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

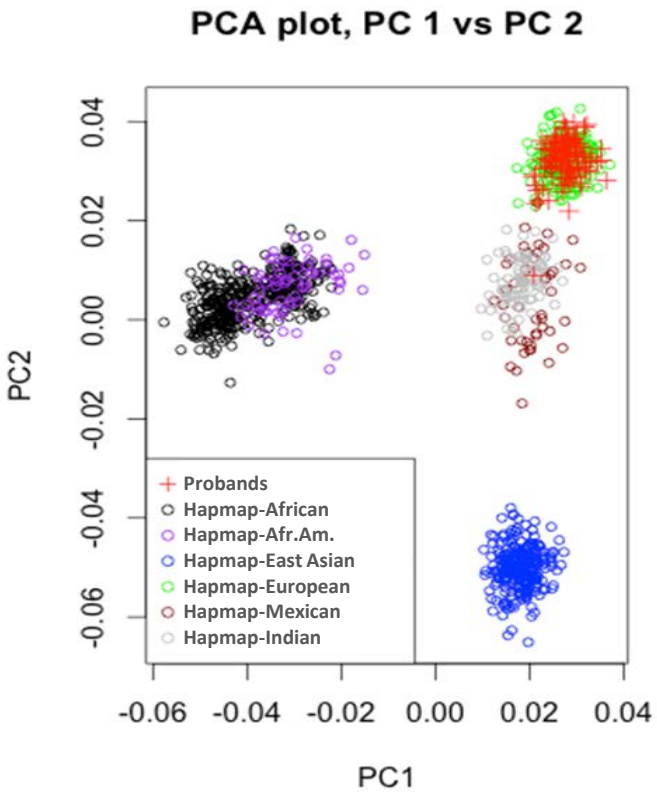


Figure S1 Results of Principle Component Analysis (PCA). The genotypes of the 100 probands who met coverage standards were compared to HapMap genotype data. A central European origin was indicated for all samples, with the exception of one outlier sample, which was subsequently excluded from further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

Figure S2

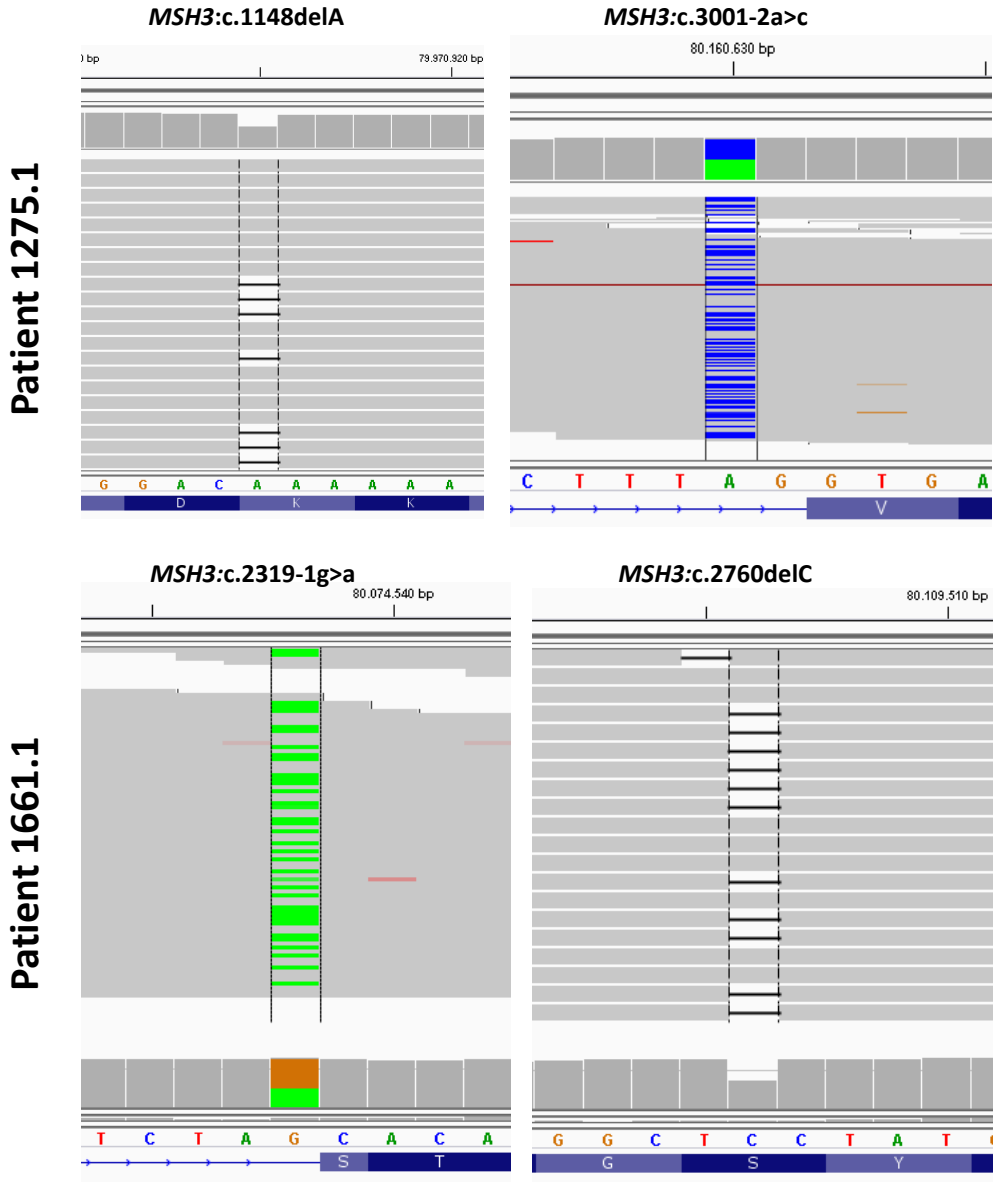


Figure S2 ***MSH3* mutations in exome sequencing data.** Biallelic loss-of-function *MSH3* germline mutations identified via exome sequencing in leukocyte DNA of patients 1275.1 and 1661.1 (visual control of reads in IGV browser).

Figure S3

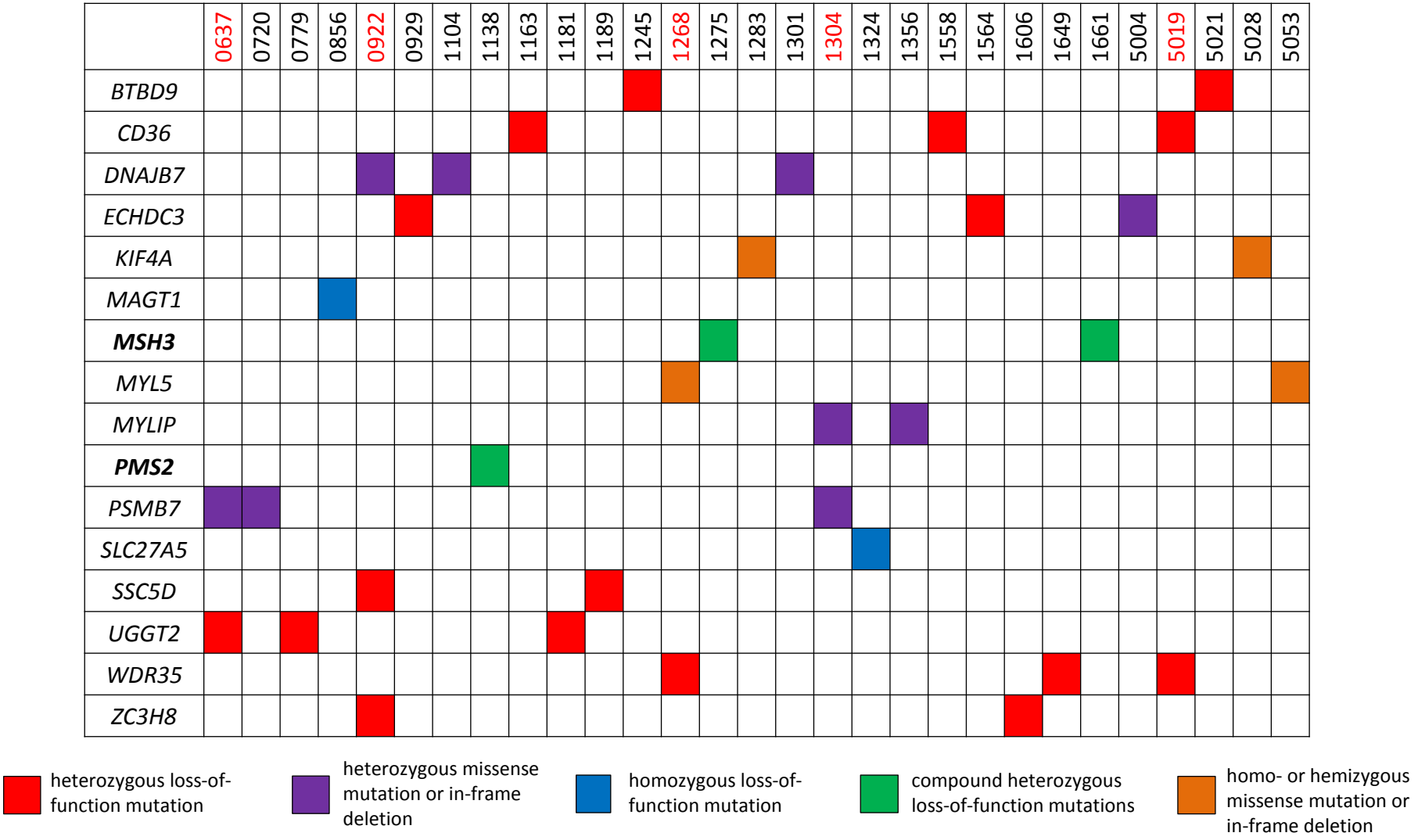
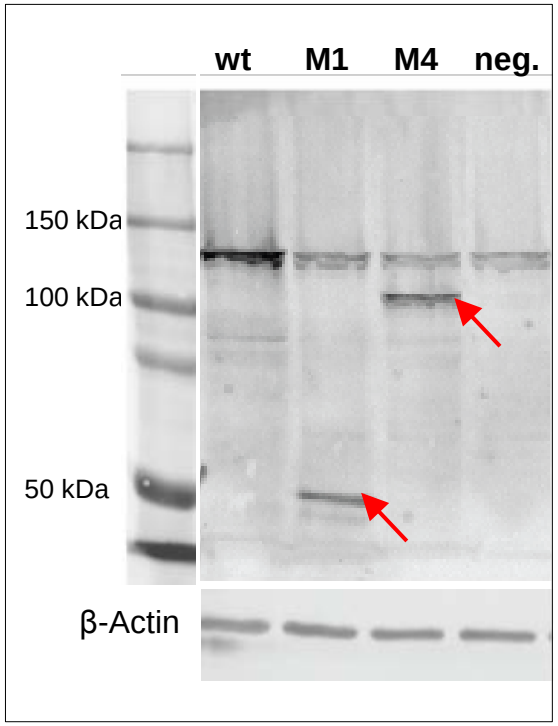


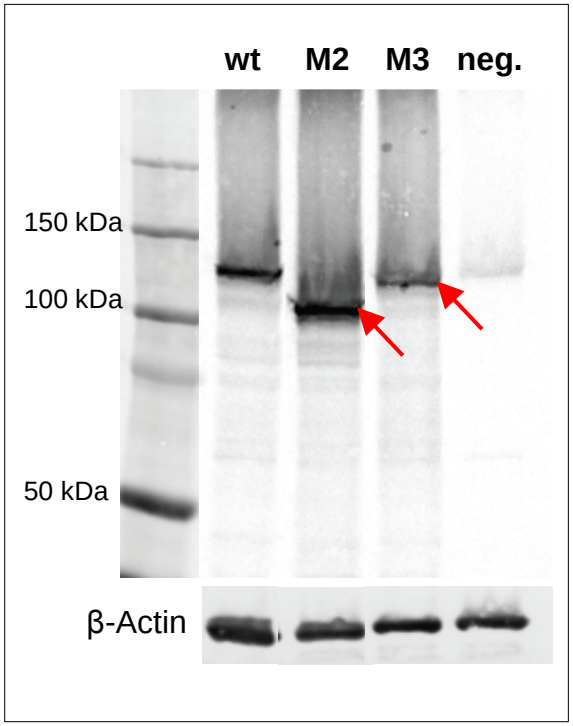
Figure S3 **Candidate genes identified in the present study.** All genes (labeled left) are affected by rare heterozygous or biallelic, potentially pathogenic variants in ≥ 2 alleles of the cohort. Patients (labeled above) harboring ≥ 2 mutated candidate genes are shown in red.

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

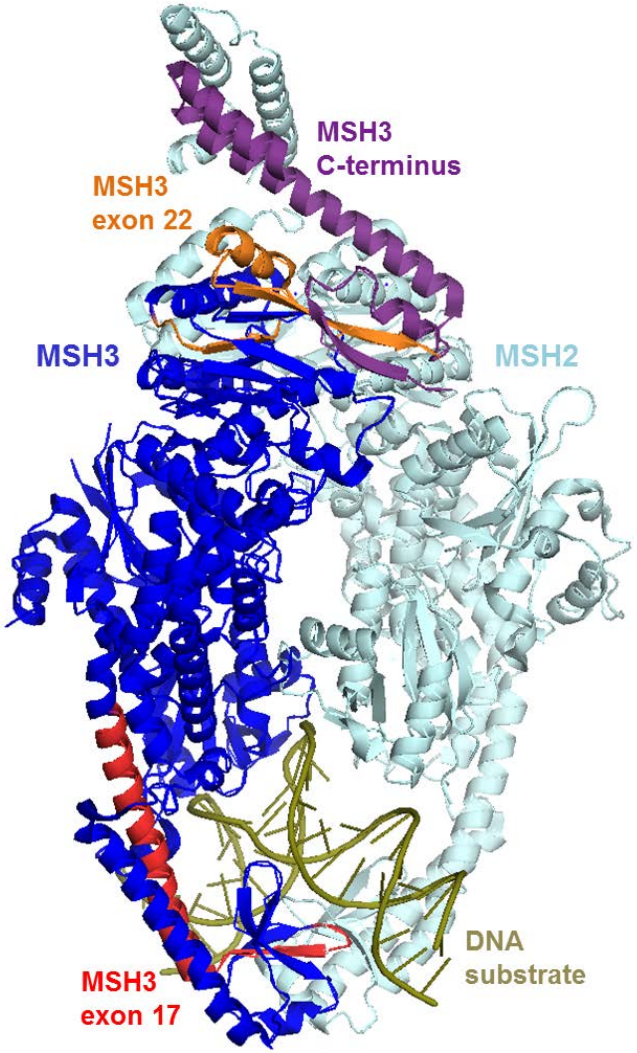


Figure S5 Mapping of deleted exons to protein structure. The human MutSβ structure published by Yang's group shows MSH2 (pale blue) and MSH3 (dark blue) binding to a dinucleotide loop (olive). We highlighted amino acids coded by regions affected by the altered splicing in red (exon 17), orange (exon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

Figure S6

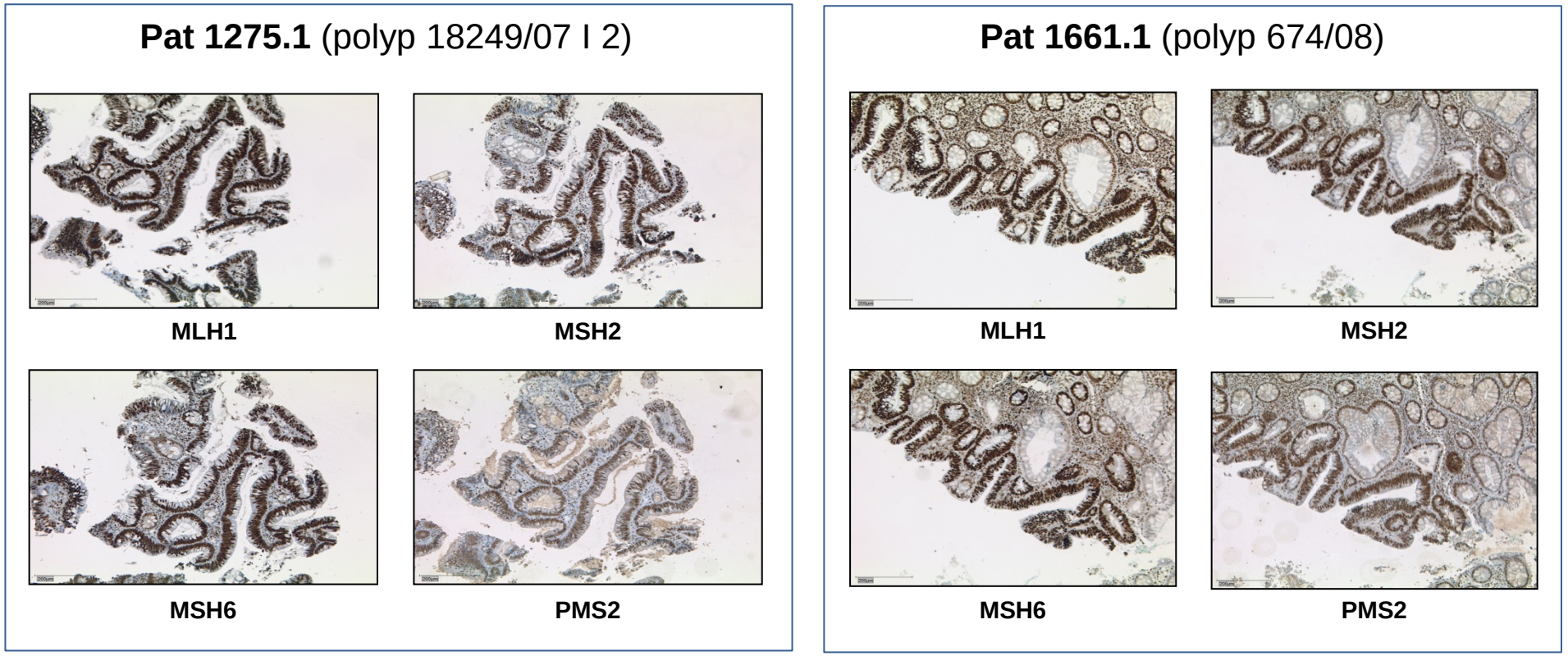


Figure S6 Expression of MMR partner proteins. Immunohistochemical staining of adenoma tissue shows normal protein levels of other MMR proteins (MLH1, MSH2, MSH6 and PMS2) for the index patients of families 1275 and 1661.

Figure S7

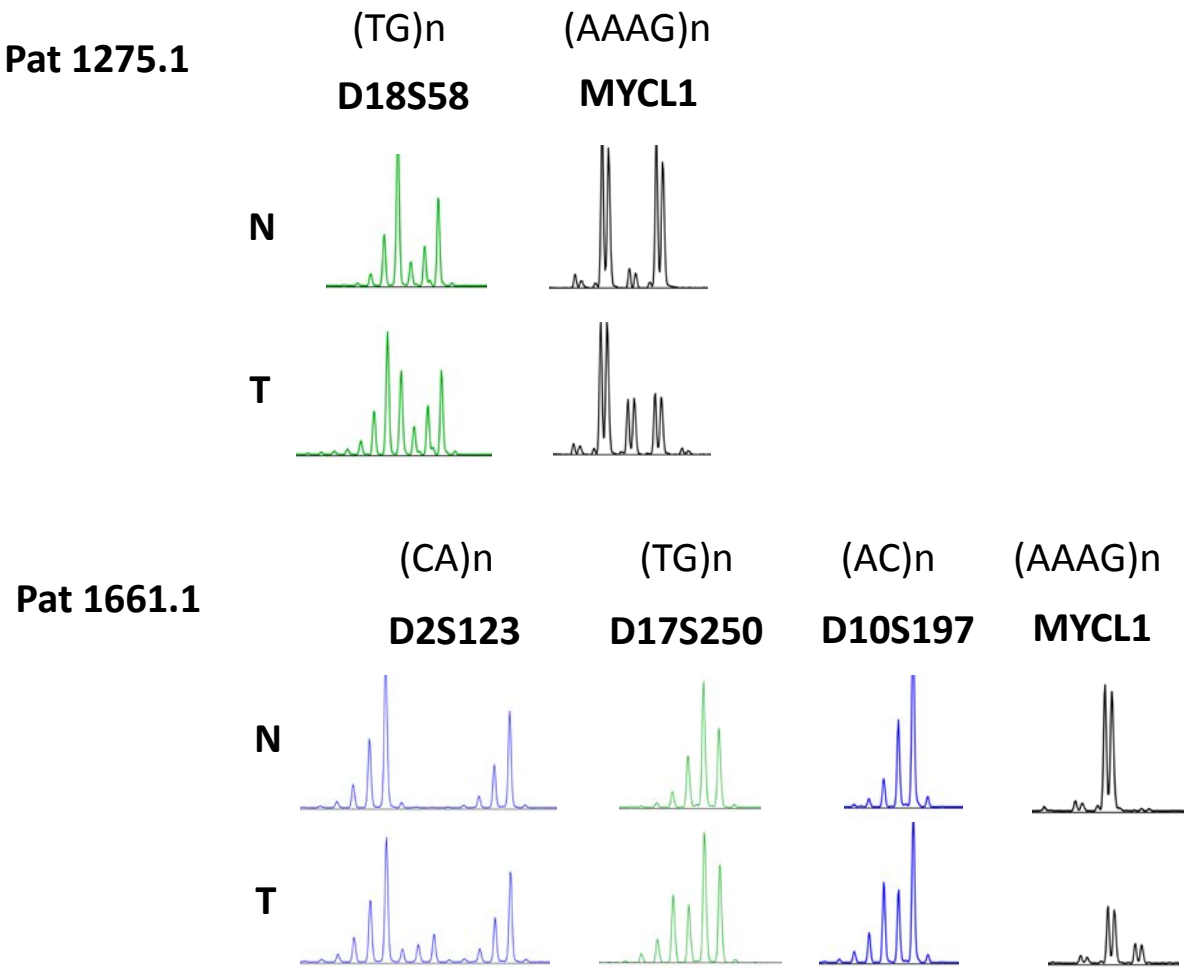
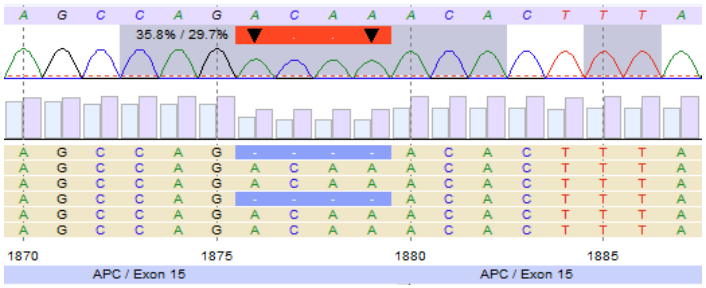


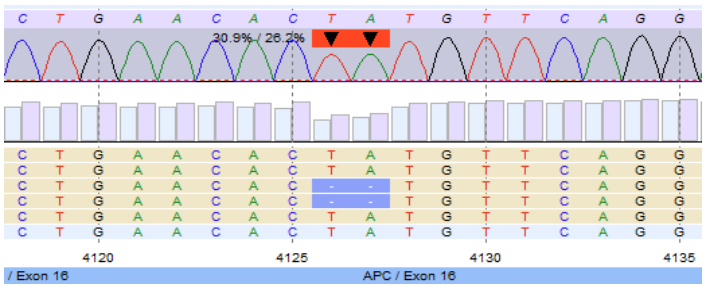
Figure S7 Results for additional microsatellite markers. Four dinucleotide repeat markers (D2S123, D17S250, D10S197, D18S58) and one further tetranucleotide repeat marker (MYCL1) were examined in normal and tumor tissue. Only those markers showing instability are depicted. All results were validated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

Polyp E 48579-08

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

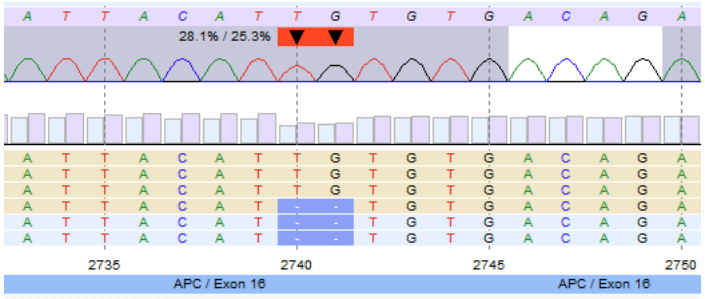


***APC*:c.4126_4127delTA** (31/26%)

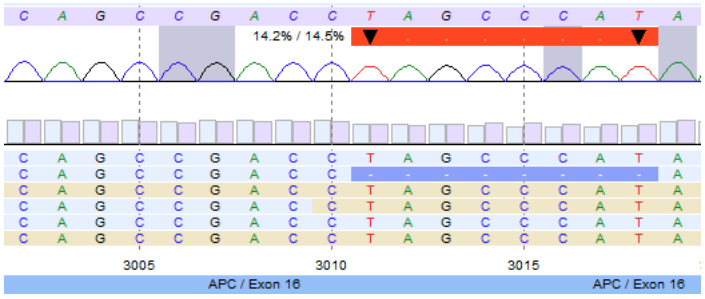


Polyp 674-08 3

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

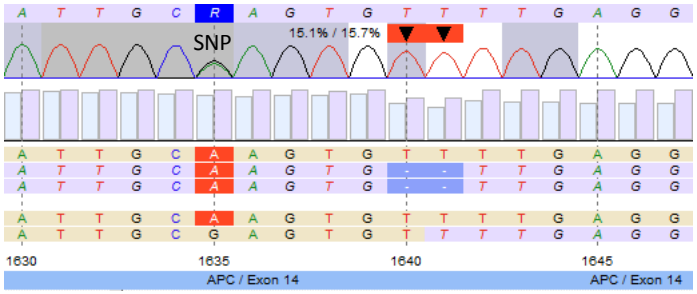


***APC*:c.3011_3018delTAGCCCAT** (14/15%)

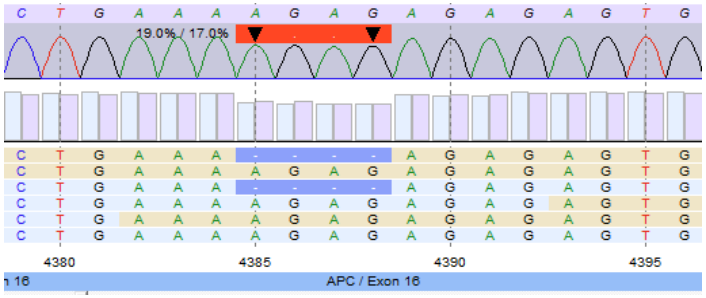


Polyp 674-08 4

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



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



Polyp 674-08 2

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

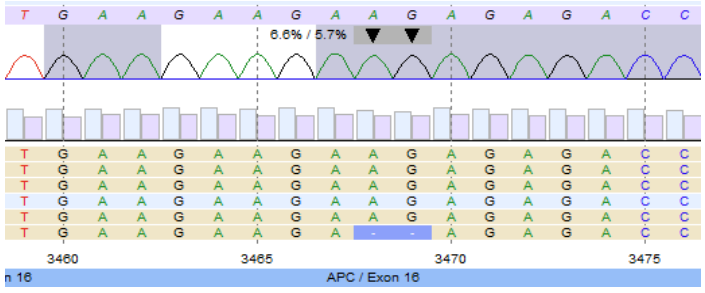


Figure S8

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

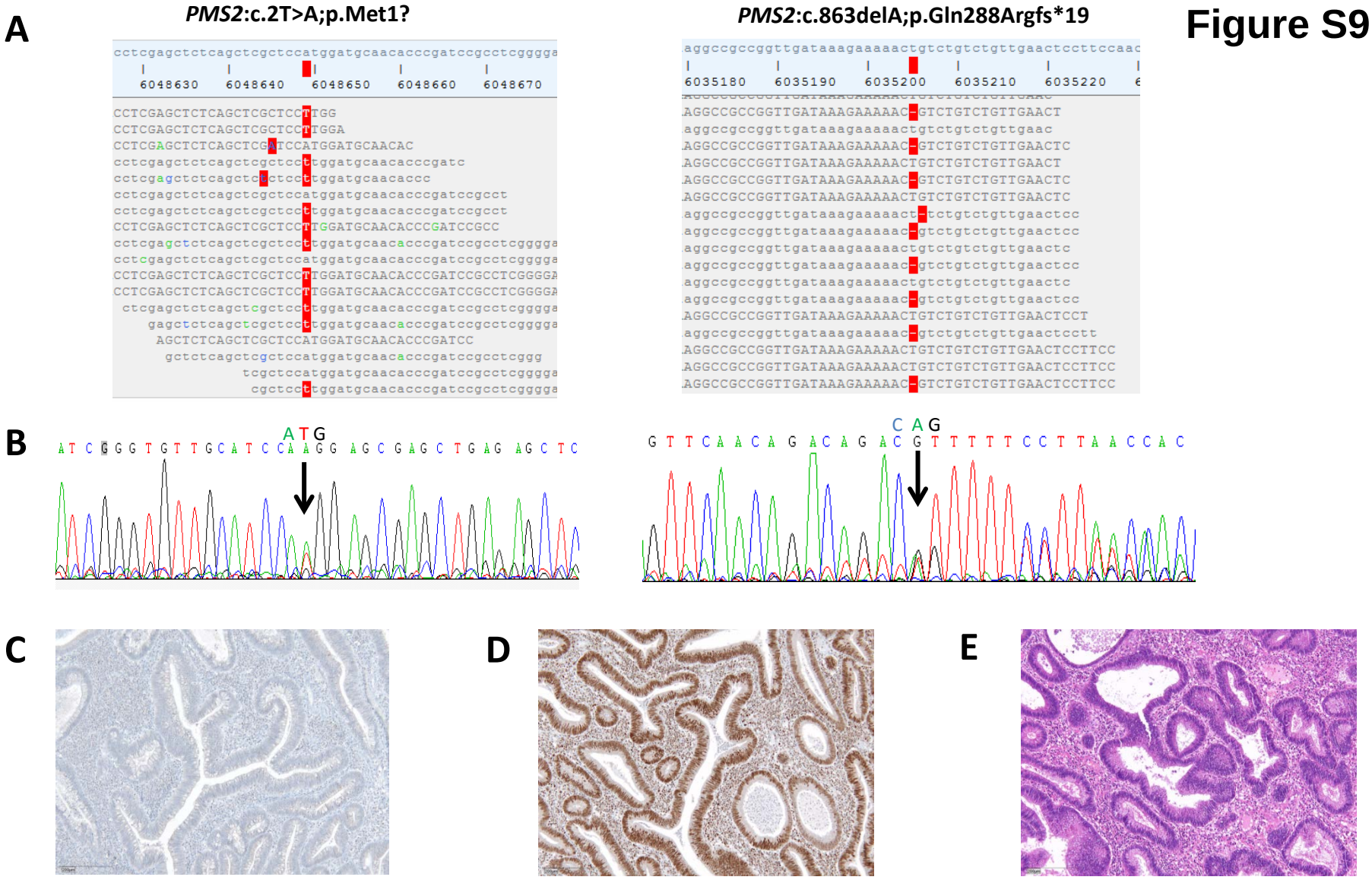


Figure S9 ***PMS2*** **germline mutations.** (A) Exome sequencing from leukocyte-derived DNA of patient 1138 revealed the mutations c.2T>A;p.Met1? and c.863delA;p.Gln288Argfs*19 in *PMS2* (reverse sequence, Varbank). (B) Sanger validation of *PMS2* exons 1 and 8 confirmed both mutations (top line shows 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.