Familial Cancer
DOI 10.1007/s10689-016-9870-z

ORIGINAL ARTICLE

Exome sequencing identifies potential novel candidate genes in patients with unexplained colorectal adenomatous polyposis

Isabel Spier[1,2] · Martin Kerick[3,4] · Dmitriy Drichel[5] · Sukanya Horpaopan[1,6] · Janine Altmüller[4,7] · Andreas Laner[8,9] · Stefanie Holzapfel[1,2] · Sophia Peters[1] · Ronja Adam[1,2] · Bixiao Zhao[10] · Tim Becker[5,11] · Richard P. Lifton[10] · Elke Holinski-Feder[8,9] · Sven Perner[12,13] · Holger Thiele[4] · Markus M. Nöthen[1,14] · Per Hoffmann[1,14,15,16] · Bernd Timmermann[17] · Michal R. Schweiger[3,4] · Stefan Aretz[1,2]

Abstract In up to 30 % of patients with colorectal adenomatous polyposis, no germline mutation in the known genes *APC,* causing familial adenomatous polyposis, *MUTYH,* causing *MUTYH*-associated polyposis, and *POLE* or *POLD1*, causing Polymerase-Proofreading-associated polyposis can be identified, although a hereditary etiology is likely. To uncover new causative genes, exome sequencing was performed using DNA from leukocytes and a total of 12 colorectal adenomas from seven unrelated patients with unexplained sporadic adenomatous polyposis. For data analysis and variant filtering, an established bioinformatics pipeline including in-house tools was applied. Variants were filtered for rare truncating point mutations and copy-number variants assuming a dominant, recessive, or tumor suppressor model of inheritance. Subsequently, targeted sequence analysis of the most promising candidate genes was performed in a validation cohort of 191 unrelated patients. All relevant variants were validated by Sanger sequencing. The analysis of exome sequencing data resulted in the identification of rare loss-of-function germline mutations in three promising candidate genes (*DSC2*, *PIEZO1*, *ZSWIM7*). In the validation cohort, further variants predicted to be pathogenic were identified in *DSC2* and *PIEZO1*. According to the somatic mutation spectra, the adenomas in this patient cohort follow the classical pathways of colorectal tumorigenesis. The

Electronic supplementary material The online version of this article (doi:10.1007/s10689-016-9870-z) contains supplementary material, which is available to authorized users.

✉ Isabel Spier
isabel.spier@uni-bonn.de

[1] Institute of Human Genetics, University of Bonn, Sigmund-Freud-Str. 25, 53127 Bonn, Germany

[2] Center for Hereditary Tumor Syndromes, University of Bonn, Bonn, Germany

[3] Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin, Germany

[4] Cologne Center for Genomics (CCG), University of Cologne, Cologne, Germany

[5] German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany

[6] Department of Anatomy, Faculty of Medical Science, Naresuan University, Phitsanulok, Thailand

[7] Institute of Human Genetics, University of Cologne, Cologne, Germany

[8] Medizinische Klinik und Poliklinik IV, Campus Innenstadt, Klinikum der Universität München, Munich, Germany

[9] Medizinisch Genetisches Zentrum, Munich, Germany

[10] Departments of Genetics, Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, CT, USA

[11] Institute of Medical Biometry, Informatics, and Epidemiology, University of Bonn, Bonn, Germany

[12] Section for Prostate Cancer Research, Center for Integrated Oncology Cologne/Bonn, Institute of Pathology, University Hospital of Bonn, Bonn, Germany

[13] Pathology Network of the University Hospital of Luebeck and Leibniz Research Center Borstel, Borstel, Germany

[14] Department of Genomics, Life and Brain Center, University of Bonn, Bonn, Germany

[15] Division of Medical Genetics, University Hospital Basel, Basel, Switzerland

[16] Department of Biomedicine, University of Basel, Basel, Switzerland

[17] Next Generation Sequencing Group, Max Planck Institute for Molecular Genetics, Berlin, Germany

Published online: 16 January 2016

present study identified three candidate genes which might represent rare causes for a predisposition to colorectal adenoma formation. Especially *PIEZO1* (*FAM38A*) and *ZSWIM7* (*SWS1*) warrant further exploration. To evaluate the clinical relevance of these genes, investigation of larger patient cohorts and functional studies are required.

Keywords Familial colorectal cancer · Adenomatous polyposis · Candidate genes · Exome sequencing · Massive parallel sequencing · Hereditary tumor syndromes

Introduction

To date, three inherited monogenic forms of colorectal adenomatous polyposis syndromes can be delineated by molecular genetic analyses: (1) autosomal dominant Familial Adenomatous Polyposis (FAP), caused by heterozygous germline mutations (www.lovd.nl/APC) in the tumor suppressor gene (TSG) *APC* [1], including *APC* mosaicism [2–4] and deep intronic *APC* mutations [5]; (2) autosomal recessive *MUTYH*-Associated Polyposis (MAP), caused by biallelic germline mutations of the base excision repair (BER) gene *MUTYH* [6]; and (3) autosomal dominant Polymerase-Proofreading-associated polyposis (PPAP), caused by specific germline missense mutations in the polymerase genes *POLE* and *POLD1* [7, 8]. Recently, a homozygous loss-of-function germline mutation in the *NTHL1* gene has been identified as rare predisposition to adenomatous polyposis and colorectal cancer (CRC) [9].

In up to 30 % of polyposis patients, no underlying germline mutation is identified, although a hereditary basis is likely. Currently, exome sequencing is considered the most powerful tool for the identification of new causative genes in Mendelian disorders of unknown etiology [10]. The underlying strategies include screening for recurrently mutated genes (overlap strategy), the biallelic hit strategy for a suspected recessive inheritance, and the tumor suppressor model in cancer predisposition syndromes (selection of genes which harbor both a heterozygous truncating germline mutation and a somatic mutation).

To uncover novel causative genes for adenomatous polyposis, the germline and tumor exomes of seven unrelated patients with unexplained disease were sequenced. Subsequently, a targeted mutation screening of the most promising candidate genes was performed in a large validation cohort. The identification of underlying genetic factors will provide insights into disease mechanisms, biological pathways, and potential therapeutic targets.

Materials and methods

Patients/data collection

All patients in the present study had unexplained colorectal adenomatous polyposis, i.e. no germline mutation in the *APC* or *MUTYH* genes was identified by Sanger sequencing of the coding regions, deletion/duplication analysis using Multiplex ligation-dependent probe amplification (MLPA), and screening for pathogenic deep intronic *APC* mutations [5, 11]. Furthermore, neither of the two hotspot mutations in *POLE* and *POLD1* was present [8]. In addition, in all patients a SNP array-based CNV analysis was performed, as described elsewhere [12].

Exome sequencing was performed in a discovery cohort of seven unrelated index patients. For targeted analysis of candidate genes, a validation cohort of 191 unrelated index polyposis patients was used. The inclusion criteria for both the discovery and validation cohort were the presence of at least 20 synchronous or 40 metachronous, histologically confirmed colorectal adenomas, irrespective of inheritance pattern and extra intestinal lesions. All patients were of central European origin according to family name and self-reporting. 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 patients provided written informed consent prior to inclusion.

High-throughput sequencing and bioinformatics workflow

Genomic leukocyte DNA was extracted from peripheral EDTA-anticoagulated blood samples using the standard salting-out procedure. Tumor DNA was extracted from punches of colorectal formalin-fixed and paraffin-embedded (FFPE) tumor tissue, as described elsewhere [13].

Library preparation and whole exome target enrichment was performed according to Agilent's SureSelect protocol (Human All Exon 50 Mb v2, 2011). Multiplexed paired-end sequencing was performed on the Illumina HiSeq 2000 platform in accordance with the manufacturer's protocol. Base calling and demultiplexing were performed using Illumina's CASAVA pipeline v1.7. Raw reads were mapped to GRCh37/hg19 using BWA v0.5.8 [14] with default parameters. Local realignment, quality value recalibration, and variant calling were performed using GATK v2.1-8 [15]. In-house tools and ANNOVAR [16] were used to annotate and filter the variants. Metrics and enrichment statistics were calculated with Picard using Agilent's SureSelect target regions.

Filtering was then performed to identify germline truncating variants (nonsense mutations, frameshift deletions/

insertions, and mutations at highly conserved splice sites) which were afterwards selected according to a recessive (biallelic mutations) or a dominant (heterozygous mutations) mode of inheritance. In addition, screening was performed for genes with both a heterozygous truncating germline variant and a second (somatic) mutation of the wildtype allele in the respective tumor sample (tumor suppressor model). Frequent alterations were excluded (minor allele frequency (MAF) $\geq$0.03 for the recessive model and $\geq$0.01 for the dominant model, based on data from dbSNP, The 1000GenomesProject, and the Exome Variant Server). In addition to truncating variants, missense variants of the candidate genes were selected for MAF <0.001 and deleterious effect, as predicted by at least two of three in silico prediction tools (PolyPhen-2, MutationTaster, and SIFT). Detailed visual inspection of the remaining variants in a read browser (Integrative Genomics Viewer) was done to exclude obvious sequencing artifacts.

To screen for point mutations of the most promising candidate genes, targeted next generation sequencing (NGS) was performed using TruSeq enrichment protocols on an Illumina HiSeq 2000 sequencer (Illumina, San Diego, USA), as described previously [8].

For data analysis, the *Varbank* pipeline version 2.6 and the Cartagenia BENCHlab NGS platform version 3.0.4 (Leuven, Belgium) were used. Splicing efficiencies of normal and mutant sequences were calculated using the splice prediction program NNSPLICE 0.9 from BDGP (the *Berkeley Drosophila Genome Project*). The expression of candidate genes in both normal and tumor colorectal tissue was checked by means of the EST profiles provided in *Unigene*. The etiological relevance of the mutations was further explored by evaluating their genetic intolerance to functional variation according to the *Residual Variation Intolerance score* (*RVIS*) [17] and the likelihood of haploinsufficiency according to the *Haploinsufficiency Score* [18].

The exome sequencing data of colorectal adenoma samples were screened for somatic mutations in known adenoma and colorectal cancer genes. The variants were selected for truncating variants (nonsense mutations, frameshift deletions/insertions, and mutations at highly conserved splice sites) and missense variants with a predicted deleterious effect by at least one of two in silico prediction tools (PolyPhen-2 and SIFT) and a MAF $\leq$ 0.01. Only variants with a read depth (coverage) of >10× and a fraction of mutated reads >10 % were considered. For filtering the *VariantStudio* software (Illumina) was used.

Sanger sequencing

The identified variants were validated by Sanger sequencing of the corresponding region using standard protocols (primer sequences available upon request). The results were analyzed with the *SeqPilot* software (JSI Medical Systems).

Frequency of colorectal tumors with somatic mutations in candidate genes

Data concerning the frequency (percentage) of colorectal tumors with somatic mutations in the 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 (https://tcga-data.nci.nih.gov/tcga/). To correct the data for the presence of passenger mutations, hypermutated tumors were excluded from the dataset. Therefore, the distribution of somatic variants in the TCGA exomes were analyzed and all tumors with >200 variants (24 % of the tumors) were excluded [12]. The remaining 295 exomes (76 % of tumors) were used to calculate the frequency of tumors with somatic mutations in the candidate genes.

Results

All seven patients in the discovery cohort presented with an attenuated colorectal phenotype and without evident extracolonic lesions or a conspicuous family history (sporadic or isolated cases). The basic clinical features of the discovery cohort are summarized in Suppl. Table S1. Our validation cohort consisted of 191 unrelated patients with the same phenotype and inclusion criteria (Suppl. Table S2).

In the discovery cohort, the mean on-target coverage of mapped reads was 57×, and 83 % of bases were covered at $\geq$10×. A total of 151,966 variants were called in the coding regions of the exome sequencing data. The overall performance of exome sequencing is shown in Suppl. Table S3.

No pathogenic germline mutation in known polyposis genes including the recently described *NTHL1* [8] was found. According to the tumor suppressor model no potential candidate gene could be identified. Afterwards, a number of stringent filter steps were applied to select for rare, truncating variants (Fig. 1), and obvious false positive results (artifacts) were excluded through detailed visual inspection. This approach identified two genes which were apparently affected by two truncating variants in at least one patient, indicating biallelic alterations (recessive model), and one gene which was affected by truncating heterozygous variants (dominant model) in at least two patients.

In the two genes consistent with a recessive model (*PIEZO1* and *ZSWIM7*), each of the variants was identified in one patient and appeared to be homozygous. In the gene

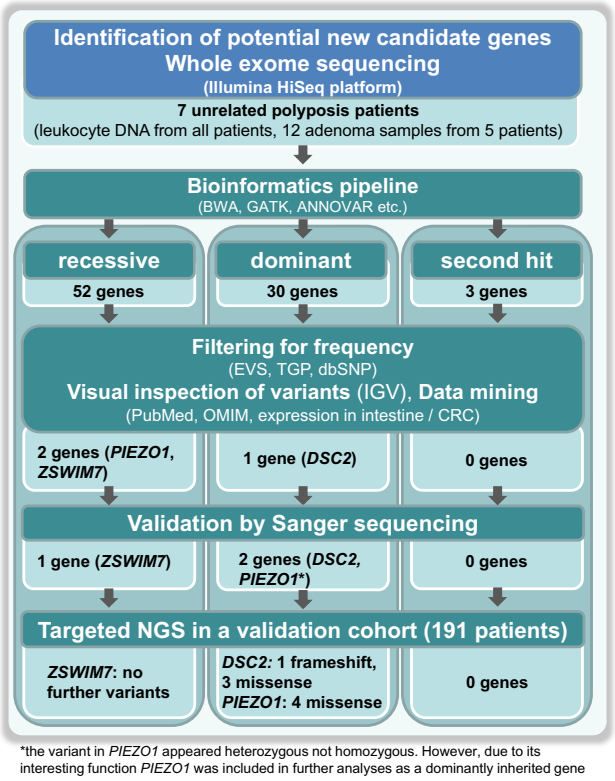

Fig. 1 Overall workflow of the exome sequencing and identification of potential new candidate genes

consistent with a dominant model (*DSC2*), both patients carried the same frameshift mutation (Table 1). Data mining according to function, pathways, and expression in colorectal tissue demonstrated that these three genes are involved in cell adhesion, proliferation, or recombination repair (Table 2). In none of these candidate genes a second hit could be detected in the corresponding colorectal tumor samples. No further candidate genes were identified through the inclusion of validated germline CNVs, i.e. if large deletions or partial duplications were considered as a second mutation in a certain gene in one patient (recessive model), or an additional heterozygous mutation in a second patient (dominant model).

Sanger sequencing of leukocyte DNA confirmed the variants in *DSC2* and *ZSWIM7* (Suppl. Fig. 1 and 2) whereas the variant in *PIEZO1* appeared heterozygous instead of homozygous which might be explained by the low read depth (7×) of this region (Suppl. Figure 3). However, in view of its interesting function, the variant was included in further analyses. In the discovery cohort, no missense variants in these three genes met the filter criteria and no putative pathogenic somatic point mutation ("second hit") was detected in tumor tissue while the germline mutation ("first hit") could be confirmed in all available colorectal adenomas of three patients (Suppl. Table S4). Unfortunately, no blood samples from relatives were available for a segregation analysis.

Subsequently, a validation cohort of 191 unrelated patients was screened for additional germline point mutations in the three promising candidate genes. In patient F807, the same *DSC2* frameshift mutation described above was detected. No other truncating mutation was found in any of the three genes. In total, six different heterozygous rare (MAF <0.001) missense variants, predicted to have deleterious effects by at least two of three in silico prediction tools, were identified (two missense variants in *DSC2* in three patients; and four unique missense variants in *PIEZO1)* (Table 1 and Suppl. Table S5). In *ZSWIM7,* no additional mutations meeting the filter criteria were found. No large deletions or partial duplications were detected in any of the three candidate genes.

In the families of two patients from the validation cohort (F386 and F807) with different *DSC2* mutations, analysis of affected and apparently unaffected relatives was possible (Suppl. Table S5). However, segregation of the variant with the phenotype was either excluded (F386), or could not be confirmed (F807).

To evaluate the functional relevance of these three candidate genes in more detail, they were examined for: (1) the occurrence of somatic mutations in colorectal tumors (TCGA exome data); (2) their genetic intolerance to functional variation (*Intolerance score* according to Petrovski [17]) and (3) the likelihood of haploinsufficiency (*haploinsufficiency score* according to Huang [18]). In none of the genes a high frequency (>3 %) of somatic mutations in 295 non-hypermutated colorectal tumors could be detected (Table 2). *DSC2* showed a low (negative) RVIS (−1.85 to −0.55, corresponding to values <25th percentile). This reflects high intolerance to genetic variation, which in turn indicates that this gene is subject to purifying selection. For *DSC2*, the *haploinsufficiency score,* which indicates dosage-sensitive genes, was also reduced (<20 %) (Table 2).

In addition, the exome sequencing data of 12 colorectal adenomas from five patients of the discovery cohort were filtered for potential pathogenic somatic mutations in 22 known adenoma and CRC driver genes (Suppl. Table S6). Pathogenic or likely pathogenic mutations could be identified in 10/12 samples (Suppl. Table S7). There was no evidence for the presence of *APC* mosaicism (i.e., in none of the patients the same *APC* mutation could be detected in different adenoma samples). In each of three adenoma samples, two different *APC* mutations could be identified and in each of two samples one *APC* mutation. In three and two samples, respectively, *CTNNB1* and *KRAS* mutations (codon 12 and 13) could be detected. In all those samples no *APC* mutation was identified. *FBXW7, MSH2,* and *TP53* mutations were found in addition to *APC* or *KRAS* mutations.

Table 1 Potential pathogenic mutations identified in exome sequencing patients (n = 7, shown in bold) and the validation sample (n = 191)

Gene	Patient ID	Mutation	Type variation	MAF (EVS, TGP, dbSNP)	Mutation prediction		
					Polyphen-2	Mutation taster	SIFT
DSC2	F386; F929	c.907G>A; p.Val303Met	Missense	0.0001 (EVS, rs145560678)	Possibly damaging	Disease causing	Damaging
DSC2	**F995; F1360**; F807	**c.2686_2687dupGA; p.Ala897Lysfs*4**	Ins	0.01 (EVS); 0.005 (TGP, rs200056085)			
DSC2	F998	c.2701A>G; p.Arg901Gly	Missense	NA	Probably damaging	Disease causing	Damaging
PIEZO1	F906	c.2104C>T; p.His702Tyr	Missense	NA	Probably damaging	Disease causing	Damaging
PIEZO1	F1899	c.3021C>G; p.Ile1007Met	Missense	NA	Probably damaging	Poly-morphism	Damaging
PIEZO1	F1909	c.5134G>A; p.Val1712Met	Missense	NA	Probably damaging	Disease causing	Damaging
PIEZO1	**F1526**	**c.5289C>G;p.Tyr1763***	Stop	NA			
PIEZO1	F1445	c.5863C>T; p.Arg1955Cys	Missense	NA	Probably damaging	Disease causing	Damaging
ZSWIM7	**F710**	**c.231_232del; p.Cys78Phefs*21 (homozygous)**	Del	0.0008 (EVS); 0.0009 (TGP, rs368517882) (heterozygous)			

With the exception of the mutation in *ZSWIM7*, all mutations were heterozygous

EVS Exome Variant Server, *MAF* minor allele frequency, *NA* not available, *TGP* 1000 Genomes Project

Table 2 Details of the candidate genes

Gene	Functions and pathways/literature	Frequency of somatic mutations, non-hypermutated tumors (%)[a]	RVIS [17] (percentile)	Haploinsufficiency score [18] (%)
DSC2 (NM_024422)	Cell adhesion; described in various cancer types (colorectal, lung, breast, urothelial, gastric, pancreatic ductal adenocarcinoma)	1 (0.3)	−1.22 (5.67)	15.5
PIEZO1 (NM_001142864)	Cell adhesion, cell migration and cell extrusion; associated with progress of lung tumors	1 (0.3)	3.74 (99.6)	NA
ZSWIM7 (NM_001042697)	Regulator of homologous recombination repair, interacts with RAD51D	NA	−0.08 (47.8)	64

[a] Data from exome sequenced colon and rectum adenocarcinomas of the Cancer Genome Atlas (TCGA) database: number of tumors with somatic mutations/total number of non-hypermutated tumors. Non-hypermutated tumors are defined as those with <200 somatic mutations per tumor (n = 295) [12]

NA not available, *RVIS* Residual Variation Intolerance Score

Discussion

In a number of patients with colorectal adenomatous polyposis, no germline mutation in the known causal genes can be identified. Although the syn- or metachronous occurrence of dozens to hundreds of adenomas is strongly suggestive of an underlying hereditary basis, it remains unclear so far, whether the predisposing genetic factors mainly act in a monogenic fashion, or contribute as low or moderately penetrant variants to a more complex, oligo/polygenic trait.

To uncover novel, potentially causative genes, exome sequencing of leukocyte and tumor DNA was performed in a discovery cohort of seven unrelated patients with histologically confirmed, genetically unexplained adenomatous polyposis (minimum of 50 adenomas). The findings were then confirmed in a large validation cohort. According to the observed somatic mutation spectra, the adenomas in this patient cohort follow the classical pathways of colorectal tumorigenesis [19].

Assuming a monogenic mode of inheritance and high penetrance, the frequency of causative germline mutations

in the general population is expected to be low. By applying an established stringent filter workflow, including comparisons with large control cohorts, we identified in four of the seven patients unique (i.e. not present in controls) or rare (i.e. frequency <1 % for dominant model or <3 % for recessive model in controls), potentially pathogenic germline variants in three protein coding genes (*DSC2*, *ZSWIM7, PIEZO1*) with molecular and cellular functions related to tumorigenesis. In two of these three genes, additional variants with predicted pathogenicity were detected in the validation cohort.

The cadherin superfamily member Desmocollin 2 (*DSC2*) is a critical component of desmosomes within the intestinal epithelium, and is thus involved in cell adhesion [20]. In colorectal cancer (CRC), decreased expression of DSC1-3 is significantly correlated with higher tumor grading [21]. Another study has shown that loss of *DSC2* results in proliferation of colonic epithelial cells through the activation of Akt/β-catenin signaling [22]. Furthermore, *DSC2* has been described in various other cancer types (lung, breast, urothelial, gastric, pancreatic ductal adenocarcinoma). The only truncating variant in *DSC2* identified in the present study was a frameshift mutation, which was detected in three patients. This and another of the three *DSC2* mutations found in our cohort have also been detected in patients with (arrhythmogenic right ventricular) cardiomyopathy (Suppl. Table S5), however, it is unknown whether these patients had any history of colonoscopy or gastrointestinal symptoms.

We identified probably pathogenic *DSC2* germline variants in 3 % of our patients. According to the known function and involved pathway, and the results of functional and in silico prediction tools, it is likely that the gene contributes in colorectal adenoma formation. However, the occurrence of 2/3 variants in normal controls at a very low frequency and the results of segregation analysis argues more in favor that *DSC2* mutations act as moderately penetrant risk factors rather than as highly penetrant mutations.

ZSWIM7 or *SWS1* is part of a complex which plays an important role in the homologous recombination pathway [23]. Within this complex, SWS1 interacts with RAD51D. The authors demonstrated that the knockdown of SWS1 reduces the number of cells with RAD51 foci, and that the SWIM domain is essential for the prorecombinogenic function of SWS1. The frameshift deletion in patient F710 is located within the SWIM-type Zinc finger domain. The very low frequency of the frameshift mutation identified in our patient in normal controls is consistent with the assumption of a recessive mode of inheritance.

Probably pathogenic *PIEZO1* germline variants were detected in 2.5 % of our patients. Knockdown of the multitransmembrane domain protein *PIEZO1* (also known as *FAM38A*) in epithelial cells has been shown to result in reduced cell adhesion and increased cell migration and metastasis in lung tumors [24]. It could be shown that a knockdown of the *PIEZO1* channel in zebrafish leads to the formation of epithelial cell masses by preventing the extrusion of supernumerary cells [25]. The authors also described the extrusion of living cells at colon surfaces. This suggests that a loss of *PIEZO1* in colon tissue could lead to polyp formation. According to a recent study, haploinsufficiency of *PIEZO1* leads to endothelial abnormality, indicating the importance of this gene for vascular structure [26]. *PIEZO1* mutations have also been described in patients with dehydrated hereditary stomatocytosis/hereditary xerocytosis [27]. However, the specific positions of these mutations differ from those observed in the present cohort, and so far, only missense and inframe variants have been reported.

Evaluation of the clinical relevance of germline variants identified in high-throughput experiments is challenging [28], particularly in the absence of recurrently mutated genes and when segregation data are inconsistent or lacking. Previous and the present data are consistent with the observation that newly identified monogenic subtypes of inherited tumor predisposition syndromes are very rare. At least some of the unexplained tumor syndromes appear to show extreme genetic heterogeneity, and large patient cohorts are therefore required to validate candidate genes through the identification of recurrent germline mutations.

The present analyses might have missed some mutations in the targeted exomes since some variants are not identified easily with currently available sequencing techniques (e.g. within repeat tracts in coding sequences) or due to low coverage of certain genomic regions. Moreover, some causative mutations might be located outside the targeted exome, e.g. in non-coding regions or unannotated genes.

In conclusion, the known functions of *DSC2, PIEZO1* and *ZSWIM7* render these genes interesting candidates, however, their causal and clinical relevance have to be further explored in larger cohorts.

Databases/URLs

APC locus-specific mutation database: www.lovd.nl/APC
COSMIC (Catalogue of somatic mutations in cancer): http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/
dbSNP (The Single Nucleotide Polymorphism database): www.ncbi.nlm.nih.gov/SNP/
Ensembl Genome Browser (release 54): http://may2009.archive.ensembl.org/index.html
EST profiles: www.ncbi.nlm.nih.gov/unigene/
EVS (Exome Variant Server): http://evs.gs.washington.edu/EVS/

HGMD (Human Gene Mutation Database): http://hgmd.org
IGV (Integrative Genomics Viewer): www.broadinstitute.org/igv/
MutationTaster: www.mutationtaster.org
MUTYH locus-specific mutation database: www.lovd.nl/MUTYH
NCBI: www.ncbi.nlm.nih.gov/
NNSPLICE 0.9: http://www.fruitfly.org/seq_tools/splice.html
Picard: http://broadinstitute.github.io/picard/
PolyPhen-2: http://genetics.bwh.harvard.edu/pph2/
Primer3 v.0.4.0: http://frodo.wi.mit.edu/primer3/input.htm
SIFT (Sorting Intolerant from Tolerant): http://sift.jcvi.org/
TCGA (The Cancer Genome Atlas): https://tcga-data.nci.nih.gov/tcga/
TGP (1000 Genomes Project): www.1000genomes.org
UCSC Genome Browser: http://genome.ucsc.edu
VARBANK: https://varbank.ccg.uni-koeln.de

Acknowledgments We thank the patients and their families for participating in the study. We are grateful to Susanne Raeder, Dietlinde Stienen, and Siegfried Uhlhaas for their excellent technical support.

Funding This work was supported by the German Cancer Aid (Deutsche Krebshilfe e.V. Bonn, Grant number 108421); the Gerok-Stipendium of the University Hospital Bonn (Grant no. O-149.0098); the NIH Centers for Mendelian Genomics (5U54HG006504); the Federal Ministry of Education and Research (0316190A); and the Volkswagenstiftung (Lichtenberg Program to M.R.S.). These 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 manuscript for publication. The corresponding author had full access to all study data, and had final responsibility for the decision to submit the manuscript for publication.

Compliance with ethical standards

Conflict of interest The authors have no conflicts of interest to declare.

References

1. Galiatsatos P, Foulkes WD (2006) Familial adenomatous polyposis. Am J Gastroenterol 101(2):385–398
2. Aretz S, Stienen D, Friedrichs N, Stemmler S, Uhlhaas S, Rahner N, Propping P, Friedl W (2007) Somatic APC mosaicism: a frequent cause of familial adenomatous polyposis (FAP). Hum Mutat 28(10):985–992
3. Hes FJ, Nielsen M, Bik EC, Konvalinka D, Wijnen JT, Bakker E, Vasen HF, Breuning MH, Tops CM (2008) Somatic APC mosaicism: an underestimated cause of polyposis coli. Gut 57(1):71–76
4. Spier I, Drichel D, Kerick M, Kirfel J, Horpaopan S, Laner A, Holzapfel S, Peters S, Adam R, Zhao B, Becker T, Lifton RP, Perner S, Hoffmann P, Kristiansen G, Timmermann B, Nothen MM, Holinski-Feder E, Schweiger MR, Aretz S (2015) Low-level APC mutational mosaicism is the underlying cause in a substantial fraction of unexplained colorectal adenomatous polyposis cases. J Med Genet. doi:10.1136/jmedgenet-2015-103468
5. Spier I, Horpaopan S, Vogt S, Uhlhaas S, Morak M, Stienen D, Draaken M, Ludwig M, Holinski-Feder E, Nothen MM, Hoffmann P, Aretz S (2012) Deep intronic APC mutations explain a substantial proportion of patients with familial or early-onset adenomatous polyposis. Hum Mutat 33(7):1045–1050
6. Mazzei F, Viel A, Bignami M (2013) Role of MUTYH in human cancer. Mutat Res 743–744:33–43
7. Palles C, Cazier JB, Howarth KM, Domingo E, Jones AM, Broderick P, Kemp Z, Spain SL, Guarino E, Salguero I, Sherborne A, Chubb D, Carvajal-Carmona LG, Ma Y, Kaur K, Dobbins S, Barclay E, Gorman M, Martin L, Kovac MB, Humphray S, Lucassen A, Holmes CC, Bentley D, Donnelly P, Taylor J, Petridis C, Roylance R, Sawyer EJ, Kerr DJ, Clark S, Grimes J, Kearsey SE, Thomas HJ, McVean G, Houlston RS, Tomlinson I (2013) Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. Nat Genet 45(2):136–144
8. Spier I, Holzapfel S, Altmuller J, Zhao B, Horpaopan S, Vogt S, Chen S, Morak M, Raeder S, Kayser K, Stienen D, Adam R, Nurnberg P, Plotz G, Holinski-Feder E, Lifton RP, Thiele H, Hoffmann P, Steinke V, Aretz S (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(2):320–331
9. Weren RDA, Ligtenberg MJL, Kets CM, de Voer RM, Verwiel ETP, Spruijt L, van Zelst-Stams WAG, Jongmans MC, Gilissen C, Hehir-Kwa JY, Hoischen A, Shendure J, Boyle EA, Kamping EJ, Nagtegaal ID, Tops BBJ, Nagengast FM, Geurts van Kessel A, van Krieken JHJM, Kuiper RP, Hoogerbrugge N (2015) A germline homozygous mutation in the base-excision repair gene NTHL1 causes adenomatous polyposis and colorectal cancer. Nat Genet 47(6):668–671
10. Gilissen C, Hoischen A, Brunner HG, Veltman JA (2012) Disease gene identification strategies for exome sequencing. Eur J Hum Genet 20(5):490–497
11. Aretz S, Stienen D, Uhlhaas S, Pagenstecher C, Mangold E, Caspari R, Propping P, Friedl W (2005) Large submicroscopic genomic APC deletions are a common cause of typical familial adenomatous polyposis. J Med Genet 42(2):185–192
12. Horpaopan S, Spier I, Zink AM, Altmuller J, Holzapfel S, Laner A, Vogt S, Uhlhaas S, Heilmann S, Stienen D, Pasternack SM, Keppler K, Adam R, Kayser K, Moebus S, Draaken M, Degenhardt F, Engels H, Hofmann A, Nothen MM, Steinke V, Perez-Bouza A, Herms S, Holinski-Feder E, Frohlich H, Thiele H, Hoffmann P, Aretz S (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(6):E578–589
13. Schweiger MR, Kerick M, Timmermann B, Albrecht MW, Borodina T, Parkhomchuk D, Zatloukal K, Lehrach H (2009) Genome-wide massively parallel sequencing of formaldehyde fixed-paraffin embedded (FFPE) tumor tissues for copy-number- and mutation-analysis. PLoS ONE 4(5):e5548
14. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25(14):1754–1760
15. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20(9):1297–1303

16. Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38(16):e164
17. Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB (2013) Genic intolerance to functional variation and the interpretation of personal genomes. PLoS Genet 9(8):e1003709
18. Huang N, Lee I, Marcotte EM, Hurles ME (2010) Characterising and predicting haploinsufficiency in the human genome. PLoS Genet 6(10):e1001154
19. Voorham QJ, Carvalho B, Spiertz AJ, Claes B, Mongera S, van Grieken NC, Grabsch H, Kliment M, Rembacken B, van de Wiel MA, Quirke P, Mulder CJ, Lambrechts D, van Engeland M, Meijer GA (2012) Comprehensive mutation analysis in colorectal flat adenomas. PLoS ONE 7(7):e41963
20. Funakoshi S, Ezaki T, Kong J, Guo RJ, Lynch JP (2008) Repression of the desmocollin 2 gene expression in human colon cancer cells is relieved by the homeodomain transcription factors Cdx1 and Cdx2. Mol Cancer Res 6(9):1478–1490
21. Knosel T, Chen Y, Hotovy S, Settmacher U, Altendorf-Hofmann A, Petersen I (2012) Loss of desmocollin 1-3 and homeobox genes PITX1 and CDX2 are associated with tumor progression and survival in colorectal carcinoma. Int J Colorectal Dis 27(11):1391–1399
22. Kolegraff K, Nava P, Helms MN, Parkos CA, Nusrat A (2011) Loss of desmocollin-2 confers a tumorigenic phenotype to colonic epithelial cells through activation of Akt/beta-catenin signaling. Mol Biol Cell 22(8):1121–1134
23. Martin V, Chahwan C, Gao H, Blais V, Wohlschlegel J, Yates JR 3rd, McGowan CH, Russell P (2006) Sws1 is a conserved regulator of homologous recombination in eukaryotic cells. EMBO J 25(11):2564–2574
24. McHugh BJ, Murdoch A, Haslett C, Sethi T (2012) Loss of the integrin-activating transmembrane protein Fam38A (Piezo1) promotes a switch to a reduced integrin-dependent mode of cell migration. PLoS ONE 7(7):e40346
25. Eisenhoffer GT, Loftus PD, Yoshigi M, Otsuna H, Chien CB, Morcos PA, Rosenblatt J (2012) Crowding induces live cell extrusion to maintain homeostatic cell numbers in epithelia. Nature 484(7395):546–549
26. Li J, Hou B, Tumova S, Muraki K, Bruns A, Ludlow MJ, Sedo A, Hyman AJ, McKeown L, Young RS, Yuldasheva NY, Majeed Y, Wilson LA, Rode B, Bailey MA, Kim HR, Fu Z, Carter DA, Bilton J, Imrie H, Ajuh P, Dear TN, Cubbon RM, Kearney MT, Prasad KR, Evans PC, Ainscough JF, Beech DJ (2014) Piezo1 integration of vascular architecture with physiological force. Nature 515(7526):279–282
27. Albuisson J, Murthy SE, Bandell M, Coste B, Louis-Dit-Picard H, Mathur J, Feneant-Thibault M, Tertian G, de Jaureguiberry JP, Syfuss PY, Cahalan S, Garcon L, Toutain F, Simon Rohrlich P, Delaunay J, Picard V, Jeunemaitre X, Patapoutian A (2013) Dehydrated hereditary stomatocytosis linked to gain-of-function mutations in mechanically activated PIEZO1 ion channels. Nat Commun 4:1884
28. MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, Adams DR, Altman RB, Antonarakis SE, Ashley EA, Barrett JC, Biesecker LG, Conrad DF, Cooper GM, Cox NJ, Daly MJ, Gerstein MB, Goldstein DB, Hirschhorn JN, Leal SM, Pennacchio LA, Stamatoyannopoulos JA, Sunyaev SR, Valle D, Voight BF, Winckler W, Gunter C (2014) Guidelines for investigating causality of sequence variants in human disease. Nature 508(7497):469–476