SUPPLEMENTARY FILE

Supplementary Methods

Study subjects: Kindreds with PMS2 c.2002A>G were identified through hereditary cancer clinics at hospitals affiliated to McGill University and the University of Manitoba. Probands were referred to medical genetics clinics by physicians. Signed informed consent was obtained from all study participants or their legal representatives. The local ethical committees of McGill University, the University of Manitoba and other collaborating centers approved this study. Bio-specimens used in laboratory investigation consisted of patient derived biological material in the form of peripheral blood lymphocytes, colon or skin biopsies. Since not every type of material was available for every patient, the number of samples used in each experimental procedure was determined by availability. Additional information for *PMS2* genotype-phenotype comparisons was retrieved from the Leiden Open Variation Database (http://www.lovd.nl/3.0/home) and Online Mendelian Inheritance in Man (http://omim.org/) and was further validated by reviewing relevant publications.

DNA, RNA and cDNA preparation: DNA and RNA from peripheral blood lymphocytes were purified using the Gentra Puregene Blood Kit (Qiagen, Toronto, Canada) and PAXgene Blood RNA Kit IVD (Qiagen, Toronto, Canada), respectively, according to manufacturer protocols. DNA and RNA concentrations were quantified using a NanoDrop™ Spectrophotometer (Thermo Scientific, Ontario, Canada). cDNA synthesis was performed using the QuantiTect reverse

transcription kit (Qiagen, Toronto, Canada) with 100 ng of total RNA as input template. cDNA was stored at -20°C for downstream applications.

Mutation identification: Immunohistochemical staining of mismatch repair proteins MLH1, MSH2, MSH6 and PMS2 was done on formalin-fixed, paraffin-embedded tissues from affected individuals using standard protocols at treating hospitals. For the protein truncation test, cDNA fragments of the *PMS2* transcript were amplified from lymphocyte RNA using reverse transcription and long-range PCR kits (Qiagen, Toronto, Canada), followed by *in vitro* translation using the TNT Quick Coupled Transcription / Translation System (Promega, Madison, United States) and ^{35}S-methionine (Amersham Biosciences, Piscataway, United States). For cDNA sequencing, transcript derived from the functional *PMS2* gene locus was amplified specifically with a forward primer located in exon 10. For gDNA sequencing, nested PCR was employed to avoid interference from pseudogenes. Briefly, the genomic region spanning exons 10, 11 and 12 was amplified using an Expand Long Template PCR System (Roche, Mississauga, Canada), the long-range PCR product was then diluted to1:40, and 2 µl of the diluted product was used as a template for amplification with an internal set of primers. The mutation was identified by sequencing the final PCR products with a reverse primer derived from exon 12 using a 3730xl DNA Analyzer (Applied Biosystems, Foster City, United States).

Prediction Algorithms used: The predicted effect of NM_000535.5:c.2002A>G on protein function was evaluated using Mutation Taster (http://mutation.taster.org/), PolyPhen-2 (Polymorphism Phenotyping V2,

http://genetics.bwh.harvard.edu/pph2/index.shtml) and PROVEAN (Protein Variation Effect Analyzer, http:/provean.jcvi.org/index.php).[1-3]

Mutation age calculation: We genotyped 17 microsatellite markers spanning a region of 8.8 megabases around *PMS2.* We used methods previously described in Hamel et al.[4] to perform the genotyping and to estimate the age of the *PMS2*:c.2002A>G mutation.

Polony Assay: Polony assay relies on solid phase PCR performed with template molecules separated and immobilized in a thin film of acrylamide gel. Each individual template molecule is then amplified, forming a homogenous molecule colony. Polonies are visualized by probe hybridization and single base extension (SBE) using fluorescein labelled dideoxynucleotides. The primer set for the solid phase PCR producing the 960 bp product was designed to unbiasedly amplify both the functional *PMS2* and pseudo *PMS2CL* loci (Figure 2A). The resulting polonies were identified by two rounds of probe hybridization coupled with SBE: one round was designed for differentiating the real/pseudo *PMS2* transcripts and the other for detecting the intact/aberrant exon 11-12 junctions (probe/primer sequences are listed in Table S2). The probes were positioned immediately before a variant distinguishing each pair of transcripts so each transcript (cDNA) polony was colored with the fluorescein of the specific dideoxynucleotides added to the pre-hybridized probe by complementary base-pairing during SBE.

Molecule-specific PCR: Locus specificity was achieved by placing the forward primer in exon 10, which is unique to the functional *PMS2* gene. Each transcript population was targeted by setting a specific nucleotide at the 3' end of the

reverse primer, complementary to either the intact or the aberrant exon 11-12 junction (Figures 2A and S4A). We performed molecule-specific PCR using the same cDNA template from the Polony assay and used a touch-down protocol.[5] The PCR products were then visualized by fragment analysis, which validated the expression of both the aberrant and intact exon 11-12 junctions from the mutant allele (Figure 2C).

Western blot and immunoprecipitation: Pierce IP Lysis Buffer (Thermal Scientific, Rockford, United States) was used for protein extraction and immunoprecipitation. Protease inhibitors (Roche, Laval, Canada) were added prior to harvesting the cells. Total cell lysates were obtained from 4×10^7 and 4×10^8 LCLs for protein analysis by direct western blot and immunoprecipitation of the MLH1-PMS2 complex, respectively. Primary antibody incubation for western blot analysis was done at a 1:200 dilution for MLH1 (BD Biosciences, Mississauga, Canada, material # 550838) and 1:500 for PMS2 (BD Biosciences, Mississauga, Canada, material # 556415). The secondary antibody was horseradish peroxidase-conjugated goat anti-mouse (Sigma Aldrich, Oakville, Canada) at a dilution of 1:20,000. For each immunoprecipitation experiment, 2 µg MLH1 antibody (same as above), 30 µl protein A sepharose beads slurry (GE Healthcare, Buckinghamshire, United Kingdom) and 1 mg cell lysate were incubated in a total volume of 1 ml in IP lysis buffer. Protein electrophoresis and western blot were performed by following the protocols for the Mini-PROTEAN® system (Biorad, Mississauga, Canada) using a 7.5% acrylamide gel. The detection system used was the Super Signal West Femto kit (Thermo Scientific, Rockford, United States)

and Amersham HyperfilmTm ECL (GE Healthcare, Buckinghamshire, United Kingdom). The MLH1-PMS2 complex was collected from the lymphoblastoid lysate via immunoprecipitation with a primary antibody against MLH1. The protein complex was then detected by the inclusion of two monocloncal antibodies (anti-MLH1 at 1:500 dilution and anti-PMS2 at 1:200 dilution) during the primary incubation of western blot.

Microsatellite Instability (MSI) Assay: A highly sensitive method is required to detect the rare variant alleles in normal tissues where MSI occurs as somatic events, a scenario quite different from detecting MSI in tumours where assessment is based upon the comparison between normal and cancer tissues.[6] Tetranucleotide microsatellite markers have been proven to have optimal sensitivity for DNA repair deficiency in the cellular context of *PMS2* null compared to mononucleotide or dinucleotide markers.[7] After pilot experiments with several markers, we chose *D17S1307*, a tetranucleotide marker located in the first intron of the Neurofibromatosis 1 (NF1) gene, to pursue quantitative MSI.

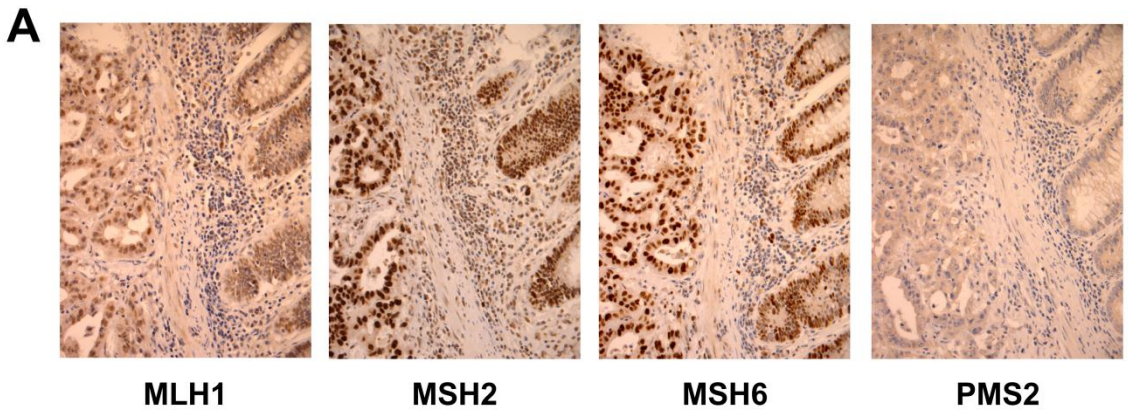
Figure S1. c.2002A>G is a coding mutation that interferes with RNA splicing.

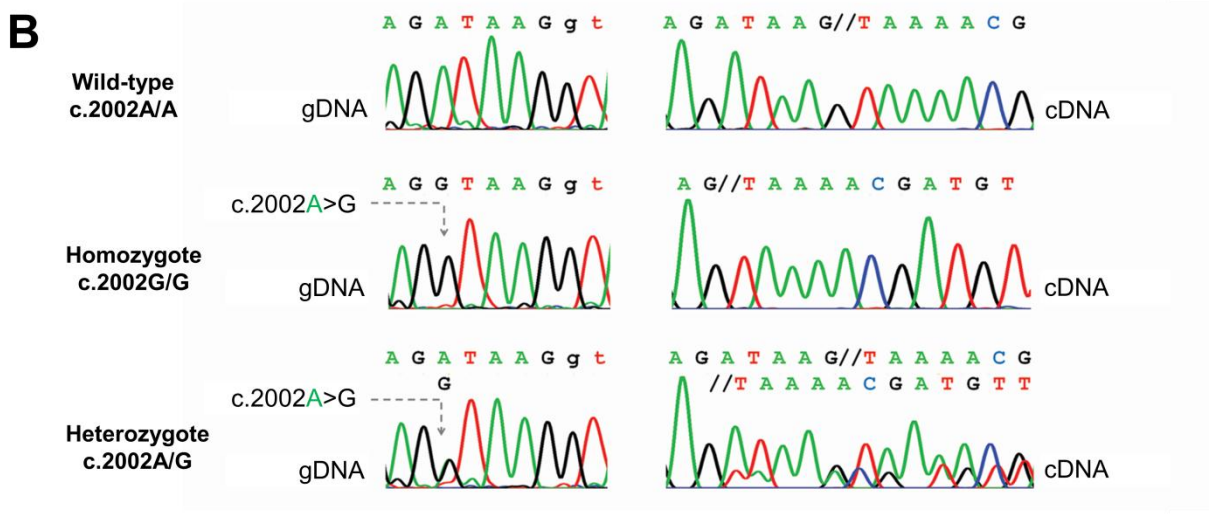
A. Representative images of immunohistochemical staining of mismatch repair proteins in the colorectal cancer from individual III-2 from Table S1, 400X magnification. Images show normal colon crypts (right) and invasive adenocarcinoma arising in the colon (left). Protein expression is identified as dark brown staining of the tumor cell nuclei. Counterstain of the nuclei with haematoxylin is blue. Note the complete absence of staining of both normal and tumour cells for PMS2.

B. Electropherograms of Sanger sequencing with gDNA and cDNA from peripheral lymphocytes. Intronic sequences in gDNA are shown in small capital letters. "//" denotes the junction between exons 11 and 12 in cDNA.

C. c.2002A>G is located at the boundary of exon 11 and intron 11, causing the substitution of Isoleucine to Valine (reference sequence NM_000535.5). Sanger sequencing with cDNA from homozygotes detected a five base-pair deletion, GTAAG, at the exon 11-12 junction of *PMS2*. The aberrant transcript results from RNA splicing at the *de novo* 5' ss generated by the "G" mutant allele.

Figure S1


A
MLH1
MSH2
MSH6
PMS2


B
Wild-type
c.2002A/A
gDNA
cDNA
Homozygote
c.2002G/G
c.2002A>G
gDNA
cDNA
Heterozygote
c.2002A/G
c.2002A>G
gDNA
cDNA

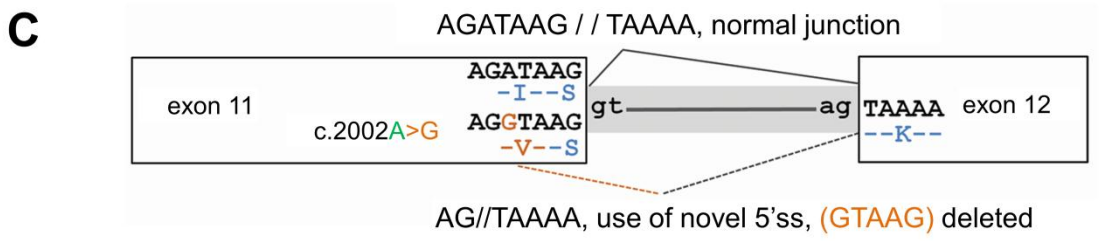
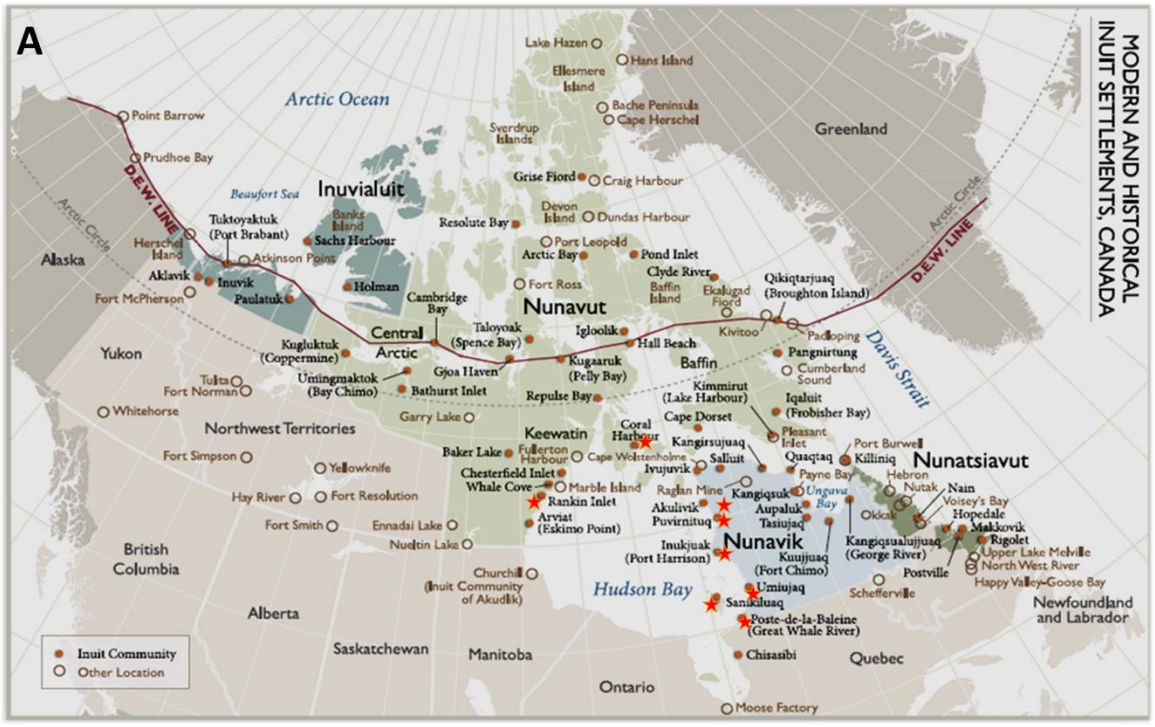

C
AGATAAG / / TAAAA, normal junction
exon 11
c.2002A>G
AGATAAG
-I--S
AGGTAAG
-V--S
gt
ag
TAAAA
--K--
exon 12
AG//TAAAA, use of novel 5'ss, (GTAAG) deleted

Figure S2. c.2002G>A is a founder mutation in the Inuit population.

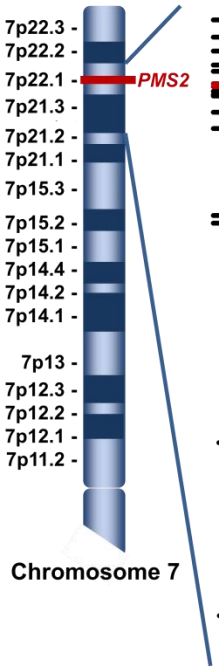
A. Geographical distribution of the c.2002A>G mutation. Communities with cancer patients homozygous for this mutation are marked with a red star. In total, eight communities were found to harbor this mutation, all located around Hudson Bay. No homo- or heterozygote individual was identified in a tested series of non-Nunavik residents (n = 6938), comprising 6435 newborns from Quebec City, 500 unselected individuals from the communities in Inuvialuit (including McKenzie inlets) and three colon cancer patients from Greenland (data not shown).

B. Shared chromosomal fragment in all c.2002A>G carriers. Genotyping of 17 short tandem repeat markers was done for 5 families where DNA was available for both heterozygous and homozygous members of the family. Numbers represent alleles. Each column of numbers represents one haplotype observed in that family. When two numbers are listed (e.g. 4/1), it indicates inconclusive status for which allele is linked to the mutation, so both alleles remain possibilities. X means no genotype information was available for this locus. Alleles in red are located within the *PMS2* gene region. The minimum conserved region is identified to be ~581 kilobases and is highlighted using a transparent box. The maximum conserved region is 1.8 megabases (alleles in bold face). The age of the mutation was estimated to be approximately 46 generations (95% CI, 24-80 generations). Assuming a generation time of 18 years for this population, we estimate the mutation first appeared late in the 11th century, Common Era.

Figure S2



B



Marker	Position hg18	Family 1		Family 2		Family 3		Family 4		Family 5		
PMS2-MS1	4899522	4	5	2	5	4	5	2	5	2	4	2/4
D7S1492	5180444	2	4	2	2	2/4	2/4	2	2	2/4	x	x
PMS2-MS2	5410784	5	8	4	8	5	8	4	8	4	4	4/5
PMS2-MS3	5597181	1	3	3	3	3	1	3	3	3	2	3/2
PMS2-MS18	5754769	x	x	1	1	1	1	1	1	1	4	1
PMS2-MS15	5843225	7	4	x	4/1	7	4	4	4	4	6	4/7
Mutation	5983772	m	m	m	m	m	m	m	m	m	m	m
PMS2-MS6	6002749	5	5	5	5	5	5	5	5	5	5	5/1
PMS2-MS10	6025122	5	5	x	5	5	5	5	5	5	5	5/10
PMS2-MS11	6051898	13	13	13	13	13	13	13	13	13	13	13/1
PMS2-MS12	6077298	1	1	x	1	1	1	1	1	1	1	1
PMS2-MS14	6366249	4	4	x	4/1	4	4	4	4	4	4	4/1
PMS2-MS19	6564678	x	x	3	3	3	3	3	3	3	3	3/4
D7S1527	7699506	3	2	4	2	2/3	2/3	2	2	2	3	2/1
D7S2514	7784387	6	9	9	8	8/9	8/9	8	8	8	6	8/11
D7S2547	9840585	3	1	3	3	3	3	3	3	4	3	4
D7S620	12154878	2	2	2	2	2	2	3	2	3	2	3/2
D7S664	13656753	8	8	5/8	5/8	7/8	7/8	8	8/7	8	8	8/7

Figure S3. The age of cancer onset in CMMRD patients inversely correlates with *PMS2* expression status

Group I consists of 43 patients carrying bi-allelic *PMS2* truncating mutations. Group II consists of 16 patients carrying bi-allelic *PMS2* mutations with laboratory evidence of residual gene expression. Patients are of various ethnic backgrounds. The X-axis indicates age and the Y-axis lists the *PMS2* genotypes of each individual. The age at diagnosis of primary (red), secondary (orange) and additional cancers (dark yellow) for each individual is plotted along the X-axis, with multiple cancers in the same person being connected by lines. Darkness indicates death. The difference in age at primary cancer onset between group I (median = 8, range = 1–16 years) and Group II (median = 20, range = 3–38 years) was statistically significant according to a Mann-Whitney test: U = 591.5, N1 = 43, N2 = 16, two-tailed P = 2.4×10^{-5}. The observation that the age for primary cancer onset in Group II is significantly delayed compared to Group I, supports that residual *PMS2* expression (shown in Group II homozygotes) contributes to genome-guarding function thus delays disease onset.

Figure S3

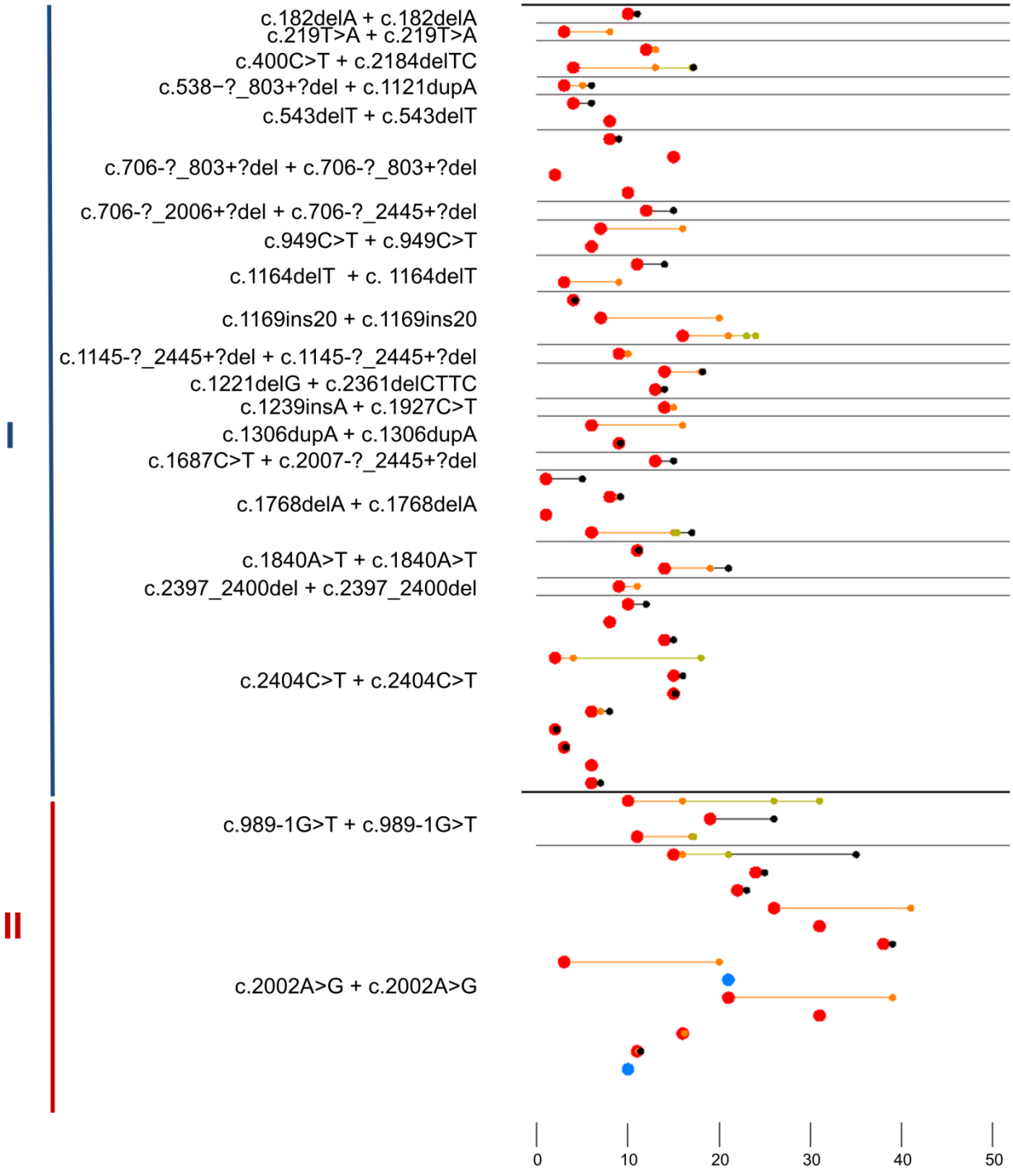

c.182delA + c.182delA
c.219T>A + c.219T>A
c.400C>T + c.2184delTC
c.538−?_803+?del + c.1121dupA
c.543delT + c.543delT
c.706-?_803+?del + c.706-?_803+?del
c.706-?_2006+?del + c.706-?_2445+?del
c.949C>T + c.949C>T
c.1164delT + c. 1164delT
c.1169ins20 + c.1169ins20
c.1145-?_2445+?del + c.1145-?_2445+?del
c.1221delG + c.2361delCTTC
c.1239insA + c.1927C>T
c.1306dupA + c.1306dupA
c.1687C>T + c.2007-?_2445+?del
c.1768delA + c.1768delA
c.1840A>T + c.1840A>T
c.2397_2400del + c.2397_2400del
c.2404C>T + c.2404C>T
c.989-1G>T + c.989-1G>T
c.2002A>G + c.2002A>G
I
II
0
10
20
30
40
50

Figure S4. Molecule-specific PCR of the exon 11-12 junction of *PMS2*

A. Graphical representation of the molecule-specific PCR designed to differentiate the intact (wild type) and aberrant exon 11-12 junction. A common forward primer is placed in exon 10 to ensure *PMS2*-specific amplification; the reverse primer is located at the exon junction, with the 3' nucleotide priming for either the intact (W for wild type) or the aberrant (M for mutant) transcript-population. W: 1016 bp amplicon with an intact exon 11-12 junction. M: 1011 bp amplicon with the aberrant exon 11-12 junction of a five-bp deletion. The PCR products are resolved in an agarose gel after 34 cycles of amplification for M and 48 cycles for W.

B. Validation of the specificity of molecule-specific PCR using cells homozygous for different *PMS2* genotypes. R802X: c.2404C>T, is a truncating mutation located at exon 14, only W junction is present in the cDNA. p.N412DfsX6 contains a deletion of exons 11-14, both W and M junctions are absent from the cDNA. These controls demonstrate that the junction anchored, molecule-specific PCR is highly specific. c.2002A>G, the mutation reported in this study, is a missense mutation interfering with RNA splicing, both W and M junctions are present in the cDNA.

C. W-PCR was performed with RNA from subjects homozygous for c.2002A>G and affected by cancer in their twenties. Lanes 1–4: PCR product with the RNA of lymphocytes from four patients. Lanes 5–6: PCR product with the RNA of primary fibroblasts derived from two patients. Lane 7: PCR product with the RNA of colon biopsy from the same patient as lane 4.

Figure S4

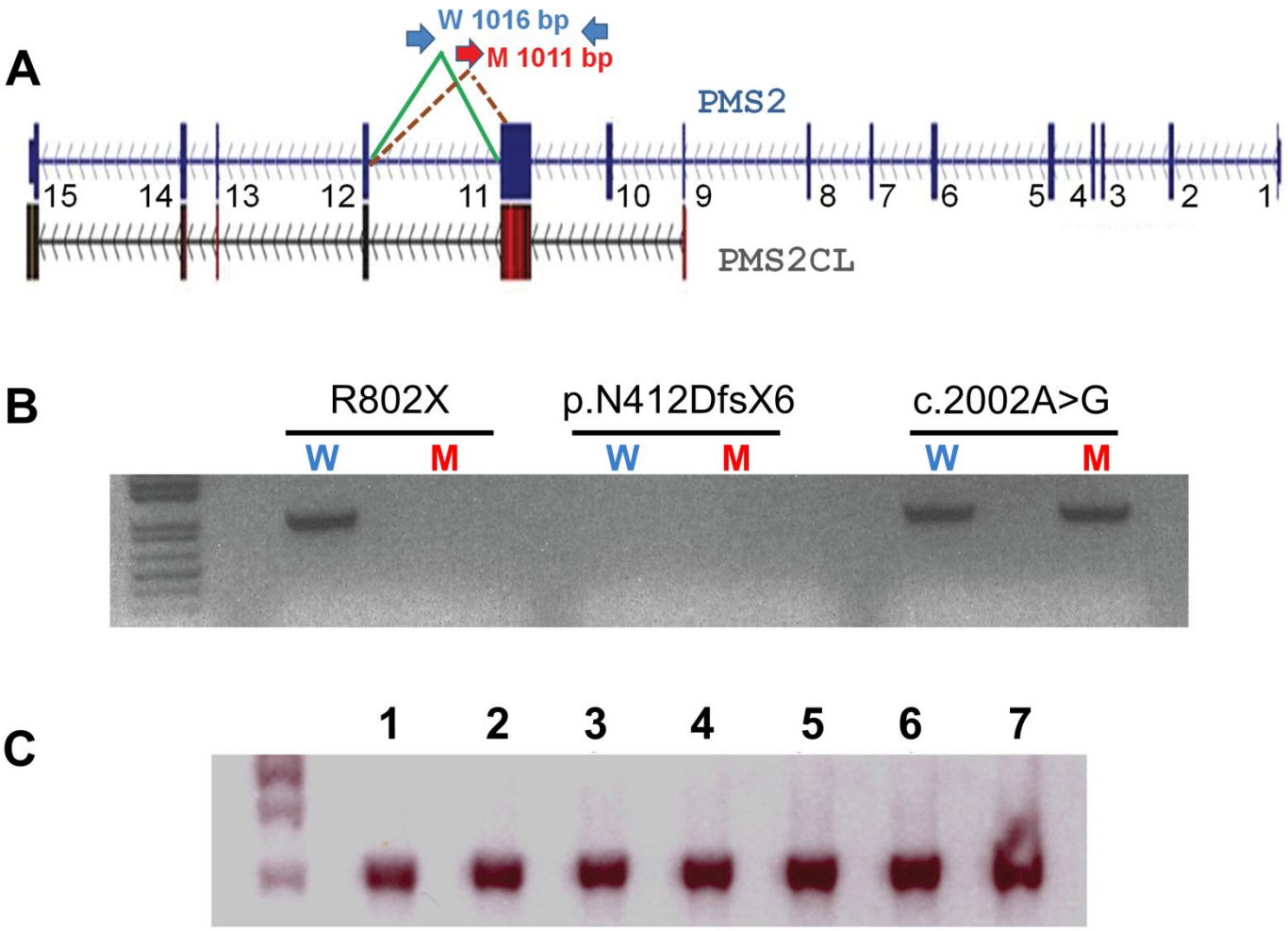

A
W 1016 bp
M 1011 bp
PMS2
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
PMS2CL
B
R802X
p.N412DfsX6
c.2002A>G
W
M
W
M
W
M
C
1
2
3
4
5
6
7

Figure S5. **Detection of low level PMS2 by Western Blot Analysis**

Full-length PMS2 protein is detected in the total lysate from lymphocytes and fibroblasts. HET: cell line established from a c.2002A>G heterozygote. HOM: cell line derived from a c.2002A>G homozygotes. The cell lines are derived from two female patients with cancer diagnosed at the age of 21 and 26 years old, respectively. CT: cell line from a healthy individual as a control of PMS2 wild type. TFIIH: Transcription factor II Human. Actin and TFIIH serve as loading controls.

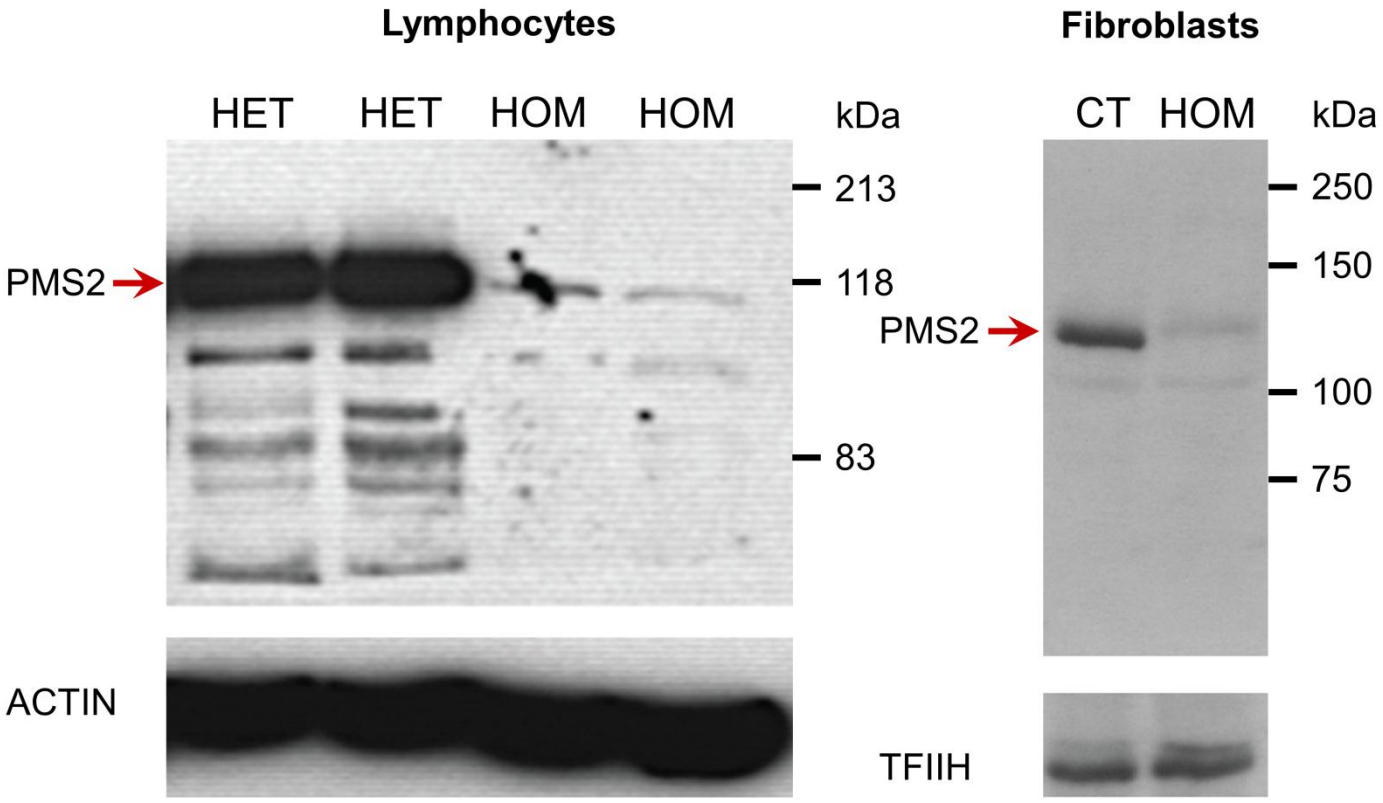


Figure S6. The MSI molecular phenotype measured with marker *D17S1307*

The assay is performed with genomic DNA purified from lymphocytes and normal colon mucosa of CMMRD patients. Individual I is homozygous for *PMS2:* c.2002A>G and Individual II is a compound heterozygote, *PMS2*: c.1221delG + c.2361delCTTC. The major alleles are determined by conventional genotyping with 0.1 ng DNA. Rare expansion allele(s) is detected by single-genome assay with 10 pg of DNA (equivalent to 3 alleles, 1.5 diploid genomes) per reaction.

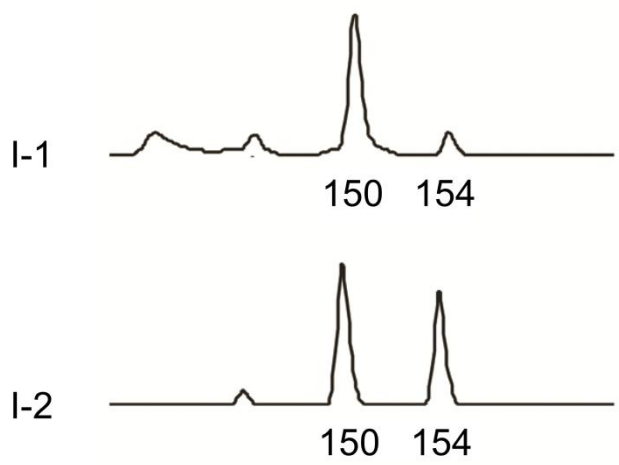
A. Representative results for individual I. I-1: a typical allele pattern observed in genotyping with a low input of 10 pg gDNA. I-2: the major alleles of *D17S1307* in individual I detected by conventional genotyping.

B. Representative results for individual II. II-1: The expansion allele (158 bp) is detected by genotyping with a low input of 10 pg gDNA. II-2: the major alleles of *D17S1307* in individual II detected by conventional genotyping.

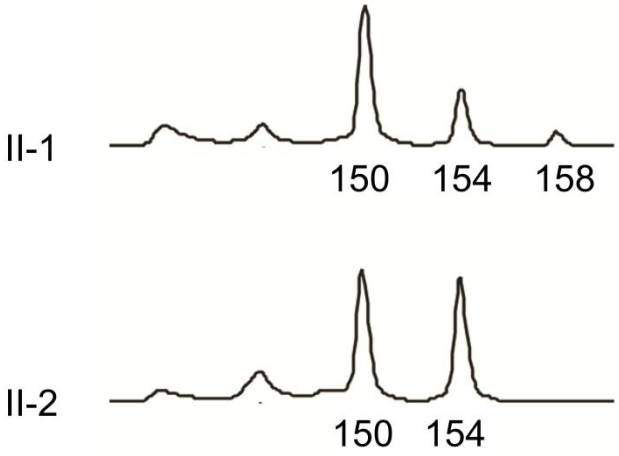
C. Summary of the detected expansion alleles at locus *D17S1307*. Fisher's exact test showed a significant difference of *D17S1307* allele expansion between the two individuals, P = 0.0037. Detailed information on the expansion alleles detected in lymphocytes and colon biopsy are presented in Table S5.

Figure S6

A



B



C

Subject	Total Alleles	Expanded Alleles
I	392	0
II	400	9

Table S1: Phenotypical characteristics of the proband's family

Individual	Alive	Polyps	CALS	Cancer Phenotypes	Age Dx (yrs)	Immunohistochemistry			
						MLH1	MSH2	MSH6	PMS2
III-2	N	Y	Y	Colon Ca	16	Present	Present	Present	Absent
				Astrocytoma	21	Present	Present	Present	Absent
III-3	Y	Y	Y	Colon Ca	26	Present	Present	Present	Absent
				Duodenal ca	40	ND	ND	ND	ND
III-5	N	Y	Y	Duodenal Ca	24	Present	Present	Present	Absent
III-7	N	Y	Y	Gastric Ca	22	ND	ND	ND	ND

CALS = Café-au-lait spots, Dx = diagnosis, Ca = cancer , ND = not determined

Table S2. Primers used in mutation identification, haplotype analysis, and gene expression characterization.

Protein truncation test (PTT)[1]	
Forward primer codons 1–863	5'-GGATCCTAATACGACTCACTATAGGGAGACCACCATGGAGCGAGCTGAGAGC-3'
Forward primer codons 415–863	5'-GGATCCTAATACGACTCACTATAGGGAGACCACCATGGTGTCCATTTCCAGACTGCG-3'
Forward primer[2] codons 332–863	5'-AATACGACTCACTATAGGGAGAGCCACCATGGTTA CTCCAGATAAAAGGCA-3'
Reverse primer (shared)	5'-AGGTTAGTGAAGACTCTGTC-3'
Forward primer codons 1–472	5'-GGATCCTAATACGACTCACTATAGGGAGACCACCATGGAGCGAGCTGAGAGC-3'
Reverse primer codons 1–472	5'-CTGAGGTCTCAGCAGGC-3'
***PMS2* cDNA sequencing**	
Forward primer located in exon 10	5'-TGTTACTCCAGATAAAAGGC-3'
Reverse primer	5'-AGGTTAGTGAAGACTCTGTC-3'
Sequencing primer	5'-TGCAGCATCTCGAAGT-3'
***PMS2* gDNA sequencing**	
Forward primer for long-range PCR	5'-GCGTTGATATCAATGTTACTCCAGA-3'
Reverse primer for long-range PCR	5'-AGTAGTCAGGGTAAAACATTCCAGT-3'
Forward primer for inner PCR	5'-TCACATAAGCACGTCCTCTCACCAT-3'
Reverse primer for inner PCR	5'-GCAACAGAGCAAGACTCTGTCTCAA-3'
Founder Mutation Linkage and Age Analysis	
PMS2-MS1 forward primer	5'-TAGGCAGAGCATCACCAAGG-3'
PMS2-MS1 reverse primer	5'-TGCTGAAGCATGAAAACTGC-3'

D7S1492 forward primer	5’-GCCTCCGAACACACTTCTTC-3’
D7S1492 reverse primer	5’-CCTGAAAATATACGTAACTACACCAA-3’
PMS2-MS2 forward primer	5’-TTCCATATGCAATCCCCATC-3’
PMS2-MS2 reverse primer	5’-GTGCTCCGCCATTTCTGTAT-3’
PMS2-MS3 forward primer	5’-GGCAATGGACAGAGGACAGT-3’
PMS2-MS3 reverse primer	5’-AGGCCAGGAGTTCAACCTG-3’
PMS2-MS18 forward primer	5’-TTTCTTAACCTTTCCTCAGCTT-3’
PMS2-MS18 reverse primer	5’-ATACTGTTGACCAATAACATGG-3’
PMS2-MS15 forward primer	5’-AAGACATTATGTAGACATTGTATGTG-3’
PMS2-MS15 reverse primer	5’-CAAGCGTGAGCTATCACGAG-3’
PMS2-MS6 forward primer	5’-AGGTTGCAGGGAGGCAGAG-3’
PMS2-MS6 reverse primer	5’-CACTTCCATAAATAGGATTGGTC-3’
PMS2-MS10 forward primer	5’-AAGAAAAAGATATGAGGCATAAA-3’
PMS2-MS10 reverse primer	5’-TGCTCATCTTCACGTTTGT-3’
PMS2-MS11 forward primer	5’-GGAACTGGGCTCCTGTTTTT-3’
PMS2-MS11 reverse primer	5’-GGATTCTCCTTCCTCCCAAG-3’
PMS2-MS12 forward primer	5’-GAGGATAAGTGGGGAATGAAA-3’
PMS2-MS12 reverse primer	5’-AACACCAGCTACGACCCATC-3’
PMS2-MS14 forward primer	5’-ATGAAAACCTGGGTGGACTG-3’
PMS2-MS14 reverse primer	5’-GCTGAGATCATGCCGTTGTA-3’
PMS2-MS19 forward primer	5’-CCGTCTCAAACGAGAAAACA-3’

PMS2-MS19 reverse primer	5'-ATTGAGCTTGAGCTGGGATG-3'
D7S1527 forward primer	5'-CCCTTGGAAAGTCTATACAT-3'
D7S1527 reverse primer	5'-CTGAAATTGTACTGTGCTTCT-3'
D7S2514 forward primer	5'-CATCAGTTGTTAAACTTTGCCAT-3'
D7S2514 reverse primer	5'-CAACCAGCCGTCATCTT-3'
D7S2547 forward primer	5'-CCGGATTTTTTTGGGACTCT-3'
D7S2547 reverse primer	5'-CCTCACATAAGTTGCCATCG-3'
D7S620 forward primer	5'-CAGGGTTCTCCAGAGAGAC-3'
D7S620 reverse primer	5'-TTATGTGAGCCAATTCTCCTC-3'
D7S664 forward primer	5'-AATTCTATCTTTCCAGGATTATCTG-3'
D7S664 reverse primer	5'-GATCAGTGCTGGTATAATAGTAGGT-3'
Polony analysis	
Forward primer[3] for solid-phase PCR	5'-/acrydite/TGCATGCAGCGGATTTGGAA-AAG-3'
Reverse primer for solid-phase PCR	5'-GTCCGTGGCATGCTGGTCCACTA-3'
Probe[4] to differentiate intact & 5-bp deletion transcripts	5'-TCCATTTCTGCAAACATCGTTTTACT-3'
Probe[5] to differentiate real & pseudo transcripts	5'-GAGCCCCTGTCCCCTGGGG-3'
Molecule-specific PCR	
Forward primer located in exon 10	5'-GTTACTCCAGATAAAAGGCA-3'
Reverse primer for the intact transcript	5'-CTGCAAACATCGTTTTACTTA-3'
Reverse primer for the transcript of the 5 bp deletion	5'-CTGCAAACATCGTTTTACTCT-3'

[1] The forward primers contain a T7 RNA polymerase promoter. [2] For the fragment spanning codons 332–863, the forward primer is located at the beginning of exon 10, which is specific for the locus of the functional *PMS2* gene. [3] The 5'-/acrydite modification mediates the attachment of primer to the gel matrix thus immobilizing the amplicons in the gel. [4] The probe is set at the position next to the base pairs involved in alternative splicing. [5] The probe is located next to a variant between the real *PMS2* gene and the transcribed pseudo *PMS2*.

Table S3. *PMS2* mutations and cancer diagnosis included in the analysis of genotype-phenotype correlation

	Central Nervous Tumour			Hematological Malignancy			Lynch Syndrome Related Cancer								
Mutation	**GLI**	**PNE**	**AST**	**ALL**	**NHL**	**OTH**	**STO**	**OV**	**END**	**INS**	**COL**	X	OT	Fig	Ref
c.2002A>G + c.2002A>G			21								16		35	1&S3	This study
												24[a]	25		
							22						23		
										41	26		A		
											16	16[b]	A		
											38		39		
		3									20		A		
												10[c]			
							39				21		A		
											31		A		
							11				11		11		
												21[c]	A		
											31		A		
c.949C>T + c.949C>T		7							16				N	S3	8
			6										N		
c.1239dupA + c.1927C>T		14						15					N	1&S3	9
c.182delA + c.182delA	10												11	S3	10
c.219T>A + c.219T>A											8	3[d]	N		11
c.1306dupA + c.1306dupA				6							16		N	1&S3	
	9												9		

c.400C>T + c.2184delTC	4				17						13		17		12
	13										12		N		13
c.706-?_803+?del + c.706-?_803+?del	8												9		
											15		N		14
				2									N		
		10											N		
c.706-?_2006 +?del + c.706?_2445+?del		12									12		15	S3	15
c.989-1G>T + c.989-1G>T	10								31	26	16		N		
											19		26		16,17
			11			17					17		N		
c.1145?_2445+?del + c.1145-?_2445 +?del						9					10		N		18
c.1169ins20											47		47	1	
c.1169ins20 + c.1169ins20				4									4		
			7								20		N	1&S3	
		24						21	23		16		N		
c.1927C>T												58[e]	N		19
c.1738A>T										49			N		
c.1939A>T											42		N		
c.1840A>T									56				N		
											74		N	1	
c.1831dupA											39		N		
											60		N		
											77		N		
											46		N		20
											66		N		
											42		N		
c.1981G>T											36		N	1	21

c.1221delG + c.2361delCTTC	14										18		18	S3	22
												13[f]	14		
c.1687C>T + c.2446?_2749+?del	13												15		23
c.1768delA + c.1768delA	9										8		N	1&S3	24
		1											5		
	6												7		
	6									15	15		17		
c.1840A>T + c.1840A>T	11												N		25
			19								14		21		
c.1882C>T											42		N	1	26
											55				
												42[g]			
c.2397_2400del + c.2397_2400del					9							11[h]	N	S3	27
c.1164delT + c.1164delT	11												14	1&S3	28
	9				3								N		
c.543delT + c.543delT		4											6	S3	29
		8											N		
c.2404C>T + c.2404C>T					10								12		
		8											N		
		14											15		
					4	2					18		N		
				15									16		
	15												15		
	7		6										8		
	2												2		
					3								3		
				6									N		
				6									7		

Each row records all cancers diagnosed in the same patient. The value in each cell is the age (in years) at the time of diagnosis. For the outcome column, numbers indicate the age at death. Fig: figure(s). Ref: reference(s). GLI: glioma; PNE: primitive neuro ectodermal tumors; AST: astrocytoma; ALL: acute lymphoblastic leukemia; NHL: non Hodgkin's lymphoma; OTH: other types of hematological cancer; STO: stomach; OV: ovary; END: endometrium; INS: intestine; COL: colon; X: other types of conditions. OT: outcome; N: data not available; A: alive; a: duodenal cancer b: pancreatic cancer; c: polyps; d: rhabdomyosarcoma; e: colorectal adenoma; f: neuroblastoma; g: adenoma, unspecified; h: mucoepidermoid carcinoma.

Table S4. Juxtaposed 5' splice sites at the exon 11 -intron 11 boundary

mutant allele	AGAG//GTAAGgtaaag		
wild-type allele	AGAGGTAAG//gtaaag		
predicting algorithm & related reference	de novo splice sequence	authentic splice sequence	consensus splice sequence
	GAG|GTAAGG	AAG|GTAAAG	CAG|GTAAGT
S&S[30]	91.82	82.74	100
NNSPLICE[31]	1.0	0.94	1
HSF[32]	96.31	84.95	100
MAXENT[33]	10.28	9.06	10.86
MDD[33]	13.38	13.78	15.08
MM[33]	10.08	7.72	12.75
WMM[33]	10.94	8.32	13.07
ΔG[34]	-8.7	-5.4	-9.6
H-bonds number[34]	7	6	9
H-bond score[35]	17.1	14.1	21.3

The mutation generates a *de novo* 5' (donor) splice site (ss) which partially overlaps with the authentic site at the boundary of exon 11 and intron 11. The splicing score for both 5'ss are predicted using various algorithms: a higher score indicates a stronger likelihood of being used by the splicing machinery. S&S:

Shapiro and Senpathy matrix score (http://esrsearch.tau.ac.il/). NNsplice: neural network splice site predictor (http://www.fruitfly.org/seq_tools/splice.html). HSF: human splicing finder (http://www.umd.be/HSF/). MAXENT: maximum entropy distribution model. MDD: maximum dependence decomposition model. MM: Markov model. WMM: weight matrix model. Scores of MAXENT, MDD, MM & WMM are calculated using MaxEntScan (http://genes.mit.edu/burgelab/maxent/Xmaxent.html). ΔG: free energy; H-bond score: strength of hydrogen bonding of 5'ss to U1 snRNA. ΔG and H-bond score are calculated with Analyzer Splice Tool (http://ibis.tau.ac.il/ssat/SpliceSiteFrame.htm). Parameters of measuring the strength of 5′ ss and U1 snRNA base-pairing are predicted using the algorithms of OligoArrayAuxo and H-Bond (http://www.uniduesseldorf.de/rna/html/hbond_score.php).

Table S5. *D17S1307* expansion alleles detected in gDNA from normal cells.

Individual	Tissue	# alleles tested	# expansion alleles detected 158 bp	# expansion alleles detected 162 bp
I	blood	189	0	0
I	colon	201	0	0
II	blood	189	5	0
II	colon	210	3	1

Individual I is individual III-3 from the index family (Table S1), who was diagnosed with colon cancer at age 26, is homozygous for c.2002A>G and tested positive for the expression of the intact exon 11-12 junction. Individual II was diagnosed with oligodendroglioma at age 14 and is a compound heterozygote for the truncating mutations c.1221delG and c.2361delCTTC.[15]

Web Resources

Leiden Open Variation Database: http://www.lovd.nl/3.0/home

Colon cancer gene variant databases: http://insight-group.org/

Mendelian Inheritance in Man: http://omim.org/

Mutation Taster: http://mutationtaster.org/

PolyPhen-2: http://genetics.bwh.harvard.edu/pph2/index.shtml

PROVEAN: http://provean.jcvi.org/index.php

Nunavik Statistics:

http://www.nunivaat.org/Statistics.aspx/Indicator/Vital_Statistics/

Programs predicting 5' splicing score:

http://esrsearch.tau.ac.il/

http://www.fruitfly.org/seq_tools/splice.html

http://www.umd.be/HSF/

http://genes.mit.edu/burgelab/maxent/Xmaxent.html

http://ibis.tau.ac.il/ssat/SpliceSiteFrame.htm

http://www.uni-duesseldorf.de/rna/html/hbond_score.php

References

1. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, and Sunyaev SR A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7:248-249
2. Ng PC and Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003:31:3812-3814
3. Schwarz JM, Rödelsperger C, Schuelke M, and Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* 2010:7:575-576
4. Hamel N, Feng BJ, Foretova L, Stoppa-Lyonnet D, Narod SA, Imyanitov E, Sinilnikova O, Tihomirova L, Lubinski J, Gronwald J, Gorski B, Hansen TO, Nielsen FC, Thomassen M, Yannoukakos D, Konstantopoulou I, Zajac V, Ciernikova S, Couch FJ, Greenwood CM, Goldgar DE, Foulkes WD. On the origin and diffusion of BRCA1 c.5266dupC (5382insC) in European populations*. Eur J Hum Genet* 2011;19:300-306
5. Korbie DJ and Mattick JS. Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nat Protoc* 2008;3:1452-1456
6. Thibodeau SN, Bren G, and Schaid D. Microsatellite instability in cancer of the proximal colon. *Science* 1993:260:816-819
7. Shah SN and Eckert KA. Human postmeiotic segregation 2 exhibits biased repair at tetranucleotide microsatellite sequences. *Cancer Res* 2009;69:1143-1149

8. Senter L, Clendenning M, Sotamaa K, Hampel H, Green J, Potter JD, Lindblom A, Lagerstedt K, Thibodeau SN, Lindor NM, Young J, Winship I, Dowty JG, White DM, Hopper JL, Baglietto L, Jenkins MA, and de la Chaple A. The clinical phenotype of Lynch syndrome due to germ-line PMS2 mutations. *Gastroenterology* 2008;135:419-428
9. Vaughn CP, Robles J, Swensen JJ, Miller CE, Lyon E, Mao R, Bayrak-Toydemir P, and Samowitz WS. Clinical analysis of PMS2: mutation detection and avoidance of pseudogenes. *Hum Mut* 2010;31:588-593
10. Etzler J, Peyrl A, Zatkova A, Schildhaus HU, Ficek A, Merkelbach-Bruse S, Kratz CP, Attarbaschi A, Hainfellner JA, Yao S, Messiaen L, Slavc I, and Wimmer K. RNA-based mutation analysis identifies an unusual MSH6 splicing defect and circumvents PMS2 pseudogene interference. *Hum Mut* 2008;29:299-305
11. Kratz CP, Holter S, Etzler J, Lauten M, Pollett A, Niemeyer CM, Gallinger S, and Wimmer K. Rhabdomyosarcoma in patients with constitutional mismatch-repair-deficiency syndrome. *J Med Genet* 2009;46:418-420
12. De Vos M, Hayward BE, Picton S, Sheridan E, and Bonthron DT. Novel PMS2 pseudogenes can conceal recessive mutations causing a distinctive childhood cancer syndrome. *Am J Hum Genet* 2004;74:954-964
13. Bonthron DT, Hayward BE, De Vos M, and Sheridan E. PMS2 mutations in childhood cancer. *Gut* 2005;54:1821

14. Tan TY, Orme LM, Lynch E, Croxford MA, Dow C, Dewan PA, and Lipton L. Biallelic PMS2 mutations and a distinctive childhood cancer syndrome. *J Pediatr Hematol Oncol* 2008;30254-257
15. Lindsay H, Jubran RF, Wang L, Kipp BR, and May WA. Simultaneous Colonic Adenocarcinoma and Medulloblastoma in a 12-Year-Old with Biallelic Deletions in *PMS2*. *J Pediatr* 2013;163:601-3
16. Sjursen W, Bjornevoll I, Engebretsen LF, Fjelland K, Halvorsen T, and Myrvold HE. A homozygote splice site PMS2 mutation as cause of Turcot syndrome gives rise to two different abnormal transcripts. *Fam Cancer* 2009;8:179-186
17. Holter S, Pollett A, Zogopoulos G, Kim H, Schwenter F, Asai K, Gallinger S, Clendenning M, Steinbach G, Jacobson A, and Boycott KM. Hepatic adenomas caused by somatic HNF1A mutations in children with biallelic mismatch repair gene mutations. *Gastroenterology* 2011;140:735-736
18. Peron S, Metin A, Gardes P, Alyanakian MA, Sheridan E, Kratz CP, Fischer A, and Durandy A. Human PMS2 deficiency is associated with impaired immunoglobulin class switch recombination. *J Exp Med* 2008;205:2465-2472
19. Trimbath JD, Petersen GM, Erdman SH, Ferre M, Luce MC, and Giardiello FM. Cafe-au-lait spots and early onset colorectal neoplasia: a variant of HNPCC? *Fam Cancer* 2001;1:101-105
20. Truninger K, Menigatti M, Luz J, Russell A, Haider R, Gebbers JO, Bannwart F, Yurtsever H, Neuweiler J, Riehle HM, Cattaruzza MS,

Heinimann K, Schar P, Jiricny J, and Marra G. Immunohistochemical analysis reveals high frequency of PMS2 defects in colorectal cancer. *Gastroenterology* 2005;128:1160-1171

21. Gulati S, Gustafson S, and Daw H. Lynch Syndrome Associated With PMS2 Mutation: Understanding Current Concepts. Gastrointest. *Cancer Res* 2011;4:188-190
22. De Rosa M, Fasano C, Panariello L, Scarano MI, Belli G, Iannelli A, Ciciliano F, and Izzo P. Evidence for a recessive inheritance of Turcot's syndrome caused by compound heterozygous mutations within the PMS2 gene. *Oncogene* 2000;19:1719-1723
23. Vaughn CP, Hart KJ, Samowitz WS, and Swensen JJ. Avoidance of pseudogene interference in the detection of 3' deletions in PMS2. *Hum Mutat* 2011;32:1063-1071
24. Gottschling S, Reinhard H, Pagenstecher C, Kruger S, Raedle J, Plotz G, Henn W, Buettner R, Meyer S, and Graf N. Hypothesis: possible role of retinoic acid therapy in patients with biallelic mismatch repair gene defects. *Eur J Pediatr* 2008;167:225-229
25. Gururangan S, Frankel W, Broaddus R, Clendenning M, Senter L, McDonald M, Eastwood J, Reardon D, Vredenburgh J, and Quinn J. Multifocal anaplastic astrocytoma in a patient with hereditary colorectal cancer, transcobalamin II deficiency, agenesis of the corpus callosum, mental retardation, and inherited PMS2 mutation. *Neuro Oncol* 2008;10:93-97

26. Hendriks Y, Jagmohan-Changur S, van der Klift HM, Morreau H, van Puijenbroek M, Tops C, van Os T, Wagner A, Ausems MG, and Gomez E. Heterozygous Mutations in PMS2 Cause Hereditary Nonpolyposis Colorectal Carcinoma (Lynch Syndrome). *Gastroenterology* 2006;130:312-322
27. Baas AF, Gabbett M, Rimac M, Kansikas M, Raphael M, Nievelstein RA, Nicholls W, Offerhaus J, Bodmer D, Wernstedt A, Krabichler B, Strasser U, Nystrom M, Zschocke J, Robertson SP, van Haelst MM, and Wimmer K. Agenesis of the corpus callosum and gray matter heterotopia in three patients with constitutional mismatch repair deficiency syndrome. *Eur J Hum Genet* 2013;21:55-61
28. Chmara M, Wernstedt A, Wasag B, Peeters H, Renard M, Beert E, Brems H, Giner T, Bieber I, Hamm H, Sciot R, Wimmer K, and Legius E. Multiple pilomatricomas with somatic CTNNB1 mutations in children with constitutive mismatch repair deficiency. *Genes Chromosomes Cancer* 2013;52:656-664
29. De Vos M, Hayward BE, Charlton R, Taylor GR, Glaser AW, Picton S, Cole TR, Maher ER, Mckeown CME, Mann JR, Yates JR, Baralle D, Rankin J, Bonthron DT, and Sheridan E. PMS2 mutations in childhood cancer. *J Nat Cancer Inst* 2006;98:358-361
30. Shapiro MB and Senapathy P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* 1987;15:7155-7174

31. Pertea M, Lin X, and Salzberg SL. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res* 2001;29:1185-1190
32. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, and Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009;37(9):e67
33. Yeo G and Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 2004;11:377-394
34. Carmel I, Tal S, Vig I, and Ast G. Comparative analysis detects dependencies among the 5G splice-site positions. *RNA* 2004;10:828-840
35. Freund M, Asang C, Kammler S, Konermann C, Krummheuer J+, Hipp M, Meyer I, Gierling W, Theiss S, and Preuss T. A novel approach to describe a U1 snRNA binding site. *Nucleic Acids Res* 2003;31:6963-6975