Exome Sequencing Identifies Biallelic *MSH3* Germline Mutations as a Recessive Subtype of Colorectal Adenomatous Polyposis

Ronja Adam,[1,2,15] Isabel Spier,[1,2,15] Bixiao Zhao,[3] Michael Kloth,[4] Jonathan Marquez,[3] Inga Hinrichsen,[5] Jutta Kirfel,[6] Aylar Tafazzoli,[1,7] Sukanya Horpaopan,[1,8] Siegfried Uhlhaas,[1] Dietlinde Stienen,[1] Nicolaus Friedrichs,[4] Janine Altmüller,[9,10] Andreas Laner,[11,12] Stefanie Holzapfel,[1,2] Sophia Peters,[1] Katrin Kayser,[1] Holger Thiele,[9] Elke Holinski-Feder,[11,12] Giancarlo Marra,[13] Glen Kristiansen,[6] Markus M. Nöthen,[1,7] Reinhard Büttner,[4] Gabriela Möslein,[14] Regina C. Betz,[1,7] Angela Brieger,[5] Richard P. Lifton,[3] and Stefan Aretz[1,2,*]

In ~30% of families affected by colorectal adenomatous polyposis, no germline mutations have been identified in the previously implicated genes *APC*, *MUTYH*, *POLE*, *POLD1*, and *NTHL1*, although a hereditary etiology is likely. To uncover further genes with high-penetrance causative mutations, we performed exome sequencing of leukocyte DNA from 102 unrelated individuals with unexplained adenomatous polyposis. We identified two unrelated individuals with differing compound-heterozygous loss-of-function (LoF) germline mutations in the mismatch-repair gene *MSH3*. The impact of the *MSH3* mutations (c.1148delA, c.2319−1G>A, c.2760delC, and c.3001−2A>C) was indicated at the RNA and protein levels. Analysis of the diseased individuals' tumor tissue demonstrated high microsatellite instability of di- and tetranucleotides (EMAST), and immunohistochemical staining illustrated a complete loss of nuclear MSH3 in normal and tumor tissue, confirming the LoF effect and causal relevance of the mutations. The pedigrees, genotypes, and frequency of *MSH3* mutations in the general population are consistent with an autosomal-recessive mode of inheritance. Both index persons have an affected sibling carrying the same mutations. The tumor spectrum in these four persons comprised colorectal and duodenal adenomas, colorectal cancer, gastric cancer, and an early-onset astrocytoma. Additionally, we detected one unrelated individual with biallelic *PMS2* germline mutations, representing constitutional mismatch-repair deficiency. Potentially causative variants in 14 more candidate genes identified in 26 other individuals require further workup. In the present study, we identified biallelic germline *MSH3* mutations in individuals with a suspected hereditary tumor syndrome. Our data suggest that *MSH3* mutations represent an additional recessive subtype of colorectal adenomatous polyposis.

Introduction

Adenomatous polyposis syndromes of the colorectum are precancerous conditions characterized by the presence of dozens to thousands of adenomatous polyps, which, unless detected early and removed, invariably result in colorectal cancer (CRC). The phenotypic spectrum ranges from an early-onset manifestation with high numbers of adenomas and a positive family history to isolated late-onset cases with a low polyp burden.

To date, two major inherited monogenic forms of colorectal adenomatous polyposis can be delineated by molecular genetic analyses: (1) autosomal-dominant familial adenomatous polyposis (FAP [MIM: 175100]), caused by heterozygous germline mutations in the tumor-suppressor gene *APC* (APC, WNT signaling pathway regulator [MIM: 611731]);[1,2] and (2) autosomal-recessive *MUTYH*-associated polyposis (MAP [MIM: 608456]), caused by biallelic germline mutations in the base-excision-repair (BER) gene *MUTYH* (mutY DNA glycosylase [MIM: 604933]).[3,4]

Currently, high-throughput sequencing approaches, in particular whole-exome sequencing (WES), are considered the most powerful tools for detecting causative variants in genes in as yet unexplained Mendelian conditions.[5,6] Very recent WES investigations have identified two rare forms of colorectal adenomatous polyposis: (1) autosomal-dominant polymerase-proofreading-associated polyposis (PPAP [MIM: 612591]), caused by specific germline missense mutations in the polymerase genes *POLE* (DNA polymerase epsilon, catalytic subunit [MIM: 174762]) and *POLD1* (DNA polymerase delta 1, catalytic subunit [MIM: 174761]);[7–9] and (2) another very rare autosomal-recessive colorectal adenomatous polyposis (MIM: 616415), caused by biallelic mutations in *NTHL1* (nth-like DNA glycosylase

[1]Institute of Human Genetics, University of Bonn, 53127 Bonn, Germany; [2]Center for Hereditary Tumor Syndromes, University of Bonn, 53127 Bonn, Germany; [3]Department of Genetics, Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, CT 06520-8005, USA; [4]Institute of Pathology, University of Cologne, 50937 Cologne, Germany; [5]Medical Clinic 1, Biomedical Research Laboratory, Goethe-University Frankfurt, 60590 Frankfurt, Germany; [6]Institute of Pathology, University of Bonn, 53127 Bonn, Germany; [7]Department of Genomics, Life & Brain Center, University of Bonn, 53127 Bonn, Germany; [8]Department of Anatomy, Faculty of Medical Science, Naresuan University, Phitsanulok, Chiang Mai 50200, Thailand; [9]Cologne Center for Genomics, University of Cologne, 50937 Cologne, Germany; [10]Institute of Human Genetics, University of Cologne, 50937 Cologne, Germany; [11]Medizinische Klinik und Poliklinik IV, Ludwig-Maximilians-University, 80336 Munich, Germany; [12]Medical Genetics Center, 80335 Munich, Germany; [13]Institute of Molecular Cancer Research, University of Zurich, CH-8057 Zurich, Switzerland; [14]HELIOS Klinikum Wuppertal, University of Witten/Herdecke, 42283 Wuppertal, Germany
[15]These authors contributed equally to this work
*Correspondence: stefan.aretz@uni-bonn.de
http://dx.doi.org/10.1016/j.ajhg.2016.06.015.

1 [MIM: 602656]).[10] The WES approach has also detected *ZSWIM7* (zinc finger SWIM-type containing 7 [MIM: 614535]) and *PIEZO1* (piezo type mechanosensitive ion channel component 1 [MIM: 611184]) as promising candidate genes carrying variants causing colorectal adenomatous polyposis.[11]

However, in around 30% of polyposis cases, no underlying germline mutation has been identified, although a genetic basis is likely. Here, classic approaches to gene identification, such as linkage analysis, are not feasible, given that most of these cases are either sporadic or characterized by an uncertain family.[12–15] Over the past two decades, a number of candidate-gene studies have been performed without convincing results.[16–19] Neither loss-of-heterozygosity (LOH) analyses nor profiling of somatic mutations has contributed to the identification of promising novel genetic causes. A fraction of cases might be explained by deep intronic *APC* mutations,[20] *APC* mutational mosaicism,[21] rare *APC* missense mutations,[22] or other genes predisposing to cancer.[23–26] In addition, rare germline copy-number variants (CNVs) and low-penetrant variants might contribute to the genetic predisposition for the formation of colorectal adenomas.[15,27,28]

Another hereditary CRC syndrome, Lynch syndrome, is not accompanied by a florid colorectal polyposis and is characterized by microsatellite instable tumors. The underlying cause is a heterozygous germline mutation in one of the mismatch-repair (MMR) genes *MLH1* (mutL homolog 1 [MIM: 120436]), *MSH2* (mutS homolog 2 [MIM: 609309]), *MSH6* (mutS homolog 6 [MIM: 600678]), *PMS2* (PMS1 homolog 2, mismatch repair system component [MIM: 600259]), or *EPCAM* (epithelial cell adhesion molecule [MIM: 185535]).[29,30] Biallelic mutations in these genes lead to constitutional MMR deficiency (CMMRD [MIM: 276300]) with multiple tumors and childhood onset.[31,32] In contrast, familial CRC without polyposis or microsatellite instability is etiologically very heterogeneous. In some of these families, germline mutations in *FAN1* (FANCD2/FANCI-associated nuclease 1 [MIM: 613534]), encoding a nuclease involved in DNA repair, were recently identified by a WES approach.[33]

To uncover additional genes with high-penetrance mutations causing colorectal polyposis, we sequenced the germline exomes of 102 unrelated individuals with unexplained adenomatous polyposis. The identification of further genetic causes will extend the knowledge of disease mechanisms, biological pathways, and potential therapeutic targets.

Material and Methods

Cohort and Data Collection

All 102 individuals were determined to have unexplained colorectal adenomatous polyposis, i.e., no germline mutation in *APC* or *MUTYH* was identified by Sanger sequencing of the coding regions or deletion and duplication analysis by multiplex ligation-dependent probe amplification (MLPA).[34] All participants were screened for *APC* mutations in the mosaic state.[21] All persons were examined for pathogenic deep intronic *APC* mutations: 67 persons were screened via transcript analysis, and 35 were tested for known intronic mutations.[20] Furthermore, the two hotspot mutations in *POLE* and *POLD1* were excluded.[9] In addition, a SNP-array-based CNV analysis was performed in all individuals, as described elsewhere.[28]

For all 102 persons included in this study, a hereditary cause of the disease was considered highly likely. The inclusion criteria were the presence of at least 20 synchronous, or 40 metachronous, histologically confirmed colorectal adenomas, irrespective of inheritance pattern or extraintestinal lesions. All participants were of central European origin according to family name and self-report. Relatives were only considered to be affected if their medical records confirmed fulfilment of the inclusion criteria. The study was approved by the local ethics review board (Medical Faculty of the University of Bonn, board no. 224/07), and all participants provided written informed consent prior to inclusion.

High-Throughput Sequencing and Bioinformatics Workflow

Genomic DNA was extracted from peripheral EDTA-anticoagulated blood samples by the standard salting-out procedure. WES was performed at the Yale Center for Genome Analysis via capture by the NimbleGen 2.1M Human Exome Array, and then paired-end sequencing was performed on a HiSeq 2000 instrument (Illumina) as described elsewhere.[35] Targeted bases were covered by a mean of 67 independent reads, and an average of 94% of all bases were covered eight or more times (Table S1). Reads were aligned to the hg19 human reference genome (UCSC Genome Browser) with ELAND (Illumina). SAMtools software was used for marking duplicated reads, performing local realignment around short indels, recalibrating base quality scores, and calling single-nucleotide variants (SNVs) and short indels.

Variant call quality was assessed with SAMtools. A minimum quality score of 100 and a minimum coverage of 10× were required. Synonymous or intronic variants other than those affecting consensus splice sites were excluded from further analysis.

The resulting variants were filtered for (1) rare truncating (loss-of-function [LoF]) alterations (nonsense mutations, frameshift indels, and mutations at highly conserved splice sites) and (2) missense variants located in highly conserved nucleotide positions and predicted to be disease causing, damaging, or deleterious by at least two of three in silico analysis tools (PolyPhen-2, MutationTaster, and SIFT).

The variants were selected according to a recessive (presumed biallelic mutations) or dominant (heterozygous mutations) mode of inheritance and an estimated disease frequency of 0.01% in the population. In the dominant and recessive disease models, variants with a minor allele frequency (MAF) of $\geq$0.01% and $\geq$1%, respectively, were considered benign polymorphisms or low-penetrance variants and excluded from further analysis. In addition, recurrent dominant (heterozygous) variants were selected with a less stringent frequency threshold (MAF = 1%). Population allele frequencies are based on data from dbSNP, the 1000 Genomes Project (TGP), the National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project (ESP) Exome Variant Server (EVS), the Exome Aggregation Consortium (ExAC) Browser, and a large in-house exome database containing all germline variants identified in 2,816 exomes of individuals without known

tumor disease and sequenced under similar conditions. To exclude obvious sequencing artifacts, we performed a detailed visual inspection of the remaining variants with a read browser (Integrative Genomics Viewer [IGV]).

We considered only genes affected by potentially pathogenic variants in at least two alleles of the cohort (heterozygous in at least two individuals or homozygous or compound heterozygous in at least one individual). Finally, we inspected all genes carrying the remaining rare variants for the presence of rare, non-polymorphic heterozygous CNVs of $\geq$10 kb in order to identify additional recurrently mutated genes or biallelic, compound-heterozygous variants.[28]

Splicing efficiencies of the normal and mutant sequences were calculated with the following splice prediction programs: Human Splicing Finder,[36] GeneSplicer (University of Maryland Center for Bioinformatics and Computational Biology), MaxEntScan,[37] and NNSPLICE 0.9 (Berkeley *Drosophila* Genome Project).

The etiological relevance of the mutations was further explored by evaluation of their genetic intolerance to functional variation, as measured by the Residual Variation Intolerance Score (RVIS),[38] and the likelihood of haploinsufficiency, as measured by haploinsufficiency scores (from dataset S2, including imputed values).[39] The expression of candidate genes was determined with the expressed-sequence-tag profiles of human colon tissue and protein detection data from human colon tissue (glandular cells) provided by UniGene and the Human Protein Atlas.

Frequency of Colorectal Tumors with Somatic Mutations in Candidate Genes

Data concerning the frequency (percentage) of colorectal tumors with somatic mutations in candidate genes were obtained from the exome database of The Cancer Genome Atlas (TCGA). Somatic variants identified in exome data from colonic (n = 273) and rectal (n = 116) adenocarcinomas were downloaded from the TCGA data portal. To correct the data for the presence of passenger mutations, we excluded hypermutated tumors from the dataset. Therefore, the distribution of somatic variants in the TCGA exomes was analyzed, and all tumors with >200 variants (24% of the tumors) were excluded. We used the remaining 295 exomes (76% of tumors) to calculate the frequency of tumors with somatic mutations in candidate genes.[28]

Sanger Sequencing and Validation

The identified truncating variants were validated via Sanger sequencing of the corresponding region according to standard protocols. We used genomic leukocyte-derived DNA to amplify the genomic region of the respective variant. PCR products were purified and sequenced on an ABI 3500xl Genetic Analyzer (Applied Biosystems). To avoid pseudogene amplification, as described previously,[40] we based Sanger sequencing of *PMS2* on long-range PCR with primers specific to *PMS2*.

Transcript Analysis

Venous blood samples were collected into PAXgene blood RNA tubes (Becton Dickinson). Total RNA was extracted with the PAXgene Blood RNA Kit (QIAGEN) in accordance with the manufacturer's protocol. First-strand cDNA was synthesized from 2–3 μg of total RNA by random hexamer-primed reverse transcription and the SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen GmbH) in accordance with the manufacturer's protocol. RT-PCR fragments were obtained according to standard PCR protocols, and different primers were used for generating the appropriate fragments. RT-PCR products were separated on 2% agarose gel and visualized with ethidium bromide with an UV imaging system (Bio-Rad). Individual bands were excised from the gel and eluted with the High Pure PCR Product Purification Kit (Roche Diagnostics GmbH). Eluted DNA was re-amplified with the same primer pairs and sequenced as described above.

Analysis of Altered *MSH3* Products

MSH3 (mutS homolog 3 [MIM: 600887]) frameshift mutations, c.1148delA and c.2760delC, were generated via site-directed mutagenesis (QuikChange II Kit, Stratagene; primers: 5′-GTTAGGGACAAAAAAAGGGCAACATT-3′ and 5′-AATGTTGCCCTTTTTTTGTCCCTAAC-3′ [c.1148delA] and 5′-GGCTCAGATTGGCTCTATGTTCCTGCAGAAG-3′ and 5′-CTTCTGCAGGAACATAGAGCCAATCTGAGCC-3′ [c.2760delC]) with the wild-type pcDNA3.1$^-$/*MSH3*-WT vector (kindly provided by Grazia Graziani, Italy).[41] Plasmids were confirmed by sequencing. To mimic the splice-site variants *MSH3* c.3001−2A>C and c.2319−1G>A, resulting cDNAs including the premature stop codons were synthesized (Gene Art) and subcloned into a pcDNA3.1$^+$ expression vector.

Transient transfection was carried out with HEK293T cells as described previously.[42] In brief, HEK293T cells were transfected at 50%–70% confluence with expression plasmid pcDNA3.1$^-$/*MSH3*-WT, pcDNA3.1$^-$/*MSH3*-c.1148delA, pcDNA3.1$^-$/*MSH3*-c.2760delC, pcDNA3.1$^+$/*MSH3*-c.3001−2A>C, or pcDNA3.1$^+$/*MSH3*-c.2319−1G>A (0.5 μg/ml, respectively) with the use of 2 μl/ml of the cationic polymer polyethylenimine (Polysciences; stock solution 1 mg/ml). 48 hr after transfection, cell extracts were prepared for western blot analysis with anti-MSH3 (H-300, Santa Cruz Biotechnologies) and anti-β-actin (Sigma). Fluorescence signals (680 and 800, Li-Cor) were detected in a FLA-9000 (Fujifilm).

The effect of whole exon deletions on protein structure was illustrated in silico. MSH3 structure was obtained from the Protein Data Bank (PDB: 3THY; MutSβ complexed with an indel loop of two bases and ADP).[43] We mapped the amino acids coded by exons 17 and 22 to the MSH3 structure by using the PyMOL Molecular Graphics System (version 1.7.0.0, Schrödinger).

Immunohistochemistry of MMR Proteins

Immunohistochemical (IHC) staining of formalin-fixed, paraffin-embedded (FFPE) tissue samples was performed according to established routine procedures on a fully automated Bond-III IHC stainer (Leica) according to the manufacturer's protocol with the following primary antibodies: MLH1, MSH2, MSH6, and PMS2 (all purchased from Leica). The amount of protein staining in tumor cells was compared to that in normal tissue. The amount of MMR protein was considered deficient if the nuclei showed no or only very weak immunostaining in relation to normal tissue.

IHC of MSH3 was performed on 2–3 μm FFPE tissue specimens with an automated staining system (480 S Autostainer, Medac). For antigen retrieval, a pre-treatment module (Medac) was used. A rabbit polyclonal antibody for MSH3, raised against an NH2-terminal polypeptide comprising amino acids 1–200, was used at a dilution of 1:100.[44,45] The reaction was developed with a horseradish-peroxidase (HRP)-conjugated detection system (C-DPVB 500 HRP, Medac) and the 3,3′-diaminobenzidine system (495192F, Medac).

Targeted Sequencing of Tumor Tissue

DNA was extracted from 10 μm FFPE tissue sections. After deparaffinization, tumor tissue was macrodissected from unstained slides. A previously marked H&E-stained slide served as a reference. Extraction of FFPE tissue DNA was carried out with the BioRobot M48 robotic workstation and the corresponding MagAttract DNA Mini M48 Kit (QIAGEN) or the Maxwell 16 FFPE Tissue LEV DNA Purification Kit (Promega) in accordance with the manufacturer's protocol. Analysis of microsatellite status was performed according to the previously described methods.[46] For examination of somatic *APC* mutations, high-coverage targeted sequencing (read depths > 1,000) was performed with the FAP MASTR Kit (Multiplicom) on a MiSeq platform (Illumina) in one person (individual 1661.1). The results were analyzed with SeqPilot software (JSI Medical Systems).

Microsatellite Analysis

Microsatellite analysis was performed on matched tumor and normal DNA samples by conventional fragment analysis or next-generation-sequencing-based analysis as previously described.[46] This involved using the National Cancer Institute panel of reference markers to evaluate microsatellite instability (MSI) in CRC. This panel consists of two mononucleotide (BAT25 and BAT26) and three dinucleotide (D2S123, D5S346, and D17S250) repeats.[47–49] Tumor DNA was extracted from microdissected tumor tissue. Normal DNA was extracted from normal tissue or peripheral-blood leukocytes. Tumors were scored as highly instable (MSI-H) if two or more of these five markers exhibited additional alleles and as stable (MSS) if none of the five markers showed instability.

A second panel of five markers was complemented. This consisted of one tetranucleotide and four dinucleotide repeats (BAT40, D10S197, D13S153, *MYCL1*, and D18S58). The tumor was classified as MSI-H if two or more of the ten markers exhibited instability and as MSI-low (MSI-L) if only one marker exhibited additional alleles.

For detecting elevated microsatellite instability at selected tetranucleotide repeats (EMAST), DNA from tumor and normal tissue was analyzed with five more tetranucleotide repeat markers (D20S82, D2S443, D21S1436, D9S747, and UTS037) as described elsewhere.[50]

Results

To identify high-penetrance germline mutations causing colorectal adenomatous polyposis and located in genes not related to polyposis so far, we performed WES of leukocyte-derived DNA in 102 unrelated individuals with unexplained adenomatous polyposis. Most of the individuals presented with an attenuated colorectal phenotype (late-onset disease and/or <100 colorectal adenomas). The mean age at diagnosis was 44 years (range = 14–73 years). The majority of individuals in the whole cohort had no evidence of extracolonic lesions and were sporadic or isolated cases. The basic clinical features of the cohort are summarized in Tables S2 and S3.

The median coverage of mapped reads was 56× (66% on-target), and 84% of bases were covered at ≥20×. The overall performance of exome sequencing is described in Table S1. A principal-component analysis demonstrated that all but one of the participants were of central European origin (Figure S1). The outlier was excluded from further analysis. Two further persons were removed as a result of low coverage. A mean of 30,152 SNVs per sample was called in the coding and flanking intronic regions. Assuming a dominant or recessive disease model, we applied a number of stringent filter steps to select for rare, non-polymorphic, truncating (LoF) variants.

We identified two unrelated individuals each carrying two different mutations in *MSH3* (Figures 1 and 2, and Figure S2). We also detected one unrelated person with two different mutations in *PMS2*. Furthermore, potentially pathogenic germline variants were identified in 14 additional protein-coding genes (Figure S3 and Table S4).

Clinical Description of Individuals Carrying Biallelic *MSH3* Mutations

Individual 1275.1 (II-4 in Figure 1A) is female and was diagnosed with colorectal adenomatous polyposis at age 36 years. She underwent a preventive sigmoidectomy at age 48 years and a right hemicolectomy at age 53 years. Histology results were available for >40 polyps, all of which were tubular or tubulovillous adenomas with low to intermediate dysplasia, often accompanied by inflammatory infiltration. Three distal hyperplastic polyps were also documented. In addition, this person has a history of proliferative disorders in other organs: thyroid adenoma at age 35 years, a small polyp of the corpus uteri and uterine leiomyomas at age 44 years, multiple small intraductal papillomas of (peripheral) mammary glands at age 44 years, and multiple adenomatous polyps in the duodenum at age 50 years. Hypertrophy of the retinal pigment epithelium was excluded by ophthalmological examination at age 50 years.

Individual 1661.1 (II-2 in Figure 1B) is also female and was diagnosed at age 32 years with colorectal tubular and tubulovillous adenomas with low-grade intraepithelial neoplasia. At age 42 years, she underwent proctocolectomy and excision of large duodenal adenomas. This individual has a striking past medical history: at age 26 years, a grade II astrocytoma was diagnosed and surgically treated. At age 27 years, she underwent oophorectomy for the presence of ovarian cysts, including one dermoid cyst. A hysterectomy was performed for a myoma at age 34 years, and a thyroidectomy was performed for follicular adenomas at age 42 years. At age 43 years, she showed a cutaneous fibrolipoma, and at age 46 years, a flat epithelial atypia, multiple peripheral small intraductal papillomas, usual ductal hyperplasias, and cysts with apocrine metaplasia were detected in the mammary glands.

Both index persons have one affected sibling, whereas their respective parents have no reported history of malignant gastrointestinal disease (Figure 1). A sister (individual 1275.2; II-1 in Figure 1A) of individual 1275.1 (II-4 in Figure 1A) was diagnosed with a rectal adenocarcinoma

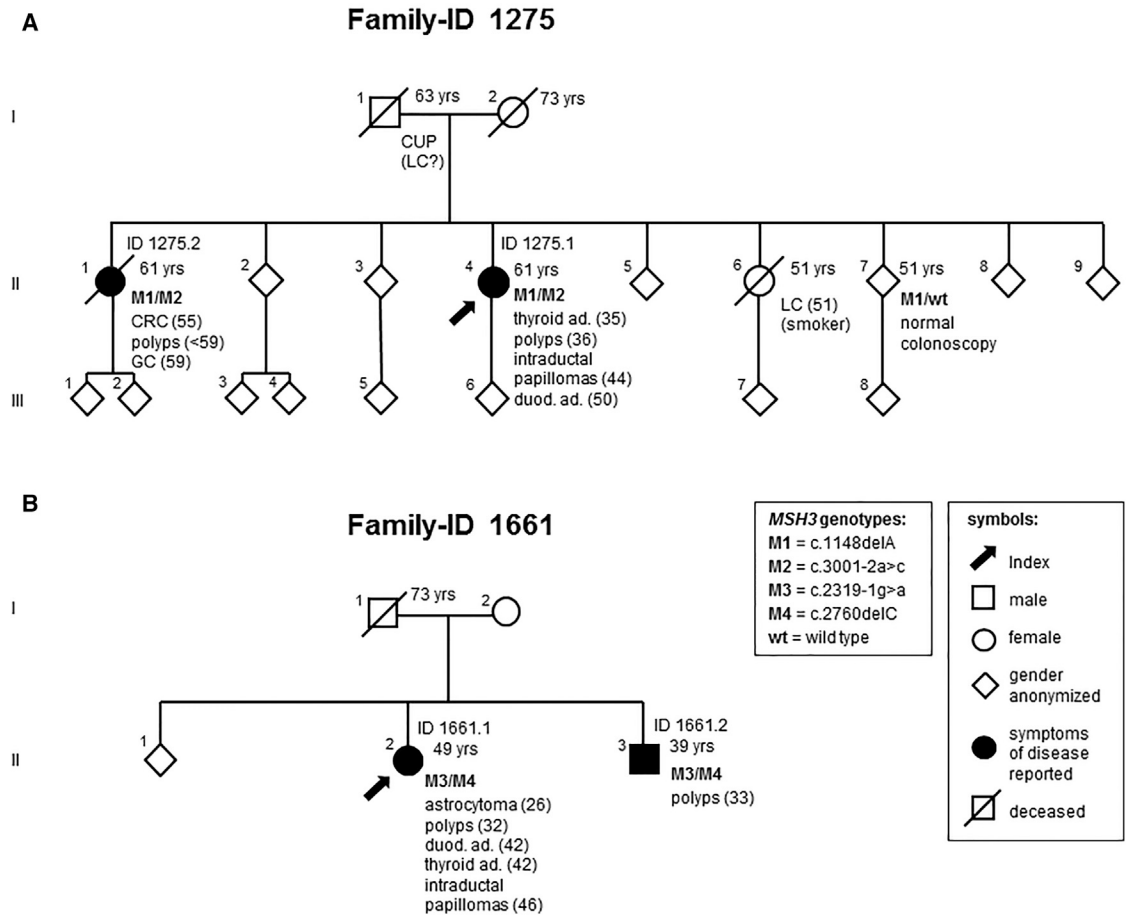


Figure 1. Pedigrees of the Two Index Individuals with Biallelic *MSH3* Germline Mutations
Pedigrees of family 1275 (A) and 1661 (B). The index persons are indicated by arrows (see main text for details). Above the symbols, identifiers are given for affected individuals. The number on the upper right side of a symbol displays the age at death, or in living persons, the age at last contact. On the lower right, genotype and phenotype information is displayed. The numbers following a disease represent the age at first diagnosis. Abbreviations are as follows: ad, adenomas; CRC, colorectal carcinoma; CUP, cancer of unknown primary; duod, duodenal; GC, gastric cancer; LC, lung cancer; polyps, multiple colorectal adenomatous polyps; and yrs, years.

at age 56 years and with a signet cell gastric carcinoma at age 59 years. The available histology reports describe multiple tubulovillous adenomas of the entire colon and proximal duodenum, with up to high-grade intraepithelial neoplasia, and two hyperplastic polyps of the transverse colon. Small bilateral renal cysts were reported as a secondary finding. The brother (individual 1661.2; II-3 in Figure 1B) of individual 1661.1 (II-2 in Figure 1B) was diagnosed with colorectal polyps at age 33 years; he underwent colectomy at age 37 years.

Characterization of the *MSH3* Mutations

In total, the two index persons harbor four different *MSH3* variants, all with a putative LoF effect. *MSH3* (GenBank: NM_002439.4) on 5q14.1 is one of the six MMR genes identified to date in eukaryotic cells.[51] It consists of 24 exons and encodes a protein composed of 1,137 amino acids, including several functional domains (Figure 2). In both families, each affected individual carries one frameshift and one splice-site mutation (Figures 1 and 2). All mutations were validated by Sanger sequencing (Figure 2).

The frameshift mutation c.1148delA (p.Lys383Argfs*32; chr5: g.79970921delA) in exon 7 in family 1275 is predicted to result in a premature stop codon after 31 amino acids. It corresponds to a known somatic cancer mutation located in a poly-A(8) tract of exon 7. The frameshift mutation c.2760delC (p.Tyr921Metfs*36; chr5: g.80109507delC) in exon 20 in family 1661 is predicted to result in a premature stop codon after 35 amino acids. After transfection of *MSH3* plasmids with each frameshift mutation in HEK293T cells, we demonstrated via western blot that the altered proteins are shortened by the expected length (Figures S4A and S4C).

The splice-site mutations c.2319−1G>A (chr5: g.80074538G>A) in intron 16 in family 1661 and c.3001−2A>C (chr5: g.80160630A>C) in intron 21 in family 1275 are both located in highly conserved splice acceptor sites and are predicted to alter splicing in four of four prediction tools. To demonstrate their functional effect, we performed a transcript analysis with primers located in flanking exons (Figures 3A and 3B). RT-PCR products obtained from cDNA of individual 1661.1, who carries the c.2319−1G>A mutation, showed two bands

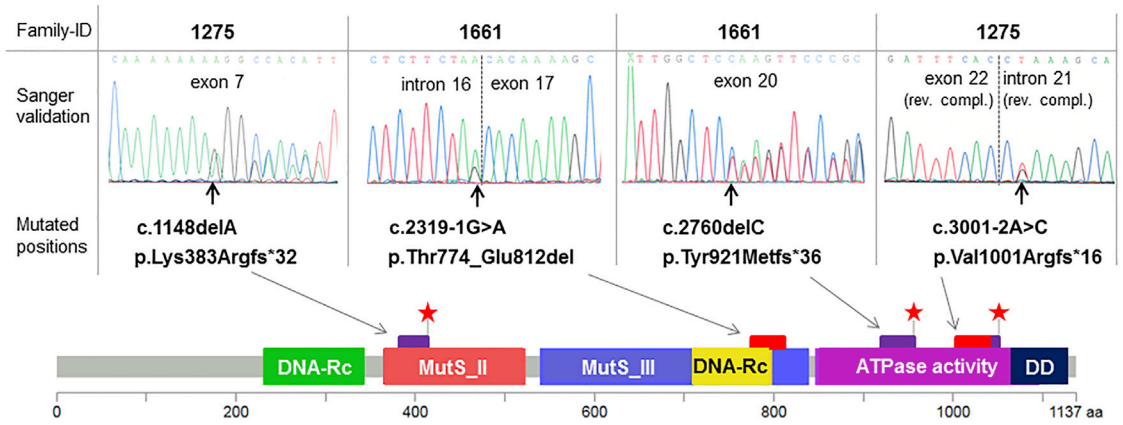


Figure 2. Germline Mutations in *MSH3*
All four *MSH3* mutations identified in the present study were validated by Sanger sequencing of the respective exons with flanking intronic regions. Their protein-level effects are depicted below in an adapted lollipop plot of MSH3 (created with MutationMapper). A premature stop is symbolized by a red star, a purple bar denotes altered amino acids, and a red bar denotes lost amino acids. Regions with conserved functional residuals[43] are highlighted. Abbreviations are as follows: DD, dimerization domain; DNA-Rc, DNA-recognition; and MutS_II-III, PFAM domains of the MutS family.

on an agarose gel. Sequencing of the shortened transcript confirmed a loss of exon 17, which is predicted to result in an in-frame loss of 39 amino acids (774–812) at the protein level. These amino acids are involved in DNA recognition (Figures 2 and S5).[43] Sequencing of the short RT-PCR fragment from individual 1275.1 (II-4 in Figure 1A), who has the c.3001−2A>C mutation, confirmed a loss of exon 22. This is predicted to result in a frameshift mutation with a premature stop codon after 16 amino acids, altering the dimerization domain (Figures 2 and S5). In addition, we demonstrated via western blot that *MSH3* cDNA variants lacking the respective exon lead to altered proteins shortened by the expected length (Figures S4B–S4C).

In control individuals (ExAC Browser), the variant c.1148delA is listed ten times (as "chr5:79970914CA/C"), and the variant c.2760delC is listed twice (as "chr5:80109505TC/T") in a heterozygous state, corresponding to an allele frequency of 0.008% and 0.0016%, respectively. The two other *MSH3* variants are not reported in the general population, and none of the variants are listed in the Human Gene Mutation Database (HGMD). To determine the frequency of *MSH3* LoF mutations in the normal population, we queried large exome datasets from control individuals (TGP, EVS, and ExAC Browser). Although these datasets listed heterozygous LoF *MSH3* mutations with a MAF of <0.2%, no homozygous mutation was reported.

We were able to obtain paraffin-embedded tumor and adjacent normal tissue from the affected sister 1275.2 (II-1 in Figure 1A) and a blood sample from the affected brother 1661.2 (II-3 in Figure 1B) and found the same two *MSH3* mutations as those present in the index individual (data not shown).

To confirm compound heterozygosity, we gathered a leukocyte-derived DNA sample from one unaffected sibling (II-7 in Figure 1A) of individual 1275.1 (II-4 in Figure 1A). Examination of both regions mutated in the index individual revealed that the unaffected sibling carries only the frameshift mutation in the heterozygous state and not the splice-site mutation (Figures 1 and 3C). For individual 1661.1 (II-2 in Figure 1B), the biallelic genotype was confirmed via transcript analysis: by creating an amplicon spanning both mutated regions, we could demonstrate that the transcript of normal length carries the single-nucleotide deletion resulting in a frameshift mutation. In contrast, this single-nucleotide deletion was not detected in the shortened product transcribed from the allele carrying the splice-site mutation (Figure 3D).

To assess potential additional causal alterations, we performed supplementary sequencing and MLPA of genes with influence on *MSH3*: neither of the two index individuals showed a germline mutation in other MMR genes (*MLH1*, *MSH2*, *MSH6*, or *PMS2*), *EPCAM*, or *TP53* (tumor protein p53 [MIM: 191170]), any evidence of *APC* mosaicism in leukocyte DNA, or somatic *MSH2* or *MSH6* mutations in tumor tissue. Furthermore, IHC staining of two adenomas per index person demonstrated a strong presence of the MMR proteins MLH1, MSH2, MSH6, and PMS2 (Figure S6).

Effect of MSH3 Deficiency on Colorectal Tumor Tissue

IHC staining with a rabbit polyclonal antibody proved complete nuclear loss of MSH3 in normal colon mucosa of individual 1275.1 and in adenomas of both index persons (Figure 4). In the control samples, non-tumorous mucosa of an independent person with CRC showed MSH3 predominantly located in the nuclei.

In adenoma-derived DNA from both index persons, we found stability of mononucleotide repeats. Complementarily, we examined several dinucleotide and tetranucleotide

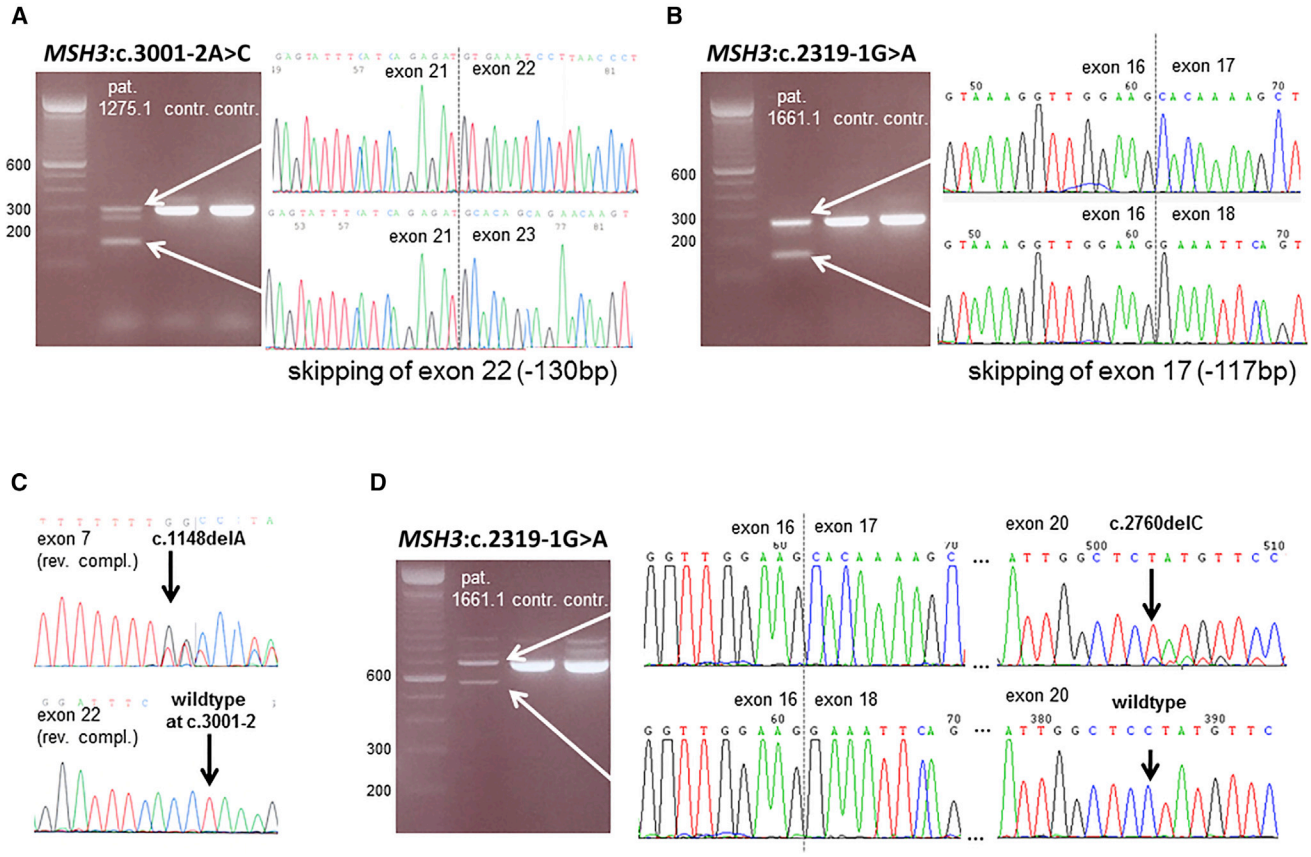


Figure 3. Transcript Analyses of *MSH3* Splice-Site Mutations and Confirmation of Compound Heterozygosity
(A) An agarose gel shows RT-PCR products obtained from mRNA of individual 1275.1 (lane 1) and control samples (lanes 2–3); primers are localized in exon 21 (forward) and exon 23 (reverse). In the affected individual's sample, a shorter band of around 165 bp is visible in addition to the expected band of 295 bp. Sequencing of the shorter fragment demonstrated a loss of exon 22 caused by the c.3001−2A>C mutation.
(B) An agarose gel shows RT-PCR products obtained from mRNA of individual 1661.1 (lane 1) and control samples (lanes 2–3); primers are localized in exon 16 (forward) and exon 18 (reverse). In the affected individual's sample, a shorter band of around 162 bp is visible in addition to the expected band of 279 bp. Sequencing of the shorter fragment demonstrated a loss of exon 17 caused by the c.2319−1G>A mutation.
(C) The mutation sites in family 1275 were amplified and sequenced from lymphocyte DNA of an unaffected sibling. Only one of the mutations was detectable in the heterozygous state, confirming that the two mutations are located on separate alleles.
(D) Agarose gel shows RT-PCR products obtained from mRNA of individual 1661.1 (lane 1) and control samples (lanes 2–3); primers are localized in exon 16 (forward) and exon 21 (reverse) in order to capture the effects of the splice-site mutation and the frameshift mutation within the same amplicon. Sequencing the shorter band proved that the product of aberrant splicing does not carry frameshift mutation c.2760delC in exon 20, whereas the frameshift mutation is present in the longer band. This confirms that the two mutations are located on distinct transcripts and thus separate alleles (compound-heterozygous state).

markers in order to focus on lesions that are processed by *MSH3*. Individuals 1275.1 and 1661.1 exhibited instability of one of four (MSI-L) and three of four (MSI-H) dinucleotide markers, respectively. In addition, individuals 1275.1 and 1275.2 showed instability of three of six and three of five tetranucleotide markers, respectively, whereas individual 1661.1 displayed instability of four of six tetranucleotide markers (Figures 5 and S7). Thus, we demonstrated EMAST in tumors of all three examined individuals.

In addition, we used the adenoma-derived DNA of individual 1661.1 to screen for somatic *APC* mutations. Using targeted deep sequencing, we compared four independent polyps with leukocyte DNA and found seven different somatic *APC* mutations (one to two per polyp) in 6%–36% of the reads (Figure S8). All mutations were small deletions of two to eight nucleotides. In four of seven mutations, the sequence context proved to be di- or trinucleotide repeats.

Identification of a Biallelic *PMS2* Mutation

Individual 1138 harbors the *PMS2* germline mutations c.2T>A (p.Met1?) in exon 1 and c.863delA (p.Gln288Argfs*19) in exon 8 (GenBank: NM_000535.5) (Figure S9A). Both mutations were validated by Sanger sequencing (Figure S9B). Compound heterozygosity could be confirmed given that the healthy mother only carries mutation c.863delA in exon 8. The start-loss mutation c.2T>A (p.Met1?) was predicted to be pathogenic or damaging by two of three in-silico tools. In accordance with the assumed protein truncation caused by the two

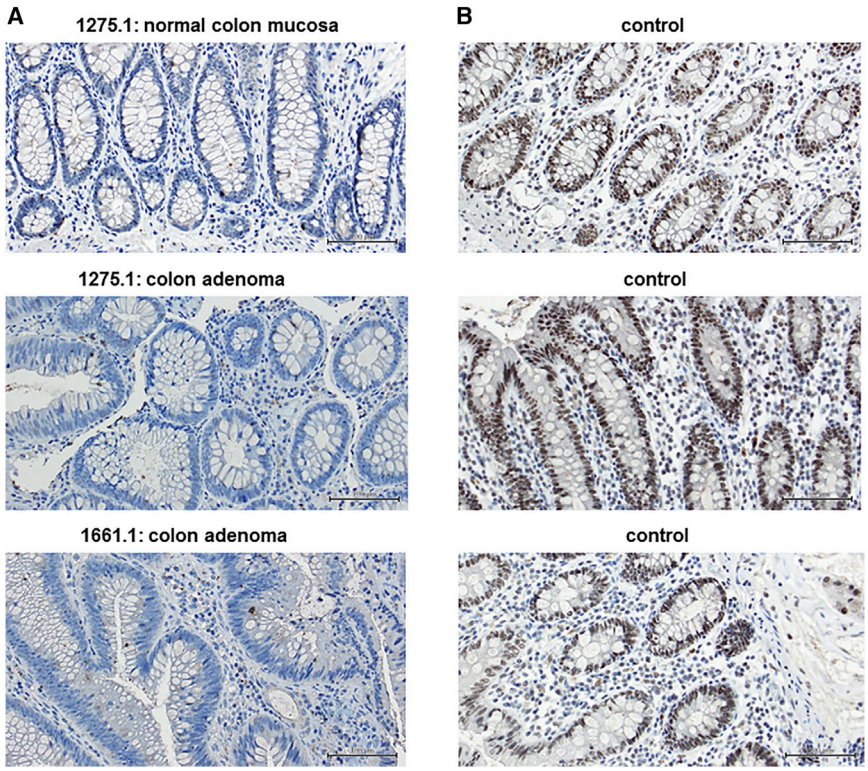


Figure 4. IHC Staining of MSH3 in Tumor and Normal Tissue
MSH3 was stained with rabbit polyclonal antibody targeting N-terminal human MSH3.
(A) MSH3 was nearly undetectable in FFPE normal colon mucosa and colorectal adenoma samples from index person 1275.1 and a colorectal adenoma sample from index person 1661.1, who has a complete loss of nuclear MSH3.
(B) In contrast, control samples, which were taken from FFPE normal mucosa of an independent subject with colon cancer and were processed in parallel, show a strong nuclear MSH3 staining. The same results were obtained in an examination of a second independent colorectal adenoma from both index persons (data not shown).
Scale bars represent 100 μm.

germline mutations, IHC staining showed complete loss of PMS2 in both tumor and normal tissue (Figure S9C).

Additional clinical information and careful re-evaluation of the medical history revealed that individual 1138 had been diagnosed with early-onset colorectal polyposis with 20–25 adenomas at age 14 years and had undergone proctocolectomy with pouch-anal anastomosis at age 16 years. In addition, a primitive neuroectodermal tumor of the cerebellum had been diagnosed at 4 years, and a history of a pilomatrixoma, thyroid cysts, and three café-au-lait spots had been reported. The family history was unremarkable.

Mutations in Further Candidate Genes

In the remaining individuals with unexplained polyposis, 29 different rare mutations in 14 protein-coding genes were found in 26 other individuals (Table S4 and Figure S3). All genes, apart from one (*DNAJB7* [DnaJ heat shock protein family (Hsp40) member B7 (MIM: 611336)]), are reported to be expressed in colon tissue. Two genes (*MAGT1* [magnesium transporter 1 (MIM: 300715)] and *SLC27A5* [solute carrier family 27 member 5 (MIM: 603314)]) were affected by a homozygous LoF mutation in one individual each, and seven genes (*BTBD9* [BTB domain containing 9 (MIM: 611237)], *CD36* [CD36 molecule (MIM: 173510)], *ECHDC3* [enoyl-CoA hydratase domain containing 3], *SSC5D* [scavenger receptor cysteine rich family member with 5 domains], *UGGT2* [UDP-glucose glycoprotein glucosyltransferase 2 (MIM: 605898)], *WDR35* [WD repeat domain 35 (MIM: 613602)], and *ZC3H8* [zinc finger CCCH-type containing 8]) were recurrently affected by heterozygous LoF mutations. Of these, three genes (*CD36*, *WDR35*, and *ZC3H8*) have been implicated in cell adhesion or apoptosis. Five individuals were found to carry more than one heterozygous mutation. On the basis of the CNV data, we identified no further heterozygous or additional biallelic large duplication or deletion in these genes.

Discussion

Most cases of colorectal adenomatous polyposis are attributable to heterozygous germline mutations in the tumor-suppressor gene *APC* and are thus diagnosed as FAP. However, the few novel subtypes delineated in recent years are caused by genes involved in DNA repair. Whereas heterozygous mutations affecting the proofreading domain encoded by DNA polymerase genes *POLE* and *POLD1* lead to the rare, dominantly inherited PPAP, recessive *MUTYH*-associated polyposis and *NTHL1*-associated polyposis are caused by biallelic germline mutations in BER genes.[3,7,10] After causal variants in those genes were initially described in only a few families, identification of additional cases expanded the mutation spectrum and allowed refinement of the respective phenotypes.[8,9,52,53]

In a number of individuals with colorectal adenomatous polyposis, however, no germline mutation in the established genes has been identified. Although the synchronous or metachronous occurrence of dozens to hundreds of adenomas is strongly suggestive of an underlying genetic basis, so far it remains unclear whether the predisposing genetic factors mainly act in a monogenic fashion or contribute as low- or moderate-penetrance variants to a more complex, oligo- or polygenic trait.

Interestingly, increasing evidence suggests that biallelic germline mutations in MMR genes can result in a phenotype

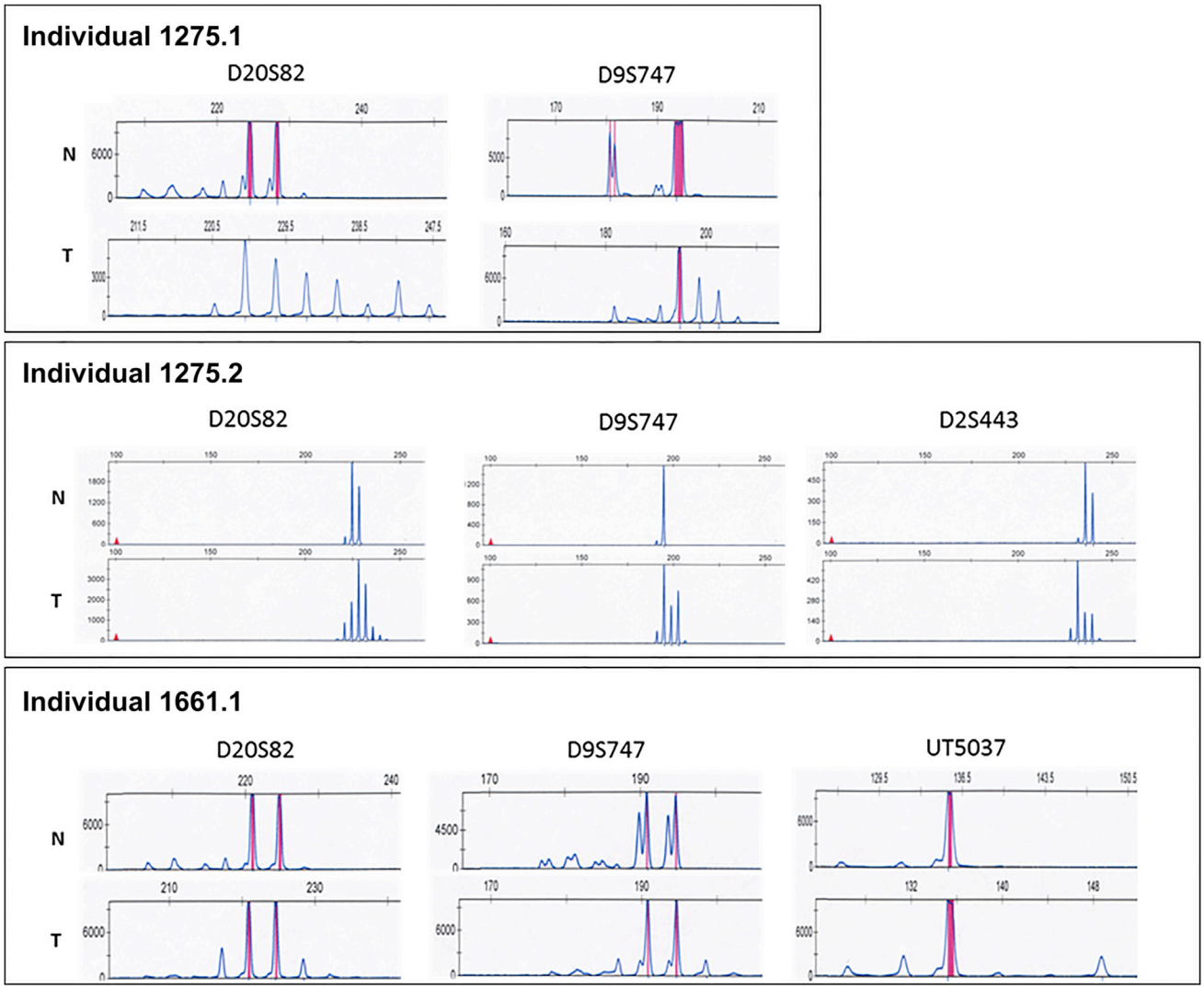


Figure 5. Results of Microsatellite Analysis
Five tetranucleotide repeat markers (D20S82, D2S443, D21S1436, D9S747, and UT5037) were examined in normal (N) and tumor (T) tissue. In each tumor sample, two or three out of five markers showed clear instability and thus demonstrated elevated microsatellite instability at selected tetranucleotide repeats (EMAST). The x axes denote peak positions related to fragment lengths. Please note that only x axes within tumor-normal pairs allow direct comparison. The y axes quantifying peak heights were generally not standardized. The results of both index persons were confirmed in a second colorectal adenoma (data not shown).

with features overlapping those of colorectal polyposis. Typically, these conditions are designated as CMMRD or biallelic MMR deficiency and are characterized by early-onset CRC, brain tumors, hematological malignancies, and café-au-lait skin macules.[31,32] Nonetheless, several individuals with homozygous or compound-heterozygous *PMS2*, *MSH2*, or *MSH6* germline mutations and an early-onset colorectal adenomatous polyposis in the second or first decade of life have been described;[23,31,54] the majority of these cases were until then misclassified as mutation-negative FAP. In some individuals, particularly those with biallelic *PMS2* mutations, however, the colorectal phenotype manifests not before the third or even fourth decade of life, resembling the clinical presentation in the present polyposis cohort.

To uncover further monogenic causes, we performed exome sequencing of leukocyte DNA in a cohort of 102 unrelated individuals with histologically confirmed, genetically unexplained adenomatous polyposis. The clinical and family characteristics of the participants are consistent with published data from other mutation-negative polyposis cohorts.[14,15,19,55] Using this approach, we identified two families affected by biallelic LoF germline mutations in the MMR gene *MSH3*, a genotype that has not yet been described as causative for a polyposis phenotype. In addition, we found one individual with biallelic *PMS2* mutations and several persons harboring homo- and heterozygous LoF variants.

Biallelic *MSH3* Germline Mutations

The genotypes and pedigrees of the two unrelated persons with compound-heterozygous *MSH3* germline mutations are in full agreement with a recessively inherited trait.

Interestingly, unlike the majority of the examined cohort, these two individuals do have affected siblings and documented extraintestinal neoplasias. Neither index person has a germline mutation in any of the known genes associated with gastrointestinal polyposis nor any further mutation in other MMR genes or *EPCAM*. In addition, IHC staining of MLH1, MSH2, MSH6, and PMS2 was normal. The haploinsufficiency score (0.486) of MSH3 indicated a rather low probability of haploinsufficiency (16.2%).[39] In large control sets, none of the LoF *MSH3* germline mutations were identified in the homozygous state, and the frequency of heterozygosity is compatible with a rare recessive disease.

Confirming compound heterozygosity is critical to demonstrating recessive inheritance; in both families, we have clearly shown that the two mutations are located on different alleles either by examination of an unaffected sibling who is heterozygous for just one mutation (family 1275) or by transcript analysis indicating that both mutations are located on different alleles (family 1661). Taken together, these data strongly support the hypothesis that deleterious *MSH3* mutations follow a recessive mode of inheritance.

The MMR system is a critical pathway that corrects base-base and indel mispairs occurring as a result of replication errors, thus increasing the fidelity of DNA replication.[56] Defects in the MMR system result in a mutator phenotype, which manifests as MSI in the DNA of affected cells. In tumors with MSI, microsatellite loci containing mono-, di-, tri-, and tetranucleotide repeats can be affected.[57]

During DNA repair, mispaired bases are recognized by two heterodimers of MutS homologs (DNA mismatch recognition complex)—MSH2 and MSH6 (MutSα) and MSH2 and MSH3 (MutSβ)—with partially overlapping mispair-recognition specificities.[56] In humans, MutSα efficiently binds single-base substitutions and small (single-base) indel mispairs, whereas MutSβ has a stronger affinity for larger base-indel loops with up to ten unpaired nucleotides.[57,58] Thus, loss of MutSβ due to *MSH3* inactivation in human cells not only results in MSI at loci containing dinucleotide repeats but also results in MSI at certain loci with tetranucleotide repeats, termed EMAST.[57,59] It is known that di- and tetranucleotide repeats are affected in the majority of CRC with MSI-L.[58] MMR deficiency can also result from an imbalance in the relative amounts of MSH3 or MSH6.[60]

All four *MSH3* germline mutations detected in the present cohort are strongly predicted to have a LoF effect. According to previous work, somatic *MSH3* frameshift mutations at the (A)8 repeat in exon 7 result in a loss of MSH3.[45,61] To evaluate the pathophysiological consequences of the four mutations in more detail, we performed several experiments. Using transcript analysis, we confirmed aberrant splicing caused by the two mutations located within the conserved consensus splice motifs. This would affect regions relevant for dimerization and for DNA recognition, according to the MSH3 structure described by Yang's group (Figure S5).[43]

Three of the four identified *MSH3* mutations are predicted to result in premature stop codons and thus might lead to nonsense-mediated mRNA decay (NMD). However, mRNA analysis, performed with fresh blood samples not treated with NMD inhibitors, demonstrated that the affected transcripts are expressed. This might be due to NMD escape, a phenomenon that is known from several mutations in other polyposis and MMR genes and that was also described for *MSH3*. You et al. found that *MSH3* transcripts with a frameshift mutation at the (A)8 repeat in exon 7 are not degraded by NMD but instead experience repression of protein translation.[62]

The western blot experiments illustrated that the altered MSH3 proteins are shortened by the expected length (Figure S4), which would lead to a loss of the conserved C-terminal dimerization domain (Figures 2 and S5).[43] Because the stable proteins were obtained in vitro in a human embryonic kidney cell line and a strong promoter for high-level expression, this observation is not per se transferable to the in vivo situation. In fact, IHC staining clearly demonstrated loss of nuclear MSH3 in both normal and colorectal tumor tissues of the affected individuals, confirming the expected MSH3 deficiency. Different mechanisms such as repressed protein translation or blocked nuclear transport by hampered dimerization or changes in protein conformation might explain the nuclear absence of MSH3.

Microsatellite analysis of adenoma-derived DNA demonstrated EMAST, high and low instability at dinucleotide markers, and no instability at mononucleotide repeats in any of the examined tumors. These findings further confirm the functional relevance of the *MSH3* mutations. In addition, presumed effects of the MSH3 deficiency are well reflected by the inflammatory infiltration, a characteristic feature of MSI colorectal tumors, and the somatic *APC* mutation spectrum observed in the adenomas.

Several lines of evidence support the causal relevance of MSH3 deficiency to the initiation of genetic instability and tumorigenesis. Around 50% of MSI tumors contain somatic frameshift mutations in the (A)8 tract in codons 381–383 of *MSH3*.[57,63,64] The detection of LOH in some of these tumors supports the role of *MSH3* and *MSH6* as primary mutators. In CRC and human colon epithelial cells, MSH3 deficiency is associated with EMAST ($[AAAG]_n$ repeats) and MSI-L at dinucleotide repeats and results in the formation of double-strand breaks and significant changes in the proteome.[57,65]

In yeast and extracts of $Msh3^{-/-}$ cells, *Msh3* deficiency leads to a partial MMR defect and MSI.[66–68] In mouse models, elimination of either *Msh3* or *Msh6* alone still maintains some functional MMR activity, which is consistent with the persistence of the MutSα or MutSβ heterodimer, respectively. Of all MMR-knockout mice, *Msh3*-deficient mice exhibited the lowest, yet still significantly elevated, mutation frequencies in comparison to wild-type mice.[68]

Whereas *MLH1*, *MSH2*, *MSH6*, and *PMS2* are established genes associated with Lynch syndrome, the causal

relevance of *MSH3* germline variants in cancer predisposition has remained uncertain until now. To date, *MSH3* mutations have neither been consistently linked to a Lynch-like phenotype nor described in polyposis cases. In several previous studies, common *MSH3* polymorphisms were significantly associated with CRC and prostate cancer as low-penetrance risk alleles.[69–72] In contrast, a potentially high-penetrance pathogenic *MSH3* germline mutation has very rarely been identified in persons with a suspected predisposition to cancer.[73,74] *Msh3*-deficient mice develop late-onset MSI gastrointestinal cancers. However, given the small number of reported tumors, the significance of this finding remains unclear, and survival did not differ significantly from that of wild-type control animals.[67,75,76]

In a Chinese cohort with suspected familial breast cancer, Yang et al. found three heterozygous *MSH3* germline variants (two in-frame deletions and one frameshift mutation; Table S5). They examined eight tumor samples from three families, and all showed MSS in the standard marker panel (BAT25, BAT26, D2S123, D5S346, and D17S250) and, on average, two instable loci in nine additional dinucleotide or EMAST markers. Two individuals showed less MSH3 staining in tumor tissue (breast and ovary) than was shown in normal tissue.[74] One family showed no evidence of Lynch syndrome; the *MSH3* in-frame variant segregated incompletely with the disease, and MSH3 staining showed no relevant deficiency in the tumors. The second family met the clinical criteria for Lynch syndrome, and the *MSH3* in-frame germline variant segregated well with the disease in three generations. However, the tumor spectrum was broad and included breast, ovarian, renal, and colon cancer. It was not reported whether genetic causes of Lynch syndrome had been excluded systematically. The third family carried the *MSH3* frameshift variant, which incompletely segregated with the disease. A comparison with currently available frequency data in the general population (ExAC Browser) suggests that both in-frame deletions are likely to be polymorphisms (Table S5).

In a family with suspected but genetically unexplained Lynch syndrome, Duraturo et al. found that two brothers each with three metachronous CRCs had a compound-heterozygous *MSH3* genotype comprising a potentially pathogenic missense variant and a silent variant.[73] However, no functional data to confirm the pathogenicity of the variants was reported. Moreover, the silent *MSH3* variant is meanwhile listed as a frequent polymorphism (rs1805355; Table S5).

In our study, all carriers of biallelic *MSH3* mutations have attenuated colorectal and duodenal involvement and no or late-onset cancer. This is similar to the phenotype observed in persons with MAP or attenuated FAP and is consistent with the phenotype described in *MSH3*-knockout mice. Two of the four carriers are reported to have extraintestinal tumors: whereas various thyroid neoplasias also occur in FAP and MAP individuals, the early-onset astrocytoma fits well in the tumor spectrum observed in CMMRD.

A high frequency of EMAST was also observed in a wide range of extraintestinal sporadic malignancies, such as skin, bladder, kidney, lung, ovarian, head, and neck cancer,[59,77,78] although the underlying mechanism remained unclear, and an association with MSH3 impairment was not proven. Recent studies, however, provide strong evidence that EMAST formation is driven by MSH3 deficiency either as a result of *MSH3* mutations or, e.g., by a nuclear-to-cytosol shift induced by oxidative stress.[79,80] Thus, it can be speculated that *MSH3*-induced EMAST is more common than previously thought and might occur in different tumor types. Consequently, the tumor spectrum in individuals with biallelic *MSH3* germline mutations might include a much broader extraintestinal tumor spectrum than observed in the persons identified in the present study.

Although the clinical information and underlying molecular changes point to a broader tumor spectrum and some degree of overlap with CMMRD, exploring the whole oncologic phenotype will require further individuals with biallelic *MSH3* mutations.

Biallelic *PMS2* Germline Mutations

The identification of one individual with a biallelic *PMS2* mutation demonstrates that CMMRD is a rare but important cause of adenomatous polyposis. The c.2T>A (p.Met1?) start-loss mutation is listed twice in ClinVar and is considered pathogenic. According to data from the ExAC Browser, the allele frequency of this mutation in the European population is 0.003%. Apart from the individual reported here, we recently identified another person with a CMMRD phenotype (B cell lymphoma, acute lymphatic leukemia, carcinoma of the rectum, and a multifocal grade III-IV astrocytoma occurring between the ages of 9 and 15 years) in a multiple-tumor cohort. This person carries similar start-loss (c.1A>T [p.Met1?]) and frameshift (c.2117delA [p.Lys706Serfs*19]) mutations in *PMS2*. A similar potential founder mutation in the start codon (c.1A>G) was found in three unrelated CMMRD individuals, all of whom have a compound-heterozygous *PMS2* genotype and isolated loss of PMS2 on IHC staining.[81] The *PMS2* locus-specific database (Leiden Open Variation Database [LOVD]) lists a fourth family with the same genotype. Although the person identified in the present study has extracolonic features suggestive of CMMRD, these manifestations are often unspecific and could remain unreported or unrecognized (e.g., café-au-lait spots). This suggests that CMMRD is an underdiagnosed condition and should be included in the differential diagnosis of any unexplained early-onset case of adenomatous polyposis.

Further Candidate Genes

Under the assumption of a monogenic mode of inheritance with high penetrance, the frequency of causative germline mutations in the general population is expected to be low. In 26 of the 96 remaining individuals (excluding three samples after quality control and three resolved

cases), we identified unique (i.e., not present in control individuals) or rare (i.e., frequency < 0.01% for the dominant or 1% for the recessive model in control individuals) potentially pathogenic germline variants in 14 protein-coding genes. The causative relevance of these interesting candidate genes awaits exploration in larger cohorts and via functional analysis.

In the present study, mutations might have been overlooked, e.g., in low-coverage regions or within repeat tracts in coding sequences. Moreover, some causative mutations might be located beyond the exome, e.g., in non-coding regions or in unannotated genes.

Conclusions

In conclusion, this study describes the identification of biallelic pathogenic *MSH3* germline mutations as a cause of an inherited tumor syndrome. Specifically, biallelic LoF *MSH3* germline mutations appear to cause an additional rare recessively inherited subtype of colorectal adenomatous polyposis, which was present in 2% of the study participants. Data from the present and previous studies consistently show that mutations in newly identified genes associated with inherited tumor predisposition syndromes are very rare (0.3%–0.5% in unexplained polyposis cohorts with familial cancer).[7,82] At least some of these syndromes appear to show extreme genetic heterogeneity, and identifying recurrently mutated genes will require large cohorts.

Preliminary experiments indicate that MSH3-deficient cells are more sensitive to cisplatin treatment or platinum-based adjuvant treatment for CRC.[57] Thus, MSH3 deficiency might also be of therapeutic relevance for individuals with *MSH3*-associated polyposis.

Accession Numbers

The accession numbers for the variant data reported in this paper are LOVD: 00074677 and 00074678.

Supplemental Data

Supplemental Data include nine figures and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2016.06.015.

Conflicts Of Interest

M.M.N. is managing scientific director of Life & Brain GmbH.

Acknowledgments

We thank the individuals and their families for participating in the study and Prof. Grazia Graziani (University of Rome Tor Vergata, Italy) for generously providing the pcDNA3.1^{-}/*MSH3*-WT vector. This work was supported by the German Cancer Aid (grant no. 108421, Deutsche Krebshilfe, Bonn), the Gerok-Stipendium of the University Hospital Bonn (grant no. O-149.0098), and NIH Centers for Mendelian Genomics (5U54HG006504). R.C.B. and M.M.N. are members of the Excellence Cluster ImmunoSensation, funded by the German Research Foundation (Deutsche Forschungsgemeinschaft). The funding sources had no involvement in the study design; the collection, analysis, or interpretation of data; the writing of the report; or the decision to submit the paper for publication. The corresponding author (S.A.) had full access to all data in the study and had final responsibility for the decision to submit the manuscript for publication.

Received: February 25, 2016
Accepted: June 14, 2016
Published: July 28, 2016

Web Resources

1000 Genomes, http://www.1000genomes.org
Berkeley *Drosophila* Genome Project NNSplice 0.9, http://www.fruitfly.org/seq_tools/splice.html
ClinVar, http://www.ncbi.nlm.nih.gov/clinvar/
dbSNP, www.ncbi.nlm.nih.gov/SNP/
Ensembl (release 54), http://may2009.archive.ensembl.org/index.html
ExAC Browser, http://exac.broadinstitute.org/
HGMD, http://www.hgmd.cf.ac.uk
IGV, http://www.broadinstitute.org/igv/
International HapMap Project, http://hapmap.ncbi.nlm.nih.gov/
LOVD 2.0, *APC*, http://www.lovd.nl/APC
LOVD 2.0, *MUTYH*, http://www.lovd.nl/MUTYH
LOVD 2.0, *PMS2*, http://www.lovd.nl/PMS2
LOVD 3.0, *MSH3*, http://databases.lovd.nl/shared/genes/MSH3
MutationMapper, http://www.cbioportal.org/mutation_mapper.jsp
MutationTaster, http://www.mutationtaster.org/
NCBI, http://www.ncbi.nlm.nih.gov/
NHLBI Exome Sequencing Project (ESP) Exome Variant Server, http://evs.gs.washington.edu/EVS/
OMIM, http://www.omim.org
PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/
Primer3, http://bioinfo.ut.ee/primer3-0.4.0/primer3/input.htm
Protein Data Bank, http://www.rcsb.org/pdb/home/home.do
RefSeq, http://www.ncbi.nlm.nih.gov/RefSeq
SIFT, http://sift.jcvi.org/
TCGA, https://tcga-data.nci.nih.gov/tcga/
The Human Protein Atlas, http://www.proteinatlas.org/
UCSC Genome Browser, http://genome.ucsc.edu
UMD-*APC* Mutations Database, http://www.umd.be/APC/
UMD-*MUTYH* Mutations Database, http://www.umd.be/MUTYH/
UniGene, http://www.ncbi.nlm.nih.gov/unigene/

References

1. Galiatsatos, P., and Foulkes, W.D. (2006). Familial adenomatous polyposis. Am. J. Gastroenterol. *101*, 385–398.
2. Grover, S., Kastrinos, F., Steyerberg, E.W., Cook, E.F., Dewanwala, A., Burbidge, L.A., Wenstrup, R.J., and Syngal, S. (2012). Prevalence and phenotypes of APC and MUTYH mutations in patients with multiple colorectal adenomas. JAMA *308*, 485–492.
3. Al-Tassan, N., Chmiel, N.H., Maynard, J., Fleming, N., Livingston, A.L., Williams, G.T., Hodges, A.K., Davies, D.R., David, S.S., Sampson, J.R., and Cheadle, J.P. (2002). Inherited variants

of MYH associated with somatic G:C–>T:A mutations in colorectal tumors. Nat. Genet. *30*, 227–232.
4. Mazzei, F., Viel, A., and Bignami, M. (2013). Role of MUTYH in human cancer. Mutat. Res. *743-744*, 33–43.
5. Krawitz, P.M., Schweiger, M.R., Rödelsperger, C., Marcelis, C., Kölsch, U., Meisel, C., Stephani, F., Kinoshita, T., Murakami, Y., Bauer, S., et al. (2010). Identity-by-descent filtering of exome sequence data identifies PIGV mutations in hyperphosphatasia mental retardation syndrome. Nat. Genet. *42*, 827–829.
6. Gilissen, C., Hoischen, A., Brunner, H.G., and Veltman, J.A. (2012). Disease gene identification strategies for exome sequencing. Eur. J. Hum. Genet. *20*, 490–497.
7. Palles, C., Cazier, J.B., Howarth, K.M., Domingo, E., Jones, A.M., Broderick, P., Kemp, Z., Spain, S.L., Guarino, E., Salguero, I., et al.; CORGI Consortium; WGS500 Consortium (2013). Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. Nat. Genet. *45*, 136–144.
8. Valle, L., Hernández-Illán, E., Bellido, F., Aiza, G., Castillejo, A., Castillejo, M.I., Navarro, M., Seguí, N., Vargas, G., Guarinos, C., et al. (2014). New insights into POLE and POLD1 germline mutations in familial colorectal cancer and polyposis. Hum. Mol. Genet. *23*, 3506–3512.
9. Spier, I., Holzapfel, S., Altmüller, J., Zhao, B., Horpaopan, S., Vogt, S., Chen, S., Morak, M., Raeder, S., Kayser, K., et al. (2015). Frequency and phenotypic spectrum of germline mutations in POLE and seven other polymerase genes in 266 patients with colorectal adenomas and carcinomas. Int. J. Cancer *137*, 320–331.
10. Weren, R.D., Ligtenberg, M.J., Kets, C.M., de Voer, R.M., Verwiel, E.T., Spruijt, L., van Zelst-Stams, W.A., Jongmans, M.C., Gilissen, C., Hehir-Kwa, J.Y., et al. (2015). A germline homozygous mutation in the base-excision repair gene NTHL1 causes adenomatous polyposis and colorectal cancer. Nat. Genet. *47*, 668–671.
11. Spier, I., Kerick, M., Drichel, D., Horpaopan, S., Altmüller, J., Laner, A., Holzapfel, S., Peters, S., Adam, R., Zhao, B., et al. (2016). Exome sequencing identifies potential novel candidate genes in patients with unexplained colorectal adenomatous polyposis. Fam. Cancer *15*, 281–288.
12. Knudsen, A.L., Bisgaard, M.L., and Bülow, S. (2003). Attenuated familial adenomatous polyposis (AFAP). A review of the literature. Fam. Cancer *2*, 43–55.
13. Renkonen, E.T., Nieminen, P., Abdel-Rahman, W.M., Moisio, A.L., Järvelä, I., Arte, S., Järvinen, H.J., and Peltomäki, P. (2005). Adenomatous polyposis families that screen APC mutation-negative by conventional methods are genetically heterogeneous. J. Clin. Oncol. *23*, 5651–5659.
14. Thirlwell, C., Howarth, K.M., Segditsas, S., Guerra, G., Thomas, H.J., Phillips, R.K., Talbot, I.C., Gorman, M., Novelli, M.R., Sieber, O.M., and Tomlinson, I.P. (2007). Investigation of pathogenic mechanisms in multiple colorectal adenoma patients without germline APC or MYH/MUTYH mutations. Br. J. Cancer *96*, 1729–1734.
15. Hes, F.J., Ruano, D., Nieuwenhuis, M., Tops, C.M., Schrumpf, M., Nielsen, M., Huijts, P.E., Wijnen, J.T., Wagner, A., Gómez García, E.B., et al. (2014). Colorectal cancer risk variants on 11q23 and 15q13 are associated with unexplained adenomatous polyposis. J. Med. Genet. *51*, 55–60.
16. Venesio, T., Balsamo, A., Rondo-Spaudo, M., Varesco, L., Risio, M., and Ranzani, G.N. (2003). APC haploinsufficiency, but not CTNNB1 or CDH1 gene mutations, accounts for a fraction of familial adenomatous polyposis patients without APC truncating mutations. Lab. Invest. *83*, 1859–1866.
17. Fearnhead, N.S., Wilding, J.L., Winney, B., Tonks, S., Bartlett, S., Bicknell, D.C., Tomlinson, I.P., Mortensen, N.J., and Bodmer, W.F. (2004). Multiple rare variants in different genes account for multifactorial inherited susceptibility to colorectal adenomas. Proc. Natl. Acad. Sci. USA *101*, 15992–15997.
18. Dallosso, A.R., Dolwani, S., Jones, N., Jones, S., Colley, J., Maynard, J., Idziaszczyk, S., Humphreys, V., Arnold, J., Donaldson, A., et al. (2008). Inherited predisposition to colorectal adenomas caused by multiple rare alleles of MUTYH but not OGG1, NUDT1, NTH1 or NEIL 1, 2 or 3. Gut *57*, 1252–1255.
19. Mongin, C., Coulet, F., Lefevre, J.H., Colas, C., Svrcek, M., Eyries, M., Lahely, Y., Fléjou, J.F., Soubrier, F., and Parc, Y. (2012). Unexplained polyposis: a challenge for geneticists, pathologists and gastroenterologists. Clin. Genet. *81*, 38–46.
20. Spier, I., Horpaopan, S., Vogt, S., Uhlhaas, S., Morak, M., Stienen, D., Draaken, M., Ludwig, M., Holinski-Feder, E., Nöthen, M.M., et al. (2012). Deep intronic APC mutations explain a substantial proportion of patients with familial or early-onset adenomatous polyposis. Hum. Mutat. *33*, 1045–1050.
21. Spier, I., Drichel, D., Kerick, M., Kirfel, J., Horpaopan, S., Laner, A., Holzapfel, S., Peters, S., Adam, R., Zhao, B., et al. (2016). Low-level APC mutational mosaicism is the underlying cause in a substantial fraction of unexplained colorectal adenomatous polyposis cases. J. Med. Genet. *53*, 172–179.
22. Azzopardi, D., Dallosso, A.R., Eliason, K., Hendrickson, B.C., Jones, N., Rawstorne, E., Colley, J., Moskvina, V., Frye, C., Sampson, J.R., et al. (2008). Multiple rare nonsynonymous variants in the adenomatous polyposis coli gene predispose to colorectal adenomas. Cancer Res. *68*, 358–363.
23. Will, O., Carvajal-Carmona, L.G., Gorman, P., Howarth, K.M., Jones, A.M., Polanco-Echeverry, G.M., Chinaleong, J.A., Günther, T., Silver, A., Clark, S.K., and Tomlinson, I. (2007). Homozygous PMS2 deletion causes a severe colorectal cancer and multiple adenoma phenotype without extraintestinal cancer. Gastroenterology *132*, 527–530.
24. Rio Frio, T., Lavoie, J., Hamel, N., Geyer, F.C., Kushner, Y.B., Novak, D.J., Wark, L., Capelli, C., Reis-Filho, J.S., Mai, S., et al. (2010). Homozygous BUB1B mutation and susceptibility to gastrointestinal neoplasia. N. Engl. J. Med. *363*, 2628–2637.
25. Lefevre, J.H., Bonilla, C., Colas, C., Winney, B., Johnstone, E., Tonks, S., Day, T., Hutnik, K., Boumertit, A., Soubrier, F., et al. (2012). Role of rare variants in undetermined multiple adenomatous polyposis and early-onset colorectal cancer. J. Hum. Genet. *57*, 709–716.
26. de Voer, R.M., Geurts van Kessel, A., Weren, R.D., Ligtenberg, M.J., Smeets, D., Fu, L., Vreede, L., Kamping, E.J., Verwiel, E.T., Hahn, M.M., et al. (2013). Germline mutations in the spindle assembly checkpoint genes BUB1 and BUB3 are risk factors for colorectal cancer. Gastroenterology *145*, 544–547.
27. Tomlinson, I.P., Webb, E., Carvajal-Carmona, L., Broderick, P., Howarth, K., Pittman, A.M., Spain, S., Lubbe, S., Walther, A., Sullivan, K., et al.; CORGI Consortium; EPICOLON Consortium (2008). A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. Nat. Genet. *40*, 623–630.
28. Horpaopan, S., Spier, I., Zink, A.M., Altmüller, J., Holzapfel, S., Laner, A., Vogt, S., Uhlhaas, S., Heilmann, S., Stienen, D., et al. (2015). Genome-wide CNV analysis in 221 unrelated patients and targeted high-throughput sequencing reveal novel

causative candidate genes for colorectal adenomatous polyposis. Int. J. Cancer *136*, E578–E589.
29. Lynch, H.T., Lynch, P.M., Lanspa, S.J., Snyder, C.L., Lynch, J.F., and Boland, C.R. (2009). Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. Clin. Genet. *76*, 1–18.
30. Niessen, R.C., Hofstra, R.M., Westers, H., Ligtenberg, M.J., Kooi, K., Jager, P.O., de Groote, M.L., Dijkhuizen, T., Olderode-Berends, M.J., Hollema, H., et al. (2009). Germline hypermethylation of MLH1 and EPCAM deletions are a frequent cause of Lynch syndrome. Genes Chromosomes Cancer *48*, 737–744.
31. Durno, C.A., Sherman, P.M., Aronson, M., Malkin, D., Hawkins, C., Bakry, D., Bouffet, E., Gallinger, S., Pollett, A., Campbell, B., and Tabori, U.; International BMMRD Consortium (2015). Phenotypic and genotypic characterisation of biallelic mismatch repair deficiency (BMMR-D) syndrome. Eur. J. Cancer *51*, 977–983.
32. Lavoine, N., Colas, C., Muleris, M., Bodo, S., Duval, A., Entz-Werle, N., Coulet, F., Cabaret, O., Andreiuolo, F., Charpy, C., et al. (2015). Constitutional mismatch repair deficiency syndrome: clinical description in a French cohort. J. Med. Genet. *52*, 770–778.
33. Seguí, N., Mina, L.B., Lázaro, C., Sanz-Pamplona, R., Pons, T., Navarro, M., Bellido, F., López-Doriga, A., Valdés-Mas, R., Pineda, M., et al. (2015). Germline Mutations in FAN1 Cause Hereditary Colorectal Cancer by Impairing DNA Repair. Gastroenterology *149*, 563–566.
34. Aretz, S., Stienen, D., Uhlhaas, S., Pagenstecher, C., Mangold, E., Caspari, R., Propping, P., and Friedl, W. (2005). Large submicroscopic genomic APC deletions are a common cause of typical familial adenomatous polyposis. J. Med. Genet. *42*, 185–192.
35. Choi, M., Scholl, U.I., Ji, W., Liu, T., Tikhonova, I.R., Zumbo, P., Nayir, A., Bakkaloğlu, A., Ozen, S., Sanjad, S., et al. (2009). Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc. Natl. Acad. Sci. USA *106*, 19096–19101.
36. Desmet, F.O., Hamroun, D., Lalande, M., Collod-Béroud, G., Claustres, M., and Béroud, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res. *37*, e67.
37. Yeo, G., and Burge, C.B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J. Comput. Biol. *11*, 377–394.
38. Petrovski, S., Wang, Q., Heinzen, E.L., Allen, A.S., and Goldstein, D.B. (2013). Genic intolerance to functional variation and the interpretation of personal genomes. PLoS Genet. *9*, e1003709.
39. Huang, N., Lee, I., Marcotte, E.M., and Hurles, M.E. (2010). Characterising and predicting haploinsufficiency in the human genome. PLoS Genet. *6*, e1001154.
40. Vaughn, C.P., Robles, J., Swensen, J.J., Miller, C.E., Lyon, E., Mao, R., Bayrak-Toydemir, P., and Samowitz, W.S. (2010). Clinical analysis of PMS2: mutation detection and avoidance of pseudogenes. Hum. Mutat. *31*, 588–593.
41. Tentori, L., Muzi, A., Dorio, A.S., Dolci, S., Campolo, F., Vernole, P., Lacal, P.M., Praz, F., and Graziani, G. (2013). MSH3 expression does not influence the sensitivity of colon cancer HCT116 cell line to oxaliplatin and poly(ADP-ribose) polymerase (PARP) inhibitor as monotherapy or in combination. Cancer Chemother. Pharmacol. *72*, 117–125.
42. Brieger, A., Plotz, G., Zeuzem, S., and Trojan, J. (2007). Thymosin beta 4 expression and nuclear transport are regulated by hMLH1. Biochem. Biophys. Res. Commun. *364*, 731–736.
43. Gupta, S., Gellert, M., and Yang, W. (2011). Mechanism of mismatch recognition revealed by human MutSβ bound to unpaired DNA loops. Nat. Struct. Mol. Biol. *19*, 72–78.
44. Kleczkowska, H.E., Marra, G., Lettieri, T., and Jiricny, J. (2001). hMSH3 and hMSH6 interact with PCNA and colocalize with it to replication foci. Genes Dev. *15*, 724–736.
45. Plaschke, J., Krüger, S., Jeske, B., Theissig, F., Kreuz, F.R., Pistorius, S., Saeger, H.D., Iaccarino, I., Marra, G., and Schackert, H.K. (2004). Loss of MSH3 protein expression is frequent in MLH1-deficient colorectal cancer and is associated with disease progression. Cancer Res. *64*, 864–870.
46. Kloth, M., Ruesseler, V., Engel, C., Koenig, K., Peifer, M., Mariotti, E., Kuenstlinger, H., Florin, A., Rommerscheidt-Fuss, U., Koitzsch, U., et al. (2016). Activating ERBB2/HER2 mutations indicate susceptibility to pan-HER inhibitors in Lynch and Lynch-like colorectal cancer. Gut *65*, 1296–1305.
47. Boland, C.R., Thibodeau, S.N., Hamilton, S.R., Sidransky, D., Eshleman, J.R., Burt, R.W., Meltzer, S.J., Rodriguez-Bigas, M.A., Fodde, R., Ranzani, G.N., and Srivastava, S. (1998). A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res. *58*, 5248–5257.
48. Suraweera, N., Duval, A., Reperant, M., Vaury, C., Furlan, D., Leroy, K., Seruca, R., Iacopetta, B., and Hamelin, R. (2002). Evaluation of tumor microsatellite instability using five quasimonomorphic mononucleotide repeats and pentaplex PCR. Gastroenterology *123*, 1804–1811.
49. Bellizzi, A.M., and Frankel, W.L. (2009). Colorectal cancer due to deficiency in DNA mismatch repair function: a review. Adv. Anat. Pathol. *16*, 405–417.
50. Burger, M., Burger, S.J., Denzinger, S., Wild, P.J., Wieland, W.F., Blaszyk, H., Obermann, E.C., Stoehr, R., and Hartmann, A. (2006). Elevated microsatellite instability at selected tetranucleotide repeats does not correlate with clinicopathologic features of bladder cancer. Eur. Urol. *50*, 770–775, discussion 776.
51. Fujii, H., and Shimada, T. (1989). Isolation and characterization of cDNA clones derived from the divergently transcribed gene in the region upstream from the human dihydrofolate reductase gene. J. Biol. Chem. *264*, 10057–10064.
52. Rivera, B., Castellsagué, E., Bah, I., van Kempen, L.C., and Foulkes, W.D. (2015). Biallelic NTHL1 Mutations in a Woman with Multiple Primary Tumors. N. Engl. J. Med. *373*, 1985–1986.
53. Sieber, O.M., Lipton, L., Crabtree, M., Heinimann, K., Fidalgo, P., Phillips, R.K., Bisgaard, M.L., Orntoft, T.F., Aaltonen, L.A., Hodgson, S.V., et al. (2003). Multiple colorectal adenomas, classic adenomatous polyposis, and germ-line mutations in MYH. N. Engl. J. Med. *348*, 791–799.
54. Herkert, J.C., Niessen, R.C., Olderode-Berends, M.J., Veenstra-Knol, H.E., Vos, Y.J., van der Klift, H.M., Scheenstra, R., Tops, C.M., Karrenbeld, A., Peters, F.T., et al. (2011). Paediatric intestinal cancer and polyposis due to bi-allelic PMS2 mutations: case series, review and follow-up guidelines. Eur. J. Cancer *47*, 965–982.
55. Tieu, A.H., Edelstein, D., Axilbund, J., Romans, K.E., Brosens, L.A., Wiley, E., Hylind, L., and Giardiello, F.M. (2016). Clinical Characteristics of Multiple Colorectal Adenoma Patients

Without Germline APC or MYH Mutations. J. Clin. Gastroenterol. *50*, 584–588.
56. Srivatsan, A., Bowen, N., and Kolodner, R.D. (2014). Mispair-specific recruitment of the Mlh1-Pms1 complex identifies repair substrates of the Saccharomyces cerevisiae Msh2-Msh3 complex. J. Biol. Chem. *289*, 9352–9364.
57. Haugen, A.C., Goel, A., Yamada, K., Marra, G., Nguyen, T.P., Nagasaka, T., Kanazawa, S., Koike, J., Kikuchi, Y., Zhong, X., et al. (2008). Genetic instability caused by loss of MutS homologue 3 in human colorectal cancer. Cancer Res. *68*, 8465–8472.
58. Plaschke, J., Preußler, M., Ziegler, A., and Schackert, H.K. (2012). Aberrant protein expression and frequent allelic loss of MSH3 in colorectal cancer with low-level microsatellite instability. Int. J. Colorectal Dis. *27*, 911–919.
59. Hile, S.E., Shabashev, S., and Eckert, K.A. (2013). Tumor-specific microsatellite instability: do distinct mechanisms underlie the MSI-L and EMAST phenotypes? Mutat. Res. *743-744*, 67–77.
60. Marra, G., Iaccarino, I., Lettieri, T., Roscilli, G., Delmastro, P., and Jiricny, J. (1998). Mismatch repair deficiency associated with overexpression of the MSH3 gene. Proc. Natl. Acad. Sci. USA *95*, 8568–8573.
61. Miquel, C., Jacob, S., Grandjouan, S., Aimé, A., Viguier, J., Sabourin, J.C., Sarasin, A., Duval, A., and Praz, F. (2007). Frequent alteration of DNA damage signalling and repair pathways in human colorectal cancers with microsatellite instability. Oncogene *26*, 5919–5926.
62. You, K.T., Li, L.S., Kim, N.G., Kang, H.J., Koh, K.H., Chwae, Y.J., Kim, K.M., Kim, Y.K., Park, S.M., Jang, S.K., and Kim, H. (2007). Selective translational repression of truncated proteins from frameshift mutation-derived mRNAs in tumors. PLoS Biol. *5*, e109.
63. Ohmiya, N., Matsumoto, S., Yamamoto, H., Baranovskaya, S., Malkhosyan, S.R., and Perucho, M. (2001). Germline and somatic mutations in hMSH6 and hMSH3 in gastrointestinal cancers of the microsatellite mutator phenotype. Gene *272*, 301–313.
64. Harrington, J.M., and Kolodner, R.D. (2007). Saccharomyces cerevisiae Msh2-Msh3 acts in repair of base-base mispairs. Mol. Cell. Biol. *27*, 6546–6554.
65. Campregher, C., Schmid, G., Ferk, F., Knasmüller, S., Khare, V., Kortüm, B., Dammann, K., Lang, M., Scharl, T., Spittler, A., et al. (2012). MSH3-deficiency initiates EMAST without oncogenic transformation of human colon epithelial cells. PLoS ONE *7*, e50541.
66. Risinger, J.I., Umar, A., Boyd, J., Berchuck, A., Kunkel, T.A., and Barrett, J.C. (1996). Mutation of MSH3 in endometrial cancer and evidence for its functional role in heteroduplex repair. Nat. Genet. *14*, 102–105.
67. Edelmann, W., Umar, A., Yang, K., Heyer, J., Kucherlapati, M., Lia, M., Kneitz, B., Avdievich, E., Fan, K., Wong, E., et al. (2000). The DNA mismatch repair genes Msh3 and Msh6 cooperate in intestinal tumor suppression. Cancer Res. *60*, 803–807.
68. Hegan, D.C., Narayanan, L., Jirik, F.R., Edelmann, W., Liskay, R.M., and Glazer, P.M. (2006). Differing patterns of genetic instability in mice deficient in the mismatch repair genes Pms2, Mlh1, Msh2, Msh3 and Msh6. Carcinogenesis *27*, 2402–2408.
69. Orimo, H., Nakajima, E., Yamamoto, M., Ikejima, M., Emi, M., and Shimada, T. (2000). Association between single nucleotide polymorphisms in the hMSH3 gene and sporadic colon cancer with microsatellite instability. J. Hum. Genet. *45*, 228–230.
70. Berndt, S.I., Platz, E.A., Fallin, M.D., Thuita, L.W., Hoffman, S.C., and Helzlsouer, K.J. (2007). Mismatch repair polymorphisms and the risk of colorectal cancer. Int. J. Cancer *120*, 1548–1554.
71. Hirata, H., Hinoda, Y., Kawamoto, K., Kikuno, N., Suehiro, Y., Okayama, N., Tanaka, Y., and Dahiya, R. (2008). Mismatch repair gene MSH3 polymorphism is associated with the risk of sporadic prostate cancer. J. Urol. *179*, 2020–2024.
72. Jafary, F., Salehi, M., Sedghi, M., Nouri, N., Jafary, F., Sadeghi, F., Motamedi, S., and Talebi, M. (2012). Association between mismatch repair gene MSH3 codons 1036 and 222 polymorphisms and sporadic prostate cancer in the Iranian population. Asian Pac. J. Cancer Prev. *13*, 6055–6057.
73. Duraturo, F., Liccardo, R., Cavallo, A., De Rosa, M., Grosso, M., and Izzo, P. (2011). Association of low-risk MSH3 and MSH2 variant alleles with Lynch syndrome: probability of synergistic effects. Int. J. Cancer *129*, 1643–1650.
74. Yang, X., Wu, J., Lu, J., Liu, G., Di, G., Chen, C., Hou, Y., Sun, M., Yang, W., Xu, X., et al. (2015). Identification of a comprehensive spectrum of genetic factors for hereditary breast cancer in a Chinese population by next-generation sequencing. PLoS ONE *10*, e0125571.
75. de Wind, N., Dekker, M., Claij, N., Jansen, L., van Klink, Y., Radman, M., Riggins, G., van der Valk, M., van't Wout, K., and te Riele, H. (1999). HNPCC-like cancer predisposition in mice through simultaneous loss of Msh3 and Msh6 mismatch-repair protein functions. Nat. Genet. *23*, 359–362.
76. Edelmann, L., and Edelmann, W. (2004). Loss of DNA mismatch repair function and cancer predisposition in the mouse: animal models for human hereditary nonpolyposis colorectal cancer. Am. J. Med. Genet. C. Semin. Med. Genet. *129C*, 91–99.
77. Arai, H., Okudela, K., Oshiro, H., Komitsu, N., Mitsui, H., Nishii, T., Tsuboi, M., Nozawa, A., Noishiki, Y., Ohashi, K., et al. (2013). Elevated microsatellite alterations at selected tetra-nucleotide (EMAST) in non-small cell lung cancers–a potential determinant of susceptibility to multiple malignancies. Int. J. Clin. Exp. Pathol. *6*, 395–410.
78. Danaee, H., Nelson, H.H., Karagas, M.R., Schned, A.R., Ashok, T.D., Hirao, T., Perry, A.E., and Kelsey, K.T. (2002). Microsatellite instability at tetranucleotide repeats in skin and bladder cancer. Oncogene *21*, 4894–4899.
79. Tseng-Rogenski, S.S., Chung, H., Wilk, M.B., Zhang, S., Iwaizumi, M., and Carethers, J.M. (2012). Oxidative stress induces nuclear-to-cytosol shift of hMSH3, a potential mechanism for EMAST in colorectal cancer cells. PLoS ONE *7*, e50616.
80. Carethers, J.M., Koi, M., and Tseng-Rogenski, S.S. (2015). EMAST is a Form of Microsatellite Instability That is Initiated by Inflammation and Modulates Colorectal Cancer Progression. Genes (Basel) *6*, 185–205.
81. Senter, L., Clendenning, M., Sotamaa, K., Hampel, H., Green, J., Potter, J.D., Lindblom, A., Lagerstedt, K., Thibodeau, S.N., Lindor, N.M., et al. (2008). The clinical phenotype of Lynch syndrome due to germ-line PMS2 mutations. Gastroenterology *135*, 419–428.
82. Meindl, A., Hellebrand, H., Wiek, C., Erven, V., Wappenschmidt, B., Niederacher, D., Freund, M., Lichtner, P., Hartmann, L., Schaal, H., et al. (2010). Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. Nat. Genet. *42*, 410–414.

The American Journal of Human Genetics, Volume *99*

Supplemental Data

Exome Sequencing Identifies Biallelic *MSH3* Germline Mutations as a Recessive Subtype of Colorectal Adenomatous Polyposis

Ronja Adam, Isabel Spier, Bixiao Zhao, Michael Kloth, Jonathan Marquez, Inga Hinrichsen, Jutta Kirfel, Aylar Tafazzoli, Sukanya Horpaopan, Siegfried Uhlhaas, Dietlinde Stienen, Nicolaus Friedrichs, Janine Altmüller, Andreas Laner, Stefanie Holzapfel, Sophia Peters, Katrin Kayser, Holger Thiele, Elke Holinski-Feder, Giancarlo Marra, Glen Kristiansen, Markus M. Nöthen, Reinhard Büttner, Gabriela Möslein, Regina C. Betz, Angela Brieger, Richard P. Lifton, and Stefan Aretz

Figure S1

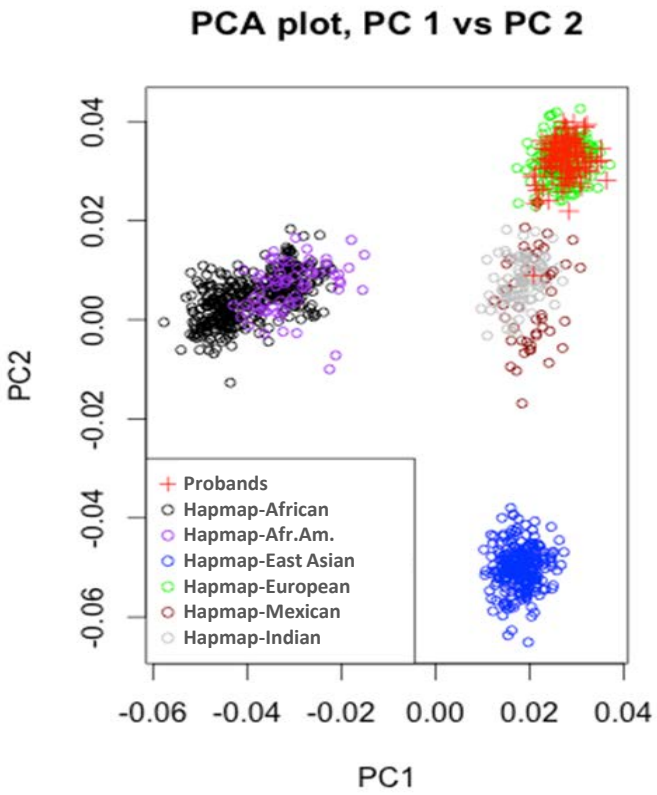


Figure S1 Results of Principle Component Analysis (PCA). The genotypes of the 100 probands who met coverage standards were compared to HapMap genotype data. A central European origin was indicated for all samples, with the exception of one outlier sample, which was subsequently excluded from further analysis.

Figure S2

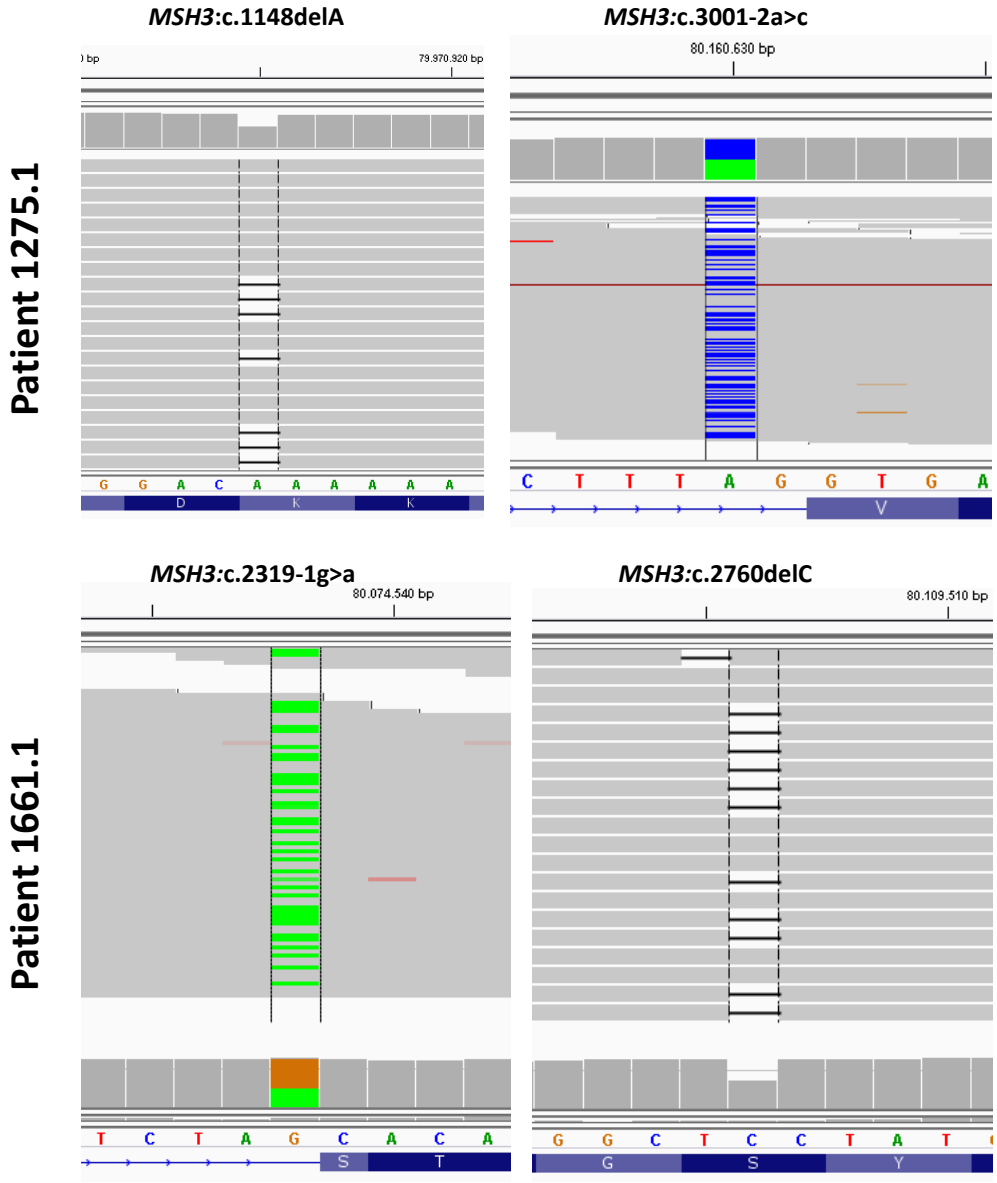


Figure S2 ***MSH3* mutations in exome sequencing data.** Biallelic loss-of-function *MSH3* germline mutations identified via exome sequencing in leukocyte DNA of patients 1275.1 and 1661.1 (visual control of reads in IGV browser).

Figure S3

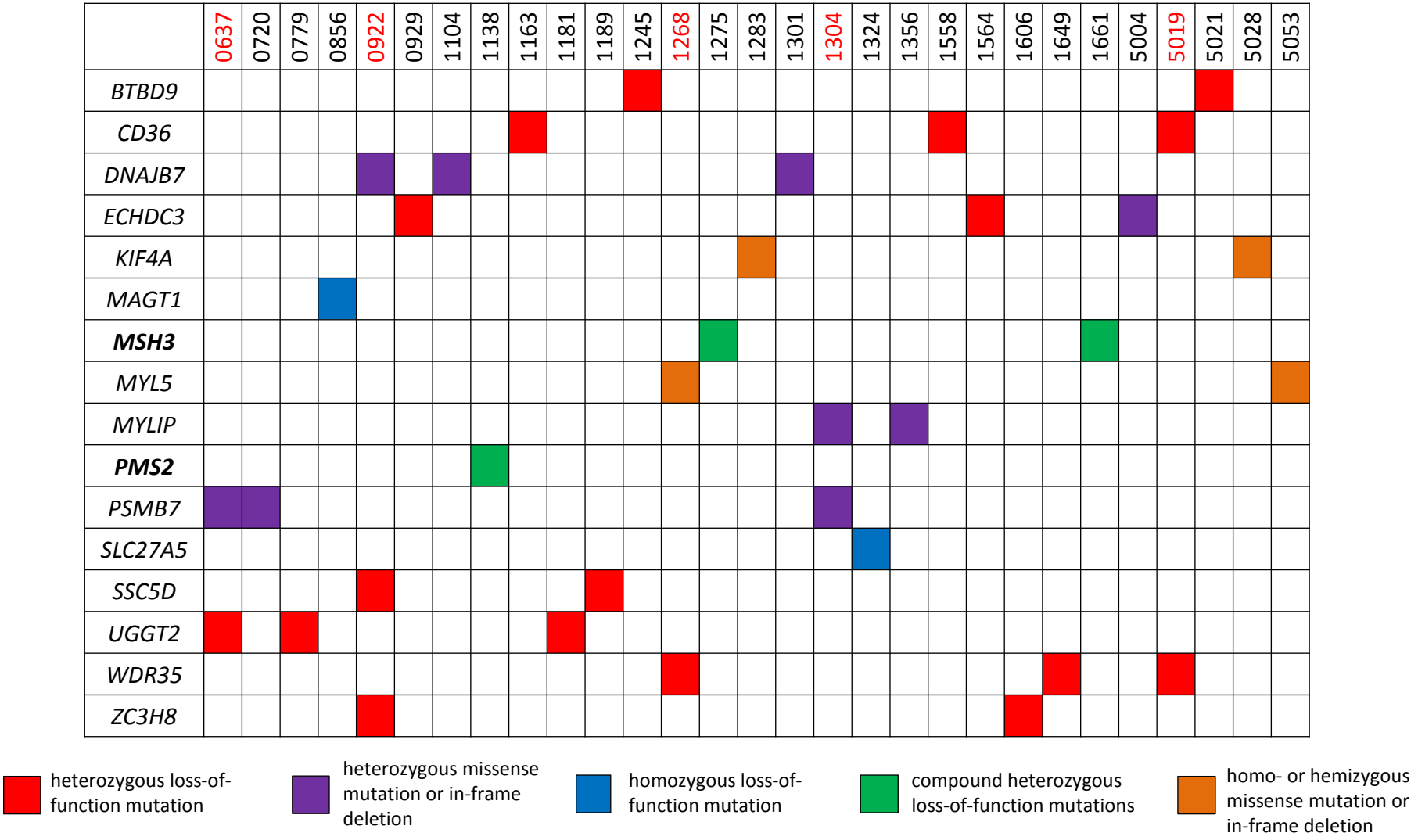
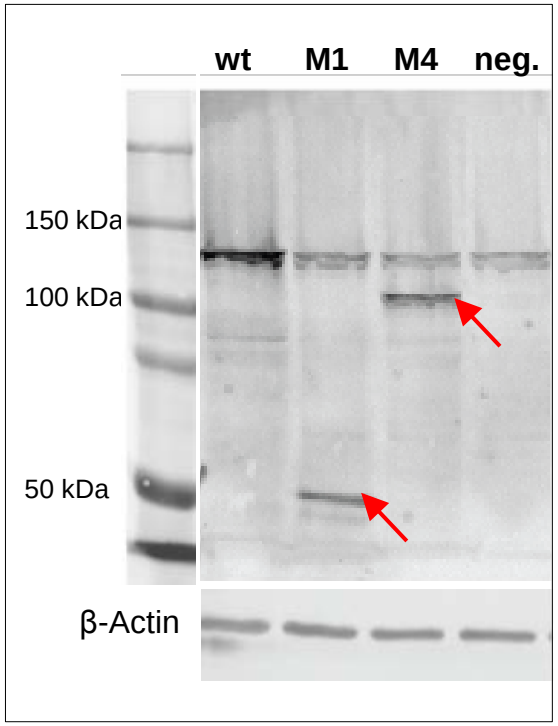


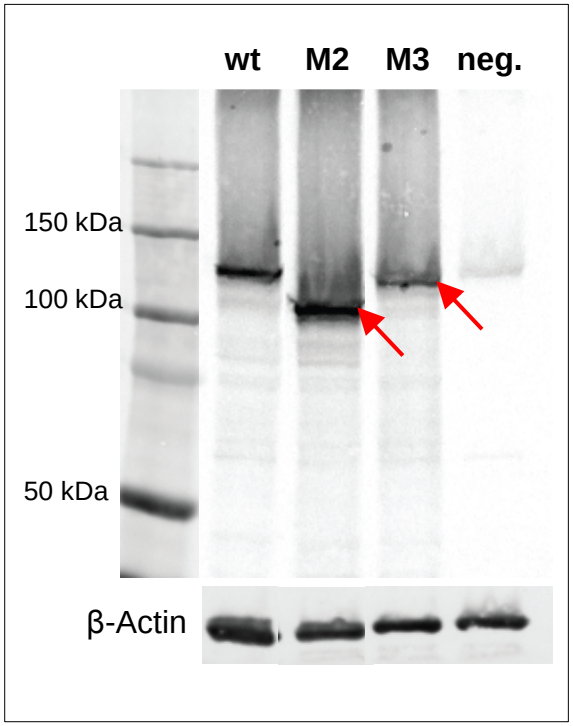
Figure S3 **Candidate genes identified in the present study.** All genes (labeled left) are affected by rare heterozygous or biallelic, potentially pathogenic variants in ≥ 2 alleles of the cohort. Patients (labeled above) harboring ≥ 2 mutated candidate genes are shown in red.

Figure S4

A



B



C

lane	*MSH3* mutation	Predicted effect on protein and *MSH3* size
wt	wild type	wild type: 130 kDa
M1	c.1148delA	p.Lys383Argfs*32: 46 kDa
M2	c.3001-2a>c	p.Val1001Argfs*16: 113 kDa
M3	c.2319-1g>a	p.Thr774_Glu812del: 123 kDa
M4	c.2760delC	p.Tyr921Metfs*36: 107 kDa
neg.	none	endogenous MSH3: 130kDa

Figure S4 Effects of *MSH3* mutations on protein level. A vector carrying *MSH3* wild type (wt) or a *MSH3* variant was over-expressed in HEK293T cells. The frameshift variants (A) were constructed by site directed mutagenesis of *MSH3* wt. The splice site variants (B) were mimicked by vectors lacking the exon 22 (M2) or exon 17 (M3), respectively. From the predicted effects on protein level, the expected sizes were calculated (C). The Western blot shows MSH3 sizes congruent with the predicted protein alterations. Endogenous wild type MSH3 is found at a size of 130 kDa in all samples, including the negative (neg.) control transfected with a vector not coding MSH3. β-Actin staining was used to assess equal lane loading.

Figure S5

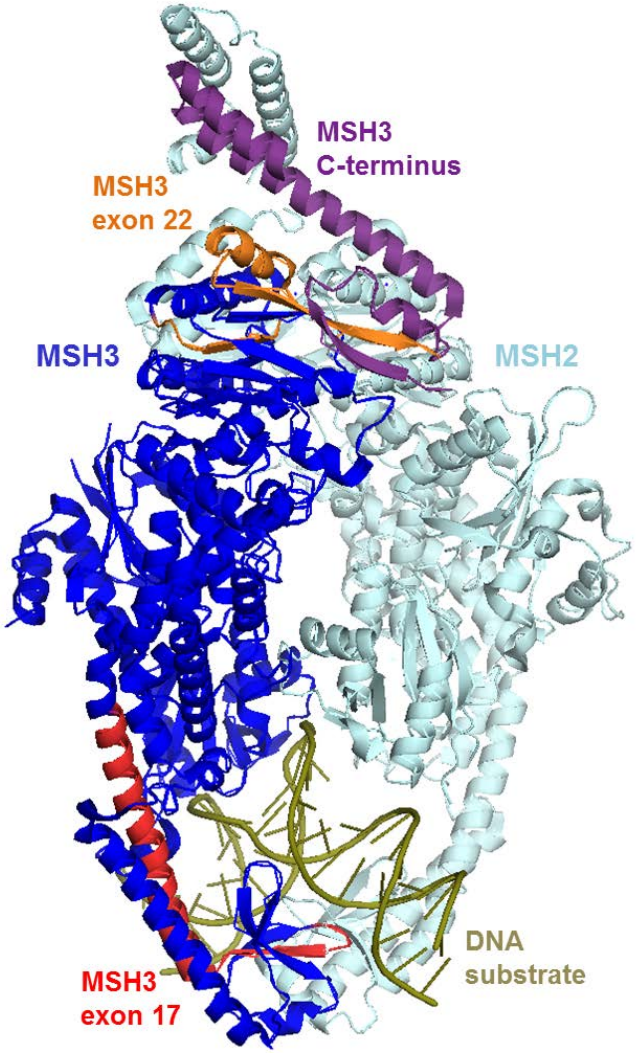


Figure S5 Mapping of deleted exons to protein structure. The human MutSβ structure published by Yang's group shows MSH2 (pale blue) and MSH3 (dark blue) binding to a dinucleotide loop (olive). We highlighted amino acids coded by regions affected by the altered splicing in red (exon 17), orange (exon 22) and purple (C-terminus altered through frameshift).

Figure S6

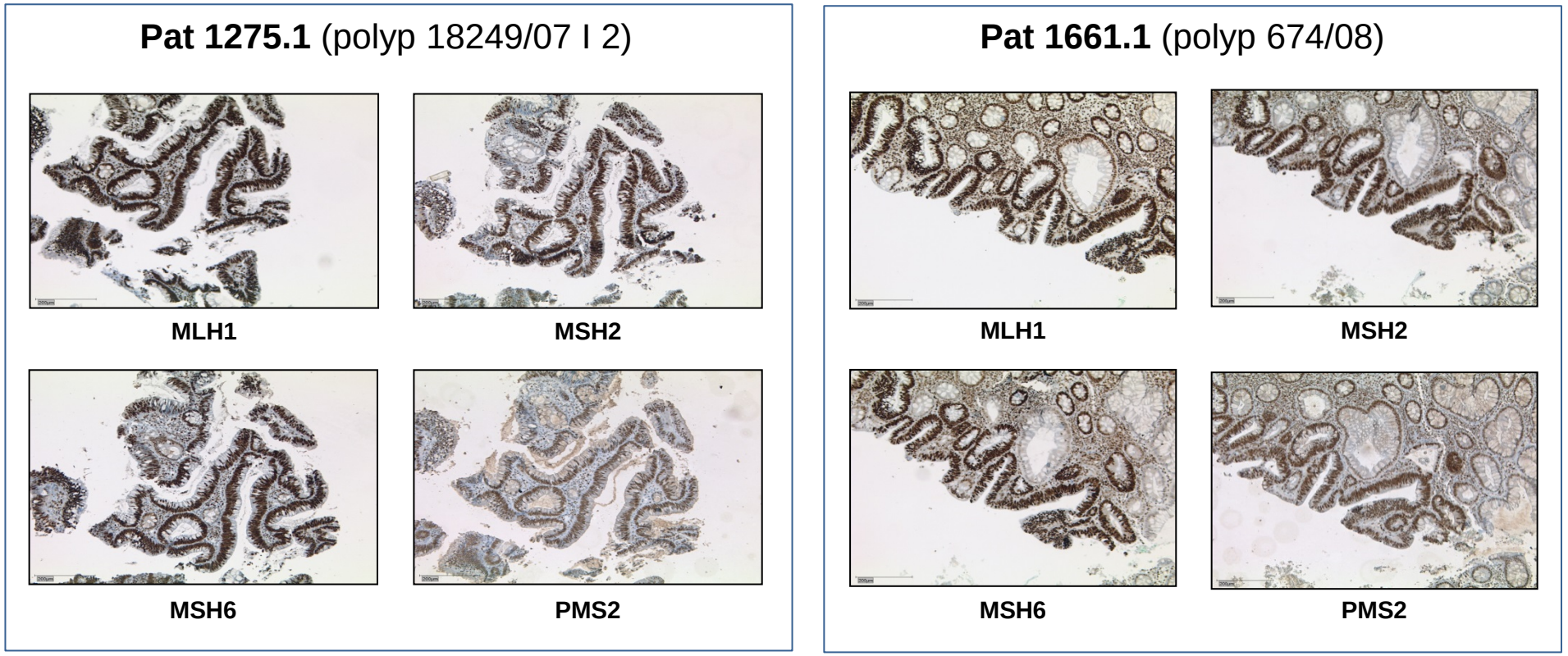


Figure S6 Expression of MMR partner proteins. Immunohistochemical staining of adenoma tissue shows normal protein levels of other MMR proteins (MLH1, MSH2, MSH6 and PMS2) for the index patients of families 1275 and 1661.

Figure S7

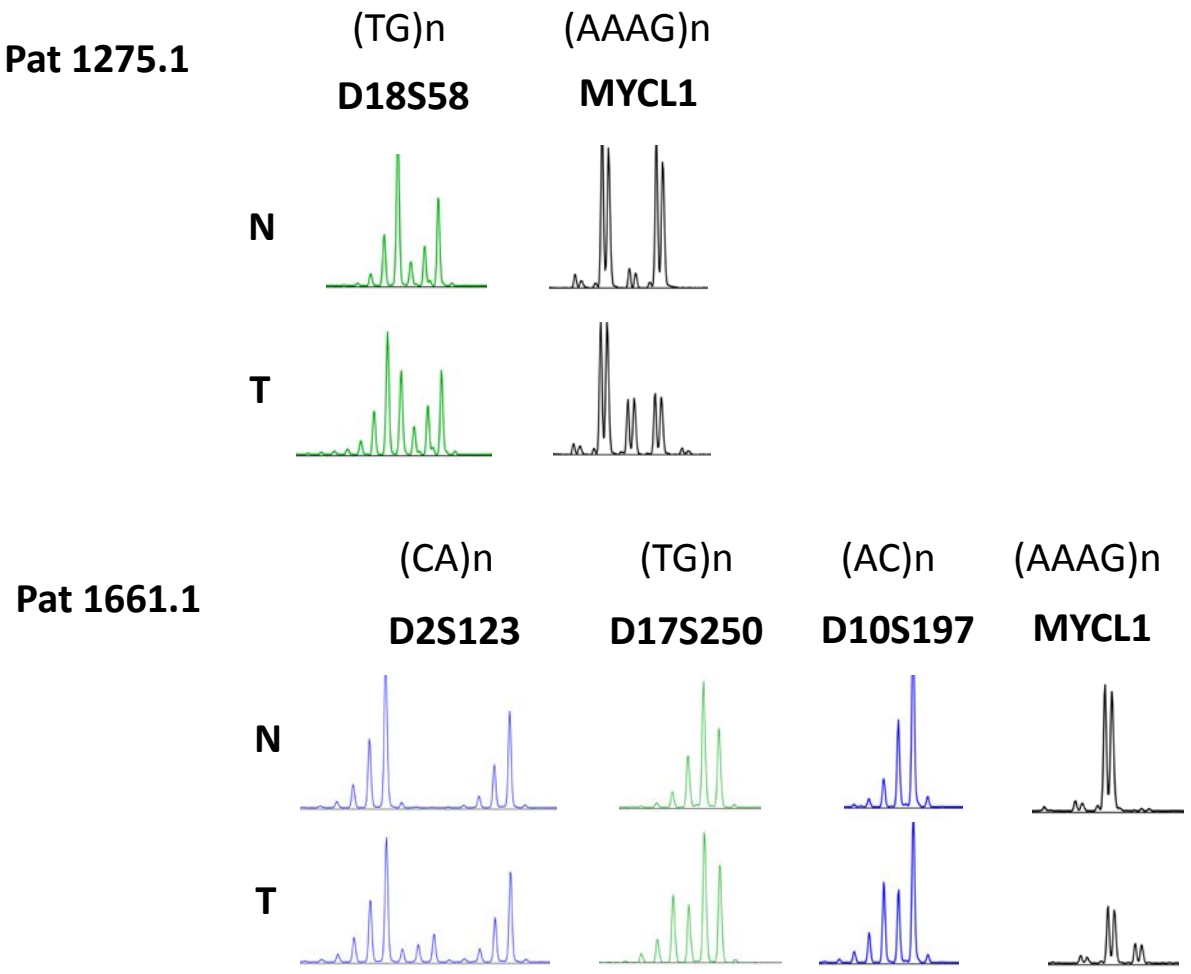
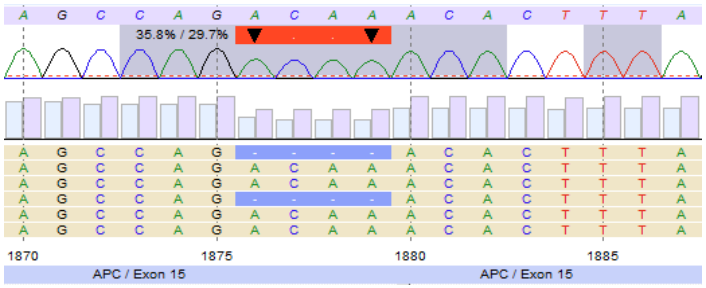


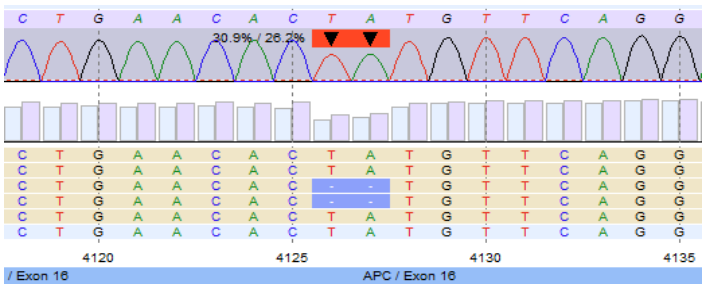
Figure S7 Results for additional microsatellite markers. Four dinucleotide repeat markers (D2S123, D17S250, D10S197, D18S58) and one further tetranucleotide repeat marker (MYCL1) were examined in normal and tumor tissue. Only those markers showing instability are depicted. All results were validated in a second tumor sample.

Polyp E 48579-08

***APC*:c.1876_1879delACAA** (36/30%)

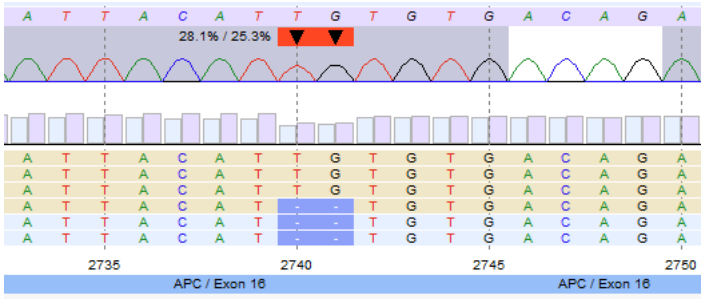


***APC*:c.4126_4127delTA** (31/26%)

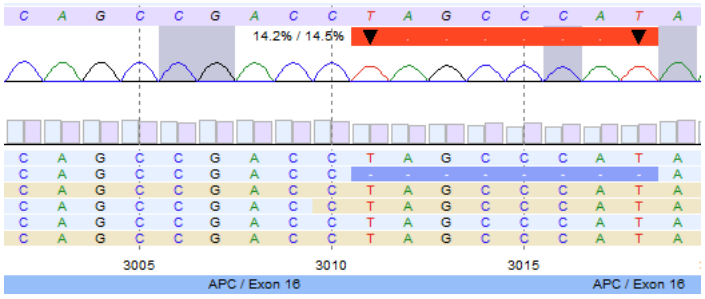


Polyp 674-08 3

***APC*:c.2740_2741delTG** (28/25%)

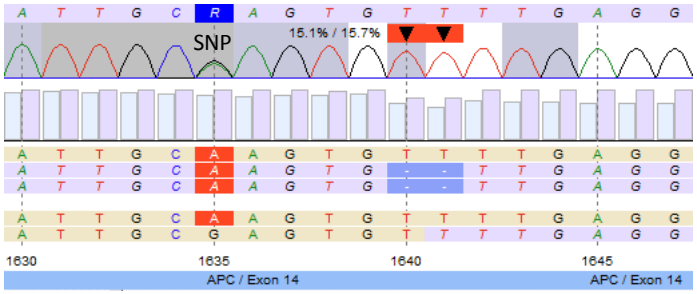


***APC*:c.3011_3018delTAGCCCAT** (14/15%)

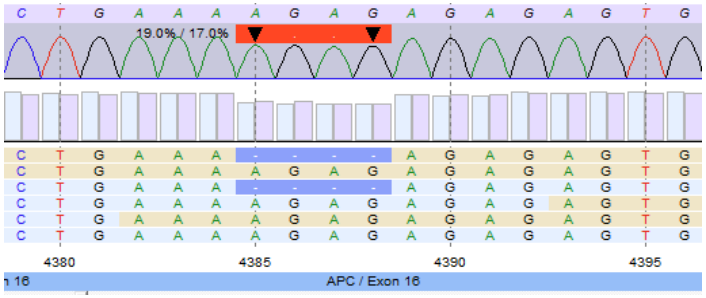


Polyp 674-08 4

***APC*:c.1640_1641delTT** (15/16%)



***APC*:c.4385_4388delAGAG** (19/17%)



Polyp 674-08 2

***APC*:c.3468_3469delAG** (7/6%)

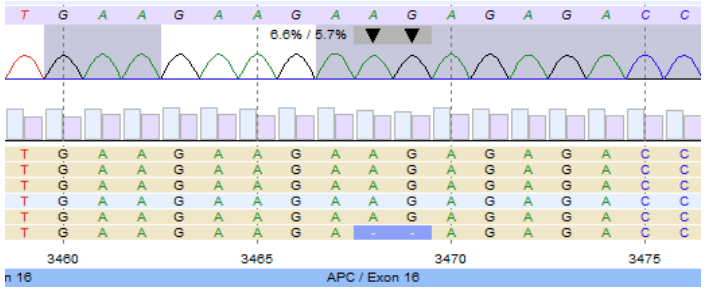


Figure S8

Figure S8 Somatic *APC* mutations. Targeted deep sequencing results for the *APC* gene in adenoma-derived DNA from patient 1661.1: Comparison of four independent adenomas with leukocyte DNA identified seven different somatic *APC* mutations (1-2 per polyp) in 6-36% of the reads (the percentage of mutated reads in the forward and reverse reads respectively is shown in brackets). All seven mutations were small deletions of 2-8 nucleotides. In 4/7 mutations, the sequence context proved to be di- or trinucleotide repeats.

Figure S9

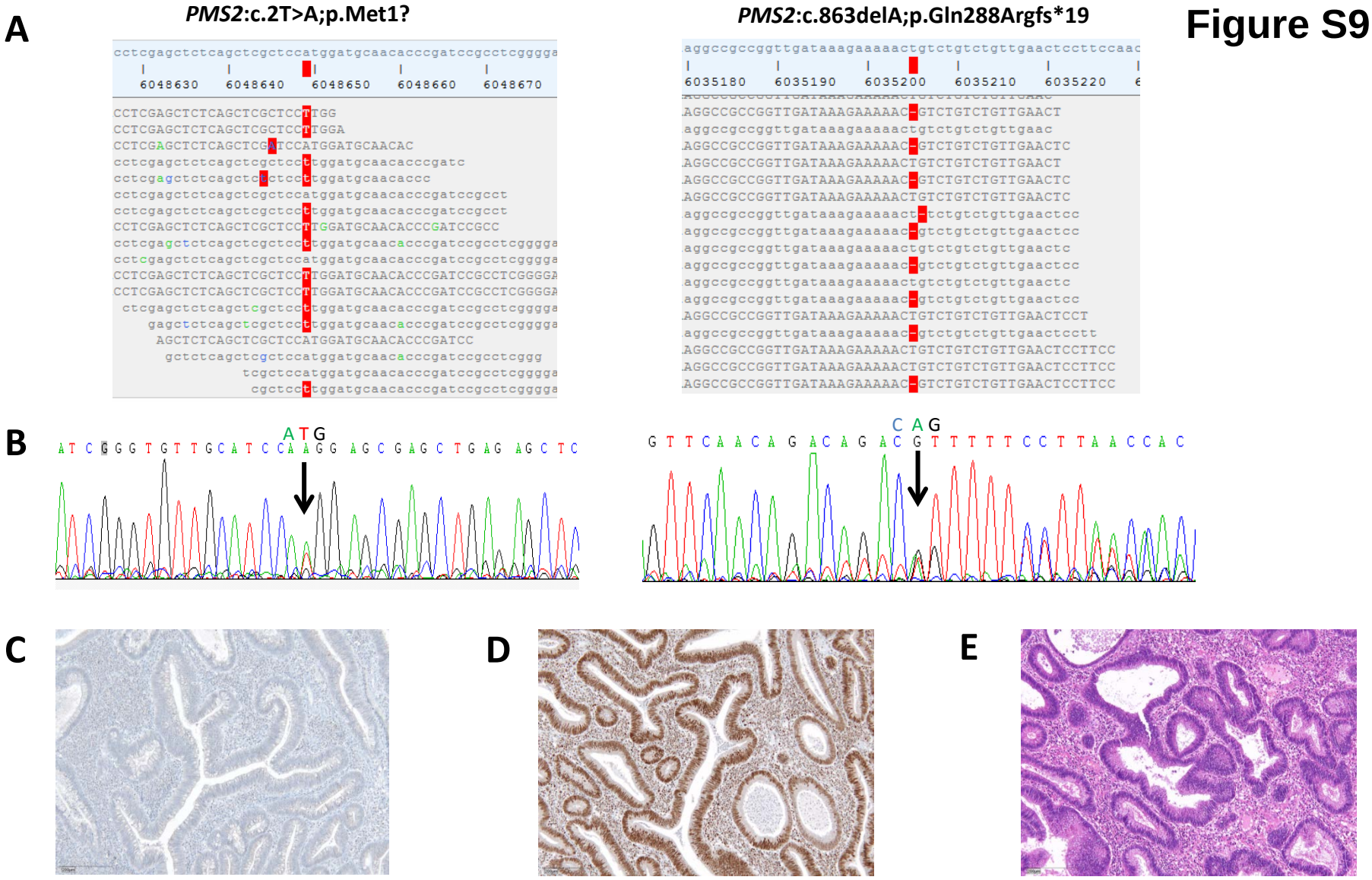


Figure S9 ***PMS2*** **germline mutations.** (A) Exome sequencing from leukocyte-derived DNA of patient 1138 revealed the mutations c.2T>A;p.Met1? and c.863delA;p.Gln288Argfs*19 in *PMS2* (reverse sequence, Varbank). (B) Sanger validation of *PMS2* exons 1 and 8 confirmed both mutations (top line shows reference). (C) Immunohistochemical staining of both tumor and normal tissue shows complete loss of PMS2 protein, whereas (D) MLH1 expression is normal. (E) H-E staining.

Table S1. Summary statistics of exome sequencing performance and calling of variants.

PARAMETER	MEAN	MEDIAN	ST DEV
# of Reads (M)	65.7	64.6	11.0
Median Coverage (X)	55.6	54.5	9.4
Mean Coverage (X)	66.6	65.3	10.8
% on genome	91.8%	91.9%	0.5%
% on target	66.2%	66.4%	1.7%
% of bases covered at least 4x	96.6%	96.8%	0.6%
% of bases covered at least 8x	94.1%	94.3%	1.1%
% of bases covered at least 20x	84.1%	84.4%	2.8%
Mean error rate	0.5%	0.5%	0.1%
% of PCR duplicate	4.9%	4.7%	1.0%
# of hom. SNVs	11606.8	11650.0	301.1
% of novel hom. SNVs	0.2%	0.2%	0.1%
# of het. SNVs	18474.9	18502	603.7
% of novel het. SNVs	2.3%	2.1%	0.8%

het. = heterozygous; hom = homozygous; SNV= single nucleotide variants; st dev = standard deviation

Table S2. Clinical characteristics of the cohort (n=102).

No. of patients	102
Gender (female / male)	45 / 57
Age at diagnosis (years)	
Mean	45
Standard deviation	14
Range	14-73
No. of colorectal adenomas	
21-50	36 (35 %)
51-100	22 (22 %)
> 100	16 (16 %)
"numerous" / "multiple" [a]	28 (27 %
Colorectal cancer	29 (28 %)
Duodenal polyps	13 (13 %)
Extracolonic tumors (details see Table S3)	14 (14 %)
Family history [b]	
Familial	20 (20 %)
Simplex	45 (44 %)
Unclear	37 (36 %)

[a] "numerous" / "multiple" refers to cases where no exact numbers of polyps were mentioned in the clinical reports. However, based on all available information it is very likely that the inclusion criteria are met in these individuals.

[b] Family history refers to the presence of a colorectal polyposis in the relatives of the index polyposis patients; "Familial" is defined as the presence of at least one further first or second degree relative with multiple colorectal polyps; "Simplex" is defined as an isolated case with complete unconspicious family history regarding colorectal polyposis or early onset colorectal cancer, and "Unclear" is defined as the situation where a distinction between simplex and familial is impossible, e.g. a first or second degree relative with non-early-onset colorectal cancer and no known colorectal polyposis.

Table S3. Extracolonic tumor spectrum (malignant tumors and rare benign tumors, sorted by location) of the cohort (index patients), including one of the patients with biallelic *MSH3* mutations (ID 1661) and the patient with biallelic *PMS2* mutations (ID 1138).

ID	**Extracolonic tumors (age at diagnosis)**
	Malignant tumors
1661	astrocytoma (26 y)
1138	primitive neuroectodermal tumor (PNET) of the cerebellum (4 y), pilomatrixoma, three café-au-lait spots
1300	M. Hodgkin (21 y), basalioma (50 y)
666	five basaliomas (34 y and 45 y)
1356	skin cancer (~45 y)
1375	malignant melanoma (31 y and 47 y)
568	breast cancer (48 y)
696	breast cancer (66 y)
1189	duodenal neuroendocrine, gastrin producing tumour (48 y)
1481	renal cell carcinoma (65 y)
5007	renal cell carcinoma (65 y)
736	prostate cancer (51 y)
	Rare benign tumors
807	retroperitoneal fibrosis (29 y)
1294	adrenal gland adenoma (27 y)

Table S4 Candidate genes affected by rare potentially pathogenic variants identified in the present study.

Pat. ID	Gene	Mutation Type	Geno-type	Coding DNA Change	Predicted Prot. Change	Result of RNA Analysis	Ref. Cov.	Non-Ref. Cov.	San-ger Valid.	Phylo P	dbSNP Accession Number	mRNA Accession Number	TCGA	mRNA Expr.[a]	Prot. Expr.[a]	RVIS[b] (Per-centile)	HIS[c] (%)	Gene Function
5021	*BTBD9*	nonsense	het.	c.1480C>T	p.Arg494*	na	36	23	yes	1.09	rs377402489	NM_001099272.1	0	T	F	-0.96 (9.1)	0.267 (33.3)	Potential involvement in synaptic plasticity and vesicle recycling
1245	*BTBD9*	nonsense	het.	c.1080G>A	p.Trp360*	na	30	25	yes	5.94								
1558	*CD36*	frameshift	het.	c.79dupA	p.Met27Asnfs*13	na	82	42	yes			NM_001001548.2	1	T	T	-1.7 (2.6)	0.847 (4.7)	Receptor for thrombospondins; *cell adhesion* molecule; transports / regulates long chain fatty acids; reported involvement in *apoptosis* and *cancer stem cell* maintenance
5019	*CD36*	frameshift	het.	c.708_709del	p.Ser237Leufs*10	na	139	74	yes		rs146885545							
1163	*CD36*	nonsense	het.	c.971C>G	p.Ser324*	na	37	23	na	-0.12								
0922	*DNAJB7*	in-frame deletion	het.	c.122_124del	p.Glu41del	na	48	29	yes	1.01		NM_145174.1	0	F	F	0.69 (85.1)	0.106 (70.4)	Likely activity as a cochaperone
1104	*DNAJB7*	in-frame deletion	het.	c.122_124del	p.Glu41del	na	41	22	yes	1.01								
1301	*DNAJB7*	missense	het.	c.205G>A	p.Gly69Ser	na	27	25	na	4.16								
0929	*ECHDC3*	splice site	het.	c.591+1G>A	p.Val131_Lys197del	in-frame loss of exon 4	17	20	yes	4.08	rs200347426	NM_024693.4	1	T	T	0.42 (77.2)	0.102 (72.1)	Mitochondrially expressed, may be involved in dehalogenase, hydratase, and isomerase activities etc.
1564	*ECHDC3*	splice site	het.	c.591+1G>A	p.Val131_Lys197del	in-frame loss of exon 4	16	12	yes	4.08	rs200347426							
5004	*ECHDC3*	missense	het.	c.667G>A	p.Val223Met	na	12	18	yes	5.20	rs375091889							
5028	*KIF4A*	missense	hemiz.	c.2387G>A	p.Arg796Gln	na	0	34	na	5.11		NM_012310.4	6	T	T	-0.75 (13.6)		Microtubule motor prot.: molecular transport; possibly involved in *chromosomal stability* during mitosis and *cell cycle control*
1283	*KIF4A*	in-frame deletion	hemiz.	c.2427_2429del	p.Ser810del	na	11	35	na									
0856	*MAGT1*	nonsense	homo.	c.71C>G	p.Ser24*	na	0	25	yes	0.52		NM_032121.5	0	T	T	0.1 (61.5)	0.104 (71.1)	Magnesium cation transporter, may have a role in N-glycosylation, mutations in this gene cause MRX95 and *XMEN-syndrome*
1275	*MSH3*	frameshift	comp.-het.	c.1148delA	p.Lys383Argfs*32	na	67	39	yes			NM_002439.4	1	T	F	-0.08 (47.3)	0.486 (16.2)	*Mismatch repair prot.*
1275	*MSH3*	splice site		c.3001-2a>c	p.Val1001Argfs*16	loss of exon 22	57	39	yes	3.40								
1661	*MSH3*	splice site	comp.-het.	c.2319-1g>a	p.Thr774_Glu812del	in-frame loss of exon 17	74	74	yes	5.61								
1661	*MSH3*	frameshift		c.2760delC	p.Tyr921Metfs*36	r.2760delC	62	48	yes									
1268	*MYL5*	missense	homo.	c.263T>C	p.Phe88Ser	na	1	69	na	3.67	rs2228354	NM_002477.1	0	T	T	1.37 (94.5)	0.174 (48.2)	Component of myosin, involved in calcium binding for regulation of muscle contraction
5053	*MYL5*	missense	homo.	c.263T>C	p.Phe88Ser	na	0	48	na	3.67	rs2228354							

1304	*MYLIP*	missense	het.	c.515A>G	p.Glu172Gly	na	41	24	na	4.79		NM_013262.3	0	T	na	0.35 (74.5)	0.163 (50.8)	E3 ubiquitin-prot. ligase, mediates degradation of myosin regulatory light chain and lipoprotein receptors
1356	*MYLIP*	missense	het.	c.779C>T	p.Ala260Val	na	20	15	na	5.57	rs201781624							
5042	*MYLIP*	missense	het.	c.850G>C	p.Val284Leu	na	23	31	na	6.42								
1138	*PMS2*	start loss	comp.-het.	c.2T>A	p.Met1?	na	6	12	yes	2.47		NM_000535.5	2	T	T	1.48 (95.3)	0.786 (6.1)	*Mismatch repair prot.*
1138	*PMS2*	frameshift		c.863delA	p.Gln288Argfs*19	na	68	68	yes									
0637	*PSMB7*	missense	het.	c.635C>G	p.Ser121Cys	na	36	38	na	5.98		NM_002799.3	0	T	T	0.1 (61.45)	0.528 (14.2)	Proteasome subunit responsible for trypsin-like activity; affects anti-cancer drug responses
0720	*PSMB7*	missense	het.	c.125C>G	p.Thr42Ser	na	70	57	na	4.35								
1304	*PSMB7*	missense	het.	c.74T>G	p.Leu25Trp	na	34	41	na	5.01								
1324	*SLC27A5*	nonsense	homo.	c.1375C>T	p.Arg459*	na	0	43	yes	0.72		NM_012254.2	1	T	F	0.29 (71.6)	0.067 (87.4)	Fatty acid metabolism
1189	*SSC5D*	nonsense	het.	c.2568G>A	p.Trp856*	na	16	10	yes	1.68		NM_001195267.1	5	T	T	3.78 (99.6)		Scavenger receptor activity, immune response
0922	*SSC5D*	nonsense	het.	c.1540C>T	p.Gln514*	na	12	18	yes	0.78		NM_001144950.1						
0637	*UGGT2*	frameshift	het.	c.3245delC	p.Thr1082Lysfs*6	na	71	46	yes			NM_020121.3	8	T	T	1.28 (93.7)		Quality control for prot. export from endoplasmatic reticulum
0779	*UGGT2*	frameshift	het.	c.2156dupA	p.Ser720Glufs*16	na	91	36	yes									
1181	*UGGT2*	frameshift	het.	c.390dupC	p.Pro131Thrfs*6	na	52	32	yes									
5019	*WDR35*	frameshift	het.	c.2238delG	p.Val747Leufs*29	na	73	55	yes			NM_001006657.1	2	T	T	0.48 (78.9)	0.117 (65.6)	Required for ciliogenesis and ciliary transport; reports of connection to CASP3, NF-κB and CaMKK/AMPK/p38-MAPK pathways and *apoptosis*
1649	*WDR35*	nonsense	het.	c.2089C>T	p.Arg697*	na	51	61	yes	1.17								
1268	*WDR35*	nonsense	het.	c.1922T>G	p.Leu641*	na	59	74	yes	4.48	rs199952377							
0922	*ZC3H8*[d]	frameshift	het.	c.859_863del	p.Lys287Valfs*3	na	62	18	na			NM_032494.2	0	T	T	0.22 (67.9)	0.393 (21.7)	Transcriptional repressor of GATA3, induces *apoptosis* when overexpressed in thymocytes
1606	*ZC3H8*[d]	frameshift	het.	c.859_863del	p.Lys287Valfs*3	na	85	20	na									

comp.-het. = compound-heterozygous; Cov. = coverage; expr. = expression; F = false; hemiz. = hemizygous; het. = heterozygous; HIS = Haploinsufficiency Score; homo.= homozygous; MRX95 = mental retardation X-linked type 95; nd = not available/applicable; Pat. = Patient; prot. = protein; Ref. = reference allele; RVIS = Residual Variation Intolerance Score; T = true; TCGA = The Cancer Genome Atlas: number of LoF mutations in non-hypermutated CRC; valid. = validation; XMEN = X-linked immunodeficiency with magnesium defect, Epstein-Barr virus infection, and neoplasia

Italic gene functions have possible implication for tumorigenesis.

[a] = The expression of candidate genes was determined using the EST profiles of colon tissue and protein expression data from colon glandular cells provided by UniGene and Human Protein Atlas.

[b] = A low (negative) RVIS (-1.85 to -0.55, corresponding to values below the 25th percentile) reflects high intolerance to genetic variation, indicating that these genes are subject of purifying selection.

[c] = The HIS, which indicates dosage-sensitive genes, is considered as very low, if the percentage is <10%, and moderately reduced, if the percentage is <40%.

[d] = identified in dominant analysis mode with relaxed cutoff (1%) of allele frequency in reference databases for recurrent variants

Table S5 Published *MSH3* germline variants in patients with suspected hereditary tumors.

Phenotype	No. of Patients Screened	*MSH3* mutation	Exon	Genotype	Predicted Conse-quence	Causal Relevance	Frequency in Controls (ExAC data)	Phenotype Mutation Carriers / Family	Reference
Familial breast cancer	99	c.162_179del18AGTGAG; p.A57_A62del	1	heterozygous	in-frame	Polymorphism?	a number of in-frame deletions are described in this region		Yang et al. 2015
		c.199_207del9;p.P67_P69del	1	heterozygous	in-frame	Polymorphism?	p.Pro66_Ala68del has an allele frequency of 57%	clinical criteria Lynch, broad tumor spectrum	
		c.2305delG;p.V769*	16	heterozygous	frameshift	Loss-of-function?			
Suspected Lynch syndrome	79	c.2732T>G;p.Leu911Trp	20	Compound-heterozygous	missense	Rare variant?	allele frequency 0.2%		Duraturo et al. 2011
		c.693G>A;p.Pro	4		silent	Polymorphism	allele frequency in normal controls 15%		
Unselected CRC	152	c.2785A>T;p.Ile929Phe	20	heterozygous	missense	VUS		late-onset CRC	Kraus et al. 2015
		c.3130+3A>G	Intron 22	heterozygous	splice?	VUS			

CRC = colorectal cancer; VUS = variant of unknown significance

SUPPLEMENTAL REFERENCES

Duraturo, F., Liccardo, R., Cavallo, A., De Rosa, M., Grosso, M., and Izzo, P. (2011). Association of low-risk MSH3 and MSH2 variant alleles with Lynch syndrome: probability of synergistic effects. Int J Cancer 129, 1643-1650.

Kraus, C., Rau, T.T., Lux, P., Erlenbach-Wunsch, K., Lohr, S., Krumbiegel, M., Thiel, C.T., Stohr, R., Agaimy, A., Croner, R.S., et al. (2015). Comprehensive screening for mutations associated with colorectal cancer in unselected cases reveals penetrant and nonpenetrant mutations. Int J Cancer 136, E559-568.

Yang, X., Wu, J., Lu, J., Liu, G., Di, G., Chen, C., Hou, Y., Sun, M., Yang, W., Xu, X., et al. (2015). Identification of a comprehensive spectrum of genetic factors for hereditary breast cancer in a Chinese population by next-generation sequencing. PLoS One 10, e0125571.