiple and/or large serrated polyps in the colon and a high risk of CRC. Although previously considered an uncommon entity, recent evidence estimates that is one of the most frequent polyposis syndromes (5). The following criteria were established by the World Health Organization in 2010 (6) to help identifying this clinical entity(1): at least 5 serrated polyps proximal to the sigmoid colon with 2 or more of these being >10 mm (2), any number of serrated polyps proximal to the sigmoid colon in an individual who has a first-degree relative with serrated polyposis, or (3) >20 serrated polyps of any size, but distributed throughout the colon. Accordingly, SPS is a very heterogeneous entity, with different phenotypes. Very recently, a new set of SPS criteria has been published, which do not consider any longer criterion number 2 (7).

Smoking and obesity have been associated as environmental factors with an increased risk of developing serrated polyps (4). Actually, most patients with SPS are diagnosed over the age of 60 years, with no significant family history of CRC or SPS. Family aggregation of SPS has been described in <5% of cases. Regarding SPS germline predisposition, both dominant and recessive models have been suggested, although the hereditary genetic basis for this clinical entity remains mostly unidentified. Rarely, serrated polyposis has been found concomitantly with hereditary forms caused by mutations in *APC* and *MUTYH* (8,9). The only proposed gene so far for germline predisposition to SPS has been *RNF43* although with controversy (10,11).

Recently, there have been several efforts to identify additional genetic factors that predispose to CRC using linkage analysis. Such studies in affected families were able to pinpoint chromosomal regions of interest such as 9q22 and 3q22, but no clear CRC predisposition genes were identified after screening for interesting candidates within these areas (12,13). A linkage analysis study specifically for SPS concluded that 2q32.2-q33.3 was the most likely region to contain a predisposition gene for this disease (14).

On the other hand, next-generation sequencing (NGS) technologies have been recently established as a new approach to identify genes involved in disease predisposition (15), and whole-exome sequencing (WES) directed to genome coding regions (exons) has become the most fruitful application of NGS in translational research (16). Several studies using WES on familial CRC cohorts have proposed several candidate genes for germline predisposition but also have evidenced that the number of candidate variants is still too high and should be reduced by other means (17). Also, linkage analysis conducted with common polymorphisms present in the already available WES data has been postulated as a cost-effective strategy to tackle regions of linkage and focus on genes located on those areas (18,19).

In the present study, we combined WES and linkage analysis in 16 families with unaffiliated SPS aggregation. Available sequencing data from 39 patients were used to identify the more plausible candidate genes in the regions of positive linkage. By doing so, we aim to help identifying the rare, high-penetrance germline predisposition cause in these families and to facilitate genetic counseling and prevention strategies.

MATERIALS AND METHODS

Study participants

Sixteen families with at least 2 affected relatives with unaffiliated SPS/CRC aggregation compatible with an autosomal dominant pattern of inheritance were selected (Figure 1). The included patients with SPS fulfilled at least 1 of the World Health Organization (WHO) SPS criteria, except for 3 patients (in families 3, 10, and 11) with CRC but without information regarding the SPS criteria. Accordingly, 39 patients from these 16 SPS families were included. Their clinical characteristics are summarized in Table 1. This study was approved by the institutional ethics committee, and written informed consent was obtained in all cases.

WES

The entire cohort of 39 patients underwent germline WES similarly to previous reported studies performed by our research group in familial CRC (20–22). Briefly, WES was characterized using the HiSeq2000 platform (Illumina, San Diego, CA) and SureSelectXT Human All Exon V5 for exon enrichment (Agilent, Santa Clara, CA). Indexed libraries were pooled and massively parallel sequenced using a paired-end 2 × 75 bp read length protocol. Mean coverage was >95× in all samples, and 51 Mb was the target size that required about 4 Gb of sequencing per sample. Burrows-Wheeler Aligner was used for read mapping to the human reference genome (build hs37d5, based on NCBI GRCh37) (23). Polymerase chain reaction duplicates were discarded using the MarkDuplicates tool from Picard, and then indel realignment and base quality score recalibration were performed with the Genome Analysis Toolkit. The HaplotypeCaller Genome Analysis Toolkit tool was used for variant calling (24).

Linkage analysis procedures

Genotypes were called from aligned WES reads using SAMtools pileup and filtered to include haplotype-informative markers (HapMap CEU population) using LINKDATAGEN (25). WES-derived genotypes were used to confirm familial relationships by pairwise identity by descent using PLINK (26), and Z0, Z1, and Z2 values were obtained. All genotyped-derived genetic relationships were consistent with the clinical records.

A linkage study was performed using 5,726 WES-derived single nucleotide polymorphisms (SNPs) from 22 autosomal chromosomes (N = 5,566 SNPs) and the X chromosome (N = 160 SNPs) across all 16 families. Nonparametric linkage analyses were performed using the "all" statistic implemented in Merlin, under the Kong and Cox linear (logarithm of the odds

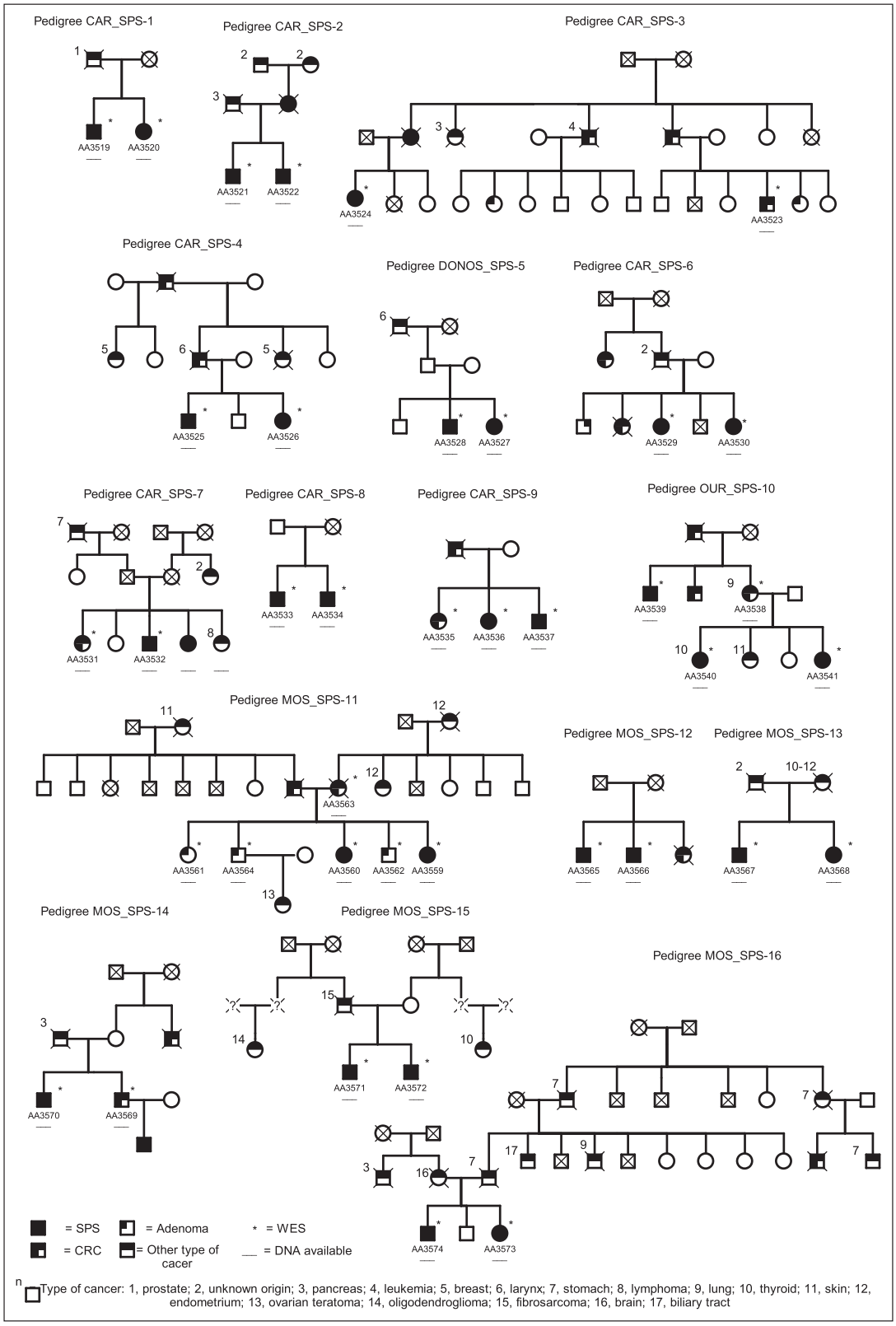

Figure 1. Pedigree structure of the 16 SPS multiplex and extended families examined in our study. Males are indicated with squares, females with circles, and diagnosis is shown by dark shading (full, patients diagnosed with SPS, white right quarter, patients diagnosed with CRC; left quarter, patients diagnosed with adenoma; half, patients diagnosed with any other type of cancer detailed in the figure legend; unshaded, unaffected individuals or unknown). Patients analyzed by WES are indicated by an asterisk, and all subjects with DNA available are underlined. CRC, colorectal cancer; SPS, serrated polyposis syndrome; WES, whole-exome sequencing.

Table 1. Clinical characteristics of 39 patients from 16 SPS families included in the reported study

FAM-ID	Ind-ID	Diagnosis	SPS criteria[a]	Sex	Age at onset (y.o.)	CRC family history
1	AA3519	SPS	1,3	M	58	–
1	AA3520	SPS	2	F	62	–
2	AA3521	SPS	1,3	M	62	+
2	AA3522	SPS	2	M	59	+
3	AA3523	CRC	–	M	32	+
3	AA3524	SPS	3	F	64	+
4	AA3525	SPS	3	M	58	+
4	AA3526	SPS	2	F	53	+
5	AA3527	SPS	3	F	22	–
5	AA3528	SPS	2	M	20	–
6	AA3529	SPS	3	F	61	+
6	AA3530	SPS	2	F	53	+
7	AA3531	SPS/CRC	1,3	F	57	+
7	AA3532	SPS	2	M	56	+
8	AA3534	SPS	2	M	65	–
8	AA3533	SPS	3	M	62	–
9	AA3537	SPS	3	M	64	+
9	AA3536	SPS	2	F	70	+
9	AA3535	SPS/CRC	2,3	F	65	+
10	AA3541	SPS	1,3	F	40	+
10	AA3538	CRC	–	F	58	+
10	AA3539	SPS	2	M	60	+
10	AA3540	SPS	2	F	53	+
11	AA3559	SPS	2	F	39	+
11	AA3560	SPS	1	F	46	+
11	AA3561	SPS	2	F	38	+
11	AA3562	SPS	2	M	44	+
11	AA3563	CRC	–	F	54	+
11	AA3564	SPS	2	M	48	+
12	AA3566	SPS	1,2,3	M	59	–
12	AA3565	SPS	2,3	M	55	–
13	AA3568	SPS	1,2,3	F	55	–
13	AA3567	SPS	2,3	M	59	–
14	AA3569	SPS/CRC	2,3	M	68	+
14	AA3570	SPS	2,3	M	74	+
15	AA3572	SPS	1,3	M	38	–
15	AA3571	SPS	2	M	67	–
16	AA3573	SPS	2,3	F	47	–
16	AA3574	SPS	2	M	51	–

CRC, colorectal cancer; Fam-ID, family identification number; F, female; Ind-ID, individual identification number; M, male; SPS, serrated polyposis syndrome; y.o., years old; +, presence of colorectal cancer in the family; -, no colorectal cancer family history.
[a]All families included a member who fulfilled either criterion 1 or 3. Therefore, the new WHO SPS classification criteria will not change considering these families as having 1 member diagnosed with SPS (7).

[LOD]) and exponential (ExLOD) models (27). All genotyped individuals in our study were affected and, where possible, from the most distant branches of each pedigree. Pedigree structures and diagnoses are detailed in Figure 1. The results of the linkage genome scan under both linear and exponential models were plotted using the "lodplot" R package (https://cran.r-project.org/src/contrib/Archive/lodplot).

Fine mapping of linkage peaks

An LOD score threshold greater than 2 was considered suggestive evidence of linkage. A fine mapping study was performed including additional polymorphic SNPs in flanking regions of a linkage peak to increase allelic informativeness and where intermarker interval was less than 1 cM. Three additional SNPs with high heterozygosity in white Europeans (http://www.internationalgenome.org/) were selected in 3p25.2-p22.3 (rs13078867, rs17016865, and rs7634752). The fine mapping linkage analysis included 340 total markers for chromosome 3 within a 1-LOD drop interval. Their allelic frequencies were extracted from a Spanish control population of 629 individuals (28). After fine mapping, the relative family contribution to the overall linkage was computed using the –perFamily option in Merlin. Further examination of the robustness of the linkage peak was assessed through a replicative analysis using 10 sets of randomly selected WES-derived markers from chromosome 3 (2,847 SNPs) that were nonmonomorphic in the HapMap CEU population.

Rare variant selection

Different parameters were considered for variant annotation including population frequency (1000 Genomes, Exome Variant Server, Exome Aggregation Consortium, and Collaborative Spanish Variant Server), functional consequences, pathogenicity, and position (SnpEff, ANNOtate VARiation, and dbNSFP).

Variant filtering was performed with an in-house R language pipeline, already described in previous studies (20,21). Briefly, parameters taken into account were sequencing quality (coverage $\geq$10× and genotype quality $\geq$50), germline allelic frequency ($\leq$0.1% in Exome Aggregation Consortium database), internal cohort frequency ($\leq$25%), and functional effect (truncating or predicted disrupting missense variants). Missense pathogenicity prediction was assessed with PhyloP (score $\geq$ 1.6), Sorting Intolerant from Tolerant (damaging), PolyPhen2-HVAR (probably or possibly damaging), MutationTaster (disease causing), likelihood ratio test (deleterious), and Combined Annotation Dependent Depletion (score $\geq$15). Missense variants predicted to be pathogenic in a least 3 of 6 predictor tools were selected. Variants were also visually inspected with the Integrative Genomics Viewer and discarded if any sequencing artifact was detected (29).

Family-based association analysis for rare variants

A total of 11 rare and potentially disruptive variants regardless their segregation status in the families from genes spanning the detected linkage interval were included for a family-based association test, using the gene-based segregation test (GESE) package implemented in R (https://cran.r-project.org/web/packages/GESE/) (30). Segregation of rare variants was assessed including all individuals diagnosed with SPS, CRC, or polyposis. *P* values for statistical significance were calculated after 100,000 simulations. Per-family weights were included in the analysis to assess a relative symptom severity variable based on the number of patients with SPS and CRC per family, earlier age at onset, and presence of precursor lesions and other extracolonic neoplasms.

Selection of novel candidate genes for SPS

The prioritization process was completed with selection of the putative candidate genes arising from the GESE family-based association analysis and their interaction partners. This selection process was performed using the Ingenuity Pathway Analysis (IPA) software program (Ingenuity Systems Redwood City, CA), which can identify significant networks using a build-in scientific literature–based database (31). The associated genes with unadjusted *P* value $<$0.05 from the family-based analysis were used as input for IPA and combined with well-known hereditary CRC genes (*APC*, *MLH1*, *MSH2*, *MSH6*, *PMS2*, *MUTYH*, *BMPR1A*, *BMP4*, *PTPRJ*, *GALNT12*, *EPHB2*, *AXIN2*, *UNC5C*, *GREM1*, *STK11*, *SMAD4*, *PTEN*, *KLLN*, *POLE*, *POLD1*, *BUB3*, *BUB1*, *BUB1B*, *RNF43*, *ATM*, *PALB2*, *SEMA4A*, *RPS20*, *NTHL1*, *FAN1*, *MCM9*, *BLM*, *LRP6*, *SMAD9*, *MSH3*, *EPCAM*, *SETD6*, and *BRF1*) and 115 cancer predisposition genes (32). Networks containing any of the GESE genes were considered of interest.

Colon tissue expression from the Human Protein Atlas (Genotype-Tissue Expression data set and protein expression) was evaluated to consider only expressed candidate genes (reads per kilo base per million mapped reads $>$1). The final candidate variant was confirmed by Sanger sequencing (GATC Biotech, Germany).

RESULTS

Linkage analysis

A genome-wide nonparametric linkage analysis was performed using WES-derived genotype data from 16 families with unaffiliated SPS aggregation (Figure 1). The highest peak LOD score was identified on chromosome 3p25.2-p22.3 with LOD_{linear} = 2.311 and LOD_{exp} = 2.11 at rs2293787 (46.484 cM) (Figure 2). When additional markers were added to fine map the region and reduce intermarker intervals, evidence for linkage at the 3p25.2-p22.3 locus increased to LOD_{linear} = 2.4 (*P* value = 4.4E-04) and LOD_{exp} = 2.25 (*P* value = 6.4E-04) with rs2293787 remaining the peak marker (Figure 3, see Table S1, Supplementary Digital Content 1, http://links.lww.com/CTG/A114).

We examined the robustness of the observed linkage signal by a replicative analysis performing linkage analysis in 10 replicate SNP sets using random markers from chromosome 3 (see Figure S1, Supplementary Digital Content 1, http://links.lww.com/CTG/A114). The linkage peak previously identified was replicated in all the 10 data sets with an LOD $>$2 under either the linear or the exponential model or both. These results suggest that the linkage to these regions is not being driven by a particular set of SNPs and is robust to SNP selection. A formal permutation analysis to exclude false-positive signals and to determine empirical significance was not possible, as all subjects were affected and permuting the subjects' phenotypes would be uninformative.

The SPS linkage interval, as defined by a 1-LOD drop interval, spanned a genetic distance of 28.4 cM (3p25.2-p22.3) and a physical interval of 20,874 Mb. Per-family linkage analysis showed that 12 SPS families were contributing positively to the overall LOD score at the 3p25.2-p22.3 linkage peak ranging from an LOD score of 0.06–0.37 (see Table S2, Supplementary Digital Content 1, http://links.lww.com/CTG/A114), with the remaining 4 families contributing only marginally with negative LOD scores ranging from −0.002 to −0.099.

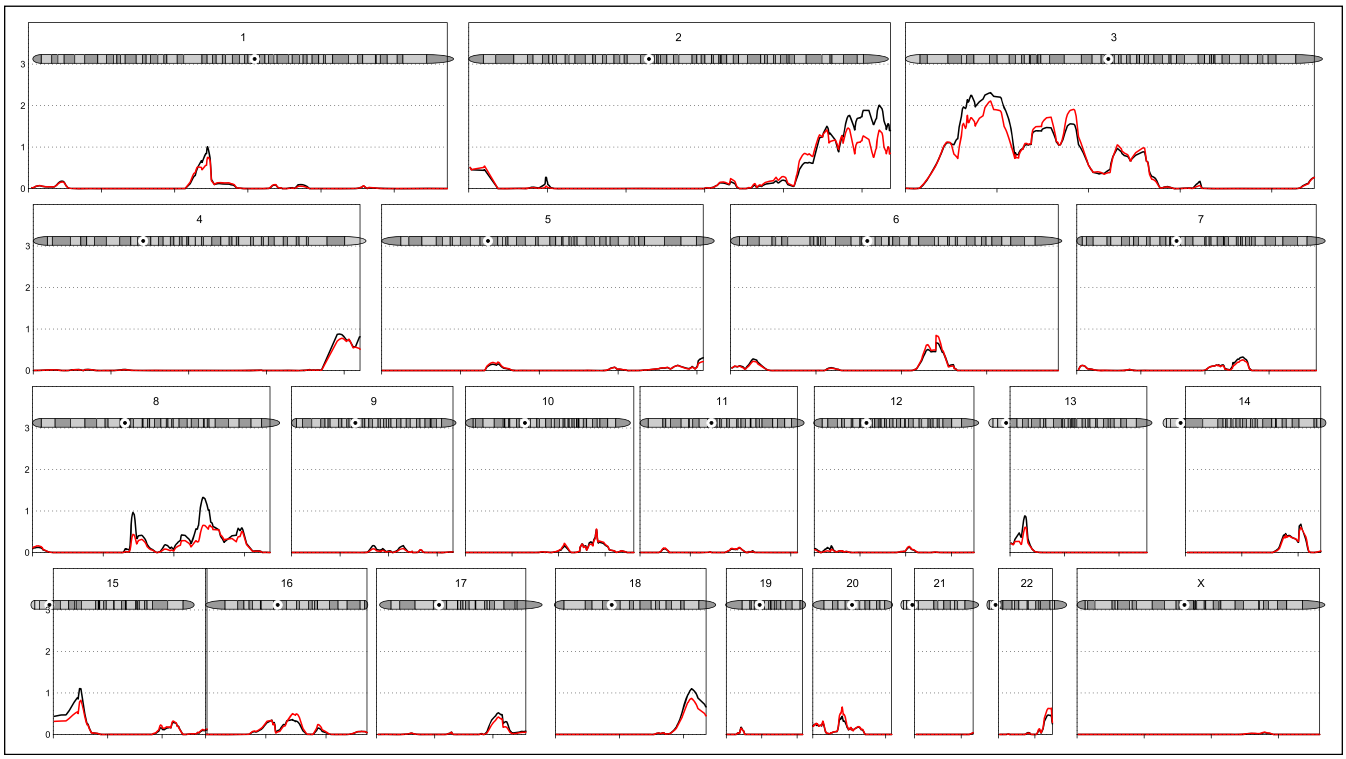

Figure 2. Results of the genome-wide linkage analysis. Nonparametric linkage analysis was performed under the linear (black line) and exponential (red line) models in 16 multiplex/extended serrated polyposis syndrome families. Each chromosome is represented in a plot, including the X chromosome. A linkage signal with LOD > 2 was observed at chromosome 3p25.2-p22.3 with a maximum linear LOD score at marker rs2293787 of 2.311 (linear model). Additional markers were subsequently added to fine map this linkage peak.

Family-based association analysis for variants under specific linkage intervals

Next, we explored the possibility that rare alleles with higher penetrance effects explained the linkage peaks at this locus. We extracted 74 single nucleotide variants (SNVs) from WES data within 42 protein-coding genes under the linkage peak interval (Figure 3), and after quality control filters and more stringent selection criteria, 11 SNVs from 10 genes were selected for segregation analysis (see Table S3, Supplementary Digital Content 1, http://links.lww.com/CTG/A114).

Then, we performed a family GESE of the 11 rare variants using the GESE package. Allele-frequency weighted segregation analysis revealed significant rare variant segregation with SPS in 4 genes (Table 2).

Novel candidate gene for SPS

Genes with evidence of familial segregation for rare pathogenic variants in patients with CRC were further inspected using a protein-protein network analysis (IPA) to identify the most plausible candidate genes from the linkage peaks involved in hereditary CRC. We pooled together the 4 genes with segregating variants from the GESE analysis with established genes for hereditary CRC (38 genes) and germline predisposition to cancer (115 genes) to investigate specific CRC networks. The resulting networks should contain those candidate genes more likely to interact with hereditary CRC and germline cancer predisposition genes and, therefore, more likely to be themselves better candidates for our study. The *FBLN2* gene was present in the only network produced containing at least 1 of the GESE candidate genes (see Figure S2, Supplementary Digital Content 1, http://links.lww.com/CTG/A114). Colon tissue expression was evaluated for this gene (RPKM = 19), whereas the final candidate variant in *FBLN2* (c.3145C>G, p.Pro1049Ala) was validated in the Integrative Genomics Viewer and further confirmed by Sanger sequencing. This rare missense variant is predicted to be pathogenic and falls within an EGF-like calcium domain. Accordingly, *FBLN2* was considered in our approach as the more plausible candidate to be involved in germline predisposition to SPS.

DISCUSSION

It is considered that up to 15%–30% of all CRC tumors have serrated polyps as their precursor lesions. These lesions are characterized by the presence of somatic mutations in *BRAF*, hypermethylation of the promoter regions of tumor suppressor genes, and microsatellite instability (33). SPS is a heterogeneous syndrome defined nowadays by the WHO criteria (7) that is linked with an increased CRC risk (16%) (34). It must be noted that the recently established new SPS criteria would not change the classification of the individuals included in the present study, so obtained results would not be altered.

Although smoking and obesity are linked to serrated polyps, it is believed that germline predisposition to SPS can be also of relevance (35). Whereas several studies have focused in germline CRC predisposition, limited efforts have been developed seeking for hereditary SPS genes. On the other hand, it is plausible that some overlap at the clinical and molecular levels may exist between SPS and another clinical entity, hereditary mixed polyposis syndrome (HMPS). HMPS is a hereditary condition associated with an increased risk of developing polyps in the digestive tract, most commonly in the colon and/or rectum. Polyps detected in patients with HMPS can be of mixed histological type including

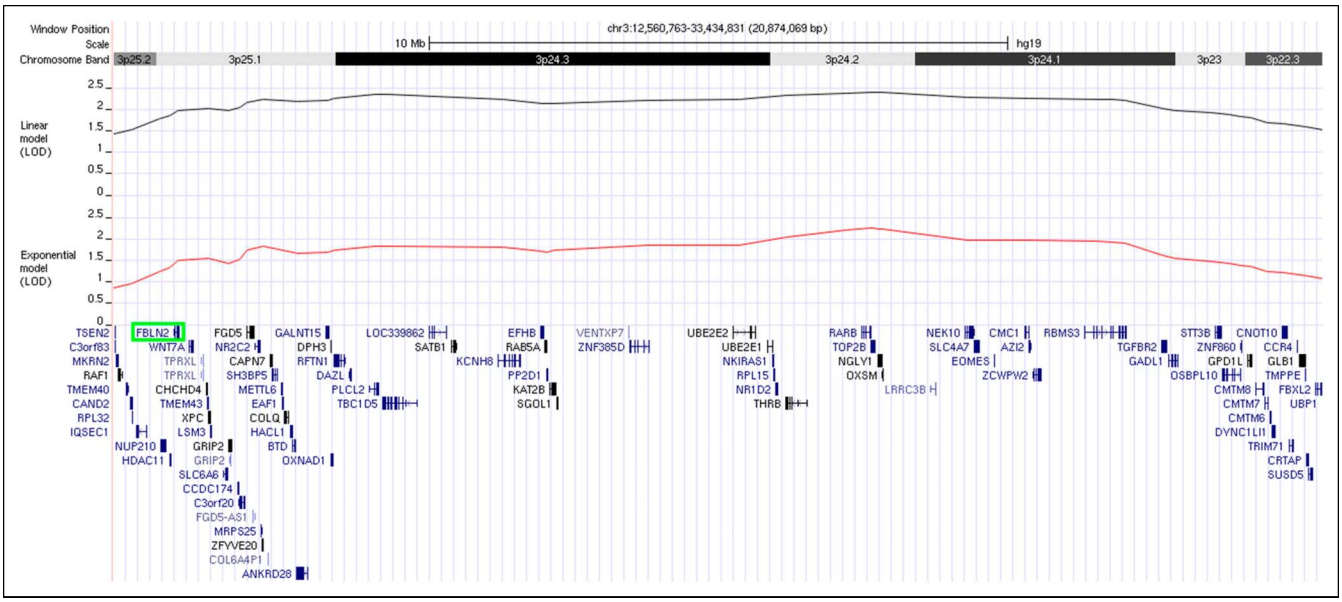

Figure 3. Schematic of the linkage interval and the gene content between the proximal and distal boundaries on chromosome 3p25.2-p22.3 after fine mapping with 3 additional single nucleotide polymorphisms. The maximum LOD score under linear and exponential models are shown at each locus. The locations of known protein-coding genes in the linkage interval are provided in the images below, which were generated using the UCSC genome browser (https://genome.ucsc.edu). Final candidate gene for serrated polyposis syndrome, after gene network analysis and colon gene expression evaluation, is highlighted with a green box.

adenomatous, hyperplastic, hamartomatous/juvenile, or serrated (35). Some linkage studies were developed in a HMPS family of Ashkenazi descent locating positive linkage at chromosome 6q16-q21 (36). New clinical and molecular data from additional members showed that linkage signal to be incorrect and significant evidence was found for chromosomal region 15q13-q14 (37). In addition, a genome-wide linkage search on 32 members of 2 HMPS families detected positive linkage on chromosome 10q23 and found a mutation in *BMPR1A* in 1 family (38). Recently, a duplication spanning the 3' end of the *SCG5* gene and a region upstream of the *GREM1* gene was found to be the mutational event behind the HMPS positive linkage to 15q13-q14 (39).

There are very few NGS studies for SPS germline predisposition. The first study was performed using WES in 20 SPS families and identified 2 independent families with a germline deleterious variant in *RNF43* (10). Additional patients with SPS carrying putative pathogenic variants in this same gene have been identified by subsequent studies (40–42), although its implication in SPS germline predisposition is still controversial (11).

Interestingly, linkage analysis has reappeared as a successful approach to uncover genes implicated in mendelian diseases when used in combination with NGS (43,44). This approach has also been recently applied in complex disorders where susceptibility loci are examined for rare variants with higher penetrance effects that are expected to segregate among patients in large families. This combined strategy has been used in large individual families or few combined families in some complex disorders (45,46), but has never been applied in multiplex or extended families with cancer predisposition.

Segregation analysis for potentially pathogenic rare variants inherited by patients with SPS followed by a protein network analysis identified *FBLN2* as the most relevant candidate for germline SPS predisposition, mapping to 3p25.2-p22.3. According to the available information, the other 3 genes without interactions with known hereditary CRC/cancer genes seem to be not linked to SPS or CRC predisposition. *CAPN7* is a member of the calpain family. Calpains are ubiquitous, well-conserved family of calcium-dependent, cysteine proteases. The calpain proteins are heterodimers consisting of an invariant small subunit and variable large subunits. The large subunit possesses a cysteine protease domain, and both subunits possess calcium-binding domains. Calpains have been implicated in neurodegenerative processes, as their activation can be triggered by calcium influx and oxidative stress (47). As for *TMEM43*, defects in this gene are the cause of familial arrhythmogenic right ventricular dysplasia type 5, also known as arrhythmogenic right ventricular cardiomyopathy type 5. Arrhythmogenic right ventricular dysplasia is an inherited disorder, often involving both ventricles, and is characterized by ventricular tachycardia, heart failure, sudden cardiac death, and fibrofatty replacement of cardiomyocytes. This gene contains

Table 2. Family-based association test of rare variants under the linkage peaks

Gene	SNVs (Seg-SNV)	FAM-ID	*P* value	Weighted *P* value
CAPN7	1 (1)	14	4.00E-07	4.00E-07
TMEM43	1 (1)	7	5.19E-04	1.61E-04
NGLY1	2 (1)	7	7.10E-04	2.10E-04
FBLN2	1 (1)	15	0.01178	7.06E-03

Results of segregation analysis of 11 rare variants (SNVs) found in 39 patients with SPS across the 16 SPS families after simulations and weight corrections. Only significant genes ($P < 0.05$) are reported.
SNVs, number of SNVs regardless segregation in patients with CRC; Seg-SNV, number of SNVs segregating in all patients with CRC in this family.
CRC, colorectal cancer; SNV, single nucleotide variant; Fam-ID, family identification number.

a response element for PPAR gamma (an adipogenic transcription factor), which may explain the fibrofatty replacement of the myocardium, a characteristic pathological finding in ARVC (48). Regarding the *NGLY1* gene, it encodes an enzyme that catalyzes hydrolysis of an N(4)-(acetyl-beta-D-glucosaminyl) asparagine residue to N-acetyl-beta-D-glucosaminylamine and a peptide containing an aspartate residue. The encoded enzyme may play a role in the proteasome-mediated degradation of misfolded glycoproteins. Recessive mutations in this gene cause a congenital disorder of deglycosylation that can include neuromotor impairment, intellectual disability, and neuropathy (49). On the other hand, FBLN2 is an extracellular matrix protein that belongs to the fibulin family. Fibulin interactions with a variety of extracellular ligands are believed to be critical in determining cell movement, proliferation, and angiogenesis, assigning them an important role in cancer development (50). Indeed, *FBLN2* was previously identified as a candidate tumor suppressor gene in nasopharyngeal carcinoma (51), whereas it was also suggested to act as a promoter of malignant progression in lung adenocarcinoma (52). Recently, it was also demonstrated to be methylated in CRC (53) and to play a role in malignancy potential of breast cancer cells (54). Altogether, it could be suggested that *FBLN2* could act as a tumor suppressor gene and, when altered, deregulate correct cell movement and proliferation, and these alterations may lead to the SPS phenotype. Accordingly, additional functional studies including analysis of somatic alterations and germline variants in this gene are warranted to further investigate this hypothesis.

In summary, we performed a genome-wide linkage analysis using WES-derived genotype data from 16 multiplex and extended families with unaffiliated strong SPS aggregation and found suggestive risk loci on chromosome 3p25.2-p22.3. Rare variant segregation analysis and protein network analyses implicated *FBLN2* as a plausible candidate for germline SPS predisposition.

It should be noted that our study has strengths and limitations. WES data from 2 or more individuals from each family permitted to embark on a linkage analysis as a novel approach to uncover possible candidates for SPS germline predisposition. WES data availability along with the strategy of using it for a traditional linkage analysis could be considered the main strengths. On the other hand, the lack of access to somatic tissue or information on environmental factors, the fact that WES only interrogates the coding genome, or that *RNF43* was not interrogated could be listed among the study limitations. Also, we should be aware that replication in additional cohorts, including targeted sequencing of large numbers of families with SPS, and further functional studies are required to confirm this potential candidate for SPS germline predisposition.

CONFLICTS OF INTEREST

Guarantor of the article: Sergi Castellví-Bel, PhD.

Specific author contributions: Claudio Toma, PhD, and Marcos Díaz-Gay, MSc, contributed equally to this work. C.T., M.D.-G., and S.C.-B.: planning the study. C.T., M.D.-G., Y.S.L., C.A.-C., S.F.-E., J.M., B.O., L.B., T.O., M.C., and S.C.-B.: conducting the study. C.T., M.D.-G., T.O., M.C., A.D.B., A.C., L.B., J.C., F.B., D.R.-A., and S.C.-B.: collecting data. C.T., M.D.-G., M.C., J.M.F., and S.C.-B.: interpreting data. C.T., M.D.-G., J.M.F., and S.C.-B.: drafting the manuscript. The final draft of the manuscript has been approved by all authors. The contents of this manuscript are our original work and have not been published, in whole or in part, before or simultaneous with our submission of the manuscript.

Financial support: M.D.-G. was supported by a contract from Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) (Generalitat de Catalunya, 2018FI_B1_00213). S.F.-E., C.A.-C. and J.M. were supported by a contract from CIBEREHD. Y.S.L. was supported by a fellowship (LCF/BQ/DI18/11660058) from "la Caixa" Foundation (ID 100010434) funded EU Horizon 2020 Programme (Marie Skłodowska-Curie grant agreement no. 713673). LB was supported by a Juan de la Cierva postdoctoral contract (FJCI-2017-32593). CIBEREHD and CIBERONC are funded by the Instituto de Salud Carlos III. CT, BJO, and JMF were supported by Australian National Health and Medical Research (NHMRC) Project Grants 1063960 and 1066177. This research was supported by grants from Fondo de Investigación Sanitaria/FEDER (16/00766, 17/00878), Fundación Científica de la Asociación Española contra el Cáncer (GCB13131592CAST), PERIS (SLT002/16/00398, Generalitat de Catalunya), CERCA Programme (Generalitat de Catalunya), and Agència de Gestió d'Ajuts Universitaris i de Recerca (Generalitat de Catalunya, GRPRE 2017SGR21, GRC 2017SGR653). This article is based on work from COST Action CA17118, supported by COST (European Cooperation in Science and Technology). www.cost.eu.

Potential competing interests: None to report.

ACKNOWLEDGMENTS

We are sincerely grateful to the patients, CNAG, the Biobank of Hospital Clínic–IDIBAPS, Biobanco Vasco para la Investigación/O+ehun-Hospital Donostia, and the Genetics Unit at the Hospital Universitario de Móstoles. The work was conducted (in part) at the Esther Koplowitz Centre, Barcelona.

Study Highlights

WHAT IS KNOWN

✓ Germline predisposition to SPS remains largely unknown.

WHAT IS NEW HERE

✓ Combined WES and linkage analysis identified *FBLN2* as a new candidate for this disease.

TRANSLATIONAL IMPACT

✓ New knowledge about SPS will permit future genetic counseling and prevention protocols.

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68(6):394–424.
2. Vogelstein B, Fearon ER, Hamilton SR, et al. Genetic alterations during colorectal-tumor development. N Engl J Med 1988;319(9):525–32.
3. Goldman H, Ming S, Hickock DF. Nature and significance of hyperplastic polyps of the human colon. Arch Pathol 1970;89(4):349–54.
4. IJspeert JE, Vermeulen L, Meijer GA, et al. Serrated neoplasia-role in colorectal carcinogenesis and clinical implications. Nat Rev Gastroenterol Hepatol 2015;12(7):401–9.
5. Rivero-Sanchez L, Lopez-Ceron M, Carballal S, et al; Procolon Group. Reassessment colonoscopy to diagnose serrated polyposis syndrome in a colorectal cancer screening population. Endoscopy 2017;49(1):44–53.
6. Snover D, Ahnen D, Burt R, et al. Serrated polyps of the colon and rectum and serrated polyposis. In: WHO Classification of Tumours of the Digestive System,4th edn, IARC Press: Lyon, France, 2010.

7. World Health Organisation. Classification of Tumours of the Digestive Tract. Lyon: IARC Press: Lyon, France, 2019.
8. Matsumoto T, Iida M, Kobori Y, et al. Serrated adenoma in familial adenomatous polyposis: Relation to germline APC gene mutation. Gut 2002;50(3):402–4.
9. Chow E, Lipton L, Lynch E, et al. Hyperplastic polyposis syndrome: Phenotypic presentations and the role of MBD4 and MYH. Gastroenterology 2006;131(1):30–9.
10. 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(2):520–9.
11. Buchanan DD, Clendenning M, Zhuoer L, et al. Lack of evidence for germline RNF43 mutations in patients with serrated polyposis syndrome from a large multinational study. Gut 2017;66(6):1170–2.
12. Wiesner GL, Daley D, Lewis S, et al. A subset of familial colorectal neoplasia kindreds linked to chromosome 9q22.2-31.2. Proc Natl Acad Sci USA 2003;100(22):12961–5.
13. Kemp Z, Carvajal-Carmona L, Spain S, et al. Evidence for a colorectal cancer susceptibility locus on chromosome 3q21-q24 from a high-density SNP genome-wide linkage scan. Hum Mol Genet 2006;15(19): 2903–10.
14. Roberts A, Nancarrow D, Clendenning M, et al. Linkage to chromosome 2q32.2-q33.3 in familial serrated neoplasia (Jass syndrome). Fam Cancer 2011;10(2):245–54.
15. Metzker ML. Sequencing technologies—The next generation. Nat Rev Genet 2010;11(1):31–46.
16. Ng SB, Turner EH, Robertson PD, et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature 2009;461(7261):272–6.
17. Valle L. Recent discoveries in the genetics of familial colorectal cancer and polyposis. Clin Gastroenterol Hepatol 2017;15(6):809–19.
18. Ott J, Wang J, Leal SM. Genetic linkage analysis in the age of whole-genome sequencing. Nat Rev Genet 2015;16(5):275–84.
19. Gazal S, Gosset S, Verdura E, et al. Can whole-exome sequencing data be used for linkage analysis? Eur J Hum Genet 2016;24(4):581–6.
20. Esteban-Jurado C, Vila-Casadesús M, Garre P, et al. Whole-exome sequencing identifies rare pathogenic variants in new predisposition genes for familial colorectal cancer. Genet Med 2015;17(2):131–42.
21. Esteban-Jurado C, Franch-Expósito S, Muñoz J, et al. The Fanconi anemia DNA damage repair pathway in the spotlight for germline predisposition to colorectal cancer. Eur J Hum Genet 2016;24(10):1501–5.
22. Franch-Expósito S, Esteban-Jurado C, Garre P, et al; EPICOLON consortium. Rare germline copy number variants in colorectal cancer predisposition characterized by exome sequencing analysis. J Genet Genomics 2018;45(1):41–5.
23. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25(14):1754–60.
24. 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(9):1297–303.
25. Smith KR, Bromhead CJ, Hildebrand MS, et al. Reducing the exome search space for mendelian diseases using genetic linkage analysis of exome genotypes. Genome Biol 2011;12(9):R85.
26. Purcell S, Neale B, Todd-Brown K, et al. Plink: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81(3):559–75.
27. Abecasis GR, Cherny SS, Cookson WO, et al. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 2002;30(1): 97–101.
28. Fernandez-Rozadilla C, Cazier JB, Tomlinson IP, et al; EPICOLON Consortium. A colorectal cancer genome-wide association study in a Spanish cohort identifies two variants associated with colorectal cancer risk at 1p33 and 8p12. BMC Genomics 2013;14:55.
29. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics viewer (IGV): High-performance genomics data visualization and exploration. Brief Bioinformatics 2013;14(2):178–92.
30. Qiao D, Lange C, Laird NM, et al. Gene-based segregation method for identifying rare variants in family-based sequencing studies. Genet Epidemiol 2017;41(4):309–19.
31. Krämer A, Green J, Pollard J Jr, et al. Causal analysis approaches in ingenuity pathway analysis. Bioinformatics 2014;30(4):523–30.
32. Rahman N. Realizing the promise of cancer predisposition genes. Nature 2014;505(7483):302–8.
33. Snover DC. Update on the serrated pathway to colorectal carcinoma. Hum Pathol 2011;42(1):1–10.
34. Carballal S, Rodriguez-Alcalde D, Moreira L, et al; Gastrointestinal oncology group of the Spanish gastroenterological Association. Colorectal cancer risk factors in patients with serrated polyposis syndrome: A large multicentre study. Gut 2016;65(11):1829–37.
35. Valle L, de Voer RM, Goldberg Y, et al. Update on genetic predisposition to colorectal cancer and polyposis. Mol Aspects Med 2019;69:10–26.
36. Thomas HJ, Whitelaw SC, Cottrell SE, et al. Genetic mapping of hereditary mixed polyposis syndrome to chromosome 6q. Am J Hum Genet 1996;58(4):770–6.
37. Jaeger EE, Woodford-Richens KL, Lockett M, et al. An ancestral ashkenazi haplotype at the HMPS/CRAC1 locus on 15q13-q14 is associated with hereditary mixed polyposis syndrome. Am J Hum Genet 2003;72(5):1261–7.
38. Cao X, Eu KW, Kumarasinghe MP, et al. Mapping of hereditary mixed polyposis syndrome (HMPS) to chromosome 10q23 by genomewide high-density single nucleotide polymorphism (SNP) scan and identification of BMPR1A loss of function. J Med Genet 2006;43(3):e13.
39. Jaeger E, Leedham S, Lewis A, et al; HMPS Collaboration. Hereditary mixed polyposis syndrome is caused by a 40-kb upstream duplication that leads to increased and ectopic expression of the BMP antagonist GREM1. Nat Genet 2012;44(6):699–703.
40. Taupin D, Lam W, Rangiah D, et al. A deleterious RNF43 germline mutation in a severely affected serrated polyposis kindred. Hum Genome 2015;2:15013.
41. Yan HHN, Lai JCW, Ho SL, et al. RNF43 germline and somatic mutation in serrated neoplasia pathway and its association with BRAF mutation. Gut 2017;66(9):1645–56.
42. Quintana I, Mejías-Luque R, Terradas M, et al. Evidence suggests that germline RNF43 mutations are a rare cause of serrated polyposis. Gut 2018;67(12):2230–2.
43. Eggers S, Smith KR, Bahlo M, et al. Whole exome sequencing combined with linkage analysis identifies a novel 3 bp deletion in NR5A1. Eur J Hum Genet 2015;23(4):486–93.
44. Marsh AP, Heron D, Edwards TJ, et al. Mutations in DCC cause isolated agenesis of the corpus callosum with incomplete penetrance. Nat Genet 2017;49(4):511–4.
45. Norton N, Li D, Rampersaud E, et al; National heart, lung, and Blood Institute GO Exome Sequencing Project and the Exome Sequencing Project Family Studies Project Team. Exome sequencing and genome-wide linkage analysis in 17 families illustrate the complex contribution of TTN truncating variants to dilated cardiomyopathy. Circ Cardiovasc Genet 2013;6(2):144–53.
46. Toma C, Shaw AD, Allcock RJN, et al. An examination of multiple classes of rare variants in extended families with bipolar disorder. Transl Psychiatry 2018;8(1):65.
47. Sorimachi H, Saido TC, Suzuki K. New era of calpain research. Discovery of tissue-specific calpains. FEBS Lett 1994;343(1):1–5.
48. Merner ND, Hodgkinson KA, Haywood AF, et al. Arrhythmogenic right ventricular cardiomyopathy type 5 is a fully penetrant, lethal arrhythmic disorder caused by a missense mutation in the TMEM43 gene. Am J Hum Genet 2008;82(4):809–21.
49. Caglayan AO, Comu S, Baranoski JF, et al. NGLY1 mutation causes neuromotor impairment, intellectual disability, and neuropathy. Eur J Med Genet 2015;58(1):39–43.
50. Gallagher WM, Currid CA, Whelan LC. Fibulins and cancer: Friend or foe? Trends Mol Med 2005;11(7):336–40.
51. Law EW, Cheung AK, Kashuba VI, et al. Anti-angiogenic and tumor-suppressive roles of candidate tumor-suppressor gene, Fibulin-2, in nasopharyngeal carcinoma. Oncogene 2012;31(6):728–38.
52. Baird BN, Schliekelman MJ, Ahn YH, et al. Fibulin-2 is a driver of malignant progression in lung adenocarcinoma. PLoS One 2013;8(6):e67054.
53. Li D, Guo J, Wang S, et al. Identification of novel methylated targets in colorectal cancer by microarray analysis and construction of co-expression network. Oncol Lett 2017;14(3):2643–8.
54. Fontanil T, Álvarez-Teijeiro S, Villaronga MÁ, et al Cleavage of Fibulin-2 by the aggrecanases ADAMTS-4 and ADAMTS-5 contributes to the tumorigenic potential of breast cancer cells. Oncotarget 2017;8(8):13716–29.