Recurrence and variability of germline EPCAM deletions in Lynch syndrome

Roland Kuiper, Lisenka Vissers, Ramprasath Venkatachalam, Danielle Bodmer, Eveline Hoenselaar, Monique Goossens, Aline Haufe, Eveline Kamping, Renée C. Niessen, Frans B.L. Hogervorst, et al.

▶ **To cite this version:**

Roland Kuiper, Lisenka Vissers, Ramprasath Venkatachalam, Danielle Bodmer, Eveline Hoenselaar, et al.. Recurrence and variability of germline EPCAM deletions in Lynch syndrome. Human Mutation, Wiley, 2011, 32 (4), pp.407. 10.1002/humu.21446 . hal-00618183

HAL Id: hal-00618183

https://hal.archives-ouvertes.fr/hal-00618183

Submitted on 1 Sep 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Human Mutation

Recurrence and variability of germline EPCAM deletions in Lynch syndrome

Journal:	*Human Mutation*
Manuscript ID:	humu-2010-0515.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	30-Nov-2010

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

	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
Key Words:	HNPCC, TACSTD1, NAHR, Alu-mediated recombination, genomic deletion, Lynch Syndrome, NHEJ

SCHOLARONE™
Manuscripts
For Review

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]

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

For Peer Review

[15]*Institute of Pathology, Department of Applied Tumor Biology, University Hospital Heidelberg, Heidelberg, Germany;* [16]*Department of Human Genetics, Ruhr-University Bochum, Germany;* [17]*Institute of Human Genetics, Heinrich-Heine-University, Duesseldorf, Germany;* [18]*Unit of Genetics, Institut Central des Hôpitaux Valaisans, Sion, Switzerland;* [19]*Wessex Regional Genetics Laboratory, Salisbury District Hospital, Salisbury, Wiltshire, UK;* [20]*Department of Internal Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA;* [21]*Department of Clinical Cancer Genetics, City of Hope, Duarte,California, USA;* [22]*Odette Cancer Centre, Sunnybrook Health Sciences Centre, Toronto, Canada.* [23]*Hereditary Gastrointestinal Cancer Genetic Diagnosis Laboratory, Department of Pathology, The University of Hong Kong, Queen Mary Hospital, Pokfulam, Hong Kong.*

*Correspondence to :

Roland P. Kuiper, PhD, Department of Human Genetics 855, Radboud University Nijmegen Medical Centre, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands, Tel: 31-24-3610868, e-mail: r.kuiper@antrg.umcn.nl

Keywords:

Lynch syndrome; TACSTD1; NAHR; Alu-mediated recombination; genomic deletion

For Peer Review

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.

For Peer Review

INTRODUCTION

The most frequently diagnosed colorectal cancer (CRC) syndrome is Lynch syndrome, also known as hereditary non-polyposis colorectal cancer (HNPCC) (MIMs 120435, 609310), which accounts for up to 5% of CRCs. Mutation carriers exhibit a high risk to develop CRC (60-90%), endometrial cancer (20-60%), as well as several other cancers [Lynch and de la Chapelle, 2003; Watson et al., 2008]. Lynch syndrome is caused by a germline mutation in one of the DNA mismatch repair (MMR) genes *MSH2, MLH1*, *MSH6* or *PMS2* [Aaltonen et al., 1998; Lynch and de la Chapelle, 2003; Hampel et al., 2005; Barnetson et al., 2006] (MIM 120436, 609309, 600678, 600259). *MSH2* and *MLH1* account for the majority of the identified mutations, whereas *PMS2* mutations explain only a few percent of the confirmed cases [Lynch and de la Chapelle, 2003; Barnetson et al., 2006].

Increasing evidence suggests that also epigenetic modifications may play a role in cancer predisposition in Lynch syndrome. Several groups have reported the occurrence of mono-allelic methylation of the *MLH1* gene promoter in peripheral blood cells of individuals that meet the criteria for Lynch syndrome, but lack germline mutations in the *MLH1* gene [Gazzoli et al., 2002; Suter et al., 2004; Hitchins et al., 2007]. Occasionally, these so-called epimutations were found to be transmitted over several generations, but the mechanism underlying this phenomenon remains to be elucidated [Hitchins et al., 2007; Morak et al., 2008; Hesson et al., 2010]. Chan et al. [2006] for the first time reported an inherited germline *MSH2* epimutation in a family presenting with Lynch-associated tumors and a mosaic *MSH2* hypermethylation pattern in normal tissues. Recently, we demonstrated that these families carry 3' end deletions in the epithelial cell adhesion molecule gene *EPCAM* (MIM 185535), previously known as *TACSTD1*, which is located upstream of the *MSH2* gene. *EPCAM* is highly expressed in epithelial tissues and carcinomas [Winter et al., 2003], and these deletions were found to result in transcriptional read-through into the *MSH2* gene and subsequent hypermethylation of its CpG island promoter in *EPCAM*-expressing tissues [Ligtenberg et al., 2009], thereby providing an explanation for the origin of the epimutation and its mode of inheritance. The identification of several additional families with 3' *EPCAM*

For Peer Review

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.

For Peer Review

MATERIALS AND METHODS

Patients and families

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

Multiplex ligation-dependent probe amplification (MLPA)

EPCAM deletion screening was performed with MLPA using SALSA MLPA kits P072-B1 MSH6 or P008 MSH2/PMS2 (MRC-Holland, Amsterdam, The Netherlands). For fine-mapping of the identified deletions we used two custom-designed probe sets as previously described [Ligtenberg et al., 2009], in which two additional probes targeting the *EPCAM* promoter region (probe O) and intron 4 of the *EPCAM* gene (probe P) were included (Fig. 1). Primers were designed using the MeltIngeny program according to guidelines provided by MRC-Holland and are available upon request.

Long range PCR and breakpoint sequencing

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

Methylation analysis

For Peer Review

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.

For Peer Review

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

For Peer Review

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

For Peer Review

EPCAM deletions [van de Klift et al., 2005; Kovacs et al., 2009; Ligtenberg et al., 2009]. Indeed, all *EPCAM* deletion breakpoints characterized in this study were located within *Alu* elements (Table 2 and Fig. 1). Together, the 19 different deletions involved 11 *Alu* repeats at the distal intragenic breakpoints (within *EPCAM*), and 13 at the proximal breakpoints (in the intergenic region between *EPCAM* and *MSH2*), of which several were involved in different deletions (Fig. 1B and 1C). As expected, the two recombined *Alu* elements were always directed in the same orientation, being either sense (deletions 5-11) or antisense (deletions 1-4 and 12-19). For 17/19 (89%) of the deletions, sequence alignment of the distal and proximal *Alu* repeats revealed the presence of stretches with microhomology at the breakpoint, ranging from 6 to 32 bp in size (Table 2 and Supporting Fig. S1), which is in line with *Alu-Alu* mediated non-allelic homologous recombination (NAHR). Interestingly, two deletions of exactly the same size (deletions 7 and 8) appeared to originate from recombination events at different positions within the same *Alu* repeat pair with high sequence homology, further illustrating the homology-based mechanism driving these genomic deletions (Fig. 2).

The remaining two deletions (9 and 12) appear to have arisen by a mechanism different from NAHR. Deletion 9, of which the breakpoints are near those of deletions 7 and 8 (Table 1), contains a 2-nt interstitial sequence (AG) and lacks microhomology at the breakpoint junction. Similarly, the sequences surrounding the breakpoint junctions of deletion 12 do, with only 3 bases, not contain sufficient homology in order to be explained by NAHR. In these cases, classical non-homologous end-joining (NHEJ) or microhomology-mediated break-induced repair (MMBIR) may serve as better explanations for the origin of the deletion [McVety et al., 2005; Vissers et al., 2009].

Partial or complete deletion of the *MSH2* gene represents a relatively frequent cause of Lynch syndrome [van der Klift et al., 2005; Li et al., 2006]. These germline deletions appear to originate almost exclusively from *Alu*-mediated recombination, which is in accordance with the relatively high local density of repetitive *Alu* elements [Li et al., 2006]. We have extended this analysis by determining the relative *Alu* element density throughout the entire *EPCAM*-

For Peer Review

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

For Peer Review

DISCUSSION

Through detailed mapping and characterization of 3' *EPCAM* gene deletions in Lynch syndrome families, we show that these deletions explain a considerable fraction (at least 1-3%) of all families with this syndrome, thus legitimating standard clinical testing. In total, we have identified and characterized 19 different *EPCAM* deletions in 45 Lynch syndrome families. These deletions turned out to be highly variable in size and location, but always encompassed the last two exons of the *EPCAM* gene, including its polyadenylation signal. In concordance with previous studies [Ligtenberg et al., 2009; Niessen et al., 2009; Nagasaka et al., 2010], all available tumor and normal colonic tissues showed hypermethylation of the *MSH2* gene promoter, thus confirming a direct correlation between these two aberrations. Detailed localization of the deletion breakpoints at the sequence level revealed *Alu*-mediated recombination as the major mechanism underlying the occurrence of *EPCAM* deletions.

The presence of a mono-allelic *EPCAM* deletion results in a highly efficient silencing of the *MSH2* gene in target tissues such as colonic mucosa. This observation is in full agreement with the lifetime risk for colorectal cancer in these families, which appears to be similar to those observed in families with other *MSH2* alterations [Kempers et al, 2010]. This efficient *MSH2* inactivation may be associated with one or more of the following structural characteristics of this locus: (i) the close vicinity of a neighboring gene (*EPCAM*) that is oriented towards *MSH2*, and (ii) the high level of expression of EPCAM in targeted tissues instilling *MSH2* promoter methylation. Together with the relative high density of *Alu* repeat elements in this genomic region, which increases the chance of *Alu*-mediated recombination, these characteristics may explain the recurrent nature of variable *EPCAM* deletions in Lynch syndrome families.

Upon analysis of the genomic region encompassing the Lynch-associated DNA mismatch repair gene *MLH1*, we found that the above described characteristics do not apply to this locus. Consequently, we postulate that in the previously reported families with germline methylation of the *MLH1* gene promoter, which in some families was found to be transmitted

Deleted: Ockeloen

Deleted: manuscript in preparation

For Peer Review

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

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

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

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

For Peer Review

of the deletions. Considering the high sequence homology between *Alu* repeats and the microhomology observed at the breakpoint junctions, however, non-allelic homologous recombination (NAHR) appears to be the most likely mechanism for most of the deletions.

Although the exact mechanism underlying the transcription-mediated epigenetic silencing of the *MSH2* gene remains to be established, several studies have pointed at a correlation between transcription and DNA methylation. For example, maternal imprinting of the *GNAS* locus in mouse oocytes was recently shown to depend on transcription across the entire locus from the upstream *NESP* promoter [Chotalia et al., 2009], of which maternal micro-deletions cause pseudohypoparathyroidism type 1b in human [Bastepe et al., 2005]. At non-imprinted loci, epigenetic silencing by antisense transcription has been reported for the *alpha-globin* gene promoter in alpha-thalassemia as well as for the *p15* gene promoter in an in vitro system [Tufarelli et al., 2003; Yu et al., 2008]. Finally, we have recently demonstrated that a constitutional partial duplication of the protein tyrosine phosphatase gene *PTPRJ*, a tumor suppressor gene associated with colorectal cancer susceptibility in the mouse [Ruivenkamp et al., 2002], induces hypermethylation of its own promoter by transcriptional read-through in a patient with colorectal cancer [Venkatachalam et al., 2010]. A possible explanation may include the formation of RNA-DNA duplexes within the promoter region that impinge the recruitment of the DNA methylation machinery resulting in epigenetic remodeling of the promoter, similar to what has been described for antisense non-coding RNAs [Hawkins et al., 2009]. These observations by others and those reported by us indicate that DNA methylation instilled by transcriptional read-through across gene promoters may serve as a general mechanism governing health and disease.

In conclusion, we have demonstrated that 3' *EPCAM* deletions represent a common cause of Lynch syndrome. Based on this notion, the implementation of *EPCAM* deletion mapping in routine diagnostics on suspected Lynch syndrome families should be considered. Since all deletions appear to include at least the last two exons of the *EPCAM* gene, the inclusion of the corresponding *EPCAM* probes in current MLPA kits may be sufficient.

For Peer Review

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).

For Peer Review

References

Aaltonen LA, Salovaara R, Kristo P, Canzian F, Hemminki A, Peltomäki P, Chadwick RB, Kääriäinen H, Eskelinen M, Järvinen H, Mecklin JP, de la Chapelle A. 1998. Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. N Engl J Med 338:1481-1487.

Barnetson RA, Tenesa A, Farrington SM, Nicholl ID, Cetnarskyj R, Porteous ME, Campbell H, Dunlop MG. 2006. Identification and survival of carriers of mutations in DNA mismatch-repair genes in colon cancer. N Engl J Med 354:2751-2763.

Bastepe M, Fröhlich LF, Linglart A, Abu-Zahra HS, Tojo K, Ward LM, Jüppner H. 2005. Deletion of the NESP55 differentially methylated region causes loss of maternal GNAS imprints and pseudohypoparathyroidism type Ib. Nat Genet 37:25-27.

Batzer MA, Deininger PL. 2002. Alu repeats and human genomic diversity. Nat Rev Genet 3:370-379.

Chan TL, Yuen ST, Kong CK, Chan YW, Chan AS, Ng WF, Tsui WY, Lo MW, Tam WY, Li VS, Leung SY. 2006. Heritable germline epimutation of MSH2 in a family with hereditary nonpolyposis colorectal cancer. Nat Genet 38:1178-1183.

Chotalia M, Smallwood SA, Ruf N, Dawson C, Lucifero D, Frontera M, James K, Dean W, Kelsey G. 2009. Transcription is required for establishment of germline methylation marks at imprinted genes. Genes Dev 23:105-117.

Gazzoli I, Loda M, Garber J, Syngal S, Kolodner RD. 2002. A hereditary nonpolyposis 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.

Guarinos C, Castillejo A, Barberá VM, Pérez-Carbonell L, Sánchez-Heras AB, Segura A, Guillén-Ponce C, Martínez-Cantó A, Castillejo MI, Egoavil CM, Jover R, Payá A, Alenda C, Soto JL. 2010. EPCAM germ line deletions as causes of Lynch syndrome in Spanish patients. J Mol Diagn. 12:765-770.

Franke G, Bausch B, Hoffmann MM, Cybulla M, Wilhelm C, Kohlhase J, Scherer G, Neumann HP. 2009. Alu-Alu recombination underlies the vast majority of large VHL germline deletions: Molecular characterization and genotype-phenotype correlations in VHL patients. Hum Mutat 30:776-786.

Deleted:

For Peer Review

Hampel H, Frankel WL, Martin E, Arnold M, Khanduja K, Kuebler P, Nakagawa H, Sotamaa K, Prior TW, Westman J, Panescu J, Fix D, Lockman J, Comeras I, de la Chapelle A. 2005. Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). N Engl J Med 352:1851-1860.

Hawkins PG, Santoso S, Adams C, Anest V, Morris KV. 2009. Promoter targeted small RNAs induce long-term transcriptional gene silencing in human cells. Nucleic Acids Res 37:2984-2995.

Hesson LB, Hitchins MP, Ward RL. 2010. Epimutations and cancer predisposition: importance and mechanisms. Curr Opin Genet Dev 20:290-298.

Hitchins MP, Wong JJ, Suthers G, Suter CM, Martin DI, Hawkins NJ, Ward RL. 2007. Inheritance of a cancer-associated MLH1 germ-line epimutation. N Engl J Med 356:697-705.

Kempers MJE, Kuiper RP, Ockeloen CW, Chappuis PO, Hutter P, Rahner N, Schackert HK, Steinke V, Holinski-Feder E, Morak M, Kloor M, Büttner R, Verwiel ETP, van Krieken JH, Nagtegaal ID, Goossens M, van der Post RS, Niessen RC, Sijmons RH, Kluijt I, Hogervorst FBL, Leter EM, Gille JJP, Aalfs CM, Redeker EJW, Hes FJ, Tops CMJ, van Nesselrooij BPM, van Gijn ME, Gómez García EB, Eccles DM, Bunyan DJ, Syngal S, Stoffel EM, Culver JO, Palomares MR, Graham T, Velsher L, Papp J, Oláh E, Chan TL, Leung SY, Geurts van Kessel A, Kiemeney LALM, Hoogerbrugge N, Ligtenberg MJL. 2010. Risk of colorectal and endometrial cancers in *EPCAM* deletion-positive Lynch syndrome: a cohort study. Lancet Oncol 'in press'.

Kovacs ME, Papp J, Szentirmay Z, Otto S, Olah E. 2009. Deletions removing the last exon of TACSTD1 constitute a distinct class of mutations predisposing to Lynch syndrome. Hum Mutat 30:197-203.

Li L, McVety S, Younan R, Liang P, Du Sart D, Gordon PH, Hutter P, Hogervorst FB, Chong G, Foulkes WD. 2006. Distinct patterns of germ-line deletions in MLH1 and MSH2: the implication of Alu repetitive element in the genetic etiology of Lynch syndrome (HNPCC). Hum Mutat 27:388.

Ligtenberg MJL, Kuiper RP, Chan TL, Goossens M, Hebeda KM, Voorendt M, Lee TYH, Bodmer D, Hoenselaar E, Hendriks-Cornelissen SJB, Tsui WY, Kong CK, Brunner HG, Geurts van Kessel A, Yuen ST, van Krieken JH, Leung SY, Hoogerbrugge N. 2009. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. Nat Genet 41:112-117.

For Peer Review

Lynch HT, de la Chapelle A. 2003. Hereditary colorectal cancer. N Engl J Med 348:919-932.

McVety S, Younan R, Li L, Gordon PH, Wong N, Foulkes WD, Chong G. 2005. Novel genomic insertion-- deletion in MLH1: possible mechanistic role for non-homologous end-joining DNA repair. Clin Genet 68:234-238.

Morak M, Schackert HK, Rahner N, Betz B, Ebert M, Walldorf C, Royer-Pokora B, Schulmann K, von Knebel-Doeberitz M, Dietmaier W, Keller G, Kerker B, Leitner G, Holinski-Feder E. 2008. 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.

Nagasaka T, Rhees J, Kloor M, Gebert J, Naomoto Y, Boland CR, Goel A. 2010. Somatic hypermethylation of MSH2 is a frequent event in Lynch Syndrome colorectal cancers. Cancer Res 70:3098-3108.

Niessen RC, Hofstra RM, Westers H, Ligtenberg MJ, Kooi K, Jager PO, de Groote ML, Dijkhuizen T, Olderode-Berends MJ, Hollema H, Kleibeuker JH, Sijmons RH. 2009. Germline hypermethylation of MLH1 and EPCAM deletions are a frequent cause of Lynch syndrome. Genes Chromosomes Cancer 48:737-744.

Nordstrom-O'Brien M, van der Luijt RB, van Rooijen E, van den Ouweland AM, Majoor-Krakauer DF, Lolkema MP, van Brussel A, Voest EE, Giles RH. 2010. Genetic analysis of von Hippel-Lindau disease. Hum Mutat 31:521-537.

Ruivenkamp CA, van Wezel T, Zanon C, Stassen AP, Vlcek C, Csikós T, Klous AM, Tripodis N, Perrakis A, Boerrigter L, Groot PC, Lindeman J, Mooi WJ, Meijjer GA, Scholten G, Dauwerse H, Paces V, van Zandwijk N, van Ommen GJ, Demant P. 2002. Ptprj is a candidate for the mouse colon-cancer susceptibility locus Scc1 and is frequently deleted in human cancers. Nat Genet 31:295-300.

Stankiewicz P, Lupski JR 2010. Structural variation in the human genome and its role in disease. Annu Rev Med 61:437–455

Suter CM, Martin DI, Ward RL. 2004. Germline epimutation of MLH1 in individuals with multiple cancers. Nat Genet 36:497-501.

Tufarelli C, Stanley JA, Garrick D, Sharpe JA, Ayyub H, Wood WG, Higgs DR. 2003. Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. Nat Genet 34:157-165.

For Peer Review

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, Möslein 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, Wisniowiecka-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.

For Peer Review

Figures Legends

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.

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 (*AlusSx*, 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.

For Peer Review

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 Netherlands[1]			
EPCAM deletions[3]	**17**	**6.9%**	**2.8%**
EPCAM founder deletions[3]	**16**	**6.5%**	**2.6%**
MSH2 mutations[4]	**230**		**37.2%**
Explained Lynch families	**618**		
Germany[2]			
EPCAM deletions	**11**	**2.3%**	**1.1%**
MSH2 mutations[4]	**458**		**47.9%**
Explained Lynch families	**957**		

[1]includes 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. [2]includes all unique families that are known by the German HNPCC consortium. [3]all cases known thus far are reported in this study. [4]Including *MSH2* deletions and *EPCAM-MSH2* deletions. [5]The total number of families with MSH2-deficient tumors is composed of families carrying *MSH2* mutations or deletions and *EPCAM* deletions.

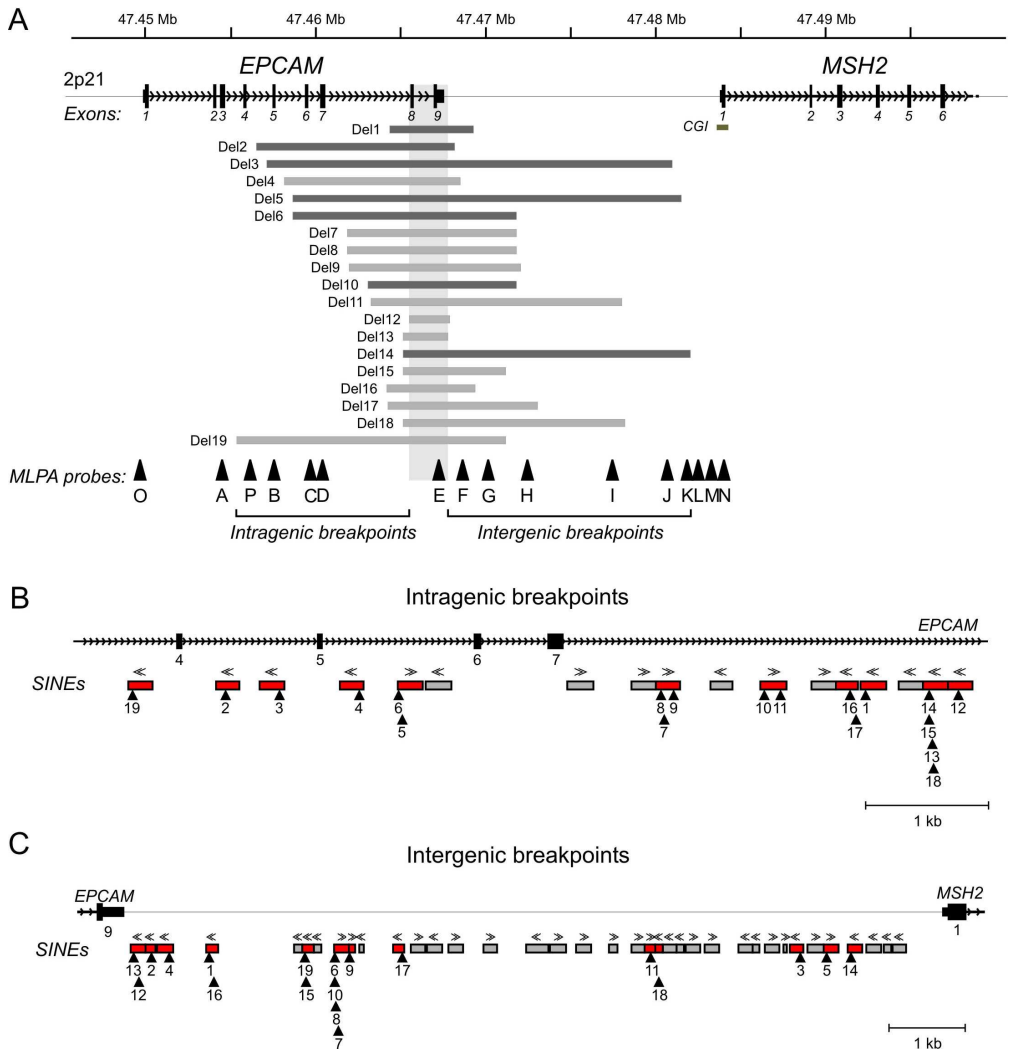
For Peer Review

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

deletion	Country of Origin[1]	number of families	deleted *EPCAM 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	*AluSx*	*AluSq*	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	*AluSg*	*AluSg/x*	78% for 156/198	yes (1)	
3	CH	2	5-9	23,829	c.492-509_*13721del	24 bp	*AluSp*	*AluSg*	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	*AluSx*	*AluSx*	77% for 241/309	NA	Kovacs et al., 2009
5	CN/USA	3	6-9	22,836	c.555+927_*14226del	32 bp	*AluY*	*AluSc*	79% for 237/300	yes (2)	Ligtenberg et al., 2009
6	USA	2	6-9	13,128	c.555+901_*4492del	15 bp	*AluY*	*AluSx*	79% for 225/282	yes (2)	
7	NL	1	8+9	9.963	c.858+1244_*4562del	18 bp	*AluSp*	*AluSx*	85% for 243/284	NA	
8	D	1	8+9	9,963	c.858+1212_*4521del	8 bp	*AluSp*	*AluSx*	82% for 240/291	yes (1)	
9	D	1	8+9	10,074	c.858+1364_*4793del	-	*AluSp*	*FLAM-C Alu*	85% for 243/284	yes (1)	
10	D/H	3	8+9	8,674	c.858+2478_*4507del	14 bp	*AluSp*	*AluSx*	83% for 232/278	NA	Kovacs et al., 2009
11	H	1	8+9	14,734	c.859-2524_*10762del	15 bp	*AluSp*	*AluSp*	86% for 137/159	NA	
12	UK	1	8+9	2,419	c.859-353_*618del	3 bp	*AluSx*	*AluSg*	78% for 222/282	yes (1)	
13	D	1	8+9	2,648	c.859-670_*530del	18 bp	*AluSx*	*AluSg*	78% for 246/312	yes (1)	
14	D[3]	4	8+9	16,834	c.859-689_*14697del	24 bp	*AluSx*	*AluSx*	82% for 246/299	yes(4)	
15	H	1	8+9	6,058	c.859-696_*3914del	19 bp	*AluSx*	*AluJo*	75% for 114/151	NA	Kovacs et al., 2009
16	D	1	8+9	5,246	c.859-1682_*2116del	13 bp	*AluJb*	*AluSq*	78% for 180/229	NA	
17	USA	1	8+9	8,879	c.859-1605_*5862del	10 bp	*AluJb*	*AluSq*	79% for 153/193	yes (1)	
18	USA	1	8+9	13,004	c.859-645_*10911del	14 bp	*AluSx*	*AluSx*	91% for 73/80	NA	Van der Klift et al., 2005
19	D	1	4-9	16,500	c.423-545_*3903del	7 bp	*AluSq*	*AluJo*	80% for 183/227	NA	

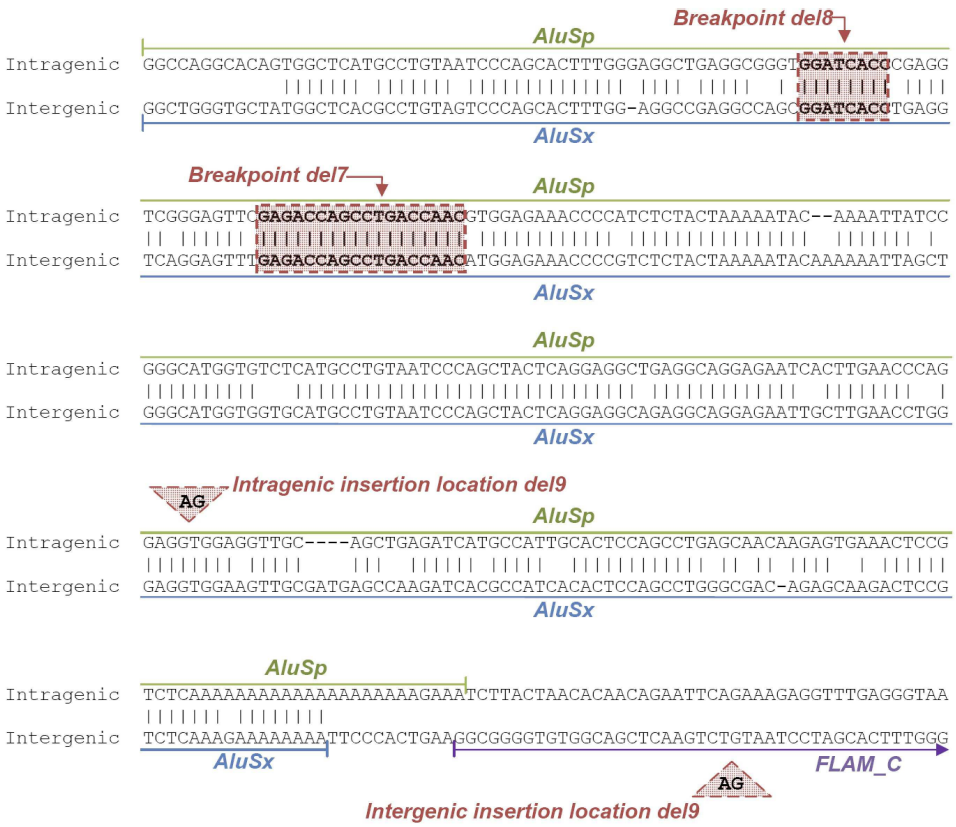
[1]Families originate from The Netherlands (NL), Germany (D), Switzerland, (CH), Hungary (H), China (CN), USA, and UK. [2]Nomenclature is based on mRNA sequence with Genebank Accession Code NM_002354. [3]Includes one family from unknown European origin [van der Klift et al., 2005]. See also Supporting Table 1. [4]Alu-subfamily identified at each each breakpoint [Batzer and Deininger, 2002]. [5]Methylation 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 (AlusSx, 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]

For Peer Review

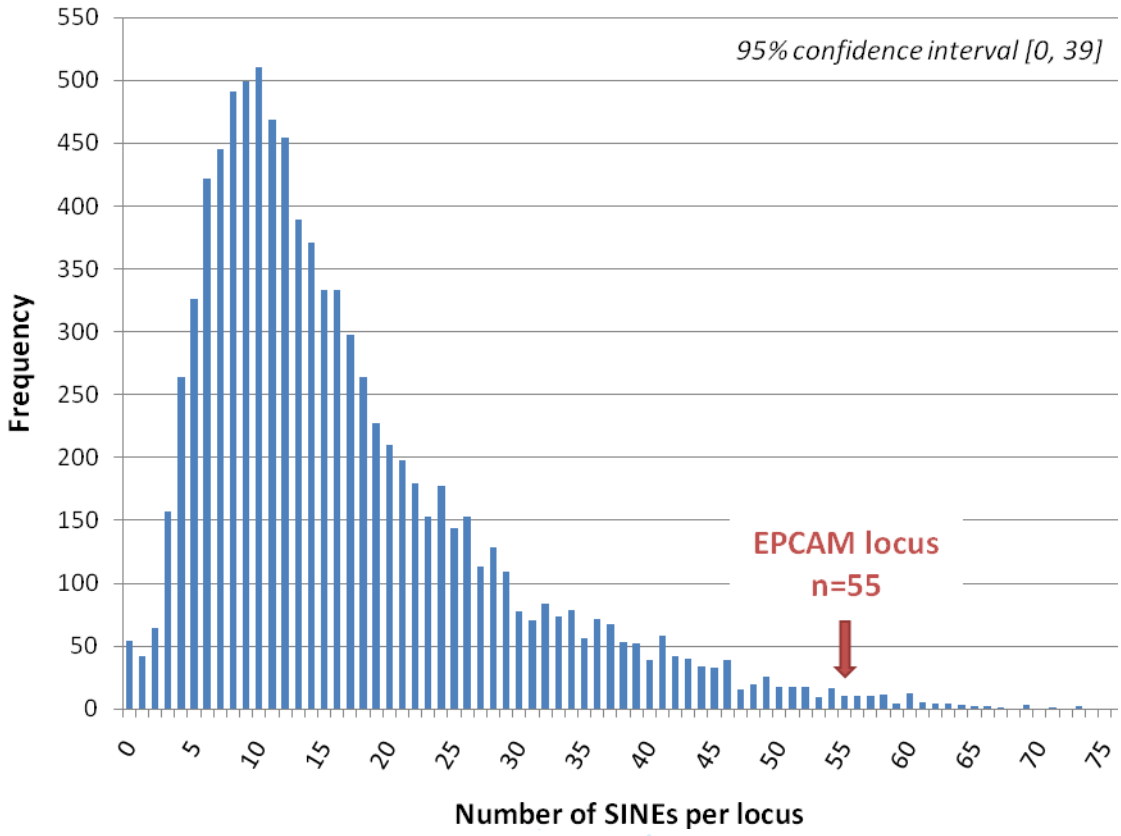


Figure S1. Distribution of short interspersed nuclear elements (SINEs), which include *Alu* repeats, for 10,000 random genomic loci of 25 kb. A total number of 10,000 loci of 25 kb in size were randomly generated from the UCSC genome browser to represent a random sampling of the genome. The size of 25 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.

Preview

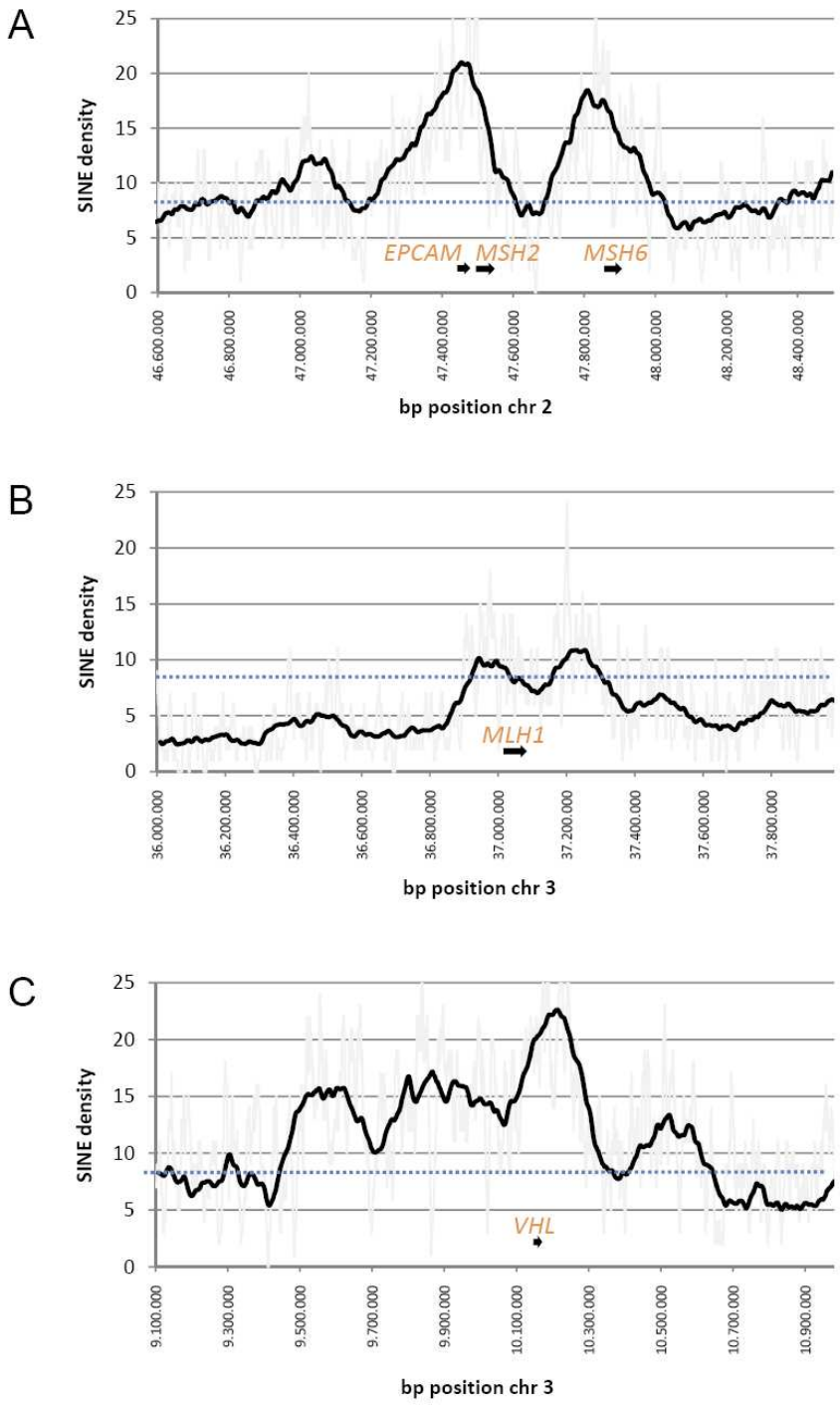


Figure S2. Local genomic SINE density surrounding the *EPCAM*/*MSH2* locus, the *MLH1* locus and the *VHL* locus. Comparison of local SINE *(Alu)* densities of the *EPCAM*/*MSH2* locus **(A)** and the *MLH1* locus **(B)**. The local SINE density of the *VHL* locus **(C)**, another locus frequently targeted 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.

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

Country of origin	Detection method[1]	*EPCAM* 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

[1]detection 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). [2]Includes clinical genetic centers from Nijmegen, Rotterdam, Leiden, Amsterdam (Netherlands Cancer Institute and University of Amsterdam), and Utrecht. [3]Includes families from Nijmegen (*n=4*), and Groningen (*n=3*).

For Peer Review

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	
F05Dld	Germany	Bonn	**8**	p	8-9	9,962	47,461,824	47,471,785	c.858+1211_*4529del	
F09Dld	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	
F02Dld	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	
F11Dld	Germany	Bonn	**13**	p	8-9	2,648	47,465,139	47,467,786	c.859-670_*530del	
F01Dld	Germany	Munich	**14**	p	8-9	16,834	47,465,120	47,481,953	c.859-689_*14697del	
F06Dld	Germany	Bonn	**14**	p	8-9	16,834	47,465,120	47,481,953	c.859-689_*14697del	
F08Dld	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)
F04Dld	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)
F10Dld	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)

For Peer Review