npg

ARTICLE

The *MLH1* c.-27C>A and c.85G>T variants are linked to dominantly inherited *MLH1* epimutation and are borne on a European ancestral haplotype

Chau-To Kwok[1,8], Ingrid P Vogelaar[2,8], Wendy A van Zelst-Stams[2], Arjen R Mensenkamp[2], Marjolijn J Ligtenberg[2,3], Robert W Rapkins[1], Robyn L Ward[1], Nicolette Chun[4], James M Ford[4], Uri Ladabaum[5], Wendy C McKinnon[6], Marc S Greenblatt[6] and Megan P Hitchins*[1,7]

Germline mutations of the DNA mismatch repair genes *MLH1*, *MSH2*, *MSH6* or *PMS2*, and deletions affecting the *EPCAM* gene adjacent to *MSH2*, underlie Lynch syndrome by predisposing to early-onset colorectal, endometrial and other cancers. An alternative but rare cause of Lynch syndrome is constitutional epimutation of *MLH1*, whereby promoter methylation and transcriptional silencing of one allele occurs throughout normal tissues. A dominantly transmitted constitutional *MLH1* epimutation has been linked to an *MLH1* haplotype bearing two single-nucleotide variants, NM_000249.2: c.−27C>A and c.85G>T, in a Caucasian family with Lynch syndrome from Western Australia. Subsequently, a second seemingly unrelated Caucasian Australian case with the same *MLH1* haplotype and concomitant epimutation was reported. We now describe three additional, ostensibly unrelated, cancer-affected families of European heritage with this *MLH1* haplotype in association with constitutional epimutation, bringing the number of index cases reported to five. Array-based genotyping in four of these families revealed shared haplotypes between individual families that extended across ≤2.6–≤6.4 megabase regions of chromosome 3p, indicating common ancestry. A minimal ≤2.6 megabase founder haplotype common to all four families was identified, which encompassed *MLH1* and additional flanking genes and segregated with the *MLH1* epimutation in each family. Our findings indicate that the *MLH1* c.−27C>A and c.85G>T variants are borne on a European ancestral haplotype and provide conclusive evidence for its pathogenicity via a mechanism of epigenetic silencing of *MLH1* within normal tissues. Additional descendants bearing this founder haplotype may exist who are also at high risk of developing Lynch syndrome-related cancers.
European Journal of Human Genetics (2014) 22, 617–624; doi:10.1038/ejhg.2013.200; published online 2 October 2013

Keywords: Lynch syndrome; *MLH1*; epimutation; founder haplotype

INTRODUCTION

Lynch syndrome, characterized by the development of colorectal, endometrial and additional cancers below 50 years of age,[1,2] is caused by dominantly inherited heterozygous mutations within one of the DNA mismatch repair genes, *MLH1*, *MSH2*, *MSH6* or *PMS2*.[3] Germline deletions of the *EPCAM* gene, which give rise to dominantly inherited epimutations of *MSH2*, have been identified in a proportion of cases.[4] An acquired functional loss of the remaining wild-type allele of the affected gene within somatic tissues gives rise to cancers exhibiting microsatellite instability (MSI).[5] A diagnosis of Lynch syndrome is confirmed by the identification of a causative mutation, allowing for accurate genetic counselling of families and targeted clinical surveillance of mutation carriers with a high-risk of cancer development.[2] However, a significant proportion of families with putative Lynch syndrome either have no identifiable mismatch repair mutation, or have

missense 'variants of uncertain significance', which confounds diagnosis and genetic counselling.[3] Classification of such variants as pathogenic or benign is based upon a compilation of clinical, genetic and epidemiological factors, as well as functional analyses.[6] A cluster of sequence variants within the promoter regions of the mismatch repair genes has also been identified in cancer-affected families;[7–10] however, the pathogenic significance of most of these has remained largely uninterpreted.

A subset of mutation-negative patients has been identified with an *MLH1* epimutation, characterized by soma-wide methylation of one allele of the promoter and transcriptional loss from this allele.[11] These cases have tended to be sporadic due to the spontaneous origination of the epimutation in carriers and its subsequent eradication in the germline.[12–15] However, a handful of familial cases with an *MLH1* epimutation have been reported in which transmission of the epimutation between generations has been shown to occur in both

non-Mendelian[14] and autosomal dominant patterns, with the latter linked to localized *cis*-acting genetic anomalies.[16,17]

In a Caucasian family from Western Australia (WA Family 16), dominant transmission of a mosaic *MLH1* epimutation was demonstrated through three successive generations linked to a particular haplotype bearing two single-nucleotide variants (SNVs) in tandem; promoter substitution c.−27C>A and missense variant c.85G>T (p.A29S) (according to coding reference sequence NM_000249.2).[17] The mosaic nature of the epimutation was observed as variable levels of somatic methylation and partial allelic losses of transcription among different tissues and carriers in the family.[17] Two other index cases have also been reported as carriers of the *MLH1* c.−27C>A and c.85G>T SNVs.[7,10] In the first familial case ever reported (Family 1744),[7] no methylation or allelic expression studies were performed and so this allele was not linked to constitutional epimutation at the time. Instead, that study focused on the potential role of the c.85G>T SNV as the disease-causing mutation in this family. However, comprehensive functional assays of the p.A29S protein variant encoded by this SNV showed normal protein activity, suggesting it is neutral.[7] The subsequent link between the c.[−27C>A; 85G>T] haplotype and *MLH1* epimutation in WA Family 16,[17] and thereafter the identification of another index case (Proband H) with this haplotype and concomitant *MLH1* epimutation,[10] strongly suggests this haplotype confers cancer susceptibility through its propensity for soma-wide epigenetic silencing. The concurrence of both SNVs in distinct cases also raises the question of whether they are borne on a founder haplotype.

Through a follow-up study of members from Family 1744, herein designated USA Family 1, as well as two newly identified cancer-affected families carrying the same c.-27C>A and c.85G>T SNVs, we provide definitive evidence that these variants are located on an ancestral haplotype and that it is associated with a mosaic, dominantly heritable form of *MLH1* epimutation.

MATERIALS AND METHODS

Patients and specimens
Patients and their family members were recruited to this study following referrals from Family Cancer Clinics due to a clinical suspicion of Lynch syndrome and the prior identification of heterozygous *MLH1* SNVs c.−27C>A and c.85G>T in a family member(s) during germline screening of the mismatch repair genes for mutations. USA Family 1 was referred from the Vermont Family Cancer Clinic by WCM and MSG. The two siblings provided fresh samples of peripheral blood, but no sample was available from their deceased parents. USA Family 2 was referred from the Stanford Family Clinic by UL and NC. The four siblings and their father each provided fresh samples of saliva, buccal mucosa and hair follicles. A formalin-fixed paraffin embedded (FFPE) block containing a biopsy of normal oesophageal tissue was available for the deceased mother. Netherlands Family 3 was diagnosed at the Radboud University Medical Centre by W.A.vZ-S. DNA samples from peripheral blood were obtained from members of the second and third generations of this family with help from F.H.M. from the VU University Medical Centre after counselling two siblings, but no samples were available from the deceased members of the first generation. Study approval was granted by the Human Research Ethics Committees from the South Eastern Sydney Local Health District, the University of New South Wales, the Radboud University Medical Centre, Stanford University and the University of Vermont. All subjects provided their written consent. Genomic DNA and total RNA were extracted from peripheral blood lymphocytes (PBL) using standard procedures. DNA was extracted from buccal mucosa and hair follicles using the Epicentre BuccalAmp™ DNA extraction kit. Genotek Oragene saliva kits were used to extract DNA and RNA from saliva samples. The Qiagen QIAamp DNA Mini Kit was used to extract genomic DNA from FFPE tissue.

Specific mutation testing
All *MLH1* sequence variants are annotated according to coding reference sequence NM_000249.2. Testing for the *MLH1* c.−27C>A and c.85G>T SNVs was performed by PCR amplification and direct sequencing of a fragment encompassing *MLH1* exon 1 as previously described, irrespective of prior genetic testing of the individuals.[18]

Exons 7, 13 and 14 of the *MUTYH* gene were PCR amplified and sequenced to detect the c.536A>G (p.Y179C), c.1187G>A (p.G396D) and c.1437_1439del (p.E480del) mutations (according to coding reference sequence NM_001128425.1), as previously described.[19]

Methylation analyses
For Netherlands Family 3, initial methylation testing was performed by methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA) using the ME011 MS-MLPA kit (MRC Holland, Amsterdam, The Netherlands). Methylation testing was performed on DNA derived from the CRC, adjacent normal colorectal mucosa and PBL of the proband (F3-II3) and from PBL of other contributing family members.

For USA Families 1 and 2, and for confirmatory testing in four members of Netherlands Family 3, methylation detection was performed by CpG pyrosequencing across five CpG sites, as previously described.[20] First, genomic DNA (1 μg) was sodium bisulphite converted using the EZ methylation Gold kit (Zymo Research, Irvine, CA, USA), and 100–200 ng was used as PCR template. To determine allelic methylation patterns, fragments within the *MLH1* CpG island encompassing either c.−27C>A or both SNVs were amplified using the forward primer (5′–3′) GGTATTGAGGTGATTGGTTGAAGG in combination with either reverse primer ATTCACCACTATCTCRTCCAAC (124 bp product) or CTATACATACCTCTACCCRAACAA (316 bp product), respectively (Figure 2a). Amplification products were cloned using the pGEMTeasy vector system (Promega, Madison, WI, USA) and the plasmid inserts from 24 individual bacterial colonies were sequenced.

Allelic expression analyses
The relative levels of transcripts derived from the two *MLH1* alleles were quantified at the r.85g>u variant site in carriers of the c.[−27C>A; 85G>T] haplotype, or at the r.655a>g site in subjects heterozygous for the benign c.655A>G SNP (both according to coding reference sequence NM_000249.2), using previously described allele quantification (AQ) pyrosequencing assays.[17,21] AQ data from mRNA was normalized to AQ data from parallel assays of genomic DNA to produce an allelic expression ratio: (allele 1_{mRNA}/allele 2_{mRNA})/(allele 1_{DNA}/allele 2_{DNA}).

Haplotyping and copy number variant analysis using the Affymetrix SNP6.0 array
Genomic DNA from five carriers of the c.[−27C>A; 85G>T] haplotype (F1-II2, F2-II1, F2-II3, F3-II2 and F3-II3) was hybridized to Affymetrix SNP6.0 arrays. The data were combined with those from prior Affymetrix SNP6.0 arrays of five carriers from WA Family 16, publicly available from the Gene Expression Omnibus (GEO) databank, accession number GSE30348,[17] representing a total of 10 carriers from four distinct families. Genotypes were generated using Birdseed analysis software implemented in the Affymetrix Genotyping Console version 4.0. Subsequently, regions across the entire genome that were shared by all carriers of the two *MLH1* SNVs were determined by defining the longest stretches of SNP genotypes that showed concordance between one or both alleles. A linkage disequilibrium plot of the chromosome 3p22 region that exhibited haplotype sharing among carriers of the *MLH1* SNVs was generated using the Haploview software.[22] Copy number analysis was performed using the Nexus Copy Number 6.1 software (BioDiscovery, El Segundo, CA, USA). The Affymetrix SNP6.0 array data from the three new families has been deposited in the GEO database (accession number GSE45149).

RESULTS

Identification of distinct cancer-affected families bearing the MLH1 c.−27C>A and c.85G>T single-nucleotide substitutions
Three Caucasian families whose proband had previously undergone selective genetic screening of the mismatch repair genes due to a

clinical suspicion of Lynch syndrome, and in whom the two SNVs c.−27C>A and c.85G>T had been identified within *MLH1*, were included in this study. Pedigrees are shown in Figure 1.

'USA Family 1' (Figure 1a) represents members of the family previously reported as 'Family 1744',[7] who met the Amsterdam I criteria for Lynch syndrome.[1] The female proband (F1-II2) and her older brother (F1-II1), who had both developed CRC exhibiting MSI and dual loss of MLH1 and PMS2 at the ages 36 and 51 years, respectively, were both carriers of the c.−27C>A and c.85G>T SNVs (Figure 1a). Their deceased father had also developed Lynch syndrome-related cancer, but no sample was available to confirm his carrier status.

'USA Family 2' had raised a clinical suspicion of Lynch syndrome (Figure 1b). The male proband (F2-II3) and his three siblings had undergone colonoscopic surveillance due to a positive family history of cancer. Their deceased mother (F2-I2) had developed CRC at the age of 59 years and the family also reported that their maternal grandmother had died of an 'abdominal cancer' at 47 years of age. Colonoscopies conducted on the proband had led to the identification of a 2-mm sessile adenoma at the age of 42 years and a 15 mm flat adenoma with high-grade dysplasia at the age of 44 years. Consistent with an early Lynch syndrome-related neoplastic lesion, molecular pathology testing of his second adenoma showed it had lost MLH1 and PMS2 expression but retained MSH2 and MSH6, and was negative for the *BRAF* NM_004333.4: c.1799T>A (p.V600E) mutation.[23] Based on these findings, targeted germline screening of the *MLH1* gene was performed, which revealed the c.−27C>A and c.85G>T SNVs. Testing of archival normal tissue from his mother showed she had also been a carrier. Clinical surveillance in his three siblings had led to the identification and removal of one adenoma in his older brother at the age of 51 years, two adenomas of <5 mm in his sister at the ages of 41 and 46 years, and two adenomas in his younger brother at 41 years of age. The *MLH1* c.−27C>A and c.85G>T SNVs were found in the older brother (F2-II1), whose adenoma was not examined by immunohistochemistry, but not in the other two siblings (F2-II2 and F2-II4) who had also both developed adenomas. Immunohistochemistry of one adenoma from the youngest brother (F2-II4) showed retention of all four mismatch repair proteins. This raised concerns that an independent genetic mutation may have contributed to the phenotype in this family and so we additionally tested the four siblings and their father for the NM_001128425.1: c.536A>G (p.Y179C), c.1187G>A (p.G396D) and c.1437_1439del (p.E480del) mutations within the *MUTYH* gene, which have been associated with polyposis and an increased risk of CRC particularly among Caucasians.[19] All were negative for these common *MUTYH* mutations.

'Netherlands Family 3' met the Amsterdam I criteria for Lynch syndrome (Figure 1c). The male proband (F3-II3) had presented with colorectal and kidney cancers at the age of 41 years. Molecular pathology testing of his CRC revealed MSI and dual absence of MLH1 and PMS2 expression. Selective germline screening of *MLH1* and *PMS2* for mutations uncovered the *MLH1* c.−27C>A and c.85G>T SNVs and his CRC demonstrated loss-of-heterozygosity (LOH) of the wild-type allele (Figure 2). Targeted screening for the two SNVs in family members identified them in both of his cancer-affected brothers, including a non-identical twin. The twin sisters were not carriers. Although one (F3-II4) had developed melanoma at a young age, this is likely to represent a sporadic cancer unrelated to Lynch syndrome. Both parents, now deceased, had developed CRC, which could have been Lynch syndrome-related in either parent. However, no samples were available to test their carrier status. One family member in the third generation (F3-III1), who was asymptomatic at 19 years of age, also carried the *MLH1* SNVs (Figure 1c).

Absence of MLH1 c.−27C>A and c.85G>T among healthy individuals suggests they are rare variants

In a previous study, we did not find either the *MLH1* c.−27C>A or the c.85G>T SNV in a screen of 304 Australian healthy control subjects.[17] To determine if these variants are represented on a wider population basis, we mined the genetic variation database generated from the 1092 subjects enrolled in the '1000 Genomes Project', which included 379 individuals of European ancestry.[24] The power to detect SNVs present at a frequency of 1% in this study population was
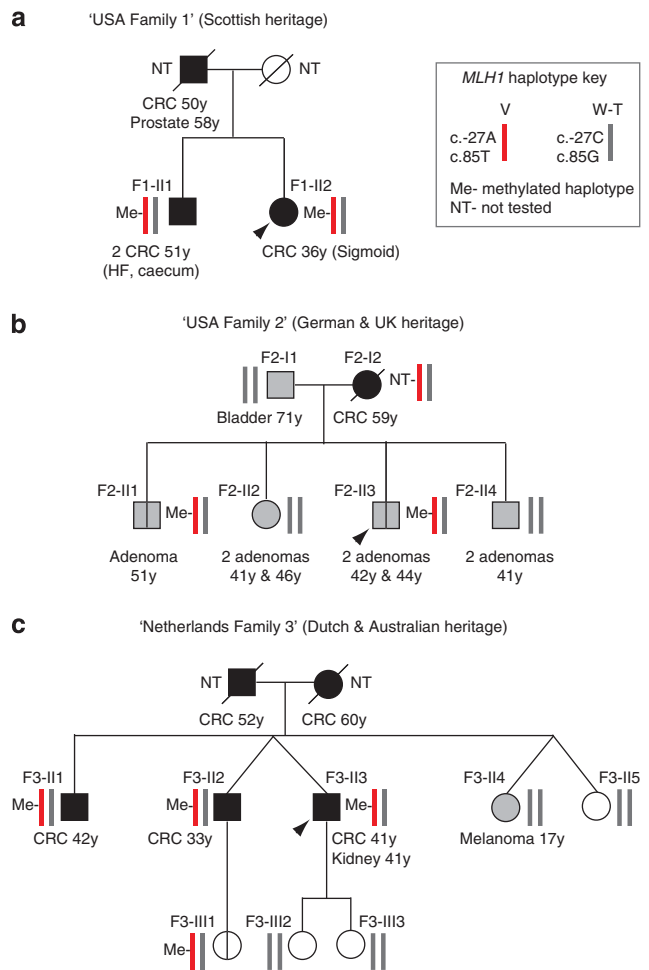
Figure 1 Pedigrees of three families harbouring the *MLH1* c.-27C>A and c.85G>T single-nucleotide variants. Pedigrees are shown for three cancer-affected families in which some members carry the variant (V) haplotype of *MLH1* bearing the c.−27C>A and c.85G>T SNVs (according to reference sequence NM_000249.2) in tandem. (a) 'USA Family 1', (b) 'USA Family 2', (c) 'Netherlands Family 3'. Alleles are shown as vertical lines; red, the variant (V) c.[−27C>A; 85G>T] haplotype identified in heterozygous carriers; grey, wild-type alleles. Segregation of constitutional *MLH1* promoter methylation with the variant haplotype is indicated (Me). NT, not tested for the presence of methylation. Allele types are shown only for those relatives whose carrier status for the c.−27C>A and c.85G>T SNVs could be tested. Circles, females; squares, males; black-filled, affected by a Lynch syndrome-type cancer; grey-filled, either affected by a cancer unlikely to be associated with Lynch syndrome, or diagnosis with a precursor lesion that could be Lynch syndrome-related. The type of neoplastic lesion and age of diagnosis in years (y) is given. HF, hepatic flexure.
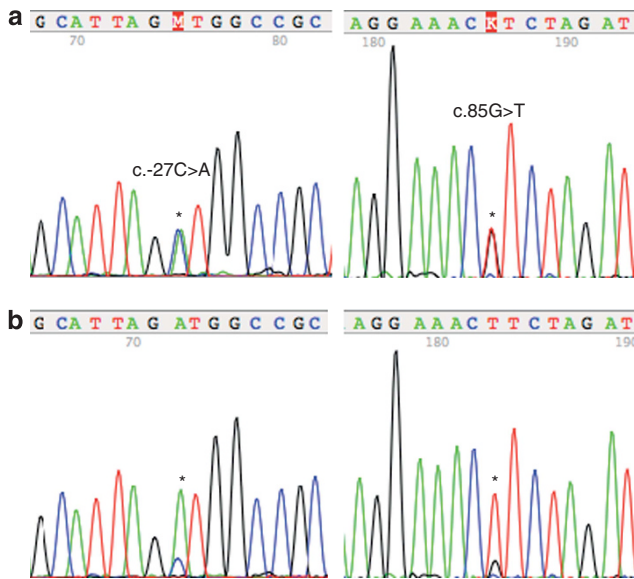
Figure 2 Acquired loss-of-heterozygosity of the normal allele in the colorectal carcinoma of the proband from Netherlands Family 3. Sequence electropherograms across the *MLH1* NM_000249.2: c.−27C>A (left) and c.85G>T (right) variants are shown for proband F3-II3, indicated by *. (**a**) Normal tissue shows heterozygosity for both variants. (**b**) CRC showing a significant reduction in the levels of the wild-type alleles, indicating somatic loss-of-heterozygosity of the normal, functional allele.

Table 1 *MLH1* methylation and relative allelic expression levels in normal tissues from family members and unrelated healthy controls

Subject	c.[−27C> A; 85G>T] carrier status	Methylation levels (%)	AER u:g at r.85g>u	AER a:g at r.655a>g
USA F1		PBL	PBL	
F1-II1 sibling	Carrier	20	0.40	
F1-II2 proband	Carrier	32	0.19	
USA F2		Saliva; hair bulb	Saliva	Saliva
F2-I1 father	Non-carrier	1; 1		
F2-I2 mother	Carrier	NT		
F2-II1 sibling	Carrier	17; 19	0.63	
F2-II2 sibling	Non-carrier	1; 1		
F2-II3 proband	Carrier	29; 30	0.22	
F2-II4 sibling	Non-carrier	1; 0		1.06
Netherlands F3		PBL; NCM; CRC	PBL	
F3-II1 sibling	Carrier	30*		
F3-II2 sibling	Carrier	21*/23		
F3-II3 proband	Carrier	23/30*; 31*; 86*	0.31	
F3-II4 sibling	Non-carrier	<1*		
F3-II5 sibling	Non-carrier	<1*		
F3-III1 niece	Carrier	35*		
F3-III2 daughter	Non-carrier	1/1*		
F3-III3 daughter	Non-carrier	1/1*		
Unrelated		PBL		
Healthy control 1	Non-carrier	0		
Healthy control 2	Non-carrier	1		

Abbreviations: CRC, colorectal cancer; F, family; PBL, peripheral blood lymphocytes; NCM, normal colorectal mucosa; NT, not tested due to lack of sample availability.
The levels of constitutional methylation and allelic expression of *MLH1* are shown in relation to the carrier status of the NM_000249.2: c.[−27C>A; 85G>T] haplotype among the members from three families. Methylation levels shown are the average across five CpG sites assayed by pyrosequencing within the C-region of the *MLH1* promoter, or by MS-MLPA (*). A background signal of 1% methylation was considered negative, since this was also found in healthy controls. CpG pyrosequencing traces are given in Supplementary Figure 2. The high level of methylation detected in the CRC of F3-II3 was consistent with loss-of-heterozygosity of the unmethylated wild-type allele. The allelic expression ratio (AER) shows the normalized relative levels of *MLH1* transcripts derived from the two alleles at expressed sequence variants NM00249.2: r.85g>u or r.655a>g in heterozygous individuals. Values are derived from allele quantification pyrosequencing data shown in Figure 3, and Supplementary Figures 3 and 4. An AER of 1.0 indicates an exactly equal level of expression between the two *MLH1* alleles, whereas an AER significantly below 1.0 for the variant allele at the r.85g>u site indicates a significant loss of expression from the variant c.85G>T allele. The source of normal tissue from which constitutional DNA or RNA was extracted is indicated.

estimated to be at least 99.3%.[24] Neither variant was found, suggesting they occur in <1% of the general population.

The haplotype bearing the c.−27C>A and c.85G>T variants is linked to constitutional methylation and transcriptional repression in all three families

The variant *MLH1* c.−27C>A and c.85G>T alleles were linked as a haplotype in each family (Supplementary Figure 1). Members from the three additional families were then tested for the presence of *MLH1* methylation within normal tissues (n = 15 subjects) and tumour tissue (n = 1 proband) by MS-MLPA and/or CpG pyrosequencing. *MLH1* methylation levels ranged from 17–32% within the normal tissues from all testable carriers of the variant *MLH1* haplotype, whereas all family members with the wild-type sequence were negative for *MLH1* methylation (Table 1; Supplementary Figure 2). Thus, substantial methylation of the *MLH1* promoter in normal somatic tissues segregated with the *MLH1* c.[−27C>A; 85G>T] haplotype in all three families (Figure 1). In Netherlands Family 3, high levels of *MLH1* methylation were first found by MS-MLPA in the CRC of the proband (F3-II3), consistent with the finding of LOH of the wild-type *MLH1* allele, and subsequently in his normal colorectal mucosa and PBL (Table 1). Notably, constitutional *MLH1* methylation was found in two generations of this family, with passage from the proband's twin brother (F3-II2) to his daughter (F3-III1), consistent with dominant transmission of the epimutation linked to this haplotype.

Allelic bisulphite sequencing across the c.−27C>A or both SNVs was performed in each proband and an additional methylation-positive subject from each family (Figure 3a). This showed that methylation was linked specifically to the haplotype bearing the variant c.−27C>A and c.85G>T alleles. Furthermore, the methylation patterns showed some copies of this haplotype were completely

or partially methylated, while others were entirely unmethylated, consistent with a mosaic epimutation (Figure 3b; Supplementary Figure 3).

Quantitative allelic expression analyses were performed at the r.85g>u site in carriers of the c.[−27C>A; 85G>T] haplotype (n = 5) to determine if the epimutation was associated with the loss of expression from the affected allele within normal tissues (Figure 3a). A partial, but significant reduction in the levels of mRNA transcripts derived from the variant allele was observed in each of the carriers tested, ranging from 19 to 63% relative to the wild-type allele (Figure 3c; Table 1; Supplementary Figure 3). By contrast, near equivalent levels of expression from each *MLH1* allele were observed at the r.655a>g site in a member of USA Family 2 (F2-II4) who was informative for the benign c.655A>G SNP but did not carry the variant c.[−27C>A; 85G>T] haplotype and was methylation-negative (Table 1; Supplementary Figure 4).
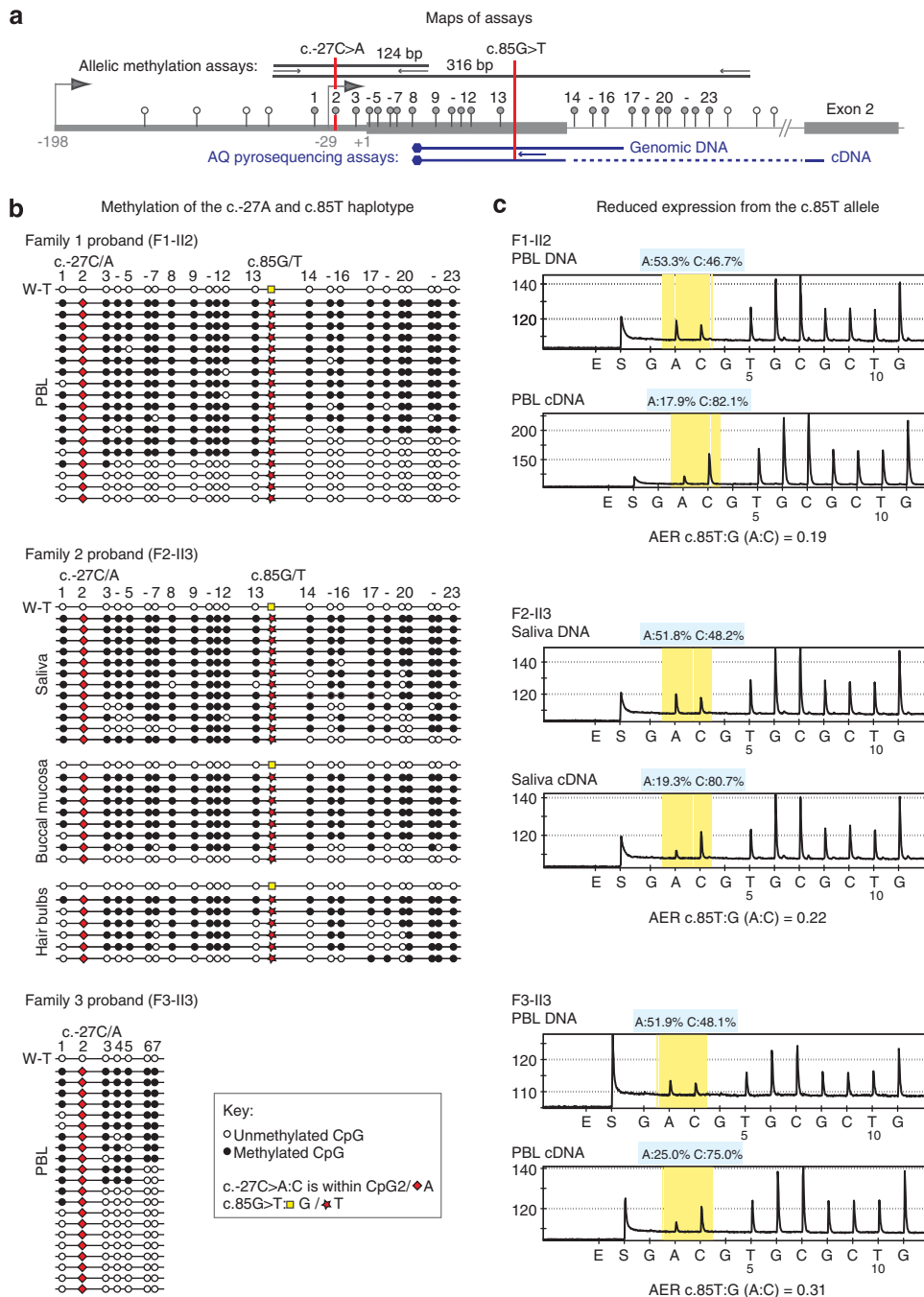
Figure 3 *MLH1* promoter methylation and transcriptional repression are linked to the c.[−27C>A; 85G>T] variant haplotype of *MLH1*. (**a**) Maps of assays used to detect allelic methylation patterns and measure the relative levels of allelic expression at the c.85G>T or r.85g>u sites, respectively. The promoter region of *MLH1* to exon 2 is depicted, with grey rectangles indicating exons as numbered. Narrow rectangles indicate 5′ untranslated sequence and wide rectangles show translated sequence. The major transcription initiation site is indicated by a large arrow. The translation start site is located at +1. Individual CpG dinucleotides are shown as lollipops, with those assessed for the presence of methylation filled in grey. The locations of the *MLH1* NM_000249.2: c.−27C>A and c.85G>T substitutions are indicated by vertical red lines. Horizontal black lines indicate PCR amplification fragments incorporating the c.−27C>A or both SNVs and a number of flanking CpGs for allelic bisulphite sequencing. The allele quantification pyrosequencing assay to determine allelic expression levels, shown in blue, has been previously published.[17] (**b**) Allelic methylation patterns are shown for the proband from each family for each normal somatic tissue type tested. PBL, peripheral blood lymphocytes. Horizontal lines represent single molecules of the DNA fragments. Circles show individual CpG sites, numbered according to the map, with black as methylated and white as unmethylated. The SNVs are indicated as shapes; with variant alleles in red. Only one representative wild-type (W-T) allele is shown for each proband as these were uniformly unmethylated. Methylation was confined to the variant haplotype, but did not affect every single copy of it, indicating methylation mosaicism. (**c**) Pyrogram traces showing the relative levels of each allele at the c.85G>T site in genomic DNA or r.85g>u in mRNA samples. Yellow shading highlights the peaks of the two nucleotides at the c.85g>u site (measured in reverse complement as c>a), from which the relative level of each allele is derived and given as a percentage of the two above. The normalized allelic expression ratios from the variant T allele relative to the W-T G allele are provided. A consistent reduction in expression from the variant allele was observed in each proband.
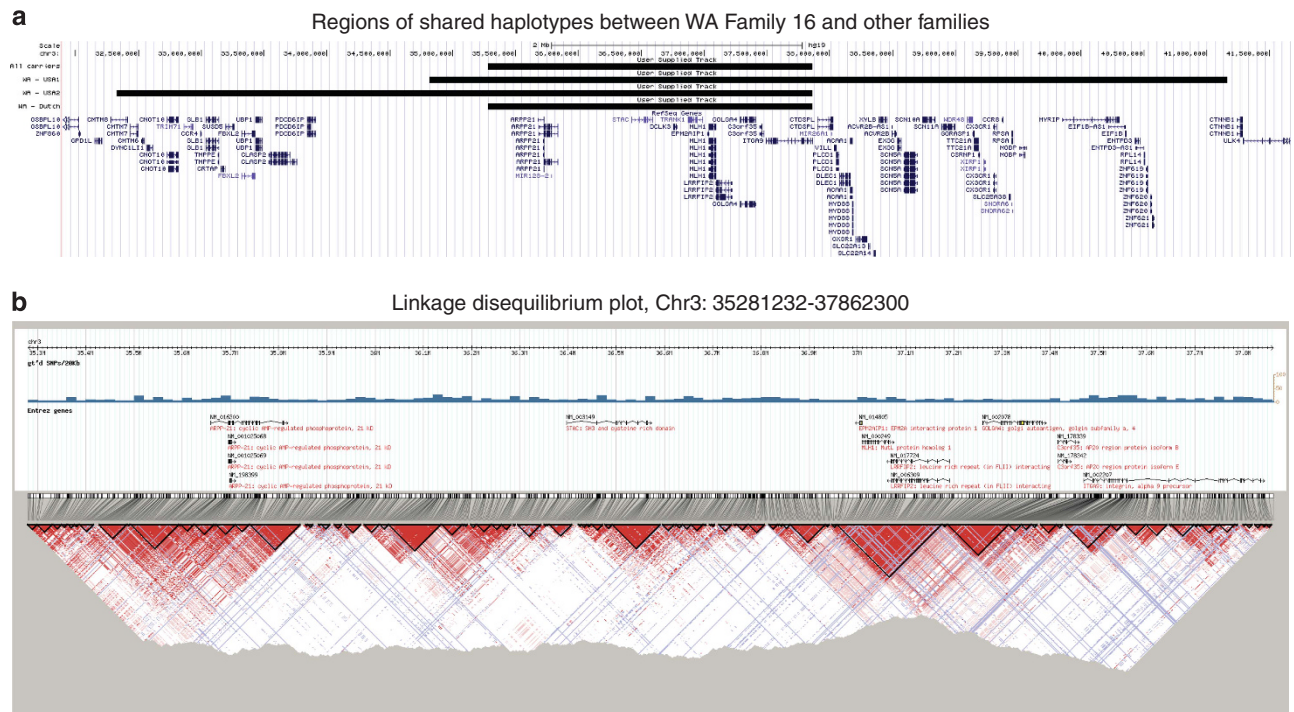
a

Regions of shared haplotypes between WA Family 16 and other families

b

Linkage disequilibrium plot, Chr3: 35281232-37862300

Figure 4 The *MLH1* c.−27C>A and c.85G>T variants are borne on a founder haplotype. (**a**) UCSC Human Genome Browser plot indicating regions of genotype overlap and genes encompassed by the shared haplotype regions. Black bars indicate the region of genotype overlap between carriers of the *MLH1* NM_000249.2: c.[−27C>A; 85G>T] haplotype from four families combined, or pairwise comparisons between carrier members of WA Family 16 and each of the three other families shown individually. Top bar, 10 carriers from all four families share a common haplotype from chromosome 3 position 35281232–37862300 spanning 2,581,068 bp and encompassing *MLH1* plus additional flanking genes from *ARPP21* to *ITGA9*. Second bar, WA Family 16 and USA Family 1 share a larger haplotype across chromosome 3 region 34814101-41164756 spanning 6,350,655 bp. Third bar, WA Family 16 and USA Family 2 share the region 32325273−37862300 encompassing 5,537,027 bp. Bottom bar, WA Family 16 and Netherlands Family 3 share the same ≤2.6 Mb region as all of the carriers in the four families. (**b**) Linkage disequilibrium plot generated using Haploview. The region containing the common ~2.6Mb shared haplotype is comprised of 89 consecutive haplotype blocks of closely linked SNPs.

Genotyping reveals an extensive shared ancestral haplotype across chromosome 3p22 between four families bearing the MLH1 c.-27C>A and c.85G>T SNVs

USA Families 1 and 2 had immigrated to the USA, but were self-reported to be of Scottish descent and mixed German and UK heritage, respectively. Netherlands Family 3 was of mixed Dutch and Australian heritage. Thus these three families and the two previously reported Australian cases were of European origin.[10,17] The occurrence of two distinct SNVs on the same *MLH1* haplotype in association with *MLH1* epimutation in five ostensibly unrelated cancer-affected European families raised the question of whether they share common ancestry. If so, the carriers in each family would be predicted to share a larger founder haplotype on chromosome 3 extending beyond *MLH1* itself.

Genome-wide array-based genotyping performed in 10 carriers of the *MLH1* c.[−27C>A; 85G>T] haplotype from among the three families and WA Family 16[17] revealed shared genotypes among 883 consecutive SNPs across chromosome 3p22. The shared genotypes spanned nearly 2.6 Mb and encompassed the *ARPP21*, *miRNA128-2*, *STAC*, *DCLK3*, *TRANK1*, *EPM2AIP1*, *MLH1*, *LRRFIP2*, *GOLGA4* and *ITGA9* genes (Supplementary Figure 5; Figure 4a). This region stretched across 89 consecutive haplotype blocks (Figure 4b), indicating a very high likelihood that the degree of genotype concordance in these families did not occur by chance. Additional shared haplotypes spanning large genomic regions were found on other chromosomes as well (data not shown). These findings strongly support our hypothesis

that the four families share a common genetic lineage and that the c.-27C>A and c.85G>T *MLH1* variants are indeed borne on a founder haplotype.

Using WA Family 16 as the reference, pair-wise genotype comparisons were made between the carriers from each of the other three families individually, to determine the degree of haplotype sharing between them. This revealed even more extensive regions of genetic overlap spanning 5.5–6.3 Mb between USA Families 1 and 2 and the WA Family 16, whereas the WA and Dutch families shared the same minimal ≤2.6 Mb region of overlap that was common to all four families (Figure 4a).

To confirm the extended ancestral haplotype segregated with the variant *MLH1* c.[−27C>A; 85G>T] haplotype and the epimutation itself, the same 3p22 SNP and STS microsatellite markers previously used to identify the epimutation-associated haplotype in WA Family 16,[17] were genotyped in both carrier and non-carrier members from the three families. Alleles segregating with the presence of constitutional methylation within each pedigree were compiled into haplotypes and compared with the 'reference' haplotype from WA Family 16 (Table 2). The epimutation-associated haplotypes were identical between the four families within the minimally shared 2.6 Mb region, and were also consistent with the distinct larger regions of overlap found between individual families (Table 2). The Dutch family and WA Family 16 were found to share an allelotype at one STS marker (D3S1277) located upstream of the ≤2.6 Mb region they both shared. However, this may simply reflect chance allelotype

Table 2 Chromosome 3p22 SNP and STS haplotypes segregating with the *MLH1* c.-27C>A and c.85G>T variants and epimutation in four families

Location	Marker	WA F16	USA F1	USA F2	Dutch F3
−2,438,039 bp	D3S1277	264	272	**264**	**264**
−609,919 bp	D3S1561	240	**240**	**240**	**240**
TRANK1 3′UTR	rs4789:A>G	G	**G**	**G**	**G**
EPM2AIP1 3′UTR	rs9311149:C>A	A	**A**	**A**	**A**
MLH1 promoter	rs1800734:G>A	G	**G**	**G**	**G**
MLH1 5′UTR	c.−27C>A	A-Me	**A-Me**	**A-Me**	**A-Me**
MLH1 exon 1	c.85G>T	T-Me	**T-Me**	**T-Me**	**T-NT**
MLH1 intron 3	rs4647224:G>A	A	**A**	**A**	NT
MLH1 exon 8	rs1799977:A>G or c.655A>G	G	**G**	**G**	**G**
MLH1 intron 9	rs4647277:A>G	G	**G**	**G**	**G**
LRRFIP2 3′UTR	rs10849:G>A	A	**A**	**A**	**A**
+1,785,165 bp	D3S1100	178	**178**	180	166

The locations of STS markers are shown with respect to the *MLH1* translation initiation site. The genotypes of chromosome 3p22 SNP and STS markers that segregated with the *MLH1* NM000249.2: c.−27C>A and c.85G>T SNVs and *MLH1* methylation within each of four families (F) are tabulated. The epimutation-associated haplotypes shown in each column were compiled from the diplotypes of the two affected siblings in USA Family 1 (F1-II1 and F2-II2), all members of USA Family 2, and two carriers (F3-II2 and F3-II3) and two non-carriers (F3-III2 and F3-III3) of Netherlands Family 3, by independent segregation analyses at each individual marker. Highlighted in bold are the haplotypes shared with the reference haplotype previously reported for WA Family 16.[17] All four families shared a common haplotype spanning at least 700 kb (from D3S1561 to *LRRFIP2* rs10849), located within the 2.6 Mb region identified by array-based genotyping. USA Family 1 shared a larger haplotype with WA Family 16 extending ∼1.8 Mb downstream of *MLH1* to include D3S1100. USA Family 2 and Netherlands Family 3 shared a more extensive haplotype >2.4 Mb upstream of *MLH1* to include D3S1277. The Dutch family and WA Family 16 share an allelotype at D3S1277 located upstream of the ≤2.6 Mb region they both shared. This may simply reflect chance allelotype sharing at this particular marker, or a double-recombination event between this upstream locus and the proximal end of their shared region. NT, not tested.

sharing at this particular marker, or a double-recombination event between this upstream locus and the proximal end of their shared region.

Dominant inheritance of the *MLH1* epimutation with the c.[−27C>A; 85G>T] haplotype strongly implicates a *cis*-acting genetic basis to this epimutation. To investigate whether any large deletions or duplications are present within the minimally shared 2.6 Mb region of chromosome 3p22, copy number variation analysis was performed in the five carriers from the three families. Similar to previous findings in WA Family 16,[17] no copy number changes within this region were consistently detected among the carriers (data not shown).

DISCUSSION

We have shown in three cancer-affected families that the *MLH1* haplotype bearing the c.−27C>A and c.85G>T SNVs is linked to a mosaic form of constitutional epimutation, which manifested as variable levels of promoter methylation accompanied by significant transcriptional loss from this haplotype within normal tissues. This study brings the number of independent cases harbouring this *MLH1* haplotype with a concomitant epimutation to five. Vertical transmission of the epimutation with this haplotype was observed in one of our families (Netherlands Family 3), consistent with the prior finding of its dominant inheritance of in WA Family 16.[17] In the other two families, sample availability from the parental generation was restricted, but segregation of the *MLH1* epimutation with the c.[−27C>A; 85G>T] haplotype was nevertheless clearly demonstrated among siblings.

These families provide collective evidence that the c.−[27C>A; 85G>T] *MLH1* haplotype is associated with a heritable *MLH1* epimutation, which in turn, confers a high risk of development of Lynch syndrome-related cancers. Four of the cases reported to date (USA Family 1, Netherlands Family 3, WA Family 16 and Proband H) met the Amsterdam I criteria due to a significant family history of syndromal cancers that included CRC below 50 years of age in at least one member. Of the cancers tested, both herein and previously,[10,17] MSI and MLH1 loss were consistently observed. Furthermore, the finding of LOH as the 'second hit' in the CRCs from the Dutch and WA probands[17] is consistent with the c.[−27C>A; 85G>T] germline haplotype serving as the 'first hit'. The role of this haplotype in disease causation in USA Family 2 is less clear. Although this family did not meet formal clinical criteria for Lynch syndrome,[2] this was likely precluded by the lack of histopathology testing in the mother's CRC and clinical intervention in the proband and siblings that may have altered their disease course. Nevertheless, absence of MLH1 was observed in the tested large adenoma from the proband, consistent with a precursor lesion with the potential to progress to Lynch-type adenocarcinoma. Furthermore, the cancer-affected mother was also a carrier. Nevertheless, the detection of adenomas in the two siblings who did not carry the variant *MLH1* haplotype is of concern. A second genetic mutation may be present in this family that has contributed to their phenotype. Immunohistochemistry of one adenoma from the younger brother showed normal mismatch repair status and hence we screened for common mutations within the *MUTYH* gene, but did not find them. Genetic studies in this family are ongoing.

The lack of representation of the *MLH1* c.−27C>A and c.85G>T variants among healthy subjects from the general population, most pertinently among Europeans, suggests they are rare and provides further epidemiological support for the pathogenicity of this haplotype. These two nucleotide substitutions most likely comprise 'private' disease-linked variants, as opposed to 'polymorphisms'.

The five families identified with the *MLH1* c.[−27C>A; 85G>T] haplotype are all of European ethnicity. We showed that carriers from four of these families share a large common haplotype that extends well beyond the two SNVs, spanning ≤2.6 Mb of chromosome 3p22 and encompassing the *MLH1* gene neighbourhood. These findings provide strong evidence that these families are descended from a common ancestor. Indeed, paired comparisons between WA Family 16 from and each of the other three families individually showed haplotype sharing across even larger regions of chromosome 3 between the Australian and USA families, suggesting they are more closely related to one-another than the Dutch family.

We confirmed that the extended founder haplotype segregated with the c.−27C>A and c.85G>T SNVs and the *MLH1* epimutation. This finding has two important implications. The first relates to the identification and diagnosis of Lynch syndrome in other carriers. The existence of additional descendants of the ancestor in whom these SNVs arose seems likely and they may be disseminated across the globe through migration. We proffer that the identification of these germline SNVs during standard mutation screening provides a firm molecular diagnosis of Lynch syndrome. In addition, any Lynch syndrome-like cases identified as carriers of a constitutional *MLH1* epimutation should be investigated for these two SNVs, since their presence indicates a 50% risk of transmission of the epimutation to offspring. The second implication is in discerning the mechanism by which this particular genetic haplotype underlies the high-penetrance epigenetic inactivation of *MLH1* throughout the soma. Faithful dominant inheritance of the *MLH1* epimutation with this genetic

haplotype provides clear evidence that its epigenetic manifestations are caused by a genetic defect located on this haplotype. However, the candidate region within which the causative genetic defect is presumably located, is now defined by the ≤ 2.6 Mb minimal region of haplotype overlap between the four families. We found no evidence for large copy number changes within this region; however, this does not rule out more subtle alterations. The $c.-27C>A$ variant remains the prime candidate underlying the *MLH1* epimutation on this haplotype. In support of this, artificial promoter reporter assays have shown this substitution results in significantly diminished transcriptional output.[10,17] However, a clear mechanistic link between this SNV and epigenetic modification of the *MLH1* promoter remains to be demonstrated. Irrespective of the precise molecular mechanism via which the *MLH1* epimutation is induced on this founder haplotype, or whether the $c.-27C>A$ variant is indeed responsible, the $c.-27C>A$ and $c.85G>T$ SNVs can serve as markers for this disease-causing genetic haplotype.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

1 Vasen HF, Mecklin JP, Khan PM *et al*: The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). *Dis Colon Rectum* 1991; **34**: 424–425.
2 Umar A, Boland CR, Terdiman JP *et al*: Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Ntl Cancer Instit* 2004; **96**: 261–268.
3 Peltomaki P, Vasen H: Mutations associated with HNPCC predisposition – update of ICG-HNPCC/INSiGHT mutation database. *Dis Markers* 2004; **20**: 269–276.
4 Ligtenberg MJ, Kuiper RP, Chan TL *et al*: Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. *Nat Genet* 2009; **41**: 112–117.
5 Kolodner RD: Mismatch repair: mechanisms and relationship to cancer susceptibility. *Trends Biochem Sci* 1995; **20**: 397–401.
6 Rasmussen LJ, Heinen CD, Royer-Pokora B *et al*: Pathological assessment of mismatch repair gene variants in Lynch syndrome: past present, and future. *Hum Mutat* 2012; **33**: 1617–1625.
7 Raevaara TE, Korhonen MK, Lohi H *et al*: Functional significance and clinical phenotype of nontruncating mismatch repair variants of MLH1. *Gastroenterology* 2005; **129**: 537–549.
8 Muller-Koch Y, Kopp R, Lohse P *et al*: Sixteen rare sequence variants of the hMLH1 and hMSH2 genes found in a cohort of 254 suspected HNPCC (hereditary non-polyposis colorectal cancer) patients: mutations or polymorphisms? *Eur J Med Res* 2001; **6**: 473–482.
9 Green RC, Green AG, Simms M *et al*: Germline hMLH1 promoter mutation in a Newfoundland HNPCC kindred. *Clin Genet* 2003; **64**: 220–227.
10 Ward RL, Dobbins T, Lindor NM *et al*: Identification of constitutional MLH1 epimutations and promoter variants in colorectal cancer patients from the Colon Cancer Family Registry. *Genet Med* 2013; **15**: 25–35.
11 Gazzoli I, Loda M, Garber J *et al*: 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* 2002; **62**: 3925–3928.
12 Hitchins M, Williams R, Cheong K *et al*: MLH1 germline epimutations as a factor in hereditary nonpolyposis colorectal cancer. *Gastroenterology* 2005; **129**: 1392–1399.
13 Hitchins MP, Ward RL: Erasure of MLH1 methylation in spermatozoa-implications for epigenetic inheritance. *Nat Genet* 2007; **39**: 1289.
14 Hitchins MP, Wong JJ, Suthers G *et al*: Inheritance of a cancer-associated MLH1 germ-line epimutation. *N Engl J Med* 2007; **356**: 697–705.
15 Pineda M, Mur P, Iniesta MD *et al*: MLH1 methylation screening is effective in identifying epimutation carriers. *Eur J Hum Genet* 2012; **20**: 1256–1264.
16 Morak M, Koehler U, Schackert HK *et al*: Biallelic MLH1 SNP cDNA expression or constitutional promoter methylation can hide genomic rearrangements causing Lynch syndrome. *J Med Genet* 2011; **48**: 513–519.
17 Hitchins MP, Rapkins RW, Kwok CT *et al*: Dominantly inherited constitutional epigenetic silencing of MLH1 in a cancer-affected family is linked to a single nucleotide variant within the 5'UTR. *Cancer Cell* 2011; **20**: 200–213.
18 Hitchins M, Owens S, Kwok CT *et al*: Identification of new cases of early-onset colorectal cancer with an MLH1 epimutation in an ethnically diverse South African cohort(dagger). *Clin Genet* 2011; **80**: 428–434.
19 Wang L, Baudhuin LM, Boardman LA *et al*: MYH mutations in patients with attenuated and classic polyposis and with young-onset colorectal cancer without polyps. *Gastroenterology* 2004; **127**: 9–16.
20 Goel A, Nguyen TP, Leung HC *et al*: De novo constitutional MLH1 epimutations confer early-onset colorectal cancer in two new sporadic Lynch syndrome cases, with derivation of the epimutation on the paternal allele in one. *Int J Cancer* 2011; **128**: 869–878.
21 Kwok CT, Ward RL, Hawkins NJ *et al*: Detection of allelic imbalance in MLH1 expression by pyrosequencing serves as a tool for the identification of germline defects in Lynch syndrome. *Fam Cancer* 2010; **9**: 345–356.
22 Barrett JC, Fry B, Maller J *et al*: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.
23 Domingo E, Laiho P, Ollikainen M *et al*: BRAF screening as a low-cost effective strategy for simplifying HNPCC genetic testing. *J Med Genet* 2004; **41**: 664–668.
24 Abecasis GR, Auton A, Brooks LD *et al*: An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; **491**: 56–65.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)