nature genetics

Heritable somatic methylation and inactivation of *MSH2* in families with Lynch syndrome due to deletion of the 3′ exons of *TACSTD1*

Marjolijn J L Ligtenberg[1,2,6], Roland P Kuiper[1,6], Tsun Leung Chan[3,4,6], Monique Goossens[2], Konnie M Hebeda[2], Marsha Voorendt[1], Tracy Y H Lee[3], Danielle Bodmer[1], Eveline Hoenselaar[1], Sandra J B Hendriks-Cornelissen[2], Wai Yin Tsui[3], Chi Kwan Kong[5], Han G Brunner[1], Ad Geurts van Kessel[1], Siu Tsan Yuen[3,4], J Han J M van Krieken[2], Suet Yi Leung[3,4] & Nicoline Hoogerbrugge[1]

Lynch syndrome patients are susceptible to colorectal and endometrial cancers owing to inactivating germline mutations in mismatch repair genes, including *MSH2* (ref. 1). Here we describe patients from Dutch and Chinese families with MSH2-deficient tumors carrying heterozygous germline deletions of the last exons of *TACSTD1*, a gene directly upstream of *MSH2* encoding Ep-CAM. Due to these deletions, transcription of *TACSTD1* extends into *MSH2*. The *MSH2* promoter *in cis* with the deletion is methylated in Ep-CAM positive but not in Ep-CAM negative normal tissues, thus revealing a correlation between activity of the mutated *TACSTD1* allele and epigenetic inactivation of the corresponding *MSH2* allele. Gene silencing by transcriptional read-through of a neighboring gene in either sense, as demonstrated here, or antisense direction[2], could represent a general mutational mechanism. Depending on the expression pattern of the neighboring gene that lacks its normal polyadenylation signal, this may cause either generalized or mosaic patterns of epigenetic inactivation.

Lynch syndrome (also known as hereditary non-polyposis colorectal cancer, HNPCC) is caused by heterozygous germline inactivation of one of the mismatch repair genes *MLH1*, *MSH2*, *MSH6* and *PMS2*. Lynch syndrome patients are susceptible to colorectal, endometrial and other cancers, which are recognized by microsatellite instability (MSI-high), a hallmark of mismatch repair defects[1]. Microsatellite instability also occurs in a subset of sporadic cancers through hypermethylation of the *MLH1* promoter[3,4]. Predisposition for this phenomenon may be associated with *MLH1* promoter variants[5]. Occasionally, epigenetic inactivation of the *MLH1* promoter has been observed in normal tissue[6–12]. Previously, we have reported one Chinese family with stable inheritance of allele-specific methylation of the *MSH2* promoter, raising the possibility of a cis-acting DNA element leading to mosaic methylation of the affected *MSH2* allele[13]. Here we report additional families of Dutch and Chinese origin, with MSH2-deficient MSI-high tumors exhibiting mosaic *MSH2* promoter methylation, in all of which we identified a causative deletion in a gene located immediately upstream of *MSH2*.

The index subject of Dutch family A (**Fig. 1**) was at risk for a germline mutation in *MSH2*, because of MSI-high colorectal cancer with loss of MSH2 and MSH6 protein. However, no germline mutation was detected in the coding regions of *MSH2*, *MLH1* or *MSH6*. Using routine diagnostic MLPA analysis of *MSH6*, a reduced signal was observed for exon 9 of *TACSTD1*, located 16 kb upstream of *MSH2*. Subsequent MLPA analyses using additional probes between *TACSTD1* exon 3 and *MSH2* exon 1, in conjunction with long range PCR, defined the deletion as encompassing the two most 3′ exons of *TACSTD1*, while leaving the promoter region of *MSH2* (refs. 14,15) intact (**Fig. 1a**,**b**). By sequence analysis the deletion breakpoints were exactly defined with an intervening deletion of 4,909 bp, which is considerably smaller than previously reported *MSH2* upstream deletions[16]. The deletion was denoted as *TACSTD1* c.859-1462_*1999del.

Subsequently, exactly the same deletion was found in three additional subjects from unrelated families B, C and D with MSH2-negative MSI-high tumors without detectable *MSH2* germline mutations. Analysis of one subject from each family on an Affymetrix SNP6.0 genotyping array revealed a stretch of 2,248 SNPs (encompassing 6.2 Mb), of which one allele was shared between all four affected subjects (**Supplementary Fig. 1** online), thus strongly suggesting that this mutation originates from a common founder. The alteration was detected in four out of ten unexplained putative Dutch Lynch syndrome families with MSH2-negative MSI-high cancer

[1]Department of Human Genetics 849 and [2]Department of Pathology, Radboud University Nijmegen Medical Centre, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands. [3]Hereditary Gastrointestinal Cancer Genetic Diagnosis Laboratory, Department of Pathology, The University of Hong Kong, Queen Mary Hospital, Pokfulam, Hong Kong and [4]Hereditary Gastrointestinal Cancer Registry, Department of Pathology, St. Paul's Hospital, No 2 Eastern Hospital Road, Causeway Bay, Hong Kong. [5]Department of Surgery, Yan Chai Hospital, Nos. 7–11, Yan Chai Street, Tsuen Wan, New Territories, Hong Kong. [6]These authors contributed equally to this work. Correspondence should be addressed to M.J.L.L (m.ligtenberg@antrg.umcn.nl) or S.Y.L. (suetyi@hkucc.hku.hk).

Received 22 May; accepted 7 October; published online 21 December 2008; doi:10.1038/ng.283

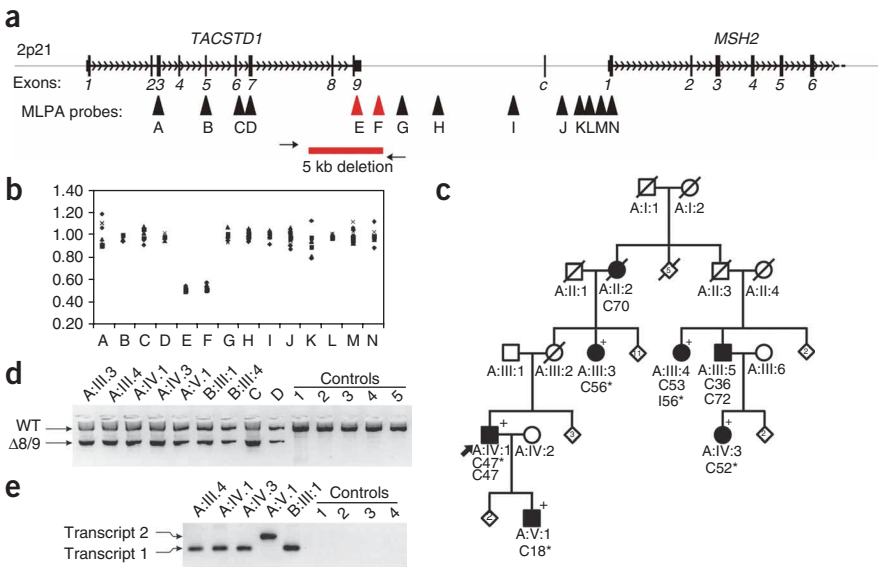

Figure 1 A constitutional microdeletion of the *TACSTD1* gene in Dutch families with *MSH2*-deficient tumors. (**a**) Structural organization of the *TACSTD1-MSH2* locus. Positions of the MLPA probes are indicated by arrowheads, of which probes E and F (red) are located within the 5 kb deletion region (red bar). Probes A and E were also present in the SALSA P008 kit used for routine *MSH6* diagnostics. Position of the long range PCR primers are indicated by horizontal arrows. A cryptic exon (**c**) located in the *MSH2* promoter region was occasionally included in *TACSTD1-MSH2* fusion transcripts (see below). (**b**) Combined representation of MLPA results of index subjects from families A to D, depicted by circles, squares, diamonds, and asterisks, respectively, analyzed in duplicate with two synthetic probe mixes (mix A probes: A, B, C, D, E, F, H, J, M; mix B probes: F, G, H, I, J, K, L, M, N). (**c**) Pedigree of family A with Lynch-associated tumors. Subjects in this family (filled symbols) were diagnosed with colorectal (C) or ileal (I) tumors at the indicated ages. Index subject A:IV:1 (arrow), family members carrying the deletion (+), and MSI-high tumors with immunohistochemical loss of MSH2 (*) are indicated. The number in the diamonds refers to the number of additional siblings. (**d**) Long range PCR, using primers flanking the deletion, detects both wild type (wt) and mutant (Δ8/9) fragments in nine affected family members of families A to D but not in five unaffected control individuals. (**e**) Nested RT-PCR on RNA from EBV-transformed cell lines of five deletion carriers from two families, in which exon 7 of *TACSTD1* is fused to exon 2 of *MSH2*. In the larger transcript 2, a 111 bp cryptic exon (see **a**) is included. In both transcripts, the translation is terminated in exon 2 of *MSH2*. For transcript, primer and probe sequences, see **Supplementary Table 2**. For sequencing chromatograms, see **Supplementary Figure 4a**,**b**.

(**Table 1**), but not in 145 other unrelated Dutch affected subjects that were tested for *MSH6* nor in 97 healthy Dutch controls. In two families with the deletion, DNA from multiple affected members was available, and co-segregation of the deletion with the presence of MSH2-negative MSI-high cancer was observed. In kindred A (**Fig. 1**) this resulted in a LOD score of 2.1. In kindred B the deletion was found in a third degree relative with such a tumor, resulting in a LOD-score of 0.6. Using MLPA analysis, loss of the wild type allele was

Table 1 Overview of Dutch subjects with a Lynch syndrome-associated MSI-high, MSH2-negative tumor without a detectable mutation in the coding region of *MSH2*

					Methylation detection					
Subject[a]	Location of carcinoma	Age at diagnosis	*TACSTD1* deletion	Loss of wt allele[b]	MSP	Pyro (%)	*Hha*I-A	*Hha*I-B	*Hha*I-C	Ep-CAM positive cells (%)[e]
A:III:3	Cecum	56	P	NA	NA	NA	NA	NA	NA	40
A:III:4	Ileum	56	P	N[c]	P	32	0.24	0.43	0.22	70
A:IV:1	Cecum	47	P	NA	NA	NA	NA	NA	NA	70
A:IV:3	Cecum	52	P	N[d]	P	37	0.25	0.58	0.26	70
A:V:1	Rectum	18	P	N	P	24	0.13	0.36	0.13	80
B:III:1	Rectum	45	P	N	P	NA	0.05	0.20	0.04	30
B:III:4	Transverse colon	50	P	N	P	NA	0.13	0.40	0.14	80
C	Ileocecum	53	P	P	P	34	0.38	0.76	0.47	5[f]
D	Cecum	40	P	N	NA	NA	NA	NA	NA	NA
E:1	Sigmoid colon	47	N		N					
E:2	Sigmoid colon	33	N		N					
F	Appendix	35	N		N					
G	Endometrium	45	N		N					
H	Rectum	58	N		N					
I	Right colon	39	N		N					
J:1	Rectum	44	N		N					
J:1	Bladder (urothelial cell carcinoma)	38	N		N					

P, positive; N, negative; NA, not assessable (no material left or DNA of too poor quality).
[a]Each subject is referred to by a character indicating a specific family and, if applicable, his or her position in the pedigree. [b]Loss of the wild-type allele as detected by MLPA analysis (see **Supplementary Fig. 2**). [c]Presence of somatic mutation c.511del (p.Arg171fs) in *MSH2*. [d]Presence of somatic mutation c.1183C>T (p.Gln395X) in *MSH2*. [e]Indicated is the percentage of Ep-CAM-positive cells in the microdissected material used for MS-MLPA; the percentage of Ep-CAM-positive cells in material microdissected for pyrosequencing can be different. [f]Most tumor cells are negative for Ep-CAM due to germline mutation of *TACSTD1* in combination with somatic deletion of the wild type allele (see also **Fig. 3d** and **Supplementary Fig. 2**).

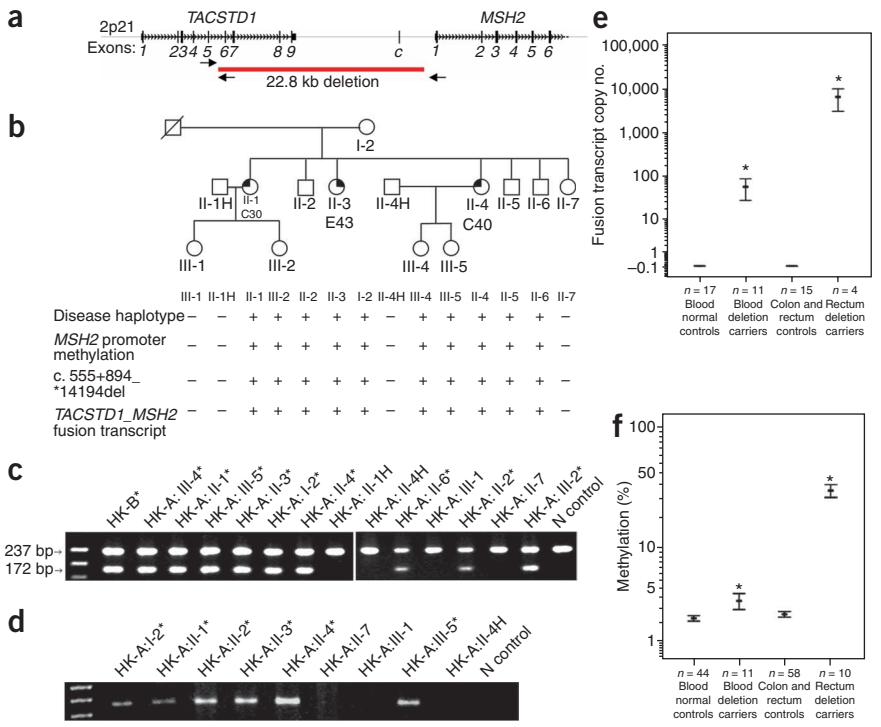

Figure 2 Deletion of the *TACSTD1* gene in Chinese families with heritable *MSH2* promoter methylation. (**a**) Location of the deletion and positions of diagnostic PCR primers. (**b**) Family pedigree of HK-family A reported previously[13]. Co-segregation of the *TACSTD1* deletion and expression of *TACSTD1-MSH2* fusion transcripts with mosaic methylation of *MSH2* promoter in this family. Subjects were diagnosed with colorectal (C) or endometrial (E) tumors at the indicated age. (**c**) Diagnostic PCR showing simultaneous amplification of both the control wild-type sequence (237 bp) and the deletion mutant (172 bp) in individuals showing *MSH2* methylation (*). Only the wild-type sequence was amplified in family members without methylation and in 108 normal individuals (representative result shown as N control). (**d**) Representative RT-PCR on blood leukocytes from HK-family A. A fusion transcript in which exon 5 of *TACSTD1* is fused to exon 2 of *MSH2* was amplified from family members with the *TACSTD1* deletion and *MSH2* methylation (denoted by *) but not in those without the deletion or in the control population (see **Supplementary Fig. 4c** for sequencing chromatogram). (**e**) Quantification of fusion transcripts in colorectal mucosae and blood leukocytes in deletion carriers and controls by real-time RT-PCR. The copy numbers are normalized to GAPDH. Error bars indicate the mean and 95% confidence interval of the respective group; $*P < 0.001$ by Mann-Whitney U-test compared across tissues and with controls. (**f**) Quantification of methylation in colorectal mucosae and blood leukocytes in deletion carriers and controls. Methylation levels were previously published and assayed by bisulfite pyrosequencing[13]. For details, see Methods.

found in a tumor of family C (**Supplementary Fig. 2** online). Moreover, somatic mutations in *MSH2* were found heterozygously in tumors of subjects A:III:4 and A:IV:3 (**Table 1**). Together, these findings suggest that the deletion in *TACSTD1* is the predisposing genetic defect leading to *MSH2* inactivation in these four families.

Concurrently, independent analysis of the Chinese family with heritable *MSH2* promoter methylation (hereby denoted as HK-family A)[13], using a combination of linkage analysis, long range PCR and sequence analysis, revealed a deletion of 22.8 kb (*TACSTD1* c.555+894_*14194del) (**Fig. 2** and **Supplementary Fig. 3** online), extending from intron 5 of *TACSTD1* to ~2.4 kb upstream of *MSH2*, which thus, again, encompasses the 3′ end of *TACSTD1*, leaving the *MSH2* promoter intact. The deletion co-segregates with the disease haplotype and methylation of the *MSH2* promoter (**Fig. 2**). Examination of other putative Lynch syndrome families revealed a second family (HK-family B) with the same deletion mutation, where *MSH2* methylation was found in two synchronous colon cancers and normal colon tissue in the index subject but not in his blood leukocytes. Haplotype analysis of these two families did not suggest a founder mutation (data not shown).

The *TACSTD1* deletions result in loss of the two or four most 3′ exons, including the polyadenylation signal, which will abolish transcription termination and could lead to transcription read-through into the downstream gene. Indeed, using RT-PCR, specific *TACSTD1-MSH2* fusion transcripts could be detected in EBV-transformed lymphocytes or blood leukocytes derived from carriers of the deletion in both ethnic groups, but not from their corresponding control populations (**Figs. 1e** and **2d**; **Supplementary Fig. 4** online). Based on these results, detailed analysis of EST databases indicating that *TACSTD1* transcription does not extend more than 35 bp beyond the polyadenylation signal in normal individuals, and the previously reported close link between polyadenylation and transcription termination[17,18], we conclude that this read-through is specific for the deletion carriers.

Similar to the Chinese families, all six MSI-high tumors from the Dutch families tested showed methylation of the *MSH2* promoter (**Table 1**) by methylation-specific PCR and subsequent bisulfite sequencing. In two subjects with a heterozygous promoter SNP (subject B:III:1 SNP rs2303425 at position c.-118; subject A:V:1 SNP rs1863332 at position c.-433), the methylation was confirmed to be allele-specific and to involve the allele segregating with the deletion (**Fig. 3b, Supplementary Table 1** online). Both alleles were present in the unmethylated fraction, suggesting that the promoter was not methylated in all cells. No methylation of the *MSH2* promoter was found in tumors of six unrelated affected subjects with a truncating germline mutation in *MSH2* or in eight as-yet-unexplained MSI-high MSH2-negative tumors of six other index subjects.

We hypothesized that the difference in the level of *MSH2* promoter methylation, observed earlier in different tissues of HK-family A[13], might correlate with the extended transcription of the aberrant *TACSTD1* gene across the CpG island of *MSH2*. Quantitative RT-PCR of *TACSTD1-MSH2* fusion transcripts revealed a >100-fold higher level of fusion transcripts in colorectal mucosa compared to blood from mutation-positive members of HK-family A (**Fig. 2e**), which correlates with the levels of *MSH2* methylation (**Fig. 2f**) and *TACSTD1* expression (**Supplementary Fig. 5** online). To further test whether methylation is limited to *TACSTD1* expressing cells, which can be identified through Ep-CAM immunohistochemistry, methylation-specific PCR on DNA isolated from microdissected Ep-CAM-positive and negative regions from one tissue block of

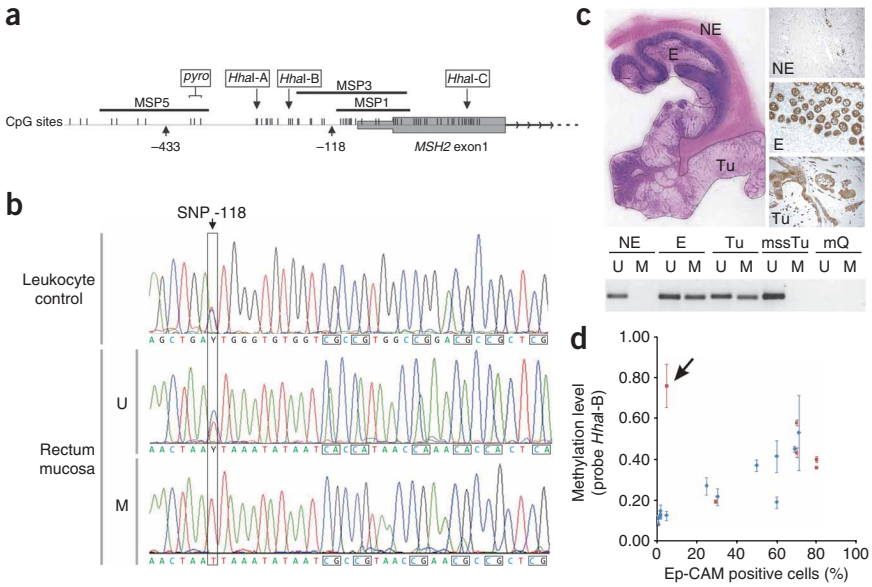

Figure 3 Allele-specific methylation of the *MSH2* promoter coincides with *TACSTD1* expression. (**a**) Schematic representation of the *MSH2* promoter, indicating the presence of two informative SNPs at positions c.-118 (rs2303425) and c.-433 (rs1863332), the positions of CpG sites analyzed by pyrosequencing of bisulfite-converted DNA (marked as pyro), the positions of three MLPA probes around methylation-sensitive *Hha*I sites (as used in **Fig. 3d** and **Tables 1** and **2)** and three methylation-specific PCR products (MSP1, 3 and 5). MSP1 and MSP3 were described earlier[3,13]. All CpG sites present in the region are indicated by bars on the baseline. (**b**) Direct sequencing of unmethylated (U) and methylated (M) MSP3 products from normal rectal mucosa in subject B:III:1 reveals that both alleles are present in the unmethylated DNA fraction, while only the T-allele of SNP c.-118, which cosegregates with the deletion (see **Supplementary Table 1**), is efficiently methylated at CpG dinucleotides (boxed). All CpG dinucleotides within the MSP3 region are affected by methylation. MSP3 primers amplify the reverse strand. Unmodified leukocyte DNA sequences were used for reference. (**c**) H&E stained ileal tissue section of subject A:III:4 (left panel) shows non-epithelial intestinal wall (NE), epithelium (E) and tumor (Tu). Immunostaining with Ep-CAM antibody is positive in epithelial tissue and tumor, whereas non-epithelial tissue is negative (right top panels). DNA isolated from microdissected sections of the three indicated tissues was subjected to methylation-specific PCR using MSP3 primers (bottom panel). The methylated *MSH2* promoter is only present in regions expressing Ep-CAM. A microsatellite-stable tumor (mssTu) and milliQ (mQ) were included as controls. (**d**) Correlation between percentage of Ep-CAM positive cells in tumors (red rectangles) or normal tissues (blue triangles), and level of methylation at position c.-198 (*Hha*I-B). The fraction of *Hha*I restriction-resistant DNA was determined in duplicate by MS-MLPA. The tumor of subject C with low Ep-CAM expression but high methylation level (arrow) was shown to have loss of the wild-type *TACSTD1* allele (see also **Table 1**).

subject A:III:4 was performed. Methylation only occurred in Ep-CAM-positive regions (**Fig. 3c**). Moreover, in both tumor and normal tissues of the Dutch mutation carriers the fraction of methylated *MSH2* promoter DNA as measured by pyrosequencing of bisulfite-converted DNA and methylation-specific MLPA correlated with the fraction of Ep-CAM positive cells (**Tables 1** and **2**, **Fig. 3d**).

In conclusion, we have identified a novel mechanism causing Lynch syndrome by showing that different deletions that disrupt the 3′ end of *TACSTD1* lead to inactivation of the adjacent *MSH2* gene through methylation induction of its promoter in tissues expressing *TACSTD1*. This mechanism explains the observed mosaic and allele-specific hypermethylation pattern of the *MSH2* promoter and transmission of the methylation over generations. Induction of methylation by transcription across a CpG island within a promoter region has been described before. In a subject with α-thalassemia, a truncating deletion in the widely expressed *LUC7L* gene was found to mediate silencing of

Table 2 Relation between the presence of hypermethylation of the *MSH2* promoter and the expression of *TACSTD1*, as measured by immunohistochemical staining of Ep-CAM

		Methylation detection					
Subject ID[a]	Normal tissue type	MSP	Pyro (%)	*Hha*I-A	*Hha*I-B	*Hha*I-C	Ep-CAM positive cells (%)[b]
A:III:4	Gastric mucosa	P	34	0.16	0.42	0.17	60
A:III:4	Ileal mucosa	P	43	0.09	0.22	0.09	30
A:III:4	Ileal lamina propria	N	NA	0.02	0.13	0.01	2
A:III:4	Lymph node	N	9	0.02	0.15	0.01	2
A:IV:3	Cecal mucosa	P	23	0.12	0.27	0.14	25
A:V:1	Rectal mucosa	P	46	0.06	0.19	0.07	60
A:V:1	Lymph node	N	2	0.02	0.11	0.01	<1
B:III:1	Rectal mucosa	P	NA	0.23	0.45	0.23	70
B:III:1	Myometrium	N	3	0.02	0.08	0.00	<1
B:III:4	Mucosa of anastomosis after colectomy	P	20	0.13	0.37	0.16	50
C	Cecal mucosa	P	44	0.22	0.53	0.22	70
C	Subcutaneous tissue	N	2	0.02	0.13	0.02	2
D	Cecal mucosa	P	NA	0.22	0.43	0.29	70
D	Ileal mucosa	P	NA	0.06	0.13	0.04	5

P, positive; N, negative; NA, not assessable (no material left or DNA of too poor quality).
[a]Each subject is referred to by a character indicating a specific family and, if applicable, his or her position in the pedigree. [b]Indicated is the percentage of Ep-CAM-positive cells in the microdissected material used for MS-MLPA; the percentage of Ep-CAM-positive cells in material microdissected for pyrosequencing can be different.

the flanking α-globin gene (*HBA2*) by antisense transcription through the *HBA2* promoter region, leading to methylation of its CpG island[2]. Similar to what we found for the *MSH2* promoter in our subjects, methylation only occurred at the promoter positioned *in cis* to the deletion[19]. Additionally, it was demonstrated that the *p15* tumor suppressor gene can be transcriptionally silenced by antisense *p15* RNA via the induction of epigenetic chromatin changes and, eventually, hypermethylation of the *p15* promoter predominantly *in cis*[20]. Taken together with our data on methylation of the *MSH2* promoter, this suggests that abolishment of transcription termination and concomitant transcriptional read-through can lead to methylation of the CpG island and silencing of a downstream gene, irrespective of the orientation of the latter.

Expression of *TACSTD1* is high in epithelial tissues[21], among which are the main target tissues in Lynch syndrome. This could explain why these families are recognized as Lynch syndrome families, although the mutant *MSH2* allele is not inactivated in all cells. Biallelic inactivating *TACSTD1* mutations lead to congenital tufting enteropathy. Heterozygous parents of these subjects, like our subjects with a *TACSTD1* 3′ end deletion, do not present such clinical features[22]. Further assessment of the predisposition to cancer in families with a *TACSTD1* 3′ end deletion is required to establish whether these mutations lead to the same tumor risk and spectrum as conventional *MSH2* mutations.

Based on our findings, transcriptional read-through due to deletion of polyadenylation signals may constitute a general mutational mechanism for the inactivation of neighboring genes. One interesting feature of this type of mutation is that it is located outside the target gene and, thus, undetectable by conventional mutation screening strategies. Moreover, the pattern of methylation of the target gene is dependent on the expression of the neighboring gene, which may lead to a somatic mosaic state of the epigenetic inactivation and restriction of the affected tissues. In addition, abrogation of polyadenylation signals due to chromosomal aberrations in cancer cells may result in aberrant promoter methylation and concomitant inactivation of tumor suppressor genes.

METHODS

Subjects and clinical samples. Index subjects in this study were referred to the outpatient clinic for hereditary tumors of the Radboud University Nijmegen Medical Centre or The Hereditary Gastrointestinal Cancer Genetic Diagnosis Laboratory based at Queen Mary Hospital, The University of Hong Kong, for genetic diagnosis (**Supplementary Note** online). Informed consent was obtained from all participating subjects. The study was performed according to the rules of the Medical Ethics Committee of the Radboud University Nijmegen Medical Centre and the Institutional Review Board of The University of Hong Kong/Hospital Authority Hong Kong West Cluster. DNA was isolated from peripheral blood leukocytes or microdissected slides from paraffin embedded tissue according to standard procedures. Normal control DNA was isolated from EBV-transformed cell lines or blood leukocytes of anonymous blood donors from the respective ethnic groups. For RT-PCR of fusion transcripts, RNA was isolated from EBV-transformed cell lines or blood leukocytes. For quantification of transcript abundance, RNA was isolated from rectal biopsies of 4 *TACSTD1* deletion carriers (HK-family A I-2, III-2, III-4, III-5), 2 family members without the deletion (II-7 and III-1), and 13 normal rectal or colonic mucosae obtained from colectomy specimen of sporadic microsatellite-stable colorectal cancers. RNA from blood leukocytes was extracted from 11 deletion carriers (10 in HK-family A and 1 in HK-family B) and 17 controls (family members without the deletion or healthy controls). For quantification of methylation level by pyrosequencing in Hong Kong families, data was previously published[13] with addition of a few new family members. The control groups differed in most cases from those used for fusion transcript quantification as RNA was not available in these cases.

The group of 145 unrelated Dutch affected subjects that was tested for *MSH6* mutations consisted of 18 subjects with a truncating *MSH2* mutation, 24 subjects with a mutation in *MLH1*, *MSH6* or *PMS2*, or with a tumor with somatic hypermethylation of *MLH1*, 18 subjects with a cancer showing microsatellite instability but with nuclear staining of MSH2, 60 subjects with a microsatellite stable tumor and 25 subjects with a Lynch syndrome-associated tumor not tested for microsatellite instability.

Microsatellite instability, mismatch repair gene mutation and immunohistochemical analyses. Microsatellite instability analysis using six markers (D2S123, D5S346, D17S250, BAT25, BAT26 and BAT40), immunohistochemistry of mismatch repair proteins and analysis of germline mutations were performed and interpreted as described before[23]. For the detection of somatic mutations, *MSH2* exons 3, 5, 7, 12 and 13 were sequenced in tumor DNA of subject A:III:4 and A:IV:3. Immunohistochemistry of Ep-CAM was performed with Ep-CAM Ab-1 (Clone VU-1D9; Thermo Fisher) using standard procedures.

MLPA, long range PCR and SNP array analyses. The diagnostic MLPA analysis of *MSH6* was performed with SALSA MLPA kit P008 MSH2/PMS2 (MRC-Holland). Synthetic MLPA primers were designed using the MeltIngeny program (Ingeny International). MLPA was performed using a SALSA MLPA reaction mixture (MRC-Holland) containing 0.5 nM of each primer, and analyzed using healthy controls as a reference (for primer sequences see **Supplementary Table 2** online). For the long range PCR across the deletion in Dutch families, a TAKARA PCR kit (Takara bio inc) was used with primers on either side of the deletion. The deletion was further specified by direct-sequencing using the forward primer in combination with an internal reverse primer. For the Chinese families, multiple long range PCRs were performed targeting the candidate chromosomal region, which yielded an aberrant amplicon suggestive of a large deletion mutation (**Supplementary Fig. 3**). The aberrant amplicon was digested with restriction enzymes, cloned, and the breakpoint identified by sequencing. Diagnostic PCR primers were designed flanking the breakpoint and used to screen the family members (see **Supplementary Table 2** for primer sequences).

Total genomic DNA from four index subjects was hybridized on Affymetrix SNP6.0 arrays, containing 906,600 SNPs throughout the genome, according to the instructions of the manufacturer (Affymetrix). Genotypes were generated using Birdseed analysis software implemented in the Affymetrix Genotyping Console version 2.1. Subsequently, genomic regions on chromosome 2 in which all four subjects shared one of the alleles were determined by defining the longest stretches of SNPs in which no opposite homozygous calls in the four subjects could be detected. For the two Chinese families, haplotype analysis was performed using microsatellite markers within and flanking *MSH2* as described[13].

RT-PCR and real-time quantitative RT-PCR. To detect fusion transcripts, a direct RT-PCR reaction (for Chinese families) or a nested RT-PCR reaction (for Dutch families) was performed. For primers see supplementary data (**Supplementary Table 2**). To analyze for transcription of *TACSTD1* beyond the polyadenylation signal in normal individuals, the entire transcript databases from GenBank were queried, including spliced and unspliced ESTs.

To quantitate the abundance of wild type *TACSTD1*, wild type *MSH2* and *TACSTD1*exon5-*MSH2*exon2 fusion transcripts in colorectal mucosae and blood leukocytes, RNA was treated with DNase I using DNA-free kit (Ambion) and reversed transcribed using the TaqMan Gold Reverse Transcription Kit (Applied Biosystems) with random hexamers. Real-time quantification was performed using the QuantiFast SYBR Green RT-PCR kit (Qiagen) on an ABI Prism 7900HT sequence detection system (Applied Biosystems). *TACSTD1* exon 5-7, *MSH2* exon 1-3 and the *TACSTD1*exon4,5-*MSH2*exon2-5 fusion transcript were cloned into pCR2.1-TOPO plasmids (Invitrogen). The plasmids were linearized and known copy numbers were used in serial dilutions to generate calibration curves for quantification. Transcript quantification was performed in triplicates and reported relative to *GAPDH*. Quantitative PCR primers for *TACSTD1* and *MSH2* were designed to detect the wild type transcripts only (see **Supplementary Table 2** for primer sequences).

Methylation analysis. To study DNA methylation by methylation-specific PCR in the Dutch families the DNA was treated with bisulfite using the EZ DNA methylation KIT™ (ZYMO Research Corp) as described earlier[23]. Methylation-specific PCR was performed by using the MSP1 and MSP3 primer sets for both methylated and unmethylated DNA as described earlier[3,13]. For the analysis of the allele specific methylation surrounding rs1863332 new primer sets were designed (MSP5; **Supplementary Table 2**). The specificity of the PCR for either unmethylated or methylated CG dinucleotides was confirmed by direct sequencing of the PCR products of all tumor samples and at least one normal sample per subject.

Pyrosequencing of bisulfite-converted DNA (EZ DNA methylation direct KIT, (ZYMO Research) was performed as described before[13]. Mean methylation levels of three CpG sites in the *MSH2* promoter (c.-366, c.-382 and c.-393) were calculated and reported. Background level of incomplete conversion was below 7.5% for all carcinomas assessed and below 5% for normal tissues.

Methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA) analyses were performed using SALSA MS-MLPA kit ME011 Mismatch Repair genes (MMR) (MRC-Holland) as described[24]. For each MS-MLPA reaction 200 ng DNA isolated from formalin fixed paraffin embedded material was used. Data were normalized using all 11 probes without an *Hha*I site as control probes. For each probe the fraction of *Hha*I restriction-resistant DNA, which is a measure for the level of methylation, was determined by dividing the normalized signal of the digested sample by the normalized signal of the undigested sample. Three *MSH2* specific probes (*Hha*I-A position c.-262; *Hha*I-B position c.-198; *Hha*I-C positions c.135 and c.150) were included. Using six formalin fixed paraffin embedded colorectal samples of subjects without a 3′ *TACSTD1* deletion background fractions of *Hha*I restriction-resistant DNA (*Hha*I-A: 0.02; *Hha*I-B: 0.08; *Hha*I-C: 0.00) were determined. Samples with known *MGMT* or *MLH1* methylation were used as positive controls.

Accession codes. GenBank: human *TACSTD1* mRNA, NM_002354.1; human *MSH2* mRNA, NM_000251.1; *TACSTD1* mutation c.859-1462_*1999del, FJ347525; *TACSTD1* mutation c.555+894_*14194del, FJ347526; human chromosome 2 genomic sequence, NT_022184.14.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank S. Wezenberg, M. Schliekelmann, E. Kamping, M. Steehouwer, R. Willems, A.S.Y. Chan, A.K.W. Chan, J.K.Y. Lau and C. Li for technical assistance, Diederik de Bruijn for advice and support, and clinicians in Hong Kong Hospital Authority for clinical care. This work was supported by research grants from the Dutch Cancer Society, the Research Grants Council of the Hong Kong Special Administrative Region (GRF HKU 7614/08M and HKU 7622/05M), the Hong Kong Cancer Fund and the Michael and Betty Kadoorie Cancer Genetics Research Programme.

AUTHOR CONTRIBUTIONS

M.J.L.L., R.P.K., T.L.C., S.T.Y. and S.Y.L. designed the study. For Dutch families, M.G. and S.J.B.H.-C. performed analyses on tumor and normal tissues; K.M.H. and J.H.J.M.v.K interpreted the histology and immunohistochemistry; D.B. and E.H. performed mutation, segregation and RT-PCR analyses; T.L.C. performed pyrosequencing; R.P.K. performed SNP-array analyses; M.V. and N.H. were responsible for patient counseling and clinical data acquisition; M.J.L.L., H.G.B., A.G.v.K., J.H.J.M.v.K. and N.H. supervised the work. For Hong Kong families, T.L.C., T.Y.H.L. and W.Y.T. performed experiments, C.K.K. provided clinical care and acquired clinical data; T.L.C., S.Y.L. and S.T.Y. analyzed and interpreted data; and M.J.L.L., R.P.K., T.L.C. and S.Y.L. wrote the manuscript, with assistance and final approval from all coauthors.

Published online at http://www.nature.com/naturegenetics/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Lynch, H.T. & de la Chapelle, A. Hereditary colorectal cancer. *N. Engl. J. Med.* **348**, 919–932 (2003).
2. Barbour, V.M. *et al.* Alpha-thalassemia resulting from a negative chromosomal position effect. *Blood* **96**, 800–807 (2000).
3. Herman, J.G. *et al.* Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc. Natl. Acad. Sci. USA* **95**, 6870–6875 (1998).
4. Kane, M.F. *et al.* Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. *Cancer Res.* **57**, 808–811 (1997).
5. Chen, H. *et al.* Evidence for heritable predisposition to epigenetic silencing of MLH1. *Int. J. Cancer* **120**, 1684–1688 (2007).
6. Gazzoli, I., Loda, M., Garber, J., Syngal, S. & Kolodner, R.D. A hereditary non-polyposis colorectal carcinoma case associated with hypermethylation of the MLH1 gene in normal tissue and loss of heterozygosity of the unmethylated allele in the resulting microsatellite instability-high tumor. *Cancer Res.* **62**, 3925–3928 (2002).
7. Suter, C.M., Martin, D.I. & Ward, R.L. Germline epimutation of MLH1 in individuals with multiple cancers. *Nat. Genet.* **36**, 497–501 (2004).
8. Hitchins, M. *et al.* MLH1 germline epimutations as a factor in hereditary nonpolyposis colorectal cancer. *Gastroenterology* **129**, 1392–1399 (2005).
9. Miyakura, Y. *et al.* Extensive but hemiallelic methylation of the hMLH1 promoter region in early-onset sporadic colon cancers with microsatellite instability. *Clin. Gastroenterol. Hepatol.* **2**, 147–156 (2004).
10. Valle, L. *et al.* MLH1 germline epimutations in selected patients with early-onset non-polyposis colorectal cancer. *Clin. Genet.* **71**, 232–237 (2007).
11. Hitchins, M.P. *et al.* Inheritance of a cancer-associated MLH1 germ-line epimutation. *N. Engl. J. Med.* **356**, 697–705 (2007).
12. Morak, M. *et al.* Further evidence for heritability of an epimutation in one of 12 cases with MLH1 promoter methylation in blood cells clinically displaying HNPCC. *Eur. J. Hum. Genet.* **16**, 804–811 (2008).
13. Chan, T.L. *et al.* Heritable germline epimutation of MSH2 in a family with hereditary nonpolyposis colorectal cancer. *Nat. Genet.* **38**, 1178–1183 (2006).
14. Scherer, S.J., Seib, T., Seitz, G., Dooley, S. & Welter, C. Isolation and characterization of the human mismatch repair gene hMSH2 promoter region. *Hum. Genet.* **97**, 114–116 (1996).
15. Iwahashi, Y. *et al.* Promoter analysis of the human mismatch repair gene hMSH2. *Gene* **213**, 141–147 (1998).
16. van der Klift, H. *et al.* Molecular characterization of the spectrum of genomic deletions in the mismatch repair genes MSH2, MLH1, MSH6, and PMS2 responsible for hereditary nonpolyposis colorectal cancer (HNPCC). *Genes Chromosom. Cancer* **44**, 123–138 (2005).
17. Proudfoot, N. New perspectives on connecting messenger RNA 3′ end formation to transcription. *Curr. Opin. Cell Biol.* **16**, 272–278 (2004).
18. Buratowski, S. Connections between mRNA 3′ end processing and transcription termination. *Curr. Opin. Cell Biol.* **17**, 257–261 (2005).
19. Tufarelli, C. *et al.* Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nat. Genet.* **34**, 157–165 (2003).
20. Yu, W. *et al.* Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature* **451**, 202–206 (2008).
21. Winter, M.J., Nagtegaal, I.D., van Krieken, J.H. & Litvinov, S.V. The epithelial cell adhesion molecule (Ep-CAM) as a morphoregulatory molecule is a tool in surgical pathology. *Am. J. Pathol.* **163**, 2139–2148 (2003).
22. Sivagnanam, M. *et al.* Identification of EpCAM as the gene for congenital tufting enteropathy. *Gastroenterology* **135**, 429–437 (2008).
23. Overbeek, L.I. *et al.* Patients with an unexplained microsatellite instable tumour have a low risk of familial cancer. *Br. J. Cancer* **96**, 1605–1612 (2007).
24. Jeuken, J.W. *et al.* MS-MLPA: an attractive alternative laboratory assay for robust, reliable, and semiquantitative detection of MGMT promoter hypermethylation in gliomas. *Lab. Invest.* **87**, 1055–1065 (2007).