Recurrence and variability of germline EPCAM deletions in Lynch syndrome

Journal:	<i>Human Mutation</i>
Manuscript ID:	humu-2010-0515.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	30-Nov-2010
Complete List of Authors:	<p>Kuiper, Roland; Radboud University Nijmegen Medical Center, Human Genetics</p> <p>Vissers, Lisenka; Radboud University Nijmegen Medical Center, Human Genetics</p> <p>Venkatachalam, Ramprasath; Radboud University Nijmegen Medical Center, Human Genetics</p> <p>Bodmer, Danielle; Radboud University Nijmegen Medical Center, Human Genetics</p> <p>Hoenselaar, Eveline; Radboud University Nijmegen Medical Center, Human Genetics</p> <p>Goossens, Monique; Radboud University Nijmegen Medical Center, Pathology</p> <p>Haufe, Aline; Technische Universitat Dresden, Surgical Research</p> <p>Kamping, Eveline; Radboud University Nijmegen Medical Center, Human Genetics</p> <p>Niessen, Renée; University Medical Center, University of Groningen, Department of Genetics</p> <p>Hogervorst, Frans; the Netherlands Cancer Institute, DNA diagnostic laboratory of the Family Cancer Clinic</p> <p>Gille, Hans; VU University Medical Center</p> <p>Redeker, Egbert; Academic Medical Centre, Clinical Genetics</p> <p>Tops, Carli; Leiden University Medical Center, Clinical Genetics</p> <p>van Gijn, Marielle; University Medical Center Utrecht, Medical Genetics</p> <p>van den Ouweland, Ans M.W.; ErasmusMC, Clinical Genetics</p> <p>Rahner, Nils; University of Bonn, Institute of Human Genetics</p> <p>Steinke, Verena; University of Bonn,, Institute of Human Genetics</p> <p>Kahl, Philip; University of Bonn, Institute of Pathology</p> <p>Holinski-Feder, Elke; University Hospital of the Ludwig-Maximilians University</p> <p>Morak, Monika; Medizinische Klinik â€œ Innenstadt, Lehrstuhl fÃ¼r Endokrinologie/Diabetologie, University Hospital of the Ludwig-Maximilian-University Munich; Center of Medical Genetics</p> <p>Kloor, Matthias; University Hospital Heidelberg, Applied Tumor</p>

	<p>Biology Stemmler, Susanne; Ruhr-University Bochum, Department of Human Genetics Betz, Beate; Heinrich-Heine University, Human Genetics Hutter, Pierre; Institute Central des Hopitaux Valaisans Bunyan, David; National Genetics Reference Laboratory (Wessex), Molecular Genetics Syngal, Sapna; Harvard Medical School Culver, Julie; City of Hope, Clinical Cancer Genetics Graham, Tracy; Odette Cancer Center Nagtegaal, Iris; Radboud University Nijmegen Medical Center, Pathology van Krieken, Johan; Radboud University Nijmegen Medical Center, Pathology Schackert, M.D., Hans; Technische Universität Dresden, Surgical Research Hoogerbrugge, Nicoline; Radboud University Nijmegen Medical Center, Human Genetics GeurtsvanKessel, Ad; Radboud University Nijmegen Medical Center, Human Genetics Ligtenberg, Marjolijn; Radboud University Nijmegen Medical Center, Human Genetics; Radboud University Nijmegen Medical Center, Pathology</p>
Key Words:	<p>HNPCC, TACSTD1, NAHR, Alu-mediated recombination, genomic deletion, Lynch Syndrome, NHEJ</p>

Formatted: Line spacing:
Double

Recurrence and variability of germline *EPCAM* deletions in Lynch syndrome

Roland P. Kuiper,^{1*} Lisenka E. L. M. Vissers,¹ Ramprasath Venkatachalam,¹ Danielle Bodmer,¹ Eveline Hoenselaar,¹ Monique Goossens,² Aline Haufe,³ Eveline Kamping,¹ Renée C. Niessen,⁴ Frans B. L. Hogervorst,⁵ Johan J. P. Gille,⁶ Bert Redeker,⁷ Carli M. J. Tops,⁸ Marielle E. van Gijn,⁹ Ans M. W. van den Ouweland,¹⁰ Nils Rahner,¹¹ Verena Steinke,¹¹ Philip Kahl,¹² Elke Holinski-Feder,¹³ Monika Morak,^{13,14} Matthias Kloor,¹⁵ Susanne Stemmler,¹⁶ Beate Betz,¹⁷ Pierre Hutter,¹⁸ David J. Bunyan,¹⁹ Sapna Syngal,²⁰ Julie O. Culver,²¹ Tracy Graham,²² Tsun L. Chan,²³ Iris D. Nagtegaal,² J.Han J.M van Krieken,² Hans K. Schackert,³ Nicoline Hoogerbrugge,¹ Ad Geurts van Kessel,¹ Marjolijn J. L. Ligtenberg^{1,2}

¹Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; ²Department of Pathology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; ³Department of Surgical Research, Universitätsklinikum Carl Gustav Carus, Technische Universität Dresden, Germany; ⁴Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; ⁵Family Cancer Clinic, The Netherlands Cancer Institute, Amsterdam, The Netherlands; ⁶Department of Clinical Genetics, VU Medical Center, Amsterdam, The Netherlands; ⁷Department of Human Genetics, Academic Medical Center, Amsterdam, The Netherlands; ⁸Center for Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands; ⁹Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; ¹⁰Department of Clinical Genetics, Erasmus Medical Center Rotterdam, Rotterdam, The Netherlands; ¹¹Institute of Human Genetics, University of Bonn, Germany; ¹²Institute of Pathology, University of Bonn, Germany; ¹³University Hospital of the Ludwig-Maximilians-University, Campus Innenstadt, Munich, Germany; ¹⁴Center of Medical Genetics, Munich, Germany;

1
2 1 ¹⁵*Institute of Pathology, Department of Applied Tumor Biology, University Hospital Heidelberg,*
3
4 2 *Heidelberg, Germany; ¹⁶Department of Human Genetics, Ruhr-University Bochum, Germany;*
5
6 3 ¹⁷*Institute of Human Genetics, Heinrich-Heine-University, Duesseldorf, Germany; ¹⁸Unit of Genetics,*
7
8 4 *Institut Central des Hôpitaux Valaisans, Sion, Switzerland; ¹⁹Wessex Regional Genetics Laboratory,*
9
10 5 *Salisbury District Hospital, Salisbury, Wiltshire, UK; ²⁰Department of Internal Medicine, Brigham and*
11
12 6 *Women's Hospital, Boston, Massachusetts, USA; ²¹Department of Clinical Cancer Genetics, City of*
13
14 7 *Hope, Duarte, California, USA; ²²Odette Cancer Centre, Sunnybrook Health Sciences Centre, Toronto,*
15
16 8 *Canada. ²³Hereditary Gastrointestinal Cancer Genetic Diagnosis Laboratory, Department of*
17
18 9 *Pathology, The University of Hong Kong, Queen Mary Hospital, Pokfulam, Hong Kong.*
19
20 10
21 11
22 12 *Correspondence to :
23 13 Roland P. Kuiper, PhD, Department of Human Genetics 855, Radboud University Nijmegen
24 14 Medical Centre, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands, Tel: 31-24-3610868,
25
26 15 e-mail: r.kuiper@antrg.umcn.nl
27
28 16
29 17 Keywords:
30
31 18 Lynch syndrome; TACSTD1; NAHR; Alu-mediated recombination; genomic deletion
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

ABSTRACT: Recently, we identified 3' end deletions in the *EPCAM* gene as a novel cause of Lynch syndrome. These truncating *EPCAM* deletions cause allele-specific epigenetic silencing of the neighboring DNA mismatch repair gene *MSH2* in tissues expressing *EPCAM*. Here we screened a cohort of unexplained Lynch-like families for the presence of *EPCAM* deletions. We identified 27 novel independent *MSH2*-deficient families from multiple geographical origins with varying deletions all encompassing the 3' end of *EPCAM*, but leaving the *MSH2* gene intact. Within the Netherlands and Germany, *EPCAM* deletions appeared to represent at least 2.8% and 1.1% of the confirmed Lynch syndrome families, respectively. *MSH2* promoter methylation was observed in epithelial tissues of all deletion carriers tested, thus confirming silencing of *MSH2* as the causative defect. In a total of 45 families, 19 different deletions were found, all including the last two exons and the transcription termination signal of *EPCAM*. All deletions appeared to originate from Alu-repeat mediated recombination events. In 17 cases regions of microhomology around the breakpoints were found, suggesting non-allelic homologous recombination as the most likely mechanism. We conclude that 3' end *EPCAM* deletions are a recurrent cause of Lynch syndrome which should be implemented in routine Lynch syndrome diagnostics.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 INTRODUCTION

2 The most frequently diagnosed colorectal cancer (CRC) syndrome is Lynch syndrome, also
3 known as hereditary non-polyposis colorectal cancer (HNPCC) ([MIMs 120435, 609310](#)),
4 which accounts for up to 5% of CRCs. Mutation carriers exhibit a high risk to develop CRC
5 (60-90%), endometrial cancer (20-60%), as well as several other cancers [Lynch and de la
6 Chapelle, 2003; Watson et al., 2008]. Lynch syndrome is caused by a germline mutation in
7 one of the DNA mismatch repair (MMR) genes *MSH2*, *MLH1*, *MSH6* or *PMS2* [Aaltonen et
8 al., 1998; Lynch and de la Chapelle, 2003; Hampel et al., 2005; Barnetson et al., 2006] ([MIM](#)
9 [120436, 609309, 600678, 600259](#)). *MSH2* and *MLH1* account for the majority of the
10 identified mutations, whereas *PMS2* mutations explain only a few percent of the confirmed
11 cases [Lynch and de la Chapelle, 2003; Barnetson et al., 2006].
12 Increasing evidence suggests that also epigenetic modifications may play a role in cancer
13 predisposition in Lynch syndrome. Several groups have reported the occurrence of mono-
14 allelic methylation of the *MLH1* gene promoter in peripheral blood cells of individuals that
15 meet the criteria for Lynch syndrome, but lack germline mutations in the *MLH1* gene [Gazzoli
16 et al., 2002; Suter et al., 2004; Hitchins et al., 2007]. Occasionally, these so-called
17 epimutations were found to be transmitted over several generations, but the mechanism
18 underlying this phenomenon remains to be elucidated [Hitchins et al., 2007; Morak et al.,
19 2008; Hesson et al., 2010]. Chan et al. [2006] for the first time reported an inherited germline
20 *MSH2* epimutation in a family presenting with Lynch-associated tumors and a mosaic *MSH2*
21 hypermethylation pattern in normal tissues. Recently, we demonstrated that these families
22 carry 3' end deletions in the epithelial cell adhesion molecule gene *EPCAM* ([MIM 185535](#)),
23 previously known as *TACSTD1*, which is located upstream of the *MSH2* gene. [EPCAM is](#)
24 [highly expressed in epithelial tissues and carcinomas](#) [Winter et al., 2003], and these
25 deletions were found to result in transcriptional read-through into the *MSH2* gene and
26 subsequent hypermethylation of its CpG island promoter [in EPCAM-expressing tissues](#)
27 [Ligtenberg et al., 2009], thereby providing an explanation for the origin of the epimutation
28 and its mode of inheritance. The identification of several additional families with 3' *EPCAM*

deletions by others [van der Klift et al., 2005; Kovacs et al., 2009; Niessen et al., 2009; Nagasaka et al., 2010; Guarinos et al., 2010] has underscored the notion that these abnormalities indeed represent a common cause of Lynch syndrome.

Here, we report the characterization of *EPCAM* deletions in 45 independent Lynch syndrome families, including hypermethylation of the *MSH2* gene promoter. The incidence of *EPCAM* deletions appeared to vary between populations and was found to represent at least 1-3% of the explained Lynch syndrome families. Detailed analysis of the *EPCAM* deletions uncovered their range of variability as well as their *Alu*-repeat-mediated origin.

1
2 1 **MATERIALS AND METHODS**

3
4 2
5
6 3 **Patients and families**

7 4 A total of 27 families with *EPCAM* deletions originating from The Netherlands (*n*=10),
8 5 Germany (*n*=11), USA (*n*=4), UK (*n*=1), and Canada (*n*=1) were identified through targeted
9 6 genomic screens in cohorts of unexplained Lynch-like families, using variable inclusion
10 7 criteria, i.e., unexplained patients with *MSH2*-deficient and/or microsatellite-unstable tumors
11 8 (Supporting Table S1). In addition, 18 *EPCAM* deletion families of various origins from earlier
12 9 studies were included in the breakpoint analyses (Supporting Table S1). All patient material
13 10 was obtained with informed consent.
14
15
16
17
18
19
20
21

22 12 **Multiplex ligation-dependent probe amplification (MLPA)**

23 13 *EPCAM* deletion screening was performed with MLPA using SALSA MLPA kits P072-B1
24 14 MSH6 or P008 MSH2/PMS2 (MRC-Holland, Amsterdam, The Netherlands). For fine-
25 15 mapping of the identified deletions we used two custom-designed probe sets as previously
26 16 described [Ligtenberg et al., 2009], in which two additional probes targeting the *EPCAM*
27 17 promoter region (probe O) and intron 4 of the *EPCAM* gene (probe P) were included (Fig. 1).
28 18 Primers were designed using the MeltIngeny program according to guidelines provided by
29 19 MRC-Holland and are available upon request.
30
31
32
33
34
35
36
37
38

39 21 **Long range PCR and breakpoint sequencing**

40 22 Based on the MLPA results, long range PCR across the deletion was applied using a
41 23 TAKARA LA PCR kit (TaKaRa Bio Inc., Otsu, Shiga, Japan) or the Expand Long Range kit
42 24 (deletions 8, 9, 13, and 14; Roche Applied Sciences, Mannheim, Germany). To identify the
43 25 exact breakpoints, the PCR products were directly sequenced at various positions in both
44 26 orientations. Primers used for these analyses are available upon request.
45
46
47
48
49
50

51 28 **Methylation analysis**

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 previously described [Ligtenberg et al., 2009], using 200 ng DNA isolated from formalin fixed paraffin embedded material. Samples with known *MGMT*, *MLH1* or *MSH2* hypermethylation levels were used as positive controls.

Bioinformatic analysis of SINE density

The density of short interspersed nuclear elements (SINEs), which include *Alu* repeats, in the maximal deletion region was compared to the remainder of the genome by random sampling of 10,000 genomic sequences of 25 kb in size. These sequences were obtained from hg18 (<http://genome.ucsc.edu/>) by random selection of autosomal chromosomes and subsequent locations. Centromeres and gaps in the sequence alignment were excluded. These 25 kb regions were annotated for the presence of all repeat masked elements, and the number of SINE elements was calculated. Next, the 95% confidence interval for the presence of SINEs within these 10,000 genomic regions was determined.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

RESULTS

Identification of novel *EPCAM* deletions in *MSH2*-deficient Lynch families

In a search for novel germline *EPCAM* deletion cases we performed a multicenter screen of unexplained Lynch-like families using multiplex ligation-dependent probe amplification (MLPA) and/or deletion PCR, which yielded 27 novel *EPCAM* deletion families (Supporting Table S1). Through the participation of all clinical genetic centers in The Netherlands, we have now identified 17 unrelated Dutch families with *EPCAM* deletions, thus representing 2.8% of all explained Lynch syndrome families and 6.9% of all explained *MSH2*-deficient families in this country, respectively (Table 1). Additionally, 11 German *EPCAM* deletion families were found in a systematic screen of 146 families with *MSH2*-deficient tumors in which no *MSH2* mutations were found (7.5%). Therefore, in Germany the frequency of *EPCAM* deletion families in explained Lynch families is at least 1.1%, which is 2.3% of all explained *MSH2*-deficient families. (Table 1).

In addition to these 27 families, we included 18 *EPCAM* deletion families that were previously reported by us and others (Supporting Table S1). Together, these screens and searches resulted in 45 independent families with *EPCAM* deletions originating from eight different countries (Supporting Table S2). Using long-range PCR we precisely localized and sequenced the breakpoints in all *EPCAM* deletion families (Table 2). In total, 19 different deletions were identified, varying in size from 2.6 to 23.8 kb. All deletions were located upstream of the *MSH2* gene promoter and encompassed at least the last two exons of the *EPCAM* gene, leaving its 5' exons intact (Fig. 1A). Our breakpoint mapping data indicate that a wide variety of *EPCAM* deletions does occur in these Lynch syndrome families.

***EPCAM* deletion carriers show *MSH2* promoter hypermethylation**

We previously showed for two different deletions (deletions 1 and 5, Table 2) that they result in allele-specific hypermethylation of the *MSH2* gene promoter in tissues expressing *EPCAM* [Ligtenberg et al., 2009]. Here, we analyzed the methylation status of the *MSH2* gene

promoter in tumor and/or normal colon mucosa tissues of at least one index patient from each of 27 different families (encompassing 11 different deletions) using methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA, Table 2, Supporting Table S2). *MSH2* promoter hypermethylation was detected in all tissues tested. One of the patients in our cohort also developed a benign dermatofibroma, which was not *MSH2*-deficient and, in contrast to the colorectal tumor, indeed was found to lack hypermethylation of the *MSH2* gene promoter. Therefore, we conclude that hypermethylation of the *MSH2* gene promoter in tissues expressing EPCAM is a general phenomenon in the deletion carriers, thereby explaining the concomitant cancer predisposition in these families.

***EPCAM* founder deletions**

Several *EPCAM* deletions appeared to be widespread both within and between different populations. The 4.9-kb *EPCAM* founder deletion, thus far observed in seven Dutch families [Ligtenberg et al., 2009; Niessen et al., 2009], was found to be present in nine out of ten additional families from The Netherlands, but in none of the families from other geographic origins, thus confirming its founder nature. Furthermore, this founder deletion appears to represent a considerable fraction (~6.5%) of the explained *MSH2*-deficient Lynch syndrome families in this population (Table 1). In addition, six *EPCAM* deletions were identified in more than one family originating from Germany (deletions 2 and 14, $n=2$ and $n=4$, respectively), Switzerland (deletion 3, $n=2$), and the USA (deletion 6, $n=2$) or from multiple origins (deletions 5 and 10, $n=3$; Table 2). Although we cannot rule out with certainty that these deletions have occurred independently, we anticipate that most of them will have an ancestral origin.

***Alu*-mediated recombination as a mechanism of origin**

It is well-established now that repetitive DNA sequences, such as *Alu* repeats, can act as facilitators of chromosomal rearrangements [Stankiewicz et al., 2010]. Previous reports have already suggested *Alu* repeat-mediated recombination as a likely mechanism for some of the

1
2 1 *EPCAM* deletions [van de Klift et al., 2005; Kovacs et al., 2009; Ligtenberg et al., 2009].
3
4 2 Indeed, all *EPCAM* deletion breakpoints characterized in this study were located within *Alu*
5
6 3 elements (Table 2 and Fig. 1). Together, the 19 different deletions involved 11 *Alu* repeats at
7
8 4 the distal intragenic breakpoints (within *EPCAM*), and 13 at the proximal breakpoints (in the
9
10 5 intergenic region between *EPCAM* and *MSH2*), of which several were involved in different
11
12 6 deletions (Fig. 1B and 1C). As expected, the two recombined *Alu* elements were always
13
14 7 directed in the same orientation, being either sense (deletions 5-11) or antisense (deletions
15
16 8 1-4 and 12-19). For 17/19 (89%) of the deletions, sequence alignment of the distal and
17
18 9 proximal *Alu* repeats revealed the presence of stretches with microhomology at the
19
20 10 breakpoint, ranging from 6 to 32 bp in size (Table 2 and Supporting Fig. S1), which is in line
21
22 11 with *Alu-Alu* mediated non-allelic homologous recombination (NAHR). Interestingly, two
23
24 12 deletions of exactly the same size (deletions 7 and 8) appeared to originate from
25
26 13 recombination events at different positions within the same *Alu* repeat pair with high
27
28 14 sequence homology, further illustrating the homology-based mechanism driving these
29
30 15 genomic deletions (Fig. 2).
31
32 16 The remaining two deletions (9 and 12) appear to have arisen by a mechanism different from
33
34 17 NAHR. Deletion 9, of which the breakpoints are near those of deletions 7 and 8 (Table 1),
35
36 18 contains a 2-nt interstitial sequence (AG) and lacks microhomology at the breakpoint
37
38 19 junction. Similarly, the sequences surrounding the breakpoint junctions of deletion 12 do,
39
40 20 with only 3 bases, not contain sufficient homology in order to be explained by NAHR. In
41
42 21 these cases, classical non-homologous end-joining (NHEJ) or microhomology-mediated
43
44 22 break-induced repair (MMBIR) may serve as better explanations for the origin of the deletion
45
46 23 [[McVety et al., 2005](#); Vissers et al., 2009].
47
48 24 Partial or complete deletion of the *MSH2* gene represents a relatively frequent cause of
49
50 25 Lynch syndrome [van der Klift et al., 2005; Li et al., 2006]. These germline deletions appear
51
52 26 to originate almost exclusively from *Alu*-mediated recombination, which is in accordance with
53
54 27 the relatively high local density of repetitive *Alu* elements [Li et al., 2006]. We have extended
55
56 28 this analysis by determining the relative *Alu* element density throughout the entire *EPCAM*-
57
58
59
60

1 *MSH2* locus in a genome-wide context. To this end, we randomly sampled 10,000 genomic
2 regions of 25 kb. This yielded a median *Alu* element density of 10 [95%CI: 0-39], which is
3 significantly lower than the density of 55 *Alu* elements that we observed within the 25-kb
4 *EPCAM-MSH2* locus (Supporting Fig. S1). This local enrichment is also observed in other
5 regions with recurrent *Alu*-mediated rearrangements (e.g. the *VHL* locus in von Hippel-
6 Lindau disease patients), but is absent in the locus encompassing the DNA mismatch repair
7 gene *MLH1* (Supporting Fig. 2). These observations may explain the wide variety of
8 deletions observed within the *EPCAM-MSH2* locus.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 **DISCUSSION**

2 Through detailed mapping and characterization of 3' *EPCAM* gene deletions in Lynch
3 syndrome families, we show that these deletions explain a considerable fraction (at least 1-
4 3%) of all families with this syndrome, thus legitimating standard clinical testing. In total, we
5 have identified and characterized 19 different *EPCAM* deletions in 45 Lynch syndrome
6 families. These deletions turned out to be highly variable in size and location, but always
7 encompassed the last two exons of the *EPCAM* gene, including its polyadenylation signal. In
8 concordance with previous studies [Ligtenberg et al., 2009; Niessen et al., 2009; Nagasaka
9 et al., 2010], all available tumor and normal colonic tissues showed hypermethylation of the
10 *MSH2* gene promoter, thus confirming a direct correlation between these two aberrations.
11 Detailed localization of the deletion breakpoints at the sequence level revealed *Alu*-mediated
12 recombination as the major mechanism underlying the occurrence of *EPCAM* deletions.
13 The presence of a mono-allelic *EPCAM* deletion results in a highly efficient silencing of the
14 *MSH2* gene in target tissues such as colonic mucosa. This observation is in full agreement
15 with the lifetime risk for colorectal cancer in these families, which appears to be similar to
16 those observed in families with other *MSH2* alterations [Kempers et al., 2010]. This efficient
17 *MSH2* inactivation may be associated with one or more of the following structural
18 characteristics of this locus: (i) the close vicinity of a neighboring gene (*EPCAM*) that is
19 oriented towards *MSH2*, and (ii) the high level of expression of *EPCAM* in targeted tissues
20 instilling *MSH2* promoter methylation. Together with the relative high density of *Alu* repeat
21 elements in this genomic region, which increases the chance of *Alu*-mediated recombination,
22 these characteristics may explain the recurrent nature of variable *EPCAM* deletions in Lynch
23 syndrome families.
24 Upon analysis of the genomic region encompassing the Lynch-associated DNA mismatch
25 repair gene *MLH1*, we found that the above described characteristics do not apply to this
26 locus. Consequently, we postulate that in the previously reported families with germline
27 methylation of the *MLH1* gene promoter, which in some families was found to be transmitted

Deleted: Ockeloen
Deleted: manuscript in preparation

1 to next generations [Hitchins et al., 2007; Morak et al., 2008; Hesson et al., 2010], the
2 mechanism causing methylation is very likely to be different.

3 Previous reports have already pointed at correlations between *Alu* repeat densities and the
4 occurrence of genomic recombinations. For example, the *VHL* locus on 3p25.3 has a local
5 *Alu* element density which is comparable to that of the *MSH2* locus on 2p21, and a similarly
6 high frequency and variety of *Alu* element-mediated deletions have been observed in von
7 Hippel-Lindau disease families [Franke et al., 2009; Nordstrom-O'Brien et al., 2010].
8 Furthermore, gross chromosomal deletions in the *MSH2* gene itself are also frequently
9 observed and, in contrast to those found in the *MLH1* gene, are all mediated by *Alu* element-
10 mediated recombination [Wijnen et al., 1998; Li et al., 2006]. The intragenic region of
11 *EPCAM* contains 25 *Alu* elements, indicating that additional deletions may be encountered in
12 the future. Eight of these elements are located upstream of exon 3 and were not involved in
13 any of the deletions identified thus far, which may indicate that a minimum of three 5'
14 *EPCAM* exons are required to induce transcription-mediated silencing of the downstream
15 *MSH2* gene.

16 Despite the high variety of *EPCAM* deletions found, a relatively large proportion of the
17 affected families shares one of at least seven distinct deletions that are likely of common
18 ancestral origin, as has been demonstrated for the Dutch founder deletion [Ligtenberg et al.,
19 2009]. The relatively high frequency of *EPCAM* deletions among Lynch syndrome families in
20 the Netherlands (Supporting Table S2) may very well be explained by the frequency of the
21 founder deletion in this population.

22 Discrimination between putative molecular mechanisms involved in the formation of the
23 *EPCAM* deletions requires a distinction between (i) meiotic recombination processes such as
24 homology-dependent NAHR and homology-independent NHEJ, and (ii) mitotic processes
25 including classical NHEJ and NHEJ mediated by microhomology (alt-NHEJ or MMEJ) and
26 replication-based mechanisms such as microhomology-mediated break-induced repair
27 (MMBIR) [Vissers et al., 2009]. The overlap in molecular fingerprints between these diverse
28 molecular mechanisms makes it difficult to discern the mechanism underlying the formation

1
2 1 of the deletions. Considering the high sequence homology between *Alu* repeats and the
3
4 2 microhomology observed at the breakpoint junctions, however, non-allelic homologous
5
6 3 recombination (NAHR) appears to be the most likely mechanism for most of the deletions.
7
8 4 Although the exact mechanism underlying the transcription-mediated epigenetic silencing of
9
10 5 the *MSH2* gene remains to be established, several studies have pointed at a correlation
11
12 6 between transcription and DNA methylation. For example, maternal imprinting of the *GNAS*
13
14 7 locus in mouse oocytes was recently shown to depend on transcription across the entire
15
16 8 locus from the upstream *NESP* promoter [Chotalia et al., 2009], of which maternal micro-
17
18 9 deletions cause pseudohypoparathyroidism type 1b in human [Bastepe et al., 2005]. At non-
19
20 10 imprinted loci, epigenetic silencing by antisense transcription has been reported for the
21
22 11 *alpha-globin* gene promoter in alpha-thalassemia as well as for the *p15* gene promoter in an
23
24 12 in vitro system [Tufarelli et al., 2003; Yu et al., 2008]. Finally, we have recently demonstrated
25
26 13 that a constitutional partial duplication of the protein tyrosine phosphatase gene *PTPRJ*, a
27
28 14 tumor suppressor gene associated with colorectal cancer susceptibility in the mouse
29
30 15 [Ruivenkamp et al., 2002], induces hypermethylation of its own promoter by transcriptional
31
32 16 read-through in a patient with colorectal cancer [Venkatachalam et al., 2010]. A possible
33
34 17 explanation may include the formation of RNA-DNA duplexes within the promoter region that
35
36 18 impinge the recruitment of the DNA methylation machinery resulting in epigenetic remodeling
37
38 19 of the promoter, similar to what has been described for antisense non-coding RNAs [Hawkins
39
40 20 et al., 2009]. These observations by others and those reported by us indicate that DNA
41
42 21 methylation instilled by transcriptional read-through across gene promoters may serve as a
43
44 22 general mechanism governing health and disease.
45
46 23 In conclusion, we have demonstrated that 3' *EPCAM* deletions represent a common cause of
47
48 24 Lynch syndrome. Based on this notion, the implementation of *EPCAM* deletion mapping in
49
50 25 routine diagnostics on suspected Lynch syndrome families should be considered. Since all
51
52 26 deletions appear to include at least the last two exons of the *EPCAM* gene, the inclusion of
53
54 27 the corresponding *EPCAM* probes in current MLPA kits may be sufficient.
55
56 28
57
58
59
60

Acknowledgements

We gratefully acknowledge the patients and their families as well as all clinical genetic centers for their cooperation. We thank E. Verwiel for expert bioinformatic support. This work was supported by grants from the Dutch Cancer Society (2009-4335 to MJL, RPK and NH), the Netherlands Organization for Health Research and Development (ZonMW 917-10-358 to RPK; ZonMW 916-86-016 to LELMV), the Sacha Swarttouw-Hijmans foundation (to NH and MJL), and the Deutsche Krebshilfe (Familial Colorectal Cancer 70-3032).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 **References**

2 2 | Aaltonen LA, Salovaara R, Kristo P, Canzian F, Hemminki A, Peltomäki P, Chadwick RB, Kääriäinen
3 3 H, Eskelinen M, Järvinen H, Mecklin JP, de la Chapelle A. 1998. Incidence of hereditary
4 4 nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. N Engl
5 5 J Med 338:1481-1487.
6 6 | Barnetson RA, Tenesa A, Farrington SM, Nicholl ID, Cetnarskyj R, Porteous ME, Campbell H, Dunlop
7 7 MG. 2006. Identification and survival of carriers of mutations in DNA mismatch-repair genes in
8 8 colon cancer. N Engl J Med 354:2751-2763.
9 9 | Bastepe M, Fröhlich LF, Linglart A, Abu-Zahra HS, Tojo K, Ward LM, Jüppner H. 2005. Deletion of the
10 10 NESP55 differentially methylated region causes loss of maternal GNAS imprints and
11 11 pseudohypoparathyroidism type Ib. Nat Genet 37:25-27.
12 12 [Batzer MA, Deininger PL. 2002. Alu repeats and human genomic diversity. Nat Rev Genet 3:370-379.](#)
13 13 | Chan TL, Yuen ST, Kong CK, Chan YW, Chan AS, Ng WF, Tsui WY, Lo MW, Tam WY, Li VS, Leung
14 14 SY. 2006. Heritable germline epimutation of MSH2 in a family with hereditary nonpolyposis
15 15 colorectal cancer. Nat Genet 38:1178-1183.
16 16 | Chotalia M, Smallwood SA, Ruf N, Dawson C, Lucifero D, Frontera M, James K, Dean W, Kelsey G.
17 17 2009. Transcription is required for establishment of germline methylation marks at imprinted
18 18 genes. Genes Dev 23:105-117.
19 19 | Gazzoli I, Loda M, Garber J, Syngal S, Kolodner RD. 2002. A hereditary nonpolyposis colorectal
20 20 carcinoma case associated with hypermethylation of the MLH1 gene in normal tissue and loss
21 21 of heterozygosity of the unmethylated allele in the resulting microsatellite instability-high tumor.
22 22 Cancer Res 62:3925-3928.
23 23 [Guarinos C, Castillejo A, Barberá VM, Pérez-Carbonell L, Sánchez-Heras AB, Segura A, Guillén-](#)
24 24 [Ponce C, Martínez-Cantó A, Castillejo MI, Egoavil CM, Jover R, Payá A, Alenda C, Soto JL.](#)
25 25 [2010. EPCAM germ line deletions as causes of Lynch syndrome in Spanish patients. J Mol](#)
26 26 [Diagn. 12:765-770,](#)
27 27 | Franke G, Bausch B, Hoffmann MM, Cybulla M, Wilhelm C, Kohlhase J, Scherer G, Neumann HP.
28 28 2009. Alu-Alu recombination underlies the vast majority of large VHL germline deletions:
29 29 Molecular characterization and genotype-phenotype correlations in VHL patients. Hum Mutat
30 30 30:776-786.

Deleted:

- 1 Hampel H, Frankel WL, Martin E, Arnold M, Khanduja K, Kuebler P, Nakagawa H, Sotamaa K, Prior
- 2 TW, Westman J, Panescu J, Fix D, Lockman J, Comeras I, de la Chapelle A. 2005. Screening
- 3 for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). *N Engl J Med* 352:1851-
- 4 1860.
- 5 Hawkins PG, Santoso S, Adams C, Anest V, Morris KV. 2009. Promoter targeted small RNAs induce
- 6 long-term transcriptional gene silencing in human cells. *Nucleic Acids Res* 37:2984-2995.
- 7 Hesson LB, Hitchins MP, Ward RL. 2010. Epimutations and cancer predisposition: importance and
- 8 mechanisms. *Curr Opin Genet Dev* 20:290-298.
- 9 Hitchins MP, Wong JJ, Suthers G, Suter CM, Martin DI, Hawkins NJ, Ward RL. 2007. Inheritance of a
- 10 cancer-associated MLH1 germ-line epimutation. *N Engl J Med* 356:697-705.
- 11 [Kempers MJE, Kuiper RP, Ockeloen CW, Chappuis PO, Hutter P, Rahner N, Schackert HK, Steinke](#)
- 12 [V, Holinski-Feder E, Morak M, Kloor M, Büttner R, Verwiel ETP, van Krieken JH, Nagtegaal ID,](#)
- 13 [Goossens M, van der Post RS, Niessen RC, Sijmons RH, Kluij I, Hogervorst FBL, Leter EM,](#)
- 14 [Gille JJP, Aalfs CM, Redeker EJW, Hes FJ, Tops CMJ, van Nesselrooij BPM, van Gijn ME,](#)
- 15 [Gómez García EB, Eccles DM, Bunyan DJ, Syngal S, Stoffel EM, Culver JO, Palomares MR,](#)
- 16 [Graham T, Velsher L, Papp J, Oláh E, Chan TL, Leung SY, Geurts van Kessel A, Kiemeny](#)
- 17 [LALM, Hoogerbrugge N, Ligtenberg MJL. 2010. Risk of colorectal and endometrial cancers in](#)
- 18 [EPCAM deletion-positive Lynch syndrome: a cohort study. *Lancet Oncol* 'in press'.](#)
- 19 Kovacs ME, Papp J, Szentirmay Z, Otto S, Olah E. 2009. Deletions removing the last exon of
- 20 TACSTD1 constitute a distinct class of mutations predisposing to Lynch syndrome. *Hum Mutat*
- 21 30:197-203.
- 22 Li L, McVety S, Younan R, Liang P, Du Sart D, Gordon PH, Hutter P, Hogervorst FB, Chong G,
- 23 Foulkes WD. 2006. Distinct patterns of germ-line deletions in MLH1 and MSH2: the implication
- 24 of Alu repetitive element in the genetic etiology of Lynch syndrome (HNPCC). *Hum Mutat*
- 25 27:388.
- 26 Ligtenberg MJL, Kuiper RP, Chan TL, Goossens M, Hebeda KM, Voorendt M, Lee TYH, Bodmer D,
- 27 Hoenselaar E, Hendriks-Cornelissen SJB, Tsui WY, Kong CK, Brunner HG, Geurts van Kessel
- 28 A, Yuen ST, van Krieken JH, Leung SY, Hoogerbrugge N. 2009. Heritable somatic methylation
- 29 and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of
- 30 TACSTD1. *Nat Genet* 41:112-117.

- 1
- 2 1 | Lynch HT, de la Chapelle A. 2003. Hereditary colorectal cancer. *N Engl J Med* 348:919-932.
- 3
- 4 2 | [McVety S, Younan R, Li L, Gordon PH, Wong N, Foulkes WD, Chong G. 2005. Novel genomic](#)
- 5 3 | [insertion-- deletion in MLH1: possible mechanistic role for non-homologous end-joining DNA](#)
- 6 4 | [repair. Clin Genet 68:234-238.](#)
- 7
- 8 5 | Morak M, Schackert HK, Rahner N, Betz B, Ebert M, Walldorf C, Royer-Pokora B, Schulmann K, von
- 9 6 | Knebel-Doeberitz M, Dietmaier W, Keller G, Kerker B, Leitner G, Holinski-Feder E. 2008.
- 10 7 | Further evidence for heritability of an epimutation in one of 12 cases with MLH1 promoter
- 11 8 | methylation in blood cells clinically displaying HNPCC. *Eur J Hum Genet* 16:804-811.
- 12
- 13 9 | Nagasaka T, Rhees J, Kloor M, Gebert J, Naomoto Y, Boland CR, Goel A. 2010. Somatic
- 14 10 | hypermethylation of MSH2 is a frequent event in Lynch Syndrome colorectal cancers. *Cancer*
- 15 11 | *Res* 70:3098-3108.
- 16
- 17 12 | Niessen RC, Hofstra RM, Westers H, Ligtenberg MJ, Kooi K, Jager PO, de Groote ML, Dijkhuizen T,
- 18 13 | Olderode-Berends MJ, Hollema H, Kleibeuker JH, Sijmons RH. 2009. Germline
- 19 14 | hypermethylation of MLH1 and EPCAM deletions are a frequent cause of Lynch syndrome.
- 20 15 | *Genes Chromosomes Cancer* 48:737-744.
- 21
- 22 16 | Nordstrom-O'Brien M, van der Luit RB, van Rooijen E, van den Ouweland AM, Majoor-Krakauer DF,
- 23 17 | Lolkema MP, van Brussel A, Voest EE, Giles RH. 2010. Genetic analysis of von Hippel-Lindau
- 24 18 | disease. *Hum Mutat* 31:521-537.
- 25
- 26 19 | Ruivenkamp CA, van Wezel T, Zanon C, Stassen AP, Vlcek C, Csikós T, Klous AM, Tripodis N,
- 27 20 | Perrakis A, Boerrigter L, Groot PC, Lindeman J, Mooi WJ, Meijjer GA, Scholten G, Dauwerse H,
- 28 21 | Paces V, van Zandwijk N, van Ommen GJ, Demant P. 2002. Ptpnj is a candidate for the mouse
- 29 22 | colon-cancer susceptibility locus Scc1 and is frequently deleted in human cancers. *Nat Genet*
- 30 23 | 31:295-300.
- 31
- 32 24 | Stankiewicz P, Lupski JR 2010. Structural variation in the human genome and its role in disease. *Annu*
- 33 25 | *Rev Med* 61:437-455
- 34
- 35 26 | Suter CM, Martin DI, Ward RL. 2004. Germline epimutation of MLH1 in individuals with multiple
- 36 27 | cancers. *Nat Genet* 36:497-501.
- 37
- 38 28 | Tufarelli C, Stanley JA, Garrick D, Sharpe JA, Ayyub H, Wood WG, Higgs DR. 2003. Transcription of
- 39 29 | antisense RNA leading to gene silencing and methylation as a novel cause of human genetic
- 40 30 | disease. *Nat Genet* 34:157-165.
- 41
- 42
- 43
- 44
- 45
- 46
- 47
- 48
- 49
- 50
- 51
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60

- van der Klift H, Wijnen J, Wagner A, Verkuilen P, Tops C, Otway R, Kohonen-Corish M, Vasen H, Oliani C, Barana D, Moller P, Delozier-Blanchet C, Hutter P, Foulkes W, Lynch H, Burn J, Möslin G, Fodde R. 2005. 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 Chromosomes Cancer* 44:123-138.
- Venkatachalam R, Ligtenberg MJL, Hoogerbrugge N, Schackert HK, Görgens H, Hahn M-M, Kamping EJ, Vreede L, Hoenselaar E, van der Looij E, Goossens M, Churchman M, Carvajal-Carmona L, Tomlinson IPM, de Bruijn DRH, Geurts van Kessel A, Kuiper, RP. 2010. Germline epigenetic silencing of the tumor suppressor gene PTPRJ in early onset familial colorectal cancer. *Gastroenterology* "In press".
- Vissers LE, Bhatt SS, Janssen IM, Xia Z, Lalani SR, Pfundt R, Derwinska K, de Vries BB, Gilissen C, Hoischen A, Nesteruk M, Wisniewiecka-Kowalnik B, Smyk M, Brunner HG, Cheung SW, van Geurts van Kessel A, Veltman JA, Stankiewicz P. 2009. Rare pathogenic microdeletions and tandem duplications are microhomology-mediated and stimulated by local genomic architecture. *Hum Mol Genet* 18:3579-3593.
- Watson P, Vasen HF, Mecklin JP, Bernstein I, Aarnio M, Järvinen HJ, Myrhøj T, Sunde L, Wijnen JT, Lynch HT. 2008. The risk of extra-colonic, extra-endometrial cancer in the Lynch syndrome. *Int J Cancer* 123:444-449.
- Wijnen J, van der Klift H, Vasen H, Khan PM, Menko F, Tops C, Meijers Heijboer H, Lindhout D, Møller P, Fodde R. 1998. MSH2 genomic deletions are a frequent cause of HNPCC. *Nat Genet* 20:326-328.
- [Winter MJ, Nagtegaal ID, van Krieken JH, Litvinov SV. 2003. The epithelial cell adhesion molecule \(Ep-CAM\) as a morphoregulatory molecule is a tool in surgical pathology. *Am J Pathol* 163:2139-2148.](#)
- Yu W, Gius D, Onyango P, Muldoon-Jacobs K, Karp J, Feinberg AP, Cui H. 2008. Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature* 451:202-206.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 **Figures Legends**

2
3
4 2
5
6 3 **Figure 1.** *EPCAM* deletions in Lynch syndrome patients. **A:** Schematic outline of the
7 4 genomic region around *EPCAM* and *MSH2*, showing 19 different deletions (grey bars)
8
9 5 identified in 45 families. All deletions include at least exons 8 and 9 of *EPCAM*. Deletions
10 6 identified in multiple (apparently) unrelated families are indicated in dark grey. Positions of
11 7 the MLPA probes used for deletion mapping are indicated by triangles. All intragenic (**B**) and
12
13 8 intergenic (**C**) breakpoints are located in *Alu* repeats (referred to as SINEs: short
14 9 interspersed nuclear elements, red bars), of which eight are involved in several different
15 10 deletions (indicated by arrows and numbers of the deletion). Arrowheads above the bars
16 11 denote the orientation of the repeats.

17
18
19
20 12
21
22 13
23
24 14 **Figure 2.** Sequence alignment of two *Alu* repeats involved in two distinct *EPCAM* deletions.
25
26 15 A distal intragenic repeat (*AluSp*, in green) and a proximal intergenic repeat (*AlusSx*, in blue)
27 16 show high local sequence homology. The microhomology around the breakpoint in deletions
28 17 7 and 8 are marked by shaded boxes. Deletion 9 involves the same intragenic repeat,
29 18 including a directly downstream located intergenic *Alu* repeat sequence (*FLAM-C*), with a
30 19 lack of local microhomology around the breakpoint. The position of the breakpoints and the
31 20 insertion of a di-nucleotide sequence AG are indicated by triangles.

Table 1. Relative incidence of *EPCAM* deletions in The Netherlands and Germany

Cohort	no. of families	% of explained MSH2-deficient families ⁵	% of explained Lynch families
The Netherlands¹			
<i>EPCAM</i> deletions ³	17	6.9%	2.8%
<i>EPCAM</i> founder deletions ³	16	6.5%	2.6%
<i>MSH2</i> mutations ⁴	230		37.2%
Explained Lynch families	618		
Germany²			
<i>EPCAM</i> deletions	11	2.3%	1.1%
<i>MSH2</i> mutations ⁴	458		47.9%
Explained Lynch families	957		

¹includes all unique families that are known in one of the DNA diagnostic laboratories in Nijmegen, Rotterdam, Leiden, Amsterdam (Netherlands Cancer Institute, University of Amsterdam, and the Free University of Amsterdam), Utrecht, and Groningen. ²includes all unique families that are known by the German HNPCC consortium. ³all cases known thus far are reported in this study. ⁴Including *MSH2* deletions and *EPCAM-MSH2* deletions. ⁵The total number of families with MSH2-deficient tumors is composed of families carrying *MSH2* mutations or deletions and *EPCAM* deletions.

Table 2. *EPCAM* deletions in 45 MSH2-deficient Lynch syndrome families

deletion	Country of Origin ¹	number of families	deleted <i>EPCAM</i> exons	size deletion (bp)	Nomenclature ²	microhomology (bp)	repetitive element distal ⁴	repetitive element proximal ⁴	max sequence homology (%)	MSH2 methylation (families) ⁵	reference
1	NL	16	8+9	4,909	c.859-1462_*1999del	6 bp	<i>AluSx</i>	<i>AluSq</i>	84% for 211/250	yes (12)	Ligtenberg et al, 2009; Niessen et al., 2009
2	D	2	5-9	11,660	c.491+529_*874del	25 bp	<i>AluSg</i>	<i>AluSg/x</i>	78% for 156/198	yes (1)	
3	CH	2	5-9	23,829	c.492-509_*13721del	24 bp	<i>AluSp</i>	<i>AluSg</i>	79% for 232/292	yes (1)	Van der Klift et al., 2005
4	H	1	6-9	10,355	c.555+402_*1220del	12 bp	<i>AluSx</i>	<i>AluSx</i>	77% for 241/309	NA	Kovacs et al., 2009
5	CN/USA	3	6-9	22,836	c.555+927_*14226del	32 bp	<i>AluY</i>	<i>AluSc</i>	79% for 237/300	yes (2)	Ligtenberg et al., 2009
6	USA	2	6-9	13,128	c.555+901_*4492del	15 bp	<i>AluY</i>	<i>AluSx</i>	79% for 225/282	yes (2)	
7	NL	1	8+9	9,963	c.858+1244_*4562del	18 bp	<i>AluSp</i>	<i>AluSx</i>	85% for 243/284	NA	
8	D	1	8+9	9,963	c.858+1212_*4521del	8 bp	<i>AluSp</i>	<i>AluSx</i>	82% for 240/291	yes (1)	
9	D	1	8+9	10,074	c.858+1364_*4793del	-	<i>AluSp</i>	<i>FLAM-C Alu</i>	85% for 243/284	yes (1)	
10	D/H	3	8+9	8,674	c.858+2478_*4507del	14 bp	<i>AluSp</i>	<i>AluSx</i>	83% for 232/278	NA	Kovacs et al., 2009
11	H	1	8+9	14,734	c.859-2524_*10762del	15 bp	<i>AluSp</i>	<i>AluSp</i>	86% for 137/159	NA	
12	UK	1	8+9	2,419	c.859-353_*618del	3 bp	<i>AluSx</i>	<i>AluSg</i>	78% for 222/282	yes (1)	
13	D	1	8+9	2,648	c.859-670_*530del	18 bp	<i>AluSx</i>	<i>AluSg</i>	78% for 246/312	yes (1)	
14	D ³	4	8+9	16,834	c.859-689_*14697del	24 bp	<i>AluSx</i>	<i>AluSx</i>	82% for 246/299	yes(4)	
15	H	1	8+9	6,058	c.859-696_*3914del	19 bp	<i>AluSx</i>	<i>AluJo</i>	75% for 114/151	NA	Kovacs et al., 2009
16	D	1	8+9	5,246	c.859-1682_*2116del	13 bp	<i>AluJb</i>	<i>AluSq</i>	78% for 180/229	NA	
17	USA	1	8+9	8,879	c.859-1605_*5862del	10 bp	<i>AluJb</i>	<i>AluSq</i>	79% for 153/193	yes (1)	
18	USA	1	8+9	13,004	c.859-645_*10911del	14 bp	<i>AluSx</i>	<i>AluSx</i>	91% for 73/80	NA	Van der Klift et al., 2005
19	D	1	4-9	16,500	c.423-545_*3903del	7 bp	<i>AluSq</i>	<i>AluJo</i>	80% for 183/227	NA	

¹Families originate from The Netherlands (NL), Germany (D), Switzerland, (CH), Hungary (H), China (CN), USA, and UK. ²Nomenclature is based on mRNA sequence with Genebank Accession Code NM_002354. ³Includes one family from unknown European origin [van der Klift et al., 2005]. See also Supporting Table 1. ⁴Alu-subfamily identified at each each breakpoint [Batzer and Deininger, 2002]. ⁵Methylation specific-MLPA on normal colon mucosa and/or colorectal tumor tissues. NA, not assessable (no material left or DNA of too poor quality).

Kuiper et al., Figure 1

EPCAM deletions in Lynch syndrome patients. A: Schematic outline of the genomic region around EPCAM and MSH2, showing 19 different deletions (grey bars) identified in 45 families. All deletions include at least exons 8 and 9 of EPCAM. Deletions identified in multiple (apparently) unrelated families are indicated in dark grey. Positions of the MLPA probes used for deletion mapping are indicated by triangles. All intragenic (B) and intergenic (C) breakpoints are located in Alu repeats (referred to as SINEs: short interspersed nuclear elements, red bars), of which eight are involved in several different deletions (indicated by arrows and numbers of the deletion). Arrowheads above the bars denote the orientation of the repeats.

186x208mm (600 x 600 DPI)

Kuiper et al., Figure 2

Sequence alignment of two Alu repeats involved in two distinct EPCAM deletions. A distal intragenic repeat (AluSp, in green) and a proximal intergenic repeat (AluSx, in blue) show high local sequence homology. The microhomology around the breakpoint in deletions 7 and 8 are marked by shaded boxes. Deletion 9 involves the same intragenic repeat, including a directly downstream located intergenic Alu repeat sequence (FLAM-C), with a lack of local microhomology around the breakpoint. The position of the breakpoints and the insertion of a di-nucleotide sequence AG are indicated by triangles.

121x114mm (600 x 600 DPI)

Supporting information

Recurrence and variability of germline *EPCAM* deletions in Lynch syndrome

Roland P. Kuiper,^{1*} Lisenka E. L. M. Vissers,¹ Ramprasath Venkatachalam,¹ Danielle Bodmer,¹ Eveline Hoenselaar,¹ Monique Goossens,² Aline Haufe,³ Eveline Kamping,¹ Renée C. Niessen,⁴ Frans B. L. Hogervorst,⁵ Johan J. P. Gille,⁶ Bert Redeker,⁷ Carli M. J. Tops,⁸ Marielle E. van Gijn,⁹ Ans M. W. van den Ouweland,¹⁰ Nils Rahner,¹¹ Verena Steinke,¹¹ Philip Kahl,¹² Elke Holinski-Feder,¹³ Monika Morak,^{13,14} Matthias Kloor,¹⁵ Susanne Stemmler,¹⁶ Beate Betz,¹⁷ Pierre Hutter,¹⁸ David J. Bunyan,¹⁹ Sapna Syngal,²⁰ Julie O. Culver,²¹ Tracy Graham,²² Tsun L. Chan,²³ Iris D. Nagtegaal,² J.Han J.M van Krieken,² Hans K. Schackert,³ Nicoline Hoogerbrugge,¹ Ad Geurts van Kessel,¹ Marjolijn J. L. Ligtenberg^{1,2}

Figure S1. Distribution of short interspersed nuclear elements (SINEs), which include *Alu* repeats, for 10,000 random genomic loci of 25 kb. A total number of 10,000 loci of 25 kb in size were randomly generated from the UCSC genome browser to represent a random sampling of the genome. The size of 25 kb was chosen to mimic the genomic size interval in which all *EPCAM* deletions occur (total locus: 25,484 bp). For all regions, SINE elements were annotated to determine their frequency per genomic region of 25 kb. The 95% confidence interval indicates 0 to 39 SINE elements per 25 kb region. The *EPCAM* locus contains 55 such SINE elements, indicating that the number of SINEs located in the *EPCAM* locus is significantly higher than elsewhere in the genome.

Figure S2. Local genomic SINE density surrounding the *EPCAM/MSH2* locus, the *MLH1* locus and the *VHL* locus. Comparison of local SINE (*Alu*) densities of the *EPCAM/MSH2* locus (A) and the *MLH1* locus (B). The local SINE density of the *VHL* locus (C), another locus frequently targeted by *Alu-Alu* mediated recombination, is shown as a reference. Local SINE density per 10 kb interval (1kb moving window) were determined for a total region of ~2 Mb surrounding these loci. The total number of SINEs (gray lines) as well as the moving averages (black lines) are plotted. The exact position of the genes and their orientations are indicated by black arrows. Blue dotted lines depict the average level of SINEs in a 10-kb region. The data indicate that the local SINE density surrounding the *MLH1* is not significantly elevated compared to the local genomic background. A drastic drop in *Alu* density within the *MSH2* gene may explain why genomic deletions are usually restricted the 5' end and rarely affect the 3' end of the gene.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table S1. Screening of *EPCAM* deletions in unexplained *MSH2*-deficient families

Country of origin	Detection method ¹	<i>EPCAM</i> deleted	Reference
The Netherlands ²	MLPA, PCR	10	
Germany	MLPA	11	
USA	MLPA	4	
UK	MLPA	1	
Canada	MLPA	1	
The Netherlands ³	MLPA	7	Ligtenberg et al., 2009; Niessen et al., 2009
Hungary	SB	5	Kovacs et al., 2009
Switzerland	SB	2	van der Klift et al., 2005
USA	SB	1	van der Klift et al., 2005
China (Hong Kong)	LR-PCR	2	Ligtenberg et al., 2009
Unknown (European)	SB	1	van der Klift et al., 2005

¹detection methods used were MLPA (P072 or P008 kit, MRC Holland and a previously described custom designed kit [Ligtenberg et al., 2009]), deletion spanning PCR (used for detection of the Dutch founder deletion), Southern blotting (SB) or linkage-based long-range PCR (LR-PCR). ²Includes clinical genetic centers from Nijmegen, Rotterdam, Leiden, Amsterdam (Netherlands Cancer Institute and University of Amsterdam), and Utrecht. ³Includes families from Nijmegen (*n*=4), and Groningen (*n*=3).

Table S2. EPCAM deletions in 45 families with Lynch-like features

FamID	Country of origin	Centre of origin	Deletion	Methylation	Deleted exons	Size (bp)	Start deletion [#]	End deletion [#]	Annotation (NM_002354)	Reference (FamID)
F01NL	Netherlands	Nijmegen	1	p	8-9	4,909	47,464,347	47,469,255	c.859-1462_*1999del	Ligtenberg et al., 2009 (A)
F02NL	Netherlands	Nijmegen	1	p	8-9	4,909	47,464,347	47,469,255	c.859-1462_*1999del	Ligtenberg et al., 2009 (B)
F03NL	Netherlands	Nijmegen	1	p	8-9	4,909	47,464,347	47,469,255	c.859-1462_*1999del	Ligtenberg et al., 2009 (D)
F04NL	Netherlands	Nijmegen	1	p	8-9	4,909	47,464,347	47,469,255	c.859-1462_*1999del	Ligtenberg et al., 2009 (C)
F05NL	Netherlands	Groningen	1	p	8-9	4,909	47,464,347	47,469,255	c.859-1462_*1999del	Niessen et al., 2009
F06NL	Netherlands	Groningen	1	p	8-9	4,909	47,464,347	47,469,255	c.859-1462_*1999del	Niessen et al., 2009
F07NL	Netherlands	Groningen	1	p	8-9	4,909	47,464,347	47,469,255	c.859-1462_*1999del	Niessen et al., 2009
F08NL	Netherlands	Amsterdam	1	p	8-9	4,909	47,464,347	47,469,255	c.859-1462_*1999del	
F09NL	Netherlands	Amsterdam	1	p	8-9	4,909	47,464,347	47,469,255	c.859-1462_*1999del	
F11NL	Netherlands	Leiden	1	NA	8-9	4,909	47,464,347	47,469,255	c.859-1462_*1999del	
F12NL	Netherlands	Utrecht	1	p	8-9	4,909	47,464,347	47,469,255	c.859-1462_*1999del	
F13NL	Netherlands	Nijmegen	1	p	8-9	4,909	47,464,347	47,469,255	c.859-1462_*1999del	
F15NL	Netherlands	Amsterdam	1	NA	8-9	4,909	47,464,347	47,469,255	c.859-1462_*1999del	
F16NL	Netherlands	Amsterdam	1	NA	8-9	4,909	47,464,347	47,469,255	c.859-1462_*1999del	
F17NL	Netherlands	Amsterdam	1	NA	8-9	4,909	47,464,347	47,469,255	c.859-1462_*1999del	
F18NL	Netherlands	Groningen	1	p	8-9	4,909	47,464,347	47,469,255	c.859-1462_*1999del	
F03DId	Germany	Munich	2	NA	5-9	11,659	47,456,471	47,468,130	c.491+529_*874del	
F07DId	Germany	Bonn	2	p	5-9	11,659	47,456,471	47,468,130	c.491+529_*874del	
F01Su	Switzerland	Sion	3	p	5-9	23,829	47,457,149	47,480,977	c.492-509_*13721del	van der Klift et al., 2005 (SA85)
F14Su	Switzerland	Sion	3	p	5-9	23,829	47,457,149	47,480,977	c.492-509_*13721del	
F04HU	Hongarije	Budapest	4	NA	6-9	10,355	47,458,122	47,468,476	c.555+402_*1220del	Kovacs et al., 2009 (HFC134)
F05HU	Hongarije	Budapest	4	NA	6-9	10,355	47,458,122	47,468,476	c.555+402_*1220del	Kovacs et al., 2009 (HFC138)
F01HK	China	Hong Kong	5	p	6-9	22,836	47,458,647	47,481,482	c.555+927_*14226del	Ligtenberg et al., 2009 (HK-A)
F02HK	China	Hong Kong	5	p	6-9	22,836	47,458,647	47,481,482	c.555+927_*14226del	Ligtenberg et al., 2009 (HK-B)
F01US	USA	Boston	5	NA	6-9	22,836	47,458,647	47,481,482	c.555+927_*14226del	
F02US	USA	Boston	6	p	6-9	13,128	47,458,621	47,471,748	c.555+901_*4492del	
F03US	USA	Boston	6	p	6-9	13,128	47,458,621	47,471,748	c.555+901_*4492del	

F10NL	Netherlands	Leiden	7	NA	8-9	9,963	47,461,856	47,471,818	c.858+1244_*4562delinsAG	
F05DId	Germany	Bonn	8	p	8-9	9,962	47,461,824	47,471,785	c.858+1211_*4529del	
F09DId	Germany	Heidelberg	9	p	8-9	10,074	47,461,976	47,472,049	c.858+1364_*4793del	
F01Can	Canada	Toronto	10	NA	8-9	8,674	47,463,090	47,471,763	c.858+2478_*4507del	
F02DId	Germany	Munich	10	NA	8-9	8,674	47,463,090	47,471,763	c.858+2478_*4507del	
F02HU	Hongarije	Budapest	10	NA	8-9	8,674	47,463,090	47,471,763	c.858+2478_*4507del	Kovacs et al., 2009 (HFC073)
F01HU	Hongarije	Budapest	11	NA	8-9	14,734	47,463,285	47,478,018	c.859-2524_*10762del	Kovacs et al., 2009 (HFC009)
F01UK	UK	Salisbury	12	p	8-9	2,419	47,465,456	47,467,874	c.859-353_*618del	
F11DId	Germany	Bonn	13	p	8-9	2,648	47,465,139	47,467,786	c.859-670_*530del	
F01DId	Germany	Munich	14	p	8-9	16,834	47,465,120	47,481,953	c.859-689_*14697del	
F06DId	Germany	Bonn	14	p	8-9	16,834	47,465,120	47,481,953	c.859-689_*14697del	
F08DId	Germany	Heidelberg	14	p	8-9	16,834	47,465,120	47,481,953	c.859-689_*14697del	
C53	not reported	not reported	14	NA	8-9	16,834	47,465,120	47,481,953	c.859-689_*14697del	van der Klift et al., 2005 (C53)
F03HU	Hongarije	Budapest	15	NA	8-9	6,058	47,465,113	47,471,170	c.859-696_*3914del	Kovacs et al., 2009 (HFC121)
F04DId	Germany	Munich	16	NA	8-9	5,246	47,464,127	47,469,372	c.859-1682_*2116del	
F04US	USA	Duarte	17	p	8-9	8,879	47,464,204	47,473,082	c.859-1605_*5862del	
F05US	USA	Omaha	18	NA	8-9	13,004	47,465,164	47,478,167	c.859-645_*10911del	van der Klift et al., 2005 (CL177)
F10DId	Germany	Munich	19	NA	4-9	15,828	47,455,333	47,471,160	c.426-544_*3904del	

start and end positions are bp position on chr 2 based on hg18; p= positive; NA, not assessable (no material left or DNA of too poor quality)