

ORIGINAL ARTICLE

An *MLH1* haplotype is over-represented on chromosomes carrying an HNPCC predisposing mutation in *MLH1*

P Hutter, J Wijnen, C Rey-Berthod, I Thiffault, P Verkuijlen, D Farber, N Hamel, B Bapat, S N Thibodeau, J Burn, J Wu, E MacNamara, K Heinemann, G Chong, W D Foulkes

J Med Genet 2002;**39**:323–327

See end of article for authors' affiliations

Correspondence to:
Dr P Hutter, Unit of Genetics, Institut Central des Hôpitaux Valaisans, Sion, Switzerland; pierre.hutter@ichv.vsn.ch or Dr W D Foulkes, Program in Cancer Genetics, Department of Oncology and Human Genetics, McGill University, Montreal, Quebec, Canada; william.foulkes@mcgill.ca

Revised version received 5 March 2002
Accepted for publication 6 March 2002

Background: The mismatch repair gene, *MLH1*, appears to occur as two main haplotypes at least in white populations. These are referred to as A and G types with reference to the A/G polymorphism at IVS14-19. On the basis of preliminary experimental data, we hypothesised that deviations from the expected frequency of these two haplotypes could exist in carriers of disease associated *MLH1* germline mutations.

Methods: We assembled a series (n=119) of germline *MLH1* mutation carriers in whom phase between the haplotype and the mutation had been conclusively established. Controls, without cancer, were obtained from each contributing centre. Cases and controls were genotyped for the polymorphism in IVS14.

Results: Overall, 66 of 119 *MLH1* mutations occurred on a G haplotype (55.5%), compared with 315 G haplotypes on 804 control chromosomes (39.2%, p=0.001). The odds ratio (OR) of a mutation occurring on a G rather than an A haplotype was 1.93 (95% CI 1.29 to 2.91). When we compared the haplotype frequencies in mutation bearing chromosomes carried by people of different nationalities with those seen in pooled controls, all groups showed a ratio of A/G haplotypes that was skewed towards G, except the Dutch group. On further analysis of the type of each mutation, it was notable that, compared with control frequencies, deletion and substitution mutations were preferentially represented on the G haplotype (p=0.003 and 0.005, respectively).

Conclusion: We have found that disease associated mutations in *MLH1* appear to occur more often on one of only two known ancient haplotypes. The underlying reason for this observation is obscure, but it is tempting to suggest a possible role of either distant regulatory sequences or of chromatin structure influencing access to DNA sequence. Alternatively, differential behaviour of otherwise similar haplotypes should be considered as prime areas for further study.

Hereditary non-polyposis colorectal cancer (HNPCC) is a dominant syndrome that affects about 1 in 1000 people. Patients with HNPCC have a family history of colorectal cancer at an early age, clinically characterised by a predominance of tumours in the proximal colon, a high frequency of synchronous and metachronous colorectal cancers, and an association with a variety of extra-colorectal tumours.^{1,2} Two to 5% of colon cancers are associated with a germline mutation in one of five mismatch repair (MMR) genes (*MLH1*, *MSH2*, *MSH6*, *PMS2*, *MLH3*).³⁻⁸ The MMR system has evolved to correct biosynthetic errors such as nucleotide misincorporations or misalignments during DNA replication. Together *MLH1* and *MSH2* account for at least 60% of all germline mutations found in families in which clinical diagnosis is based on the Amsterdam criteria. According to the database of the HNPCC consortium (<http://www.nfdht.nl/database/mdbchoice.html>), 178 mutations have been reported in *MLH1* and 133 in *MSH2*.⁹ Defects in *MLH1* also account for the majority of sporadic cancers exhibiting a characteristic tumour signature of DNA microsatellite instability, which are observed in a fraction of tumours of the colon, the endometrium, and the stomach.¹⁰⁻¹² Between 8 and 12% of all colon cancers are associated with promoter hypermethylation of both *MLH1* alleles at the somatic level.¹³ In addition to its role in DNA editing, the *MLH1* gene has been shown to participate in mitotic and meiotic recombination,^{14,15} where it plays a role in the correction of heteroduplex¹⁶ and in apoptosis.¹⁷ In a previous study,¹⁸ we reported that two major haplotypes of the *hMLH1* gene appear to segregate in white populations, typically carrying either an A or a G nucleotide at IVS14-19. In particular, we observed that among 151 chromo-

somes stemming from nine European countries that carried a G at IVS14, all 151 had 11 CA dinucleotides at microsatellite D3S1611. In contrast with this, the 192 chromosomes which had an A at IVS14-19 exhibited six sizes of CA repeats, ranging from eight to 17, but never of 11 CAs. Marker D3S1611 is separated from IVS14-19 by 15 013 bp of genomic DNA, most of which corresponds to the large 11 253 bp long intron 13. Therefore, chromosomes carrying A or G at IVS14-19 will hereafter be referred to as A and G haplotypes, respectively. Preliminary results from 19 different *MLH1* germline mutations had suggested that chromosomes carrying the G haplotype may more often harbour an *MLH1* mutation causing HNPCC than chromosomes carrying the A haplotype, although the latter chromosomes are more abundant in the general population. In the present study, we have extended the analysis of the association between the A/G polymorphism at IVS14-19 to a series of 119 HNPCC associated *MLH1* mutations stemming from Europe and North America.

MATERIALS AND METHODS

Polymorphism genotyping

In Switzerland, the A/G single nucleotide polymorphism at IVS14-19 of *MLH1* was genotyped by direct sequence analysis

Abbreviations: HNPCC, hereditary non-polyposis colorectal cancer; MMR, mismatch repair; IVS, intervening sequence; DGGE, denaturing gradient gel electrophoresis; FET, Fisher's exact test; CI, confidence interval; OR, odds ratio; M-H, Mantel-Haenszel

of PCR products from genomic DNA. Genomic DNA was purified using standard phenol-chloroform extraction methods and amplified for 30 cycles in 20 μ l (3–5 pmol of each primer, 50 μ mol/l of each dNTP, and 0.1 U of EXTRA-POL II DNA polymerase from Eurobio, Les Ulis, France). Primers used to genotype the A/G polymorphism were 5' ATTTGTCCCAACTGGTTGTA 3' (forward primer) and 5' TCAGTTGAAATGTCA GAAGTG 3' (reverse primer). The reverse primer was tailed with the M13 universal (-21) sequence, in order to use this primer labelled with IRD800 dye (Lincoln, NE) for sequencing. Cycle sequencing was performed using ThermoSequenase (Nycomed Amersham, Buckinghamshire, UK). Amplicons were usually sequenced in one direction only but occasionally the results were confirmed by bidirectional sequencing. One μ l of reaction product was denatured and electrophoresed on denaturing polyacrylamide gels (obtained by mixing 19 ml of 8% Sequagel XR from National Diagnostics, Atlanta, GA, with 952 μ l of Long Ranger, FMC, Rockland, ME). Samples were separated and analysed on a Li-Cor 4000 automated sequencer.

In Montreal, the A/G polymorphism was typed by direct sequencing using the same primers indicated above and a Cy5.5 labelled M13 primer, but using the Visible Genetics apparatus (Toronto, ON) for sequence analysis. A PCR-RFLP technique was also used as a rapid method of mutation detection for a subset of the controls. For this purpose the forward primer 5'-TCTTCTCATGCTGTCCCT-3' and the reverse primer 5'-ATAATAGAGAAGCTAAGTTAAAC-3' were used. Following amplification, the PCR products were digested with *Mae*III (Roche Diagnostics, Mannheim, Germany). The products were run on 8% polyacrylamide gels. The A allele was indicated by a lower band (53 bp), whereas the G allele was represented by a 181 bp band.

In The Netherlands, genotyping was performed both by denaturing gradient gel electrophoresis (DGGE) of exon 14 and by DNA sequencing as described previously.¹⁹

Statistical analysis

Cases and controls were compared using several statistical methods. First, we compared the frequency of the two haplotypes in the entire set of cases and controls using Fisher's exact test (FET). Secondly, we calculated the odds ratio for the association between carrying a germline *MLH1* mutation and the presence of the G haplotype, with 95% confidence intervals (CI) estimated according to the method of Gart. To ensure that the effect observed was not entirely because of one of the subsets of cases and controls, we used the Mantel-Haenszel (M-H) method of establishing a pooled odds ratio from different sets of data. We also calculated Q, which provides a p value for the probability that the subsets of data analysed are derived from a single large population. A small p value (<0.05) implies that the data should not be combined, as they are not likely to be derived from a single population. Finally, we recalculated the p values for each subgroup when compared with all controls. For all statistical tests, the Arcus Quickstat package was used (Addison Wesley Longman, UK).

Sources of the chromosomes studied

Chromosomes carrying *MLH1* mutations and control chromosomes were ascertained from several geographical areas, as shown in table 1. Ethnicity of the mutation carriers was not precisely established, but our sampling is likely to reflect the population from which the cases were drawn. Overall, *MLH1* mutation carriers from North American centres will probably be from more varied ethnic backgrounds than those from The Netherlands or Switzerland. For this reason, we established control populations from each centre, rather than using one single source of control chromosomes. The country of origin of mutation carriers and controls that were analysed at the three coordinating centres were as follows: Leiden University Medical Centre, 62 cases, The Netherlands (n=23, 37%), USA

Table 1 Haplotype frequencies in cases: ethnicity/country of origin of carriers of mutation bearing chromosomes

Country of residence	Total	A haplotype (%)	G haplotype (%)
The Netherlands	23	14 (60.9)	9 (39.1)
USA	23	11 (47.8)	12 (52.2)
Switzerland	21	7 (33.3)	14 (66.7)
Canada	15	6 (40.0)	9 (60.0)
United Kingdom	12	4 (33.3)	8 (66.7)
Germany	8	3 (37.5)	5 (62.5)
Others	17	8 (47.1)	9 (52.9)
Italy	5		
Ireland	3		
Norway	2		
Portugal	2		
Spain	1		
Denmark	1		
Yugoslavia	1		
Turkey	1		
Australia	1		
Totals	119	53 (44.5)	66 (55.5)

(n=13, 21%), UK (n=9, 15%), Germany (n=7, 11%), Ireland (n=3, 5%), Norway (n=2, 3%) and one each from Portugal, Denmark, Italy, Spain, and Australia, and 306 control chromosomes (all from The Netherlands); Sion and Basel Medical Genetics Centres, 30 cases, Switzerland (n=21, 70%), Italy (n=4, 13%), and one each from Germany, Portugal, Turkey, Yugoslavia, and USA, and 216 control chromosomes (all from Switzerland); McGill University, 27 cases, Montreal (n=11, 41%), Mayo Clinic, Minnesota (n=9, 33%), Mount Sinai Hospital, Toronto (n=4, 15%), and UK (n=3, 11%), and 282 control chromosomes (all from Montreal). Following polymorphism analysis in one of the three centres, we reassigned cases and controls on the basis of country/region of residence or origin, where appropriate (tables 1 and 2). In each centre only cases for which the phase could be established were included. This resulted in some exclusions, as we were unable to determine phase in a further 12 *MLH1* mutation carriers from North America, five cases from Switzerland, and four cases from The Netherlands. Among the 119 germline mutation carriers, 98 were predicted or known to result in a truncated protein (82.4%), whereas 21 (17.6%) were missense mutations that have been shown to be disease causing, either on the basis of segregation, or functional studies, or both. These 21 mutations were as follows: C39R, S44F, Q62K, G67R (observed three times), V77M, T117M, T117R, R265C (twice), N306K, N551T, I565F, K618A (three times), P654L, L676P, and V716M (twice). Mutations that were found more than once were all found to have distinct haplotypes with respect to markers closely linked to *MLH1*, as described by Hutter *et al.*¹⁸ No controls had cancer. In The Netherlands, we included 240 population control chromosomes and 66 chromosomes from spouses of subjects carrying *MLH1* mutations. As stated, for the North American series all the controls came from Montreal. The ethnic distribution of controls was selected to reflect the cases in this subgroup. The breakdown of the origin of the chromosomes from Montreal controls was as follows: 76 Ashkenazim (27%), 68 British (24%), 56 southern European (20%), 50 other European (18%), and 30 French Canadian (11%) (table 2).

RESULTS

Overall, the G allele at IVS14-19 was over-represented on chromosomes carrying a germline *MLH1* mutation, compared with controls (55.2% v 39.2%, p=0.001, tables 1–3). The association can also be expressed as an odds ratio (OR): the G haplotype was 1.93 times more likely to be associated with a

Table 2 Haplotype frequencies in controls: ethnicity/country of origin of control chromosomes

Country of residence	Total	A haplotype (%)	G haplotype (%)
The Netherlands	306	187 (61.1)	119 (38.9)
Switzerland	216	132 (61.1)	84 (38.9)
Canada*			
Ashkenazi Jewish	76	46 (60.5)	30 (39.5)
British	68	47 (69.1)	21 (30.9)
French/French Canadian	30	17 (56.6)	13 (43.3)
Other European†	108	60 (55.6)	48 (44.4)
Canada (total)	282	170 (60.2)	112 (39.7)
Totals	804	489 (60.8)	315 (39.2)

*For Canadian controls, ethnicity/country of origin was defined by birth place of parents or grandparents. Thus we were able to match Canadian controls to US and UK cases, as both these groups were likely to have western European origins. In the case of Ashkenazi Jewish (AJ) controls, birth place was not included. Unless indicated as AJ, Canadian controls did not report AJ origins.

†"Other European" included southern Europeans (n=56), eastern Europeans (n=34), and mixed non-French, non-British European ancestry (n=18).

Table 3 Comparison of haplotype frequencies in cases and controls

Subjects	A haplotype	G haplotype	Total
Controls	489	315	804
Cases	53	66	119
OR (95% CI)			1.93 (1.29 to 2.91)
p value			0.001*

*Two sided p value, FET, by summation.

germline mutation than was the A haplotype (table 3). The magnitude of the effect differed. The effect was as strong in North America as it was in Europe, with the possible, though not statistically significant, exception of The Netherlands. On a country by country basis, the point estimate of the risk was greatest in the Swiss series (OR 3.14, 95% CI 1.32 to 7.87, $p=0.005$) and was smallest in the Dutch group (OR 1.01, 95% CI 0.37 to 2.6, $p=0.98$). The comparison of the frequencies of the A and G haplotypes in all 923 DNA samples analysed is shown in table 3. We then pooled the ORs derived from three groups where we had roughly comparable countries of origins of cases and controls: The Netherlands (23 cases and 306 controls, tables 1 and 2), Switzerland (21 cases and 216 controls), and others (North America, UK, and other areas of continental Europe, 75 cases and 282 controls). The pooled M-H odds ratio for the association of the G haplotype with *MLH1* mutation carriers was 1.89 (95% CI 1.28 to 2.82, $p=0.0019$). A pooled estimate of the odds ratio was used because the three subsets analysed were derived from different geographical locations; furthermore, we wished to exclude the possibility that one subset of cases and controls was unduly influencing the overall result. This result suggests that the observed effect is not restricted to one subgroup and that the play of chance cannot be excluded as an explanation for the differences in the strength of the association in the different subgroups. None is statistically significantly different from each other in the magnitude of the effect observed.

The appropriateness of pooling the data in this way was based on Q calculation, which establishes an estimate of whether it is reasonable to combine the individual ORs. The derived p value reflects acceptance or rejection of the null hypothesis (that the subgroups under study are all derived from one much larger population). This analysis showed that this should not be rejected ($Q=3.19$, 2 df, $p=0.20$). Of relevance to this finding is that the haplotype frequencies for A and G were stable in the control populations, varying over 804

Table 4 Haplotype frequency by mutation type and *MLH1* mutation type frequencies by haplotype

Mutation type	Total	A haplotype (%)	G haplotype (%)	p value
<i>(A) Haplotype frequency by mutation type</i>				
Insertions	12	10 (83.3)	2 (16.7)	–
Deletions	37	13 (35.1)	24 (64.9)	0.006*
Substitutions	70	30 (42.9)	40 (57.1)	0.012*
Total	119	53 (44.5)	66 (55.5)	0.013†
<i>(B) MLH1 mutation type frequencies by haplotype</i>				
Deletion				
1 nucleotide	10	7	3	
2 nucleotides	1	0	1	
3 nucleotides	12	3	9	
4 nucleotides	3	1	2	
5 nucleotides	1	1	0	
7 nucleotides	1	0	1	
15 nucleotides	1	1	0	
Exon or greater	8	0	8	
Total deletions	37	13 (35.1)	24 (64.9)	
Insertion				
1 nucleotide	8	7	1	
2 nucleotide	1	1	0	
4 nucleotide	3	2	1	
Total insertions	12	10 (83.3)	2 (16.7)	
Base substitution	70	30 (42.9)	40 (57.1)	
Total	119	53 (44.5)	66 (55.5)	

*FET: each comparison with insertions as the base line.

†Total $\chi^2=8.72$, $p=0.013$.

The frequency of G haplotypes also varies according to the type of deletion: 1 nucleotide (nt) deletions: 7A, 3G; >1 nt deletions, 6A, 21G, $p=0.017$. All 1 bp insertions or deletions compared with >1 bp insertions/deletions: 14A, 4G; 9A, 22G, $p=0.0013$ (FET).

controls from the three centres from 60.2% to 61.1% for the A haplotype, a difference of only 0.9% (table 2). Moreover, when considering Ashkenazi Jewish controls from Montreal, the frequency of the two haplotypes was 60.5% A and 39.5% G, which further suggests that this haplotype frequency is remarkably stable in different populations in the New and Old World. Some of the differences observed in the smaller subdivisions of the control populations are likely to be random variations based on small sample sizes (table 2).

We also repeated the analyses after excluding all subjects carrying disease associated missense mutations in *MLH1*. This resulted in an OR of 1.75 (95% CI 1.13 to 2.74, $p=0.009$). These findings do not suggest that the association we have identified is limited to any one type of mutation. We also studied the effect of the position of the mutation along the coding sequence of *MLH1* on the frequency of the G haplotype. There was no evidence that the probability of the mutation occurring on the G haplotype varied according to the exon in which the mutation was situated (data not shown). However, we did notice some differences in the distribution of A and G haplotypes on the *MLH1* mutation bearing chromosomes on the basis of the type of mutation (table 4A). The total χ^2 was 8.72, $p=0.013$, implying that the observed distribution of A and G haplotypes in the subgroups is significantly different from that expected. When comparing the distribution of A/G haplotypes among chromosomes bearing deletions, insertions, and substitutions, the direction of the association between G haplotypes and mutation bearing chromosomes is positive for deletions and substitutions (ORs 2.9 and 2.1 respectively) and negative for insertions (OR 0.31), but because of the small sample sizes, these differences are not statistically significant. When compared with controls, deletion (n=37) and substitution mutations (n=70) were preferentially represented on the G haplotype ($p=0.003$ and 0.005, respectively). No significant difference was seen in the distribution of haplotypes for insertions ($p=0.14$), but only 12

MLH1 insertion mutations were analysed. The precise breakdown of the types of mutation by haplotype is shown in table 4B.

DISCUSSION

We have found that the G haplotype is over-represented on germline *MLH1* mutation bearing chromosomes from several populations of European origin that we have analysed in this study (tables 1-3). In addition, we also observed a stronger association between the G haplotype and *MLH1* mutations that were either deletions or substitutions, rather than insertions (table 4). The G haplotype was not over-represented in the Dutch population, and no simple explanation based on this population's history can account for this result. For example, although deletions were over-represented on the G haplotype in the Dutch population (5G, 2A), consistent with the overall findings (table 4A), substitutions were much more frequently seen on an A haplotype than on a G allele (4G, 11A), which is in the opposite direction to the effect shown in table 4A. Only one insertion mutation was observed (on an A haplotype). Interestingly, in the entire data set, only 22% of the 1 bp insertions/deletions were found on the G haplotype, whereas 71% of the larger insertions/deletions were found on this haplotype (table 4B). Even in the Dutch series, where overall there was no association between a germline mutation in *MLH1* and the G haplotype, insertions/deletions of >1 bp were more commonly seen on a G than an A haplotype (5G, 2A). This may support the hypothesis that poorer repair of alterations occurs on the G haplotype, even more so when larger alterations, which are possibly more difficult to correct, are involved. Nevertheless, at this time, we cannot formally exclude the possibility that the differences observed between different groups simply reflect the small size of each group. To resolve this question, further investigation of the frequency of the G haplotype in *MLH1* mutation carriers with different types of mutations is warranted.

With the current data set, we cannot exclude the possibility that the observed association is at least in part the result of incomplete adjustment for hidden population stratification: it will be necessary to study other matched populations of *MLH1* mutation carriers and non-carriers to address this question fully. However, the relative stability of the A/G alleles at *MLH1* IVS14 that we observed, in different control populations from Europe and North America, including Ashkenazim, argues against unrecognised population stratification as a complete explanation for the association. In fact, we consider that this relative stability suggests, as previously discussed,¹⁸ that A and G *MLH1* alleles may not be functionally equivalent.

The *MLH1*/*PMS2* heterodimer is required to assemble individual components of the MMR system at the S phase of the cell cycle, but expression pattern of *MLH1* strongly suggests that this protein is needed in all phases of the cycle.²⁰ Nevertheless, only *MLH1* is relatively stable in monomeric form, which suggests that *PMS2* alone should be degraded in order to avoid interference with the other functions of *MLH1*. For example, *MLH1* protein can be produced alone in a baculovirus system, whereas *PMS2* cannot be produced in the absence of *MLH1* (G Marra, personal communication). Our previous results¹⁸ showed that the G alleles are consistently found on chromosomes that also differ from those carrying A alleles, at least with respect to nearby markers D3S1611 in IVS12 and BAT-21 in IVS11. On G alleles, D3S1611 monomorphic marker always carries 11 CA dinucleotides, whereas for the A allele, at least six sizes ranging from eight to 17 (but never 11) CA are found. BAT-21 marker comprises a run of 11 TA dinucleotides directly followed by a run of 21 T mononucleotides, just 7 nt upstream of the acceptor splice junction of exon 12. This repeat is typically shortened by eight nucleotides on G alleles only.

When considering only MMR, at least four speculative interpretations of our findings could be proposed. First, one or more of the above polymorphisms at BAT-21, D3S1611, and IVS14-19 may not be completely neutral with respect to *MLH1* DNA repair function, and thus result in A/A genotypes being better than G/G genotypes at repairing replication errors occurring during the S phase of the cell cycle. This could be reminiscent of the common polymorphism N372H in the *BRCA2* gene, which confers an increased risk of breast cancer, but clearly has other, unrelated functions.²¹ This possibility was already discussed with respect to BAT-21 variation that may result in abnormal splicing of the large exon 12.¹⁸ Interestingly, four subjects have been identified who are double heterozygotes for mutations in *MLH1* (three in *cis* and one in *trans*) (PH and WDF, unpublished data). All five mutation bearing chromosomes had the G haplotype. Secondly, one or more of the above three polymorphisms may be associated with *MLH1* chromatin structure characteristics that render the *MLH1* G haplotype slightly less accessible to DNA interacting proteins. This could have an effect, for instance, on the interaction with the MSH6 protein which is the actual mismatch recognition protein of the MSH2/MSH6 complex,²² thus altering MMR efficiency. If this were the case, both neutral polymorphisms and pathogenic mutations would be expected to occur more often on G than A haplotypes, and this effect should be restricted to *MLH1*. Thirdly, a slight functional difference in *MLH1* between A and G alleles might be related to a *MLH1* linked factor, located some distance apart from the three neutral polymorphisms, but capable of influencing *MLH1* expression. In mice, distant enhancers of promoters have been identified some 250 kb away from the *Hox* gene complex.²³ In *Drosophila*, long range transcription activators of genes have been found to be mediated by proteins which bind remote enhancer sequences, located several kb from the promoter.²⁴ Between enhancers and promoters, insulator sequences have been identified, which interfere with enhancer-promoter communication. Allelic variants at such insulators can alter the physical proximity between enhancer and promoters, and a promoter on one chromosome can even be activated by an enhancer on the paired homologue.²⁵ If similar factors were involved in A/G *MLH1* variants, one would have to argue that an *MLH1* promoter is enhanced in A/A genotypes relative to G/G genotypes, resulting in a relatively increased DNA repair capacity by the former variant. This hypothetical scenario would mean that the *MLH1* differences in functionality should have genome wide consequences on overall mutation rates. One would also expect some breakdown of the above linkage disequilibrium to occur occasionally between the postulated distant factor and the intragenic *MLH1* marker polymorphisms, as a result of meiotic recombination. This would then be compatible with the incomplete linkage that we observed between the postulated factors in our data sets, possibly reflecting the present distribution of haplotypes generated by historical recombinational events that took place between two major ancient haplotypes. Finally, it is possible that differences in susceptibility to silencing of promoter elements could exist between the two haplotypes, and this in turn could result in differences in the expression levels of the two alleles.

The findings reported here require confirmation in other series of germline *MLH1* mutation carriers. Nevertheless, further work is immediately suggested. These studies might include establishing the frequency of these two ancient haplotypes in other non-European populations and the functional characterisation of A and G haplotypes.

ACKNOWLEDGEMENTS

We are very grateful to clinicians, patients, and their families who contributed to this study. P Hutter was supported by the following institutions: Recherche Suisse Contre le Cancer (AKT446), Fondation

pour la Lutte Contre le Cancer (No 101), Fonds National Suisse de la Recherche Scientifique (No 3138-051088), and Loterie Suisse Romande. W D Foulkes is a Chercheur Boursier Clinician (J2) of the Fonds de la Recherche en Santé du Québec and would like to thank the Judy Steinberg Trust, the Cancer Research Society Inc (Canada), and the Canadian Genetic Diseases Network for support. The subjects studied in Leiden were recruited with the help of world wide members of the CAPP studies (Concerted Action Polyp Prevention), following referral for mutation detection as part of the funding provided by the European Union Biomed 2 programme, Imperial Cancer Research Fund, and the Bayer Corporation.

Authors' affiliations

P Hutter, C Rey-Berthod, Unit of Genetics, Institut Central des Hôpitaux Valaisans, Sion, Switzerland

J Wijnen, P Verkuiljen, Department of Human and Clinical Genetics, Leiden University Medical Centre, Leiden, The Netherlands

I Thiffault, D Farber, E MacNamara, G Chong, W D Foulkes, Departments of Diagnostic Medicine, Medicine and Oncology, Sir M B Davis-Jewish General Hospital, McGill University, Montreal, Quebec, Canada

N Hamel, W D Foulkes, Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada

B Bapat, Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Canada

S N Thibodeau, Molecular Genetics Laboratory, Mayo Clinic, Rochester, Minnesota, USA

J Burn, Institute of Human Genetics, University of Newcastle upon Tyne, UK

J Wu, North West Regional Genetics Laboratory, St Mary's Hospital, Manchester, UK

K Heinemann, Research Group Human Genetics, Division of Medical Genetics, University Clinics, Basel, Switzerland

W D Foulkes, Program in Cancer Genetics, Department of Oncology and Human Genetics, McGill University, Montreal, Quebec, Canada

REFERENCES

- Lynch HT, Smyrk TC, Watson P, Lanspa SJ, Lynch JF, Lynch PM, Cavalieri RJ, Boland CR. Genetics, natural history, tumor spectrum and pathology of hereditary nonpolyposis colorectal cancer: an updated review. *Gastroenterology* 1993;**104**:1535-49.
- Watson P, Lynch HT. Extracolonic cancer in hereditary nonpolyposis colorectal cancer. *Cancer* 1993;**71**:677-85.
- Papadopoulos N, Nicolaides NC, Wei YF, Ruben SM, Carter KC, Rosen CA, Haseltine WA, Fleischmann RD, Fraser CM, Adams MD, Venter JC, Hamilton SR, Peterson GM, Watson P, Lynch HT, Peltomäki P, Mecklin JP, de la Chapelle A, Kinzler KW, Vogelstein B. Mutation of a multihomolog in hereditary colon cancer. *Science* 1994;**263**:1625-9.
- Bronner CE, Baker SM, Morrison PT, Warren G, Smith LG, Lescoe MK, Kane M, Earabino C, Lipford J, Lindblom A, Tannergard P, Bollag RJ, Godwin AR, Ward DC, Nordenskjöld M, Fishel R, Kolodner R, Liskay RM. Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. *Nature* 1994;**368**:258-61.
- Leach FS, Nicolaides NC, Papadopoulos N, Liu B, Jen J, Parsons R, Peltomäki P, Sistonen P, Aaltonen LA, Nyström-Lahti M, Guan XY, Zhang J, Meltzer PS, Yu JW, Kao FT, Chen DJ, Cerosaletti KM, Fournier REK, Todd S, Lewis T, Leach RJ, Naylor SL, Weisenbach J, Mecklin JP, Järvinen H, Petersen GM, Hamilton SR, Green J, Jass J, Watson P, Lynch HT, Trent JM, de la Chapelle A, Kinzler KW, Vogelstein B. Mutations of a multihomolog in hereditary nonpolyposis colorectal cancer. *Cell* 1993;**75**:1215-25.
- Palombo F, Gallinari P, Iaccarino I, Lettieri T, Hughes M, D'Arrigo A, Truong O, Hsuan JJ, Jiricny J. GTBP, a 160-kilodalton protein essential for mismatch-binding activity in human cells. *Science* 1995;**268**:1912-14.
- Nicolaides NC, Papadopoulos N, Liu B, Wei YF, Carter KC, Ruben SM, Rosen CA, Haseltine WA, Fleischmann RD, Fraser CM, Adams MD, Venter JC, Dunlop MG, Hamilton SR, Petersen GM, de la Chapelle A, Vogelstein B, Kinzler KW. Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. *Nature* 1994;**371**:75-80.
- Wu Y, Berends MJ, Sijmons RH, Mensink RG, Verlind E, Kooi KA, van Der Sluis T, Kempinga C, van Der Zee AG, Hollema H, Buys CH, Kleibeuker JH, Hofstra RM. A role for MLH3 in hereditary nonpolyposis colorectal cancer. *Nat Genet* 2001;**29**:137-8.
- ICG-HNPCC database: <http://www.nfdht.nl/database/mbchoice.htm>
- Ionov Y, Peinado MA, Malkhosyan S, Shibata D, Perucho M. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for clonic carcinogenesis. *Nature* 1993;**363**:558-61.
- Thibodeau SN, Bren G, Schaid D. Microsatellite instability in cancer of the proximal colon. *Science* 1993;**260**:816-19.
- Millar AL, Pal T, Madlensky L, Sherman C, Temple L, Mitri A, Cheng H, Marcus V, Gallinger S, Redston M, Bapat B, Narod S. Mismatch repair gene defects contribute to the genetic basis of double primary cancers of the colorectum and endometrium. *Hum Mol Genet* 1999;**8**:823-9.
- Cunningham JM, Kim CY, Christensen ER, Tester DJ, Parc Y, Burgart LJ, Halling KC, McDonnell SK, Schaid DJ, Walsh Vockley C, Kubly V, Nelson H, Michels VV, Thibodeau SN. The frequency of hereditary defective mismatch repair in a prospective series of unselected colorectal carcinomas. *Am J Hum Genet* 2001;**69**:780-90.
- Baker SM, Plug AW, Pralla TA, Bronner CE, Harris AC, Yao X, Christie DM, Monell C, Arnheim N, Bradley A, Ashley T, Liskay RM. Involvement of Mlh1 in DNA mismatch repair and meiotic crossing over. *Nat Genet* 1996;**13**:336-42.
- Anderson LK, Reeves A, Webb LM, Ashley T. Distribution of crossing over on mouse synaptonemal complexes using immunofluorescent localisation of MLH1 protein. *Genetics* 1999;**15**:1569-79.
- Paques F, Haber JE. Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* 1999;**63**:349-404.
- Zhang H, Richards B, Wilson T, Lloyd M, Cranston A, Thorburn A, Fishel R, Meuth M. Apoptosis induced by overexpression of hMSH2 or hMLH1. *Cancer Res* 1999;**59**:3021-7.
- Hutter P, Couturier A, Rey-Berthod C. Two common forms of the human MLH1 gene may be associated with functional differences. *J Med Genet* 2000;**37**:776-81.
- Wijnen J, Khan PM, Vasen H, Menko F, van der Klift H, van den Broek M, van Leeuwen-Cornelisse I, Nagengast F, Meijers-Heijboer EJ, Lindhout D, Griffioen G, Cats A, Kleibeuker J, Varesco L, Bertario L, Bisgaard ML, Mohr J, Kolodner R, Fodde R. Majority of hMLH1 mutations responsible for hereditary nonpolyposis colorectal cancer cluster at the exonic region 15-16. *Am J Hum Genet* 1996;**58**:300-7.
- Klingler H, Hemmerle C, Bannwart F, Haider R, Marra, G. Expression of the hMSH6 mismatch-repair protein in colon cancer and HeLa cells. *Sweiz Med Wochenschr* (in press).
- Healey CS, Dunning AM, Teare MD, Chase D, Parker L, Burn J, Chang-Claude J, Mannermaa A, Kataja V, Huntsman DG, Pharoah PD, Luben RN, Easton DF, Ponder BA. A common variant in BRCA2 is associated with both breast cancer risk and prenatal viability. *Nat Genet* 2000;**26**:362-4.
- Jiricny J. Mediating mismatch repair. *Nat Genet* 2000;**24**:6-8.
- Spitz F, Gonzalez F, Peichel C, Vogt TF, Duboule D, Zakany J. Large scale transgenic and cluster deletion analysis of the HoxD complex separate an ancestral regulatory module from evolutionary innovations. *Genes Dev* 2001;**15**:2209-14.
- Dorsett D. Distant liaisons: long-range enhancer-promoter interactions in *Drosophila*. *Curr Opin Genet Dev* 1999;**9**:505-14.
- Morris JR, Chen JL, Geyer PK, Wu CT. Two modes of transvection: enhancer action in trans and bypass of a chromatin insulator in cis. *Proc Natl Acad Sci USA* 1998;**95**:10740-5.

If you have a burning desire to respond to a paper published in *Journal of Medical Genetics*, why not make use of our "rapid response" option? Log on to our website (www.jmedgenet.com), find the paper that interests you, and send your response via email by clicking on the "eLetters" option in the box at the top right hand corner. Providing it isn't libellous or obscene, it will be posted within seven days. You can retrieve it by clicking on "read eLetters" on our homepage. The editors will decide as before whether to publish it in a future paper issue as well.