Type of the Paper (Article)

Germline variants of *CYBA* and *TRPM4* predispose to familial colorectal cancer

Lizhen Zhu, MD[1,2,#], Beiping Miao, PhD[1,3,4,#], Dagmara Dymerska, PhD[5], Magdalena Kuswik, MSc[5], Elena Bueno-Martínez, PhD[6], Lara Sanoguera-Miralles, MSc[6], Eladio A. Velasco, PhD[6], Nagarajan Paramasivam, PhD[7], Matthias Schlesner, PhD[8], Abhishek Kumar, PhD[1,9,10], Ying Yuan, Prof.[2], Jan Lubinski, Prof.[5,§], Obul Reddy Bandapalli, PD, PhD[1,3,4,11,§,*], Kari Hemminki, Prof.[1,12,13,§,*], Asta Försti, PhD[1,3,4,§, *]

[1]Division of Molecular Genetic Epidemiology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 580, D-69120, Heidelberg, Germany
[2]Department of Medical Oncology, The Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou 310009, Zhejiang Province, China
[3]Hopp Children's Cancer Center (KiTZ), Heidelberg, Germany
[4]Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), German Cancer Consortium (DKTK), Heidelberg, Germany
[5]Hereditary Cancer Center, Department of Genetics and Pathology, Pomeranian Medical University, Unii Lubelskiej 1, 71-252 Szczecin, Poland
[6]Splicing and genetic susceptibility to cancer, Instituto de Biología y Genética Molecular (CSIC-UVa), Valladolid, Spain
[7]Computational Oncology, Molecular Diagnostics Program, National Center for Tumor Diseases (NCT), Heidelberg, Germany
[8]Bioinformatics and Omics Data Analytics, German Cancer Research Center (DKFZ), Heidelberg, Germany
[9] Institute of Bioinformatics, International Technology Park, Bengaluru, 560066 India
[10]Manipal Academy of Higher Education (MAHE), Manipal 576104, Karnataka, India
[11]Medical Faculty Heidelberg, Heidelberg University, Heidelberg, Germany
[12]Faculty of Medicine and Biomedical Center in Pilsen, Charles University in Prague, 30605 Pilsen, Czech Republic
[13]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany
# Equal contribution.
§ Shared senior authorship
* Correspondence: ORB: bandapalli@gmail.com; AF: a.foersti@kitz-heidelberg.de; KH: k.hemminki@dkfz.de

Simple Summary: Whole genome sequencing and bioinformatics analysis on unique colorectal cancer families revealed two attractive candidate predisposition genes, *CYBA* and *TRPM4*, each with a loss-of-function variant. Supported by our functional studies, we suggest that the two gene defects mechanistically involve intestinal barrier integrity through reactive oxygen species and mucus biology, which converges in chronic bowel inflammation, a known risk factor for colorectal cancer.

Abstract: Familial colorectal cancer (CRC) is only partially explained by known germline predisposing genes. We performed whole genome sequencing in 15 Polish families of many affected individuals, without mutations in known CRC predisposing genes. We focused on loss-of-function variants and functionally characterized them. We identified a frameshift variant in the *CYBA* gene (c.246delC) in one family and a splice site variant in the *TRPM4* gene (c.25-1 G>T) in another family. While both variants were absent or extremely rare in gene variant databases, we identified four additional Polish familial CRC cases and two healthy elderly individuals with the *CYBA* variant (odds ratio 2.46, 95% confidence interval 0.48-12.69). Both variants led to a premature stop codon and to a truncated protein. Functional characterization of the variants showed that knockdown of *CYBA* or *TRPM4* depressed generation of reactive oxygen species (ROS) in LS174T and HT-29 cell lines. Knockdown of *TRPM4* resulted in decreased MUC2 protein production. *CYBA* encodes a component in the NADPH oxidase system which generates ROS and controls, e.g., bacterial colonization in the gut. Germline *CYBA* variants are associated with early onset inflammatory bowel disease, supported with experimental evidence on loss of intestinal mucus barrier function due to ROS deficiency. *TRPM4* encodes a calcium-activated ion channel, which in a human colonic cancer cell line controls calcium-mediated secretion of MUC2, a major component of intestinal mucus barrier. We

suggest that the gene defects in *CYBA* and *TRPM4* mechanistically involve intestinal barrier integrity through ROS and mucus biology, which converges in chronic bowel inflammation.

Keywords: whole genome sequencing; cancer predisposition; mucin; reactive oxygen species

1. Introduction

Colorectal cancer (CRC) is the third most commonly diagnosed cancer and the second leading cause of death from cancer [1]. Familial CRC accounts for some 15% of all cases and the twin estimate of heritability amounts up to 30%, but only 2-5% of CRCs are confirmed to be caused by inherited syndromes related to CRC [2-6]. Among these, Lynch syndrome related CRC, caused by germline mutations in mismatch repair (MMR) genes *MLH1*, *MSH2*, *MSH6*, *PMS2*, *EPCAM*, accounts for the majority of hereditary CRC. These tumors are characterized by deficient MMR (dMMR). Rarer syndromes are related to mutations in genes *APC*, *MUTYH*, *STK11*, *PTEN*, *BMPR1A* and *SMAD4* [6].

With the advancement of next-generation sequencing technologies, increasing number of genes have been reported to be possible candidates for CRC predisposition. These include *POLE*, *POLD1*, *NTHL1*, *GERM1*, *GALNT12*, *RNF43*, *RPS20*, *MLH3* and *MSH3* [3, 6-9]. According to the National Comprehensive Cancer Network clinical practice guidelines, most of these new candidates do not have well-established evidence of increased risk for CRC [3]; however, with more data, some of these genes will be proved to be CRC predisposing genes [3, 8, 10]. Epidemiological evidence, such as lacking correlation of CRC risk between spouses, suggests that most of familial aggregation in this cancer is genetic [11]. Thus it is likely that novel predisposing genes will be identified in families of affected individuals.

Most 'classical' cancer predisposing genes were found in linkage studies in families with multiple affected patients [12]. Following this paradigm, we performed germline whole genome sequencing (WGS) in 15 Mendelian type of CRC families unrelated to known CRC predisposing genes focusing on loss-of function variants. We identified a frameshift variant in *CYBA* in one family and a splice site variant in *TRPM4* in another family; while the genes encode proteins in diverse pathways their functions appeared to converge in intestinal barrier integrity and mucus biology targeting inflammatory bowel disease, a known risk factor of CRC [13].

2. Materials and Methods

Population recruitment

In several regions of Poland, population screening was performed mainly in years 2000-2014, in which questionnaires about cancer family history were collected. Persons with positive CRC family history were invited to genetic outpatient clinics all over Poland and their more detailed family histories were collected through face-to-face detailed interviews. Similarly, persons with negative cancer family history were interviewed. An average review took 20-30 minutes. Eligible individuals were asked to participate to the study and they signed an informed consent.

Colorectal cancer families

Fifteen families with strong CRC aggregation compatible with an autosomal dominant pattern of inheritance were recruited to the study. Each family had at least three pathologically confirmed CRC cases; 13 families had at least one case diagnosed below the age

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0354.v1

of 55 years. As in Poland colonoscopy is offered for the family members, many families had also members diagnosed with polyps. Twelve families had CRC in at least two generations. In the other three families one had 5 siblings diagnosed with CRC and 5 of their children had polyps at the age of 41-57 years. In another family, 3 siblings were diagnosed with CRC at the age of 53, 54 and 63 years; their father had been diagnosed with prostate cancer and had died at the age of 72 years and their children were born in 1980's and thus uninformative regarding the CRC status. The third family had 3 siblings diagnosed with CRC at the age of 50, 55 and 62 years; their father had suffered of cancer and died at the age of 55 years. All families were screened for alterations in *APC*, the mismatch repair genes *MLH1, MSH2, MSH3,* large deletions in *EPCAM* and *POLE* p. Leu424Val, *POLD1* p.Ser478Asn and *NTHL1* p.Gln90* mutations and were found to be negative.

Validation cohort

Altogether 1705 unrelated familial CRC cases and 1674 healthy elderly individuals without family history of cancer were included to the validation cohort.

DNA isolation

Peripheral blood samples were collected from affected and unaffected family members who agreed to participate in the study as well as from the validation cohort. Genomic DNA was isolated using a modified Lahiri and Schnabel method [14].

Germline whole genome sequencing (WGS)

WGS was performed in the Illumina X10 platform using DNA extracted from the blood samples as paired-end sequencing with a read length of 150 bp. Mapping of reads to the human reference genome (GRCh37 assembly version hs37d5) was performed using BWA mem (version 0.7.8) and duplicates were marked using Picard (version 1.125). Small variants, single nucleotide variants (SNVs) and indels were called using Platypus (version 0.8.1). Variants were annotated using ANNOVAR [15], dbNSFP v2.9 [16], 1000 Genomes phase III [17], dbSNP [18] and ExAC [19], as described previously [20]. On the variants that passed all the internal Platypus filters a further filtering was performed with considering the QUAL score >20 and coverage of minimum 5 reads. The sequencing coverage and quality statistics for each sample are summarized in Table S1. Minor allele frequency (MAF) of 0.1% was used with respect to population databases (the 1000 Genomes phase III [17] and non-TCGA ExAC [19] data), and the variant frequency of 5% from the local data sets was used to remove technical artefacts. Pairwise comparison of variants among the cohort was performed to check for sample swaps and family relatedness.

Variant identification

Variants were filtered based on the pedigree information considering the members diagnosed with CRC as cases. Family members who were diagnosed with polyps were considered possible variant carriers as were individuals whose parents were diagnosed with CRC and who had not yet reached the age of diagnosis of the youngest CRC case in the family. Unaffected family members were considered as controls. All variants segregating with the disease were filtered for loss-of-function (stop gain, frameshift and splicing) variants. All variants with a MAF of <0.1% in the gnomAD (https://gnomad.broadinstitute.org/) non-Finnish European population were further screened for their location in the gene. Variants located in the last exon or in the non-protein coding transcript of the gene were excluded. Sequencing data were visually inspected using the Integrative Genomic Viewer (IGV [21]) to exclude false positive variants. Combined Annotation-Dependent Depletion (CADD [22]) score was used to evaluate the deleteriousness of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0354.v1

the variants; the scores >20 and >30 are indicative of the top 1% and top 0.1% of deleterious variants, respectively. The final variant selection was based on a literature search.

Copy number variants

Structural variants were analyzed using the SmallPedigree-WGS workflow of Canvas (version 1.40.0.1613 [23]) separately to detect larger copy number variants as described earlier [24]. Variants that affected known cancer predisposing genes were manually inspected using IGV.

Variant confirmation

Candidate variants and their segregation with the disease in the families were confirmed in all available family members by Sanger sequencing with the primers *CYBA* F 5'-GGAGCTTGGTTTCTCACTTGG-3', R 5'-GGAGCTCCTCGGATTTGGA-3' and *TRPM4* F 5'-GTGGCTCTGTGTCCCATAGG-3', R 5'-TCTACACAGACCCAAACGCT-3'. The variants were checked for frequency in the 1705 familial CRC cases and 1674 healthy elderly individuals using custom-made Taqman assays and the existence of the heterozygous variants were confirmed by Sanger sequencing.

TRPM4 minigene assay

Splicing predictions

In silico analysis of the wild type (wt) and mutant sequences was made with the algorithm Max Ent Scan [25] that is integrated into the splicing software of Human Splicing Finder version 3.1 (HSF, http://www.umd.be/HSF3/) [26]. Variant and transcript descriptions were according to the Human Genome Variation Society (HGVS) guidelines on basis of the *TRPM4* GenBank sequence NM_017636.4.

Construction of the minigene mgTRPM4_ex1-2.

A single fragment with exons 1 and 2 of the *TRPM4* gene (728 bp) was amplified from a patient (III-5) DNA with the variant with Kapa High Fidelity polymerase (Kapa Biosystems, Wilmington, MA) and the primers mgTRPM4_ex1-fw 5' AGTCACCTGGACAACCTCAAAGGCACCTTT GAAGCAGAGCCGGCGGAGGG 3' and mgTRPM4_ex2-rv 5' ATAAGCTTGATATCGAATTCCTGCAGCCCG AGCACATAGAAGAGACATCGG 3' (cloning tails are underlined). This insert has the following structure: EX1 [90 bp] – ivs1 [301 bp] – EX2 [68 bp] – ivs2 [269bp]. This fragment was cloned into the splicing vector pSAD v9.0 (Patent P201231427, CSIC) [27] by Overlap Extension PCR [28], to generate a chimeric fusion of vector exon V1-*TRPM4* exon 1. Several colonies were sequenced at the Macrogen facility (Macrogen, Madrid, Spain) to confirm the presence of both wt and mutant minigenes (c.25-1G>T).

Transfection of eukaryotic cells

Transfections were carried out as previously described [29]. Cells were transfected with 1 µg of each construct using low-toxicity Lipofectamine (Life Technologies, Carlsbad, CA). To inhibit nonsense mediated decay (NMD), a 4h incubation with cycloheximide 300 µg/mL (Sigma-Aldrich, St. Louis, MO) was carried out.

Reverse transcription PCR of minigene RNA

RNA was purified with the Genematrix Universal RNA Purification Kit (EURx) including on-column DNAse I digestion. Reverse transcription was carried out with 400 ng of RNA and the RevertAid First Strand cDNA Synthesis Kit (Life Technologies), using the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0354.v1

vector-specific primer RTPSPL3-RV (5′-TGAGGAGTGAATTGGTCGAA-3′). Samples were incubated at 42°C for 1h, followed by 5min at 70°C. Then, 40 ng of cDNA (final volume of 50 µl) were amplified with RT-pMAD502-FW (5′-GAGGTTCTTCGAG-TCCTTT-3′) and RTpSAD-RV (Patent P201231427) (full-length transcript 394 nt) using Platinum-Taq DNA polymerase (Life Technologies). Samples were denatured at 94°C for 2min, followed by 35 cycles x [94°C, 30 sec, 60°C, 30 sec, 72°C, 30 sec], and a final extension step at 72°C for 5min. RT-PCR products were run on 1.2% agarose gels and sequenced at the Macrogen facility.

Capillary electrophoresis of fluorescent RT-PCR

In order to relatively quantify all transcripts, semi-quantitative fluorescent RT-PCRs were performed in triplicate with primers RT-pMAD502-FW and FAM-RTpSAD-RV as described earlier [29].

Cell culture

Human colorectal adenocarcinoma cell line HT-29 (RRID:CVCL_0320) (a kind gift from Peter Krammer´s lab, DKFZ) was cultured in RPMI 1640 media (21875091, Life Technologies) supplemented with 10% fetal bovine serum (FBS) (10500064, Life Technologies). Mucin secreting colorectal adenocarcinoma cell line LS174T (RRID:CVCL_1384) (AddexBio, C0009013) was cultured in RPMI 1640 media supplemented with 10% FBS supplemented with 2mM L-glutamine (51411C, Sigma-Aldrich). HEK293T (RRID: CVCL_0063) cells were a gift from Andreas Trump (DKFZ, Heidelberg). The cells were maintained in DMEM high glucose supplemented with 10% FBS (Gibco), penicillin (50 U/ml, Life Technology), and streptomycin (50 µg/ml, Life Technology). The cells have been authenticated using SNP or STR profiling within the last 3 years and all experiments were performed with mycoplasma-free cells.

Site directed mutagenesis

pcDNA4TO-HA-*TRPM4* was a generous gift from Hugues Abriel [30]. Gateway clone of pENTR221-*CYBA* was obtained from DKFZ´s genomics and proteomics core facility (GPCF). We used LR Clonase Enzyme Mix (11791-019, Invitrogen) to make pENTR221-*CYBA* into pDEST26-*CYBA*. Variants identified in the WGS were introduced into respective plasmids using QuikChange II XL Site-Directed Mutagenesis Kit (200521, Agilent Technologies, Inc). Mutated cDNA clones were confirmed by Sanger sequencing before using them in further experiments. Plasmids were scaled up by transforming into bacteria and harvested by HiPure Plasmid Midiprep Kit (K210005, Invitrogen), and the sequences were checked by Sanger sequencing. Since the *TRPM4* variant was a splice site variant not affecting an exon, we created the pcDNATO-HA-*TRPM4* c.25delAG according to the minigene results of the original *TRPM4* variant. Primers for site directed mutagenesis to get pcDNATO-HA-*TRPM4* c.25delAG and pDEST26-*CYBA* c.246delC were *TRPM4* F 5′-CGGAGAAGGAGCAGCTGGATCCCCAAGA-3′, R 5′-TCTTGGG-GATCCAGCTGCTCCTTCTCCG-3′; *CYBA* F 5′-TAGTAATTCCTGG-TAAAGGCCCGAACAGCTTCAC-3′, R 5′-GTGAAGCTGTTCGGGCCTTTACCAG-GAATTACTA-3′. Sanger sequencing primers for mutated plasmids were *TRPM4* F 5′-CACGCTGTTTTGACCTCCAT-3′, R 5′-CGGAGGAAATTGCTGTGCTT-3′; *CYBA* F 5′-CATGTGGGCCAACGAACAG-3′, R 5′-TCAGTAGGTAGATGCCGCTC-3′.

SiRNA mediated Knockdown of *CYBA* and *TRPM4*

HT-29 and LS174T cells were transfected with human CYBA siRNA (sc-36149, Santa Cruz Biotechnology, INC) to knockdown CYBA or human TRPM4 siRNA (sc-45439, Santa Cruz Biotechnology, INC) to knockdown TRPM4, or scrambled siRNA (sc-37007, Santa Cruz Biotechnology, INC) as a control. All the siRNA products consist of pools of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0354.v1

three to five target-specific 19-25 nt siRNAs designed to knockdown gene expression. HEK293T cells were transfected with wild type (WT) or mutated (MUT) pDEST26-CYBA or pDEST26 and WT or MUT pcDNA4TO-HA-TRPM4 or pcDNA4TO. All the transfections were performed with Lipofectamine 2000 (11668027, Invitrogen).

***In vitro* cell proliferation assay**

Cell proliferation was estimated by cell counting kit 8 (96992-500TESTS-F, Sigma-Aldrich, Inc). HT-29 cells with a density of 5×10^3 and LS174T cells with a density of 7×10^3 cells per well were seeded in 96-well plates. The absorbance optical (OD) density value was measured at 450 nm using Multiskan FC (Thermo Scientific) for 5 days after transfection.

Real-time PCR

Total RNA was extracted by TRIzol (15596018, Invitrogen) and 1 μg of RNA was reverse-transcribed into cDNA using ProtoScript First Strand cDNA Synthesis Kit (E6300L, NEW ENGLAND BioLabs). Quantitative expression analysis was performed with QuantiFast SYBR Green (204056, Qiagen) with 20ng of cDNA. *HPRT1* was used as a housekeeping gene. Primers of *MUC2* and *HPRT1* were from QuantiTect Primer Assay (QT00059066, QT01004675, Qiagen). The detections were performed by Applied Biosystems 7300. The primer sequences for real-time PCR were *TRPM4* F 5′-TGGCTCTCACCTGCTTCCT-3′, R 5′-CCGCACCGTGAAAACCATG-3′; *CYBA* F 5′-ACCAGGAATTACTATGTTCGGGC-3′, R 5′-TAGGTAGATGCCGCTCGCAATG-3′; *MUC1* F 5′-CTGGTCTGTGTTCTGGTTGC-3′, R 5′-CCACTGCTGGGTTTGTGTAA-3′.

Western blot

Protein lysates were prepared and quantified using Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific, #23225). NuPAGE™ 4-12% Bis-Tris Protein Gels and the respective running buffer (Thermo Fisher Scientific; #NP0321PK2, #NP0001) were used for separation of 20 μg of each protein sample. Blotted membranes were blocked with 5 % milk for 1h, incubated overnight at 4°C with the primary antibodies (HA tag monoclonal antibody (26183, Invitrogen, 1:5,000), anti-CYBA antibody (sc-271968, Santa Cruz, 1:250), anti-TRPM4 antibody (ab123936, Abcam, 1:1000) and loading control alpha tubulin antibody (ab4074, Abcam, 0.25 μ/ml) and subsequently for 1h at room temperature with anti-rabbit IgG antibody (7074S, Cell Signaling Technology, 1:10,000), anti-mouse IgG antibody (7076S, Cell Signaling Technology, 1:2,500) diluted in 5 % milk. The membranes were washed and signals were detected by Immobilon Western Chemiluminescent HRP substrate (WBKLS0500, Millipore) and scanned by INTAS Chemostar.

Immunofluorescence

Cells were seeded on Chamber slide (C6932, Sigma-Aldrich) with the seeding density of $30\text{x}10^3$ per well. The cells were allowed to attach to the wells overnight before transfecting with siRNA. After transfection with the siRNA for 72h, slides were incubated with 4% paraformaldehyde (sc-281692, Santa Cruz) 20min at room temperature; PBS 5min 3 times, 0.1% Triton X-100 (11332481001, Sigma-Aldrich) 10min at room temperature, PBS 5min 3 times, 3% BSA block 1h at 37°C. After that the slides were incubated with anti-MUC2 antibody (ab11197, Abcam, 1:500) for 3h at 37°C, PBS 5min 3 times, and with anti-mouse IgG (ab150113, Abcam, 1:1,000) for 1h at 37°C, PBS 5min 3 times. The nucleus was dyed with 1ug/ml DAPI (62248, Thermo Scientific) for 2min at room temperature, PBS 2min 3 times. The slides were mounted with Limonene Mounting Medium (ab104141, Abcam) and covered by coverslips (7695031, Th. Geyer). Confocal microscopy was performed with Leica SP8 and images were saved as TIFF file formats. For each group in

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0354.v1

every experiment, we randomly selected three images with similar numbers of cells and uniform background.

Reactive oxygen species (ROS) detection

Cellular ROS detection was performed with DCFDA Cellular ROS Detection Assay Kit (ab113851, Abcam) in both HT-29 and LS174T cell lines according to the manufacturer´s instructions.

Statistical analysis

The comparison of proliferation, real-time PCR and ROS between the knockdown group and the control group were analyzed by non-paired Student's t test. Image J was used to calculate the fluorescence intensity and area in every group, and quantitative statistical comparison of fluorescence intensity of MUC2 was performed with non-paired Student's t test [31, 32]. Statistical analysis was performed in GraphPad Prism 5.0. Significant differences were considered when $P < 0.05$. Means and standard errors of the means of each assay were presented on the graphs. For real-time PCR assays and ROS detection, we performed three independent experiments in triplicate for each analysis, and two for western blot and immunofluorescence.

3. Results

3.1 Whole genome sequencing revealed no variants in known colorectal cancer predisposing genes

None of the 15 families had copy number variants or pathogenic or likely pathogenic variants or variants of unknown significance in any of the genes attributed to the hereditary CRC syndromes including *MLH1*, *MSH2*, *MSH6*, *PMS2*, *EPCAM*, *APC*, *MUTYH*, *STK11*, *PTEN*, *BMPR1A*, *SMAD4* and *TP53* nor in the genes suggested to be related with familial CRC including *POLE*, *POLD1*, *NTHL1*, *GERM1*, *GALNT12*, *CHEK2*, *BLM*, *AXIN2*, *ATM*, *BMPR1A*, *RECQL*, *RPS20*, *MLH3* and *MSH3*.

3.2 Identification of loss-of-function variants by whole genome sequencing

We identified altogether 14 rare loss-of-function variants (nonsense variants, indels leading to a frameshift and variants in canonical splice sites) that segregated with the disease in 9 of the 15 whole genome sequenced families (**Table S2**). CADD scores of the variants were between 24.7 and 44, supporting their deleterious nature. Literature search indicated a role in inflammation and mucin homeostasis of the intestine for two genes, *CYBA* and *TRPM4*, that were mutated in two unrelated families (**Figure 1**); these were investigated further.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0354.v1

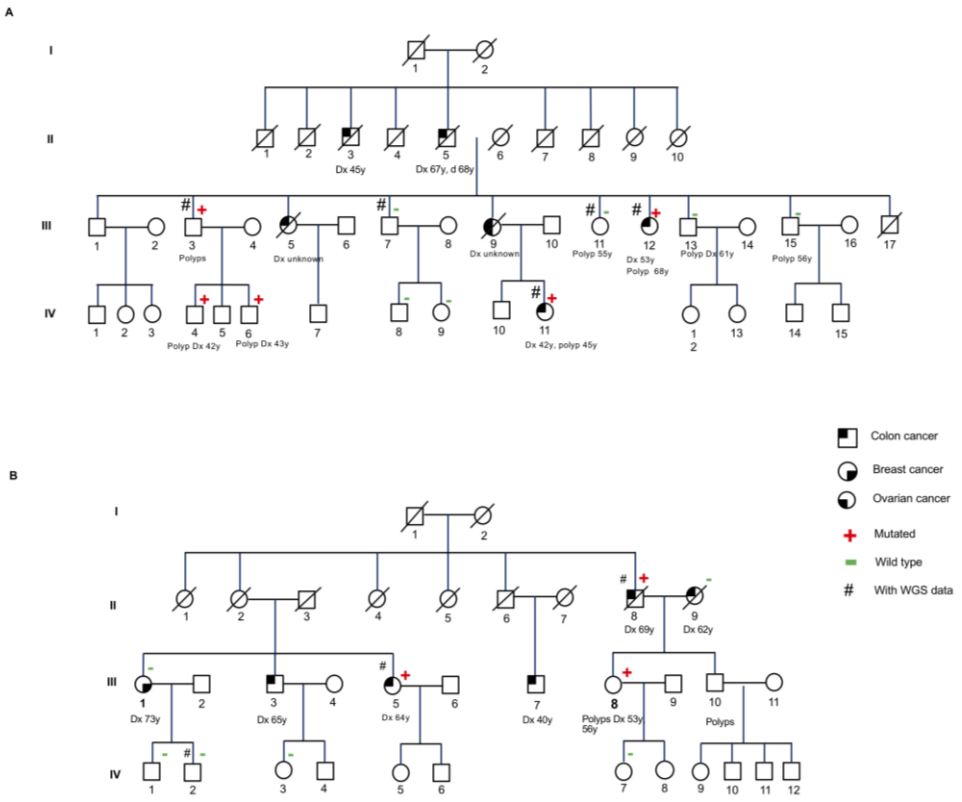


Figure 1. Pedigrees of colon cancer families. (A) Family 8 with CYBA c.246delC variant; (B) Family 11 with TRPM4 c.25-1 G>T variant.

3.3 CYBA c.246delC and *TRPM4* c.25-1 G>T variant segregation in CRC families

In Family 8, we found a heterozygous germline *CYBA* c.246delC (ENST00000261623) variant, which segregated with the disease in the pedigree (**Figure 1A** and **Figure S1**, upper panel). In this family, four family members were diagnosed with CRC at ages from 42 to 67 years (II-5, III-9, III-12 and IV-11) and two with cancer in the abdominal cavity (II-3 and III-5). Several family members had also polyps in the colon. The two sequenced family members diagnosed with CRC (III-12 and IV-11) carried the variant, as did III-3 who was diagnosed with polyps in the colon and his two sons (IV-4, IV-6) who were also diagnosed with polyps in their 40s. Of note, the mother of IV-11 (III-9) was diagnosed with both colon and ovarian cancer. In contrary, III-7 and both of his sons (IV-8 and IV-9) who did not carry the variant had no cancer, nor III-13 or III-15, although they had polyps diagnosed at the age of 61 and 56 years.

In Family 11, a heterozygous germline *TRPM4* c.25-1 G>T (ENST00000252826) variant segregated with the disease in the pedigree (**Figure 1B** and **Figure S1**, lower panel). In this family, two siblings (III-3 and III-5), their cousin (III-7) and their uncle (II-8) were diagnosed with CRC at the age ranging from 40 to 69 years. Several family members had also polyps. The two sequenced family members (II-8 and III-5) diagnosed with CRC carried the variant, as did the daughter of II-8 (III-8), who had multiple polyps. Of note, the wife of II-8 (II-9) who was also diagnosed with CRC did not carry the variant. No *TRPM4* c.25-1 G>T variant was found in a female member (III-1) diagnosed with breast cancer at the age of 73 years, whose colonoscopy was negative till the age of 74 years, alike her two sons (IV-1 and IV-2).

3.4 CYBA c.246delC and *TRPM4* c.25-1 G>T are rare variants

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0354.v1

CYBA c.246delC is absent from the population-based variant databases gnomAD including data from 76156 individuals from the main world populations (https://gnomad.broadinstitute.org/), Exome Variant Server including data from 6503 individuals from European and African American populations (http://evs.gs.washington.edu/EVS) and Leiden Open Variation Database (LOVD; https://www.lovd.nl/) but it was reported in dbSNP (hg38) (http://www.ncbi.nlm.nih.gov/snp) showing an allele frequency of 0.00004 (5/125,568, https://www.ncbi.nlm.nih.gov/snp/rs1439134665) and Clinvar (https://www.ncbi.nlm.nih.gov/clinvar/), in which it was suggested by a single submitter to be apathogenic variant of granulomatous disease (https://www.ncbi.nlm.nih.gov/clinvar/variation/619030/). *TRPM4* c.25-1 G>T is absent in all above population databases.

3.5 Screening of a large cohort of familial CRC patients

Screening of the *CYBA* c.246delC and *TRPM4* c.25-1 G>T variants in 1705 familial CRC cases and 1674 healthy elderly individuals, both from Poland, using custom-made Taqman assays confirmed the presence of the CYBA variant in family 8 and identified the *CYBA* c.246delC variant in 4 additional familial CRC cases and 2 healthy individuals (odds ratio 2.46, 95% confidence interval 0.48-12.69). Two of the CRC patients, diagnosed at ages of 59 years and 64 years, had a family history of CRC. One CRC patient, diagnosed at the age of 44 years, had no contact to family members, and another one, diagnosed at the age of 63 years, had a family history of female genital tract and larynx cancers. The sampling ages of the two healthy individuals were 94 years and 69 years, and they had no family history of any cancer. The existence of the heterozygous *CYBA* c.246delC variant was confirmed by Sanger sequencing in all positive samples. In addition to the *TRPM4* variant in family 11, no other families were found to carry the variant.

3.6 CYBA c.246delC led to loss of CYBA protein

As shown in **Figure 2A**, HEK293T cells transfected with pDEST26-*CYBA* c.246delC expressed less CYBA protein than those transfected with wild type pDEST26-*CYBA*. pDEST26 only expressed similar amount of CYBA to that of the mutant, suggesting that *CYBA* c.246delC led to a loss of the CYBA protein.

3.7 TRPM4 c.25-1G>T led to a frameshift transcript r.25_26del and loss of TRPM4 protein

In silico analysis of the *TRPM4* wild type and c.25-1G>T sequences with MaxEnt Scan (MES) showed that this variant disrupted the canonical acceptor site of TRPM4 exon 2 (MES= 11.38→2.78) but simultaneously created a strong *de novo* acceptor site two nucleotides downstream (MES= 9.9).

The minigene mgTRPM4_ex1-2 was constructed using DNA from a patient (III-5). The wild type and mutant minigenes were introduced into MCF-7 cells and RNA was isolated and reverse transcribed (**Figure 2B**). The wild type minigene produced a single transcript of the expected size (394 nt) and structure (V1- TRPM4 ex1-2 –V2). The c.25-1G>T-minigene also generated a single transcript whose sequence revealed the loss of the first two nucleotides of exon 2 (r.25_26del) using the new alternative acceptor site predicted by MES. This transcript was called Δ(E2p2) following previously suggested rules [33]. Δ(E2p2) would introduce a frameshift and a premature protein truncation 17 codons downstream (p. Ser9Leufs*17) that would inactivate the TRPM4 protein. High-resolution fragment analysis of fluorescent RT-PCR products confirmed the presence of this aberrant transcript (**Figure 2C** and **Figure S2**).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0354.v1

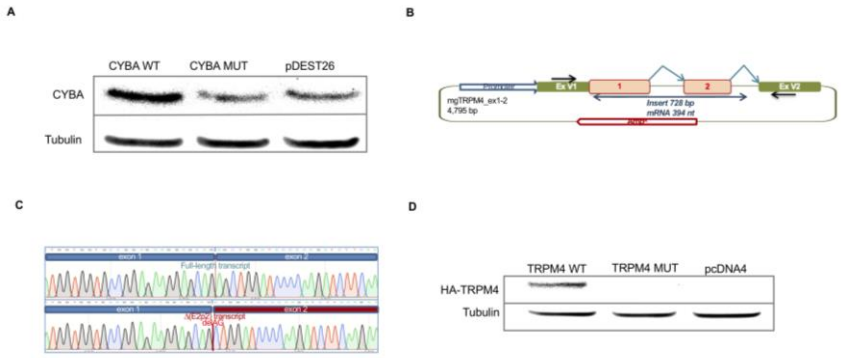


Figure 2. *CYBA* c.246delC and *TRPM4* c.25-1G>T variants led to loss of the corresponding proteins. (A) Western blot detection of the CYBA protein in HEK293T transfected with wild type (CYBA WT), mutated (CYBA MUT) or control plasmids (pDEST26). HEK293T cells transfected with the mutant CYBA plasmid expressed less protein compared to cells transfected with the wild type CYBA plasmid. CYBA c.246delC transfected cells expressed similar amount of protein to those transfected with pDEST26 only. (B) Outline of the *TRPM4* minigene construct. The black arrows in vector exons V1 and V2 indicate specific RT-PCR minigene primers; broken arrows represent the expected splicing reactions. (C) Sequencing traces of the transcripts generated by the wild type (above) and mutant (c.25-1G>T) minigenes, suggesting that TRPM4 c.25-1G>T led to a frameshift transcript r.25_26del, that would be equivalent to c.25delAG. (D) *TRPM4* c.25delAG led to loss of TRPM4 protein. Western blot of TRPM3-HA Tag antibody in HEK293T cells transfected with wild type, mutated or control plasmids. HEK293T cells transfected with pcDNA4TO-HA-TRPM4 c.25delAG (TRPM4 MUT) and pcDNA4 did not express any HA-TRPM4; only HEK293T cells transfected with wild type pcDNA4TO-HA-TRPM4 (TRPM4 WT) expressed HA-TRPM4.

As *TRPM4* c.25-1G>T was not in the cDNA, and the minigene assay showed that *TRPM4* c.25-1G>T led to r.25_26del. i.e., equivalent of c.25delAG, we created pcDNATO-HA-*TRPM4* c.25delAG. We transfected the pcDNATO-HA-*TRPM4* c.25delAG into HEK293T cells to see if the variant affects the expression of TRPM4. As shown in **Figure 2D**, only HEK293T cells transfected with wild type pcDNA4TO-HA-*TRPM4* expressed HA-TRPM4, while HEK293T cells transfected with pcDNA4TO-HA-*TRPM4* c.25delAG or pcDNA4 did not express HA-TRPM4, confirming that *TRPM4* c.25delAG resulted from *TRPM4* c.25-1G>T led to a loss of TRPM4 protein.

3.8 Knockdown of *CYBA* or *TRPM4* promoted proliferation in HT-29 cells

To study the effect of *CYBA* and *TRPM4* on cell proliferation, we knocked down *CYBA* and *TRPM4* with respective siRNAs separately in two colorectal adenocarcinoma cell lines, LS174T and HT-29; LS174T is a mucin secreting cell line. qPCR and western blot analysis confirmed the absence of CYBA or TRPM4 mRNA and protein expression in both LS174T and HT-29 cells (**Figure 3A-D**). While knockdown of *CYBA* or *TRPM4* did not seem to affect cell proliferation of LS174T cells in *in vitro* cell proliferation assay (**Figure S3A-B**), knockdown of *CYBA* or *TRPM4* in the HT-29 cells led to an increase in cell proliferation throughout the time up to 96h (**Figure S3C-D**; $P < 0.05$).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0354.v1

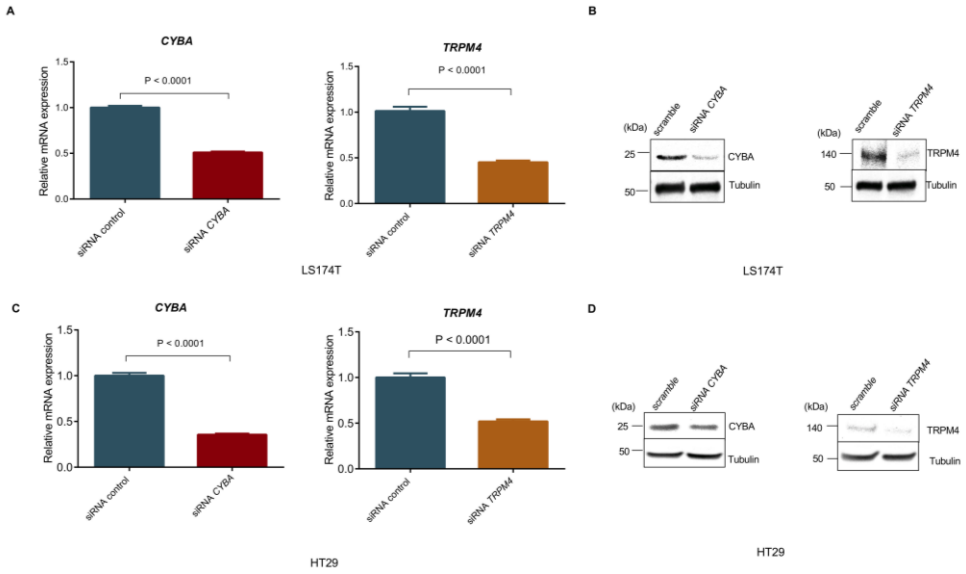
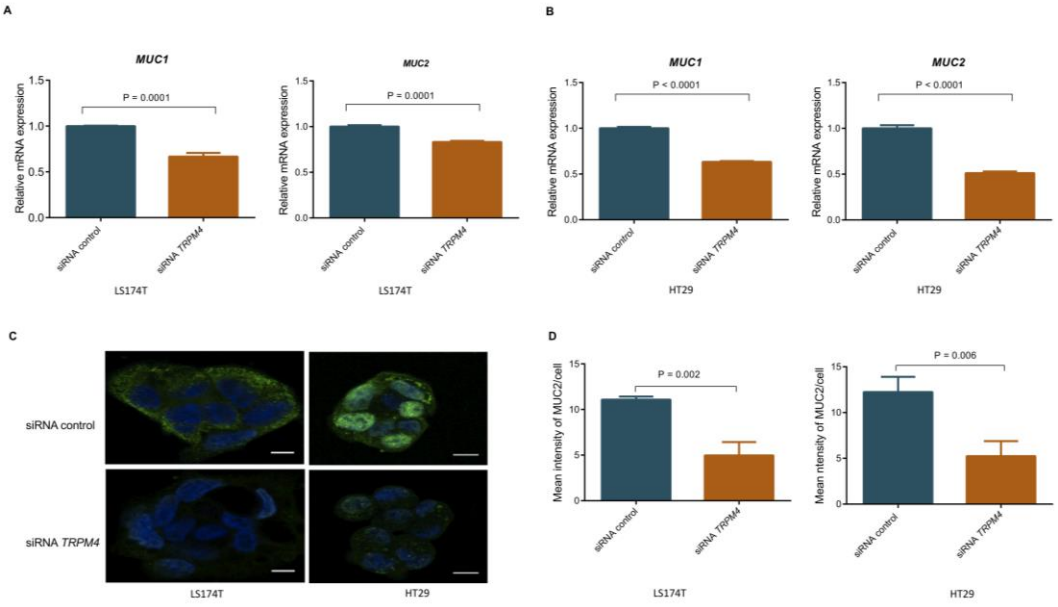


Figure 3. *SiRNA knockdown of CYBA or TRPM4 lead to reduced mRNA and protein of the respective genes: mRNA (A) and protein (B) levels of CYBA and TRPM4 decreased after siRNA knockdown of CYBA or TRPM4 in LS174T cells; mRNA (C) and protein (D) levels of CYBA and TRPM4, decreased after siRNA knockdown of CYBA or TRPM4 in HT-29 cells. For real-time PCR assays, three independent experiments in triplicate were performed and means and standard errors of the means are presented on the graphs. For western blot two independent experiments were performed.*

3.9 Knockdown of *TRPM4* decreased *MUC1* and *MUC2* in both LS174T and HT-29 cells

In order to test the hypothesis that loss of *TRPM4* disrupts mucus barrier by inhibiting mucin secretion, we measured the expression of *MUC1* and *MUC2* by qPCR after down regulating *TRPM4* in LS174T and HT-29 cell lines. The mRNA levels of *MUC1* and *MUC2* were decreased after knockdown *TRPM4* in both LS174T and HT-29 cells ($P \leq 0.0001$; **Figure 4A-B**). Besides, the protein expression of MUC2, measured by immunofluorescence was decreased in both LS174T and HT-29 cells after knockdown of *TRPM4* compared to those treated with control siRNA (LS174T: P=0.002; HT-29: P = 0.006; **Figure 4C-D**).



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0354.v1

Figure 4. RT-PCR results showed that mRNA levels of *MUC1* and *MUC2* decreased, after siRNA knockdown of *TRPM4* in LS174T (A) and HT-29 (B) cells. Three independent experiments in triplicate were performed and means and standard errors of the means are presented on the graphs. Immunofluorescence showed that MUC2 expression decreased after knockdown of *TRPM4* in LS174T (C) and HT-29 (D) cells. Two independent experiments were performed and means and standard errors of the means are presented on the graphs.

3.10 Knockdown of *CYBA* or *TRPM4* led to ROS deficiency

As *CYBA* is involved in ROS modulation, we measured the ROS activity in both LS174T and HT-29 cells after siRNA-mediated knockdown of *CYBA* or *TRPM4*. There was a small, but significant decrease in ROS activity in LS174T cells after knockdown of *CYBA* or *TRPM4* (p=0.009 and 0.014; **Figure 5A**) but no decrease in HT-29 cells (**Figure 5B**).

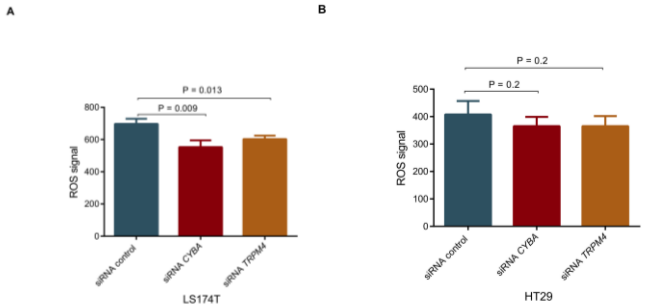


Figure 5. ROS activity upon *CYBA* or *TRPM4* knockdown. *CYBA* or *TRPM4* depletion led to significant reduction of ROS in LS174T cells (A), but no decrease in HT-29 cells (B). Three independent experiments in triplicate were performed and means and standard errors of the means are presented on the graphs.

4. Discussion

We carried out germline WGS analyses of 15 Polish CRC families each of which presented with several CRCs. Loss-of-function variant analysis in two separate families pointed out two rare variants in the genes *CYBA* and *TRPM4* that segregated with the disease in these families. The variant in *CYBA* (*CYBA* c.246delC) leads to a frameshift and a premature protein truncation 41 codons downstream (p.Phe83Asnfs*41), while the splice site variant in *TRPM4* (*TRPM4* c.25-1 G>T) leads to a frameshift and a premature protein truncation 17 codons downstream (p. Ser9Leufs*17). We showed that both variants lead to a loss of protein expression. *CYBA* encodes the alpha chain of cytochrome B-245 which is part of the NADPH oxidase (NOX) system to generate superoxide [34]. It is a component in NOX1-4, of which NOX1 is expressed in the gastrointestinal tract and NOX2 in phagocytes [35]. *TRPM4* encodes a nonselective monovalent cation channel, the upregulation of which enhances sodium entry, which in turn leads to depolarization of the membrane potential [36]. Both genes are highly expressed in the colon and are active in mucin-secreting goblet cells [37, 38]. Although the functions of these two genes appeared initially to be quite diverse, our literature search resulted in a challenging combination of pathways targeting inflammatory bowel disease, a risk factor of CRC [13]. The present study gives further insight in the complementary function of the *CYBA* and *TRPM4* pathways which converge in the protection of the colonic epithelium. The colon is covered by an inner colon mucus layer which is proteolytically converted into an outer mucus layer, both consisting of the MUC2 mucin as the main protein component [39]. The inner mucus layer acts as a barrier against bacteria; only when it is damaged are bacteria able to reach the epithelial surface, which may lead to severe inflammation [39].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0354.v1

Superoxide, generated by NOX analogues, is a member of ROS for which the optimal concentration is critical; overproduction will lead to oxidative stress and development of disease, and likewise, insufficient ROS production may be detrimental to health [37]. Germline variants in *CYBA* are associated with autosomal recessive chronic granulomatous disease, characterized by the failure of activated phagocytes (neutrophils and macrophages), to generate enough superoxide to accomplice intracellular killing of pathogens. Patients suffer from life-threatening infections and from excessive inflammatory reactions [40]. About half of chronic granulomatous disease colon patients present with granuloma formation and acute or chronic inflammation mimicking inflammatory bowel disease [41]. Germline *CYBA* variants have been associated with early onset inflammatory bowel disease [42, 43]. Bacterial penetration of colonic epithelium, the normally restricted zone, is observed in many colitis models and also in patients with inflammatory bowel disease, such as ulcerative colitis [44]. Mucus defects that allow bacteria to reach the epithelium and to stimulate an immune response can lead to the development of intestinal inflammation [45]. The comparison of inflamed ulcerative colitis patients and *MUC2*$^{-/-}$ mice revealed that bacteria in both models had been able to penetrate the colon epithelium and thus causing inflammation [45]. *MUC2*$^{-/-}$ mice have also been shown to develop adenomas in the intestine, the majority of which even progressed into adenocarcinomas [46].

An important clue about the mechanism of action of *CYBA* in inflammatory bowel disease was recently shown in *CYBA* mutant mice which generated low intestinal ROS [47]. The mice suffered from profound mucus layer disruption with bacterial penetration into crypts and from a compromised innate immune response to invading microbes, leading to mortality. The results implicated loss of mucus barrier and innate immune defense in the pathogenesis of intestinal inflammation due to ROS deficiency [47]. In another study, using a dextran sodium sulfate-induced colitis mouse model, ROS deficient Ncf1-mutant mice developed well-differentiated adenocarcinomas, while only high grade dysplasia without malignant invasion were observed in ROS proficient mice with Ncf1 wild type, suggesting that ROS deficiency may cause CRC in response to environmental risk factors [48]. Our experimental results and data from the literature suggest that the loss-of function variant in *CYBA* promotes CRC by at least two mechanisms because of depressed ROS production, first by faltering defense against intestinal bacteria at colonic epithelium and second by suppressing bacterial killing by intestinal phagocytes such as neutrophils and macrophages (Figure 6A) [37].

TRPM4 encodes a calcium-activated nonselective ion channel, the activity of which increases with increasing intracellular calcium concentration; however, this channel does not transport calcium. Disturbance of the membrane potential is deleterious to calcium homeostasis and this is suggested to contribute to carcinogenesis through uncontrolled cell migration and invasion [36]. *TRPM4* gene is highly expressed in CRC tumor buds, contributing to proliferation and invasion of tumor cells [49]. TRPM4 cooperates in the control of mucin secretion from goblet cells in response to extracellular stimuli [38]. The results in human colonic cancer goblet cell line (HT-29-18N2) showed that TRPM4 protein controls calcium-mediated secretion of MUC2 and MUC5AC, in conjunction with a Na^+/Ca^{2+} exchanger NCX [38]. Knock-down of TRPM4 in the HT29-18N2 cells blocked MUC2 secretion [38]. In support to this finding, we showed for the first time that *TRPM4* knockdown also affects *MUC2* mRNA and protein expression. These finding suggests that the loss-of-function variant in *TRPM4* leads to disruption of the mucus layer, allowing bacterial penetration into the mucin-protected epithelium, resulting in inflammation and risk of CRC (Figure 6B). In their publication, Cantero-Recasens et. al. speculated about two distinct modes of mucin secretion: In baseline mucin secretion, TRPM4 is inactive and mucin secretion is regulated by KChIP3 and intracellular Ca^{2+} oscillation [38]. Stimulated mucin secretion is caused by exogenous stimulation of the cells by e.g. ATP or IL-13, which leads to rapid Ca^{2+} release from endoplasmic reticulum, activation of TRPM4 and increase of intracellular Na^+. This in turn triggers NCX to export Na^+ and import Ca^{2+}

leading to rapid burst in mucin secretion. The authors showed similar results in goblet cells earlier with TRPM5, a homologue of TRPM4 [50]. *TRPM4* and *TRPM5* share high sequence homology and similar biophysical properties, yet they are not able to fully compensate each other [50]. In addition by affecting mucin expression and secretion, TRPM4 may influence inflammation through modulation of immune responses [30, 36]. As an alternative mechanism, TRPM4 has been implicated in regulation of the Wnt signaling pathway, which is often dysregulated in CRC [36].

The present variants were not present in the gnomAD database, which listed 8 other loss-of-function variants for *CYBA* and 54 for *TRPM4*; the database covers 125,748 exome sequences. Of note, exactly the same *CYBA* c.246delC variant was submitted to the Clinvar database as a pathogenic variant of granulomatous disease, which is a recessive condition (https://www.ncbi.nlm.nih.gov/clinvar/variation/619030/). In the CRC family, the variant was present in the heterozygous form and the phenotype would be expected to be milder with a later onset than in the case of granulomatous disease. Unfortunately, tumor samples from the patients were not available for testing this. Although the *CYBA* variant was absent from 125,748 exomes of the gnomAD database it was found in 4 additional familial CRC patients among 1705 tested Polish patients and in 2/1674 controls, of whom only one was past the common diagnostic age of CRC. As a DNA sample was only available for the index case of the four families, we could not evaluate the segregation of the variant with the disease in these families, which is a limitation of our study.

The Polish National Colorectal Cancer Screening program offers colonoscopy to all inhabitants between 50 and 66 years; individuals with family history of CRC are eligible starting at age of 40 years. Among 236,089 individuals screened from 2000 through 2011, 17.7% were diagnosed with adenomas; among individuals with family history 18.9% [51]. As individuals with adenomas are at an increased risk of CRC, we considered family members diagnosed with polyps as potential carriers of the predisposing variant. Since we did not have access to any clinical or pathological features of the patients, their tumors and polyps, we were not able to evaluate the risk category of these individuals. Thus, we can only speculate about the role of inflammation in the patients carrying the variant.

In the present study, we focused on the loss-of-function variants, as these are considered to be the most deleterious types of sequence variants. We plan to assess the less deleterious variants, such as missense variants in the future, but they require an increased focus on functional characterization. Many biological pathways are cell and tissue specific which calls for application of different cell lines. In the present study, MUC1 and MUC2 secretion was affected in both LS174T and HT-29 cell lines, while proliferation was affected only in the HT-29 cells, and the effect on ROS activity was more prominent in the LS174T cells.

5. Conclusions

Our germline sequencing efforts of familial CRC lead to the identification of two possible pathogenic variants of two likely CRC predisposing genes *CYBA* and *TRPM4*. The function of these two genes appeared to provide complementary pathways of protecting the colonic epithelium, as shown in **Figure 6** [39]. Although our data did not allow to demonstrate a direct mechanism to cancer formation, we show associations of the *CYBA* and *TRPM4* functions with ROS and mucin production with a likely link to tumorigenesis of CRC.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0354.v1

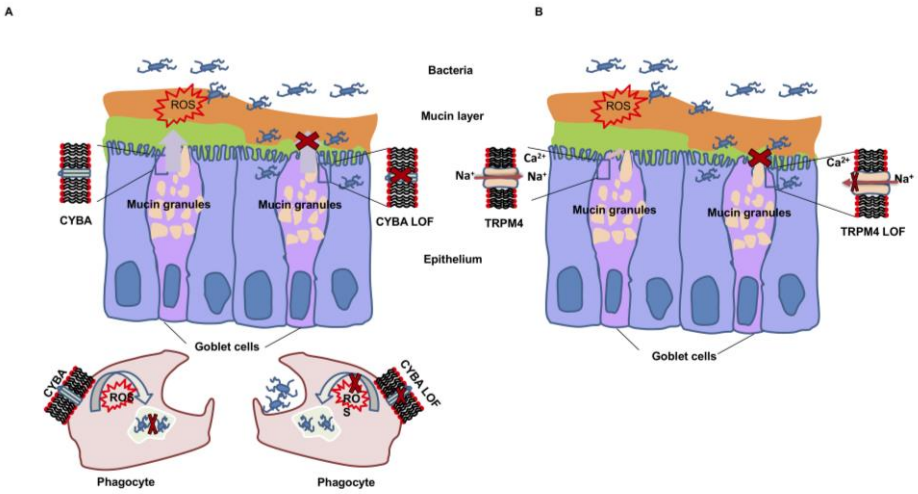


Figure 6. Schematic presentation of the suggested consequences of *CYBA* and *TRPM4* variants on colonic mucin layer integrity. (A) The loss-of-function variant in *CYBA* leads to decreased ROS production and promotes CRC by faltering defense against intestinal bacteria at colonic epithelium and by suppressing bacterial killing by intestinal phagocytes. (B) The loss-of-function variant in *TRPM4* leads to decreased mucus secretion due to inactivation of the TRPM4 channel, potentially leading to mucus layer disruption with bacterial penetration into the mucin protected epithelium, inflammation, and colorectal cancer. The function of the wild type CYBA and TRPM4 is shown in the left goblet cell; the function of CYBA and TRPM4 after the loss-of-function (LOF) variant is shown in the right goblet cell.

Supplementary Materials: Supporting information can be downloaded at: www.mdpi.com/xxx/, Supplementary.

Author Contributions: Conceptualization ORB, KH, AF; data curation NP,MS; formal analysis LZ, NP, MS, AK; funding acquisition EAV, KH; investigation LZ, BM, EBM, LSM, EAV, ORB; methodology LZ, MB, NP, MS, ORB; project administration KH, AF; resources DD, MK, NP, MS, JL; software NP, MS, AK; supervision ORB, KH, AF; validation DD, MK; visualization LZ, BM, EBM, LSM, EAV, ORB; writing – original draft LZ, ORB, KH, AF; writing – review & editing all authors.

Funding: LZ was supported by Zhejiang University Lu's Graduate Education International Exchange Fund. EAV's lab was supported by grants from the Spanish Ministry of Science, Innovation and Universities, Plan Nacional de I+D+I 2013-2016, ISCIII (PI17/00227) co-funded by FEDER from Regional Development European Funds (European Union), and grant CSI242P18 (actuación cofinanciada P.O. FEDER 2014-2020 de Castilla y León) from the Consejería de Educación, Junta de Castilla y León. LSM is supported by a predoctoral fellowship from the Spanish Association Against Cancer (AECC), Junta Provincial de Valladolid (2019-2023). KH was supported by the EU Horizon 2020 program grant No. 856620. AK is recipient of Ramalingaswami Re-Retry Faculty Fellowship (Grant; BT/RLF/Re-entry/38/2017). This article is based upon work from COST Action CA17118, supported by COST (European Cooperation in Science and Technology).

Institutional Review Board Statement: The study was approved by the Bioethics Committee of the Pomeranian Medical Academy in Szczecin No: BN-001/174/05.

Informed Consent Statement: All participating individuals signed an informed consent.

Data Availability Statement: The WGS data generated in this study is available in EGA under accession number EGAS00001005118. Other data that supports the findings of this study are available from the corresponding author upon request.

Acknowledgments: The authors thank the members of the families for participating in this study. We further thank the Genomics and Proteomics Core Facility (GPCF) of the German Cancer Research Center (DKFZ) for providing excellent library preparation and sequencing services and the Omics IT and Data Management Core Facility (ODCF) of the DKFZ for the whole genome sequencing data management.

Conflicts of Interest: The authors disclose no conflicts. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;**68**: 394-424.
2. Frank C, Fallah M, Sundquist J, Hemminki A, Hemminki K. Population Landscape of Familial Cancer. *Sci Rep* 2015;**5**: 12891.
3. Gupta S, Provenzale D, Regenbogen SE, Hampel H, Slavin TP, Hall MJ, Llor X, Chung DC, Ahnen DJ, Bray T, Cooper G, Early DS, et al. NCCN Guidelines Insights: Genetic/Familial High-Risk Assessment: Colorectal, Version 3.2017. *J Natl Compr Canc Netw* 2017;**15**: 1465-75.
4. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K. 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.
5. Mucci LA, Hjelmborg JB, Harris JR, Czene K, Havelick DJ, Scheike T, Graff RE, Holst K, Moller S, Unger RH, McIntosh C, Nuttall E, et al. Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. *JAMA* 2016;**315**: 68-76.
6. Valle L, de Voer RM, Goldberg Y, Sjursen W, Forsti A, Ruiz-Ponte C, Caldes T, Garre P, Olsen MF, Nordling M, Castellvi-Bel S, Hemminki K. Update on genetic predisposition to colorectal cancer and polyposis. *Mol Aspects Med* 2019;**69**: 10-26.
7. Bellido F, Pineda M, Aiza G, Valdes-Mas R, Navarro M, Puente DA, Pons T, Gonzalez S, Iglesias S, Darder E, Pinol V, Soto JL, et al. POLE and POLD1 mutations in 529 kindred with familial colorectal cancer and/or polyposis: review of reported cases and recommendations for genetic testing and surveillance. *Genet Med* 2016;**18**: 325-32.
8. Grolleman JE, de Voer RM, Elsayed FA, Nielsen M, Weren RDA, Palles C, Ligtenberg MJL, Vos JR, Ten Broeke SW, de Miranda N, Kuiper RA, Kamping EJ, et al. Mutational Signature Analysis Reveals NTHL1 Deficiency to Cause a Multi-tumor Phenotype. *Cancer Cell* 2019;**35**: 256-66 e5.
9. Jaeger E, Leedham S, Lewis A, Segditsas S, Becker M, Cuadrado PR, Davis H, Kaur K, Heinimann K, Howarth K, Collaboration H, East J, et al. 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**: 699-703.
10. Buchanan DD, Stewart JR, Clendenning M, Rosty C, Mahmood K, Pope BJ, Jenkins MA, Hopper JL, Southey MC, Macrae FA, Winship IM, Win AK. Risk of colorectal cancer for carriers of a germ-line mutation in POLE or POLD1. *Genetics in medicine : official journal of the American College of Medical Genetics* 2018;**20**: 890-5.
11. Frank C, Fallah M, Ji J, Sundquist J, Hemminki K. The population impact of familial cancer, a major cause of cancer. *International journal of cancer* 2014;**134**: 1899-906.
12. Rahman N. Realizing the promise of cancer predisposition genes. *Nature* 2014;**505**: 302-8.
13. Keller DS, Windsor A, Cohen R, Chand M. Colorectal cancer in inflammatory bowel disease: review of the evidence. *Tech Coloproctol* 2019;**23**: 3-13.
14. Lahiri DK, Schnabel B. DNA isolation by a rapid method from human blood samples: effects of MgCl2, EDTA, storage time, and temperature on DNA yield and quality. *Biochem Genet* 1993;**31**: 321-8.
15. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;**38**: e164.
16. Liu X, Wu C, Li C, Boerwinkle E. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Hum Mutat* 2016;**37**: 235-41.
17. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. *Nature* 2015;**526**: 68-74.
18. Smigielski EM, Sirotkin K, Ward M, Sherry ST. dbSNP: a database of single nucleotide polymorphisms. *Nucleic Acids Res* 2000;**28**: 352-5.
19. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;**536**: 285-91.
20. Kumar A, Bandapalli OR, Paramasivam N, Giangiobbe S, Diquigiovanni C, Bonora E, Eils R, Schlesner M, Hemminki K, Forsti A. Familial Cancer Variant Prioritization Pipeline version 2 (FCVPPv2) applied to a papillary thyroid cancer family. *Sci Rep* 2018;**8**: 11635.
21. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 2013;**14**: 178-92.
22. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM. A general framework for estimating the relative pathogenicity of human genetic variants 2014;**46**: 310-5.
23. Roller E, Ivakhno S, Lee S, Royce T, Tanner S. Canvas: versatile and scalable detection of copy number variants. *Bioinformatics* 2016;**32**: 2375-7.

24. Srivastava A, Giangiobbe S, Kumar A, Paramasivam N, Dymerska D, Behnisch W, Witzens-Harig M, Lubinski J, Hemminki K, Forsti A, Bandapalli OR. Identification of Familial Hodgkin Lymphoma Predisposing Genes Using Whole Genome Sequencing. *Front Bioeng Biotechnol* 2020;**8**: 179.
25. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 2004;**11**: 377-94.
26. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009;**37**: e67.
27. Acedo A, Hernandez-Moro C, Curiel-Garcia A, Diez-Gomez B, Velasco EA. Functional classification of BRCA2 DNA variants by splicing assays in a large minigene with 9 exons. *Hum Mutat* 2015;**36**: 210-21.
28. Bryksin AV, Matsumura I. Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *BioTechniques* 2010;**48**: 463-5.
29. Fraile-Bethencourt E, Diez-Gomez B, Velasquez-Zapata V, Acedo A, Sanz DJ, Velasco EA. Functional classification of DNA variants by hybrid minigenes: Identification of 30 spliceogenic variants of BRCA2 exons 17 and 18. *PLoS genetics* 2017;**13**: e1006691.
30. Bianchi B, Smith PA, Abriel H. The ion channel TRPM4 in murine experimental autoimmune encephalomyelitis and in a model of glutamate-induced neuronal degeneration. *Mol Brain* 2018;**11**: 41.
31. Tawiah A, Cornick S, Moreau F, Gorman H, Kumar M, Tiwari S, Chadee K. High MUC2 Mucin Expression and Misfolding Induce Cellular Stress, Reactive Oxygen Production, and Apoptosis in Goblet Cells. *The American journal of pathology* 2018;**188**: 1354-73.
32. Cobo ER, Kissoon-Singh V, Moreau F, Holani R, Chadee K. MUC2 Mucin and Butyrate Contribute to the Synthesis of the Antimicrobial Peptide Cathelicidin in Response to Entamoeba histolytica- and Dextran Sodium Sulfate-Induced Colitis. *Infection and immunity* 2017;**85**.
33. Fraile-Bethencourt E, Valenzuela-Palomo A, Diez-Gomez B, Caloca MJ, Gomez-Barrero S, Velasco EA. Minigene Splicing Assays Identify 12 Spliceogenic Variants of BRCA2 Exons 14 and 15. *Front Genet* 2019;**10**: 503.
34. O'Neill S, Brault J, Stasia MJ, Knaus UG. Genetic disorders coupled to ROS deficiency. *Redox Biol* 2015;**6**: 135-56.
35. Stasia MJ. CYBA encoding p22(phox), the cytochrome b558 alpha polypeptide: gene structure, expression, role and physiopathology. *Gene* 2016;**586**: 27-35.
36. Gao Y, Liao P. TRPM4 channel and cancer. *Cancer Lett* 2019;**454**: 66-9.
37. Aviello G, Knaus UG. NADPH oxidases and ROS signaling in the gastrointestinal tract. *Mucosal Immunol* 2018;**11**: 1011-23.
38. Cantero-Recasens G, Butnaru CM, Brouwers N, Mitrovic S, Valverde MA, Malhotra V. Sodium channel TRPM4 and sodium/calcium exchangers (NCX) cooperate in the control of Ca(2+)-induced mucin secretion from goblet cells. *J Biol Chem* 2019;**294**: 816-26.
39. Johansson ME, Hansson GC. Mucus and the goblet cell. *Dig Dis* 2013;**31**: 305-9.
40. Roos D. Chronic granulomatous disease. *Br Med Bull* 2016;**118**: 50-63.
41. Alimchandani M, Lai JP, Aung PP, Khangura S, Kamal N, Gallin JI, Holland SM, Malech HL, Heller T, Miettinen M, Quezado MM. Gastrointestinal histopathology in chronic granulomatous disease: a study of 87 patients. *The American journal of surgical pathology* 2013;**37**: 1365-72.
42. Ashton JJ, Andreoletti G, Coelho T, Haggarty R, Batra A, Afzal NA, Beattie RM, Ennis S. Identification of Variants in Genes Associated with Single-gene Inflammatory Bowel Disease by Whole-exome Sequencing. *Inflamm Bowel Dis* 2016;**22**: 2317-27.
43. Denson LA, Jurickova I, Karns R, Shaw KA, Cutler DJ, Okou DT, Dodd A, Quinn K, Mondal K, Aronow BJ, Haberman Y, Linn A, et al. Clinical and Genomic Correlates of Neutrophil Reactive Oxygen Species Production in Pediatric Patients With Crohn's Disease. *Gastroenterology* 2018;**154**: 2097-110.
44. Johansson ME. Mucus layers in inflammatory bowel disease. *Inflamm Bowel Dis* 2014;**20**: 2124-31.
45. Wenzel UA, Magnusson MK, Rydstrom A, Jonstrand C, Hengst J, Johansson ME, Velcich A, Ohman L, Strid H, Sjovall H, Hansson GC, Wick MJ. Spontaneous colitis in Muc2-deficient mice reflects clinical and cellular features of active ulcerative colitis. *PloS one* 2014;**9**: e100217.
46. Velcich A, Yang W, Heyer J, Fragale A, Nicholas C, Viani S, Kucherlapati R, Lipkin M, Yang K, Augenlicht L. Colorectal cancer in mice genetically deficient in the mucin Muc2. *Science (New York, NY)* 2002;**295**: 1726-9.
47. Aviello G, Singh AK, O'Neill S, Conroy E, Gallagher W, D'Agostino G, Walker AW, Bourke B, Scholz D, Knaus UG. Colitis susceptibility in mice with reactive oxygen species deficiency is mediated by mucus barrier and immune defense defects. *Mucosal immunology* 2019;**12**: 1316-26.
48. Carvalho L, Gomes JRM, Tavares LC, Xavier AR, Klika KD, Holmdahl R, Carvalho RA, Souto-Carneiro MM. Reactive Oxygen Species Deficiency Due to Ncf1-Mutation Leads to Development of Adenocarcinoma and Metabolomic and Lipidomic Remodeling in a New Mouse Model of Dextran Sulfate Sodium-Induced Colitis. *Frontiers in immunology* 2018;**9**: 701.
49. Kappel S, Stoklosa P, Hauert B, Ross-Kaschitza D, Borgstrom A, Baur R, Galvan JA, Zlobec I, Peinelt C. TRPM4 is highly expressed in human colorectal tumor buds and contributes to proliferation, cell cycle, and invasion of colorectal cancer cells. *Mol Oncol* 2019;**13**: 2393-405.
50. Guo J, She J, Zeng W, Chen Q, Bai XC, Jiang Y. Structures of the calcium-activated, non-selective cation channel TRPM4. *Nature* 2017;**552**: 205-9.
51. Wieszczy P, Kaminski MF, Franczyk R, Loberg M, Kobiela J, Rupinska M, Kocot B, Rupinski M, Holme O, Wojciechowska U, Didkowska J, Ransohoff D, et al. Colorectal Cancer Incidence and Mortality After Removal of Adenomas During Screening Colonoscopies. *Gastroenterology* 2020;**158**: 875-83 e5.