Advances in Brief

Structure of the Human *MLH1* Locus and Analysis of a Large Hereditary Nonpolyposis Colorectal Carcinoma Kindred for mlh1 Mutations[1]

Richard D. Kolodner,[2] Nigel R. Hall, James Lipford, Michael F. Kane, Paul T. Morrison, Paul J. Finan, John Burn, Pamela Chapman, Christine Earabino, Elizabeth Merchant, and D. Timothy Bishop

Division of Cell and Molecular Biology [R. D. K., J. L., M. F. K.] and Molecular Biology Core Facility [R. D. K., P. T. M., C. E., E. M.], Dana-Farber Cancer Institute, Boston, Massachusetts 02115; Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115 [R. D. K.]; Imperial Cancer Research Fund, Genetic Epidemiology Laboratory, St. James University Hospital, Leeds LS9 7TF, United Kingdom [N. R. H., D. T. B.]; Department of Surgery and Centre for Digestive Diseases, General Infirmary, Great George Street, Leeds LS1 3EX, United Kingdom [N. R. H., P. J. F.]; and Department of Human Genetics, 19/20 Claremont Place, University of Newcastle upon Tyne, Newcastle NE2 4AA, United Kingdom [J. B., P. C.]

Abstract

Hereditary nonpolyposis colorectal carcinoma is a major cancer susceptibility syndrome known to be caused by inheritance of mutations in at least four genes such as *hMSH2*, *hMLH1*, *hPMS1*, and *hPMS2* which encode components of a DNA mismatch repair system. The *hMLH1* genomic locus on chromosome 3p has been cloned and shown to cover ~58 kilobases of genomic DNA and contain 19 exons. The sequence of all of the intron-exon junctions has been determined and used to develop methods for analyzing each hMLH1 exon for mutations. Using these methods to analyze a 3p-linked hereditary nonpolyposis colorectal carcinoma kindred, we have demonstrated that cancer susceptibility in this family is due to the inheritance of a frame shift mutation in the *hMLH1* gene.

Introduction

HNPCC[3] is one of the most common cancer susceptibility syndromes known in the human population. This syndrome is characterized by dominant transmission and high penetrance, with colorectal carcinoma being the predominant neoplasm and extracolonic cancers representing up to two-fifths of the diagnoses of malignancy (1–4). Recently it has been shown that a significant proportion of HNPCC is due to inherited mutations in DNA mismatch repair genes (5–9). Tumors in HNPCC patients show a particular form of genetic instability, termed microsatellite instability or replication error, characteristic of a mismatch repair defect (10–13). This results in the accumulation of length changes in microsatellite and other short-repeat sequences and possibly also single-base changes due to failure of correction of mistakes that are made when DNA is replicated (14, 15). The implication of this observation is that lack of mismatch repair causes the accumulation of mutations at increased rates in key oncogenes or tumor suppressor genes, *e.g.*, *ras*, *APC*, *p53*, and *DCC*, thus speeding the development of malignancy (6, 7, 10, 16). In addition, the observation of microsatellite instability in a variety of sporadic tumors (17–24) suggests that somatic mutations or low penetrance germline mutations causing mismatch repair defects may be the cause of a significant proportion of many types of sporadic tumors.

Four different DNA mismatch repair genes have been implicated in HNPCC (5–9, 25–28). These include *hMSH2* which maps to chromosome 2p, *hMLH1* which maps to chromosome 3p, *hPMS1* which maps to chromosome 2q, and *hPMS2* which maps to chromosome 7p.

hMSH2 encodes a homologue of the bacterial MutS protein known to recognize mispaired bases in DNA. A combination of linkage and mutational analysis have indicated that *hMSH2* is likely the most prevalent HNPCC gene, accounting for 50 to 60% of HNPCC (25, 29–31). *hMLH1* encodes a homologue of the bacterial MutL and yeast MLH1 proteins and, based on studies of the bacterial and yeast proteins, is likely to interact with the hMSH2 protein (32). Linkage studies suggest that *hMLH1* is the second most prevalent HNPCC gene, possibly accounting for 30% of HNPCC (29, 33). *hPMS1* and *hPMS2* encode homologues of the bacterial MutL and yeast PMS1 proteins and, based on studies of the yeast protein, are also likely to interact with both the hMSH2 and hMLH1 proteins (32). *hPMS1* and *hPMS2* are thought to account for a minor fraction of *HNPCC* and consistent with this, only 1 individual with a germline *pms1* mutation and 1 individual with a germline *pms2* mutation were identified among 40 unrelated individuals with a family history of cancer (8). At present, it is not known if inherited mutations in one of these four genes account for all of HNPCC or if inherited mutations in other DNA repair genes can cause HNPCC.

Much of the present understanding of HNPCC with regard to prevalence, penetrance of HNPCC causing mutations, and associated risk of development of different tumors is based on epidemiological studies which are dependent on imperfect clinical criteria for identifying HNPCC. These types of studies are subject to some uncertainty because of lack of knowledge about the genotype of individual HNPCC family members. With the identification of the genes implicated in the majority of HNPCC, it should be possible to develop a more definitive definition of HNPCC and answer some of the many clinical questions about HNPCC. To facilitate this process we describe here the organization of the *hMLH1* locus and the development of methods for detecting mutations in the *hMLH1* gene. We also report linkage to *hMLH1* in a large HNPCC kindred which has been under investigation for over 20 years. Application of the analytical methods described here has demonstrated that cancer susceptibility in this family is due to the inheritance of an mlh1 frame shift mutation.

Materials and Methods

DNAs, Oligonucleotides, Cloning, and DNA Analysis. All of the methods for DNA isolation, oligonucleotide synthesis, screening P1 libraries, PCR and long range PCR to analyze the size of introns, and other types of DNA analysis including DNA sequencing methods are essentially as previously described and are not described in detail here (26).

DNA Sequence Analysis. Intron-exon junction sequences were determined by cycle sequence analysis of P1 *hMLH1* clones using a SequiTherm cycle sequencing kit (Epicentre Technologies, Madison, WI) followed by comparison of the resulting sequences with the *hMLH1* cDNA sequence (5). All sequences were verified by amplifying each exon along with flanking intron sequences and then sequencing the resulting PCR product on an Applied

Fig. 1. Diagram of the organization of the *MLH1* locus and *MLH1* containing genomic clones. *Boxes containing numbers 1 to 19, individual MLH1 exons. The size of each exon is given below each exon and size of each intron is given above the region between individual pairs of exons. Lines below the gene, each of the individual P1 clones obtained. Each clone is labeled with an identification number and the identification number of each exon contained in the clone. The presence of the indicated exons was determined either by direct sequence analysis or by PCR with the exon-specific primers listed in Table 1 using each clone as template.*

Biosystems 373 DNA sequencer (Foster City, CA). When the PCR products were to be sequenced with *Taq* DNA polymerase and Dye Terminators, they were purified using Qiagen QIAquick-spin PCR purification kits (Qiagen Inc., Chatsworth, CA). PCR products synthesized with one M13 forward sequencing primer labeled primer and one biotinylated primer (Table 1) were captured on magnetic beads (Dynal A. S., Oslo, Norway) and sequenced with Sequenase and M13 forward dye primers using solid-phase sequencing methods (34) on an Applied Biosystems 373 DNA sequencer using protocols and kits obtained from Applied Biosystems.

Genetic Linkage Analysis. DNA from family members was typed for up to eight microsatellite markers in the region of interest. Initially, the markers D3S1029, D3S1007, and D3S1076 (35) and Not73 and D3S1100 (kindly provided by A. Lindblom) were typed. Analysis was performed according to the protocol of Lindblom *et al.* (33) with the following modifications: an annealing temperature of 55°C was used for all markers except D3S1007 where the annealing temperature was 60°C and the extension step at 72°C during each PCR cycle was omitted.

Following the precise localization of *hMLH1*, further analysis was performed with an intragenic marker, D3S1611, and two closely flanking markers, D3S1561 and D3S1298 (9). These three markers were fluorescently labeled and coamplified in a single 20-μl PCR. Each reaction contained approximately 50 ng genomic DNA, 30 pmol D3S1561 primers, 10 pmol D3S1611 and D3S1298 primers, PCR buffer containing 1.5 mM MgCl₂, 200 μM dATP, dCTP, dGTP, and dTTP, and 1–5 units *Taq* DNA polymerase. Twenty-two cycles of amplification were carried out as described above using an annealing temperature of 61°C. For DNA derived from parafin-embedded, formalin-

fixed specimens 31–35 cycles of PCR were used. Samples were separated by denaturing PAGE and analyzed on an Applied Biosystems genescanner using 672 Genescan software. The precise order of the markers used is unclear. The following order, from telomeric to centromeric, is compatible with published maps (36, 37): D3S1007, D3S1561, D3S1611, D3S1298, D3S1029, D3S1100, Not73, and D3S1076.

Linkage analysis was performed using the LINKAGE package (38), assuming that susceptibility was due to a rare dominant gene with a lifetime penetrance for either bowel or endometrial cancer of 0.90. For this analysis, only these two cancers were considered to define the syndrome; all other cancers were ignored. Two-point analyses were conducted between the HNPCC phenotype and each marker in turn, as well as a multipoint analysis involving D3S1561, D3S1611, and D3S1298 and the occurrence of the specific *hMLH1* mutation in each family member. Haplotypes for the eight markers were constructed so as to minimize the number of required recombination events.

Mutational Analysis. Mutational analysis of individual family members was performed by DNA sequence analysis as previously described (6, 26) using a multiplex PCR method to amplify groups of exons and then reamplify individual exons from the multiplex amplification products so that they could be sequenced with a standard sequencing primer (26). The entire *hMLH1* gene from one affected family member was sequenced to detect mutations using template DNA isolated from a blood sample. Then additional family members were analyzed by DNA sequencing for the presence or absence of the detected mutation. Mutational data and haplotypes are reported so as to preserve the confidentiality of the family and prevent analysis of the status of unaffected individuals from the published data.

Results

Structure of the *hMLH1* Genomic Locus. The genomic region encoding the human *MLH1* gene was cloned by screening a P1 library using PCR primers that amplified N-terminal *hMLH1* coding sequence from genomic DNA. The resulting clones were rescreened with PCR primers that either amplified 3′-nontranslated *hMLH1* sequence or *hMLH1* exon 3 from genomic DNA. Four clones were obtained: two that contained the whole gene and two that contained only N-terminal sequences. These clones were characterized by DNA sequencing and PCR to determine the sequence of the intron-exon junctions, the lengths of individual introns, and which portions of the *hMLH1* gene were present in individual clones. Using this analysis,

Fig. 2. Sequence of the intronic region flanking each *MLH1* exon. The nucleotide sequence from the flanking PCR primer site (see Table 1 for primers) up to the intron-exon junction is given in *upper-case letters* in each case except for two places: exon 1 where the sequence is given up to the 5′ end of the mRNA sequence, and exon 19 where only the sequence from the stop codon to the primer site is given. *Underlined*, primer recognition sites for the first-stage primers. To aid in the location of the exons within the *MLH1* cDNA sequence (5), the first 5′,3 nucleotides of each exon and the last 3′,3 nucleotides of each exon are given in *lower-case letters* except for exon 1, where the entire 5′ untranslated mRNA sequence and the first 3 translated nucleotides are given. *Numbers in parentheses* between the intron sequences are the nucleotide coordinates of the exon sequences or cDNA sequences assuming the A of the ATG is nucleotide number 1. Additional intron sequence has been determined in many cases and is available on request.

243

Table 1 hMLH1 *exon amplification primers*[a]

First-stage amplification primer[b]	Second-stage amplification primer

Exon 1

N-18442 5'AGGCACTGAGGTGATTGGC   N-19295 5'TGTAAAACGACGGCCAGTCACTGAGGTGATTGGCTGAA
C-19109 5'TCGTAGCCCTTAAGTGAGC   C-19446 *5'TAGCCCTTAAGTGAGCCCG

Exon 2

N-19689 5'AATATGTACATTAGAGTAGTTG   N-18685 5'TGTAAAACGACGGCCAGTTACATTAGAGTAGTTGCAGA
C-19688 5'CAGAGAAAGGTCCTGACTC   C-19067 *5'AGGTCCTGACTCTTCCATG

Exon 3

N-19687 5'AGAGATTTGGAAAATGAGTAAC   N-18687 5'TGTAAAACGACGGCCAGTTTGGAAAATGAGTAACATGATT
C-19786 5'ACAATGTCATCACAGGAGG   C-19068 *5'TGTCATCACAGGAGGATAT

Exon 4

N-18492 5'AACCTTTCCCTTTGGTGAGG   N-19294 5'TGTAAAACGACGGCCAGTCTTTCCCTTTGGTGAGGTGA
C-18421 5'GATTACTCTGAGACCTAGGC   C-19077 *5'TACTCTGAGACCTAGGCCCA

Exon 5

N-18313 5'GATTTTCTCTTTTCCCCTTGGG   N-19301 5'TGTAAAACGACGGCCAGTTCTCTTTTCCCCTTGGGATTAG
C-18179 5'CAAACAAAGCTTCAACAATTTAC   C-19046 *5'ACAAAGCTTCAACAATTTACTCT

Exon 6

N-18318 5'GGGTTTTATTTTCAAGTACTTCTATG   N-19711 5'TGTAAAACGACGGCCAGTGTTTTATTTTCAAGTACTTCTATGAATT
C-18317 5'GCTCAGCAACTGTTCAATGTATGAGC   C-19079 *5'CAGCAACTGTTCAATGTATGAGCACT

Exon 7

N-19009 5'CTAGTGTGTGTTTTTGGC   N-19293 5'TGTAAAACGACGGCCAGTGTGTGTGTTTTTGGCAAC
C-19135 5'CATAACCTTATCTCCACC   C-19435 *5'AACCTTATCTCCACCAGC

Exon 8

N-18197 5'CTCAGCCATGAGACAATAAATCC   N-19329 5'TGTAAAACGACGGCCAGTAGCCATGAGACAATAAATCCTTG
C-18924 5'GGTTCCCAAATAATGTGATGG   C-19450 *5'TCCCAAATAATGTGATGGAATG

Exon 9

N-18765 5'CAAAAGCTTCAGAATCTC   N-19608 5'TGTAAAACGACGGCCAGTAAGCTTCAGAATCTCTTTT
C-18198 5'CTGTGGGTGTTTCCTGTGAGTGG   C-19449 *5'TGGGTGTTTCCTGTGAGTGGATT

Exon 10

N-18305 5'CATGACTTTGTGTGAATGTACACC   N-19297 5'TGTAAAACGACGGCCAGTACTTTGTGTGAATGTACACCTGTG
C-18306 5'GAGGAGAGCCTGATAGAACATCTG   C-19081 *5'GAGAGCCTGATAGAACATCTGTTG

Exon 11

N-18182 5'GGGCTTTTTCTCCCCCTCCC   N-19486 5'TGTAAAACGACGGCCAGTCTTTTTCTCCCCCTCCCACTA
C-19041 5'AAAATCTGGGCTCTCACG   C-19455 *5'TCTGGGCTCTCACGTCT

Exon 12[c]

N-18579 5'AATTATACCTCATACTAGC   N-20546 *5'CTTATTCTGAGTCTCTCC
C-18178 5'GTTTTATTACAGAATAAAGGAGG   C-20002 5'TGTAAAACGACGGCCAGTGTTTGCTCAGAGGCTGC

    N-19829 *5'GATGGTTCGTACAGATTCCCG
    C-19385 5'TGTAAAACGACGGCCAGTTTATTACAGAATAAAGGAGGTAG

Exon 13

N-18420 5'TGCAACCCACAAAATTTGGC   N-19300 5'TGTAAAACGACGGCCAGTAACCCACAAAATTTGGCTAAG
C-18443 5'CTTTCTCCATTTCCAAAACC   C-19078 *5'TCTCCATTTCCAAAACCTTG

Exon 14

N-19028 5'TGGTGTCTCTAGTTCTGG   N-19456 *5'TGTCTCTAGTTCTGGTGC
C-18897 5'CATTGTTGTAGTAGCTCTGC   C-19472 5'TGTAAAACGACGGCCAGTTGTTGTAGTAGCTCTGCTTG

Exon 15

N-19025 5'CCCATTTGTCCCAACTGG   N-19697 *5'ATTTGTCCCAACTGGTTGTA
C-18575 5'CGGTCAGTTGAAATGTCAG   C-19466 5'TGTAAAACGACGGCCAGTTCAGTTGAAATGTCAGAAGTG

Exon 16

N-18184 5'CATTTGGATGCTCCGTTAAAGC   N-19296 5'TGTAAAACGACGGCCAGTTTGGATGCTCCGTTAAAGCTTG
C-18314 5'CACCCGGCTGGAAATTTTATTTG   C-19047 *5'CCGGCTGGAAATTTTATTTGGAG

Exon 17

N-18429 5'GGAAAGGCACTGGAGAAATGGG   N-19298 5'TGTAAAACGACGGCCAGTAGGCACTGGAGAAATGGGATTTG
C-18315 5'CCCTCCAGCACACATGCATGTACCG   C-19080 *5'TCCAGCACACATGCATGTACCGAAAT

Exon 18

N-18444 *5'TAAGTAGTCTGTGATCTCCG   N-19436 *5'GTAGTCTGTGATCTCCGTTT
C-18581 5'ATGTATGAGGTCCTGTCC   C-19471 5'TGTAAAACGACGGCCAGTTATGAGGTCCTGTCCTAG

Exon 19

N-18638 5'GACACCAGTGTATGTTGG   N-19447 *5'ACCAGTGTATGTTGGGATG
C-18637 5'GAGAAAGAAGAACACATCCC   C-19330 5'TGTAAAACGACGGCCAGTGAAAGAAGAACACATCCCACA

---

[a] All sequence reads 5'-3'. Primer identification numbers are listed before each primer sequence. N, primer on the 5' side of the exon. C, primer on the 3' side of the exon. *, the 5' nucleotide is biotinylated. The sequence 5'-TGTAAAACGACGGCCAGT at the 5' end of the nonbiotinylated nested primers is the sequence of the standard M13-21 forward sequencing primer and is not homologous to *hMLH1*. Some alternate primers have been devised and are available on request.

[b] Exons 1-7, 10, 13,and 16–19 can be specifically amplified in PCR containing either 1.5 or 3 mM MgCl2. Exons 11 and 14 can only be specifically amplified in PCR containing 1.5 mM MgCl2 and exons 8, 9, 12, and 15 can only be specifically amplified in PCR containing 3 mM MgCl2.

[c] Exon 12 does not always amplify well due to the 5' intron sequence. The second-stage amplification primers have been designed so that exon 12 is reamplified in two halves. The 20546 + 20002 primer set amplifies the N-terminal half and this set is somewhat less efficient in PCR because of the unusual sequence in the 5' intron. The primer set 19829 + 19835 amplifies the C-terminal half and is more efficient in PCR. An alternative to primer 20546 is primer 20545 5'-TTTTAATACAGACTTTGC; however, this primer is not ideal as it is complementary to the 5' intron-exon junction and does not allow analysis of these sequences.
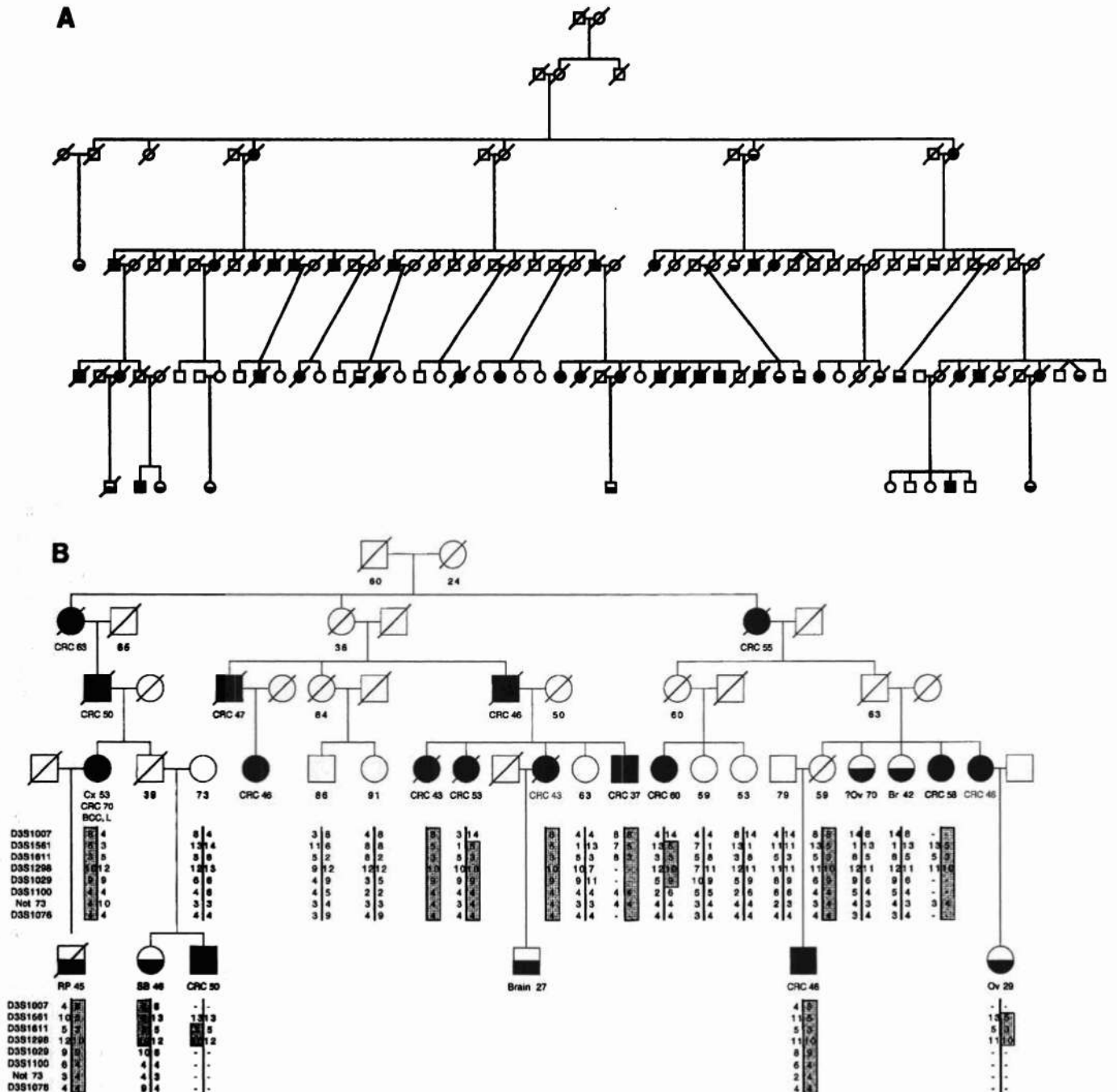
Fig. 3. Pedigree of a HNPCC family demonstrating linkage to 3p. *A*, overview of the HNPCC kindred on which linkage analysis was performed. □, males; ○, females; *oblique line*, deceased; ●, ■, colorectal cancer; *half shading*, cancer at other site. *B*, part of HNPCC kindred illustrating 3p haplotypes for eight microsatellite markers used in linkage analysis. The chromosomal region bearing the disease gene is *shaded* to show its inheritance through the family. *Brackets*, inferred haplotypes; *crosses*, recombinations. *Blanks* or *dashes*, where the allele could not be determined. Haplotypes in some individuals are not shown to protect confidentiality of unaffected family members. □, males; ○, females; *oblique line*, deceased; *full shading*, colorectal cancer; *half shading*, cancer at other site. *Under each symbol*, diagnosis and age at diagnosis if affected or current age/age of death if unaffected. *BCC*, basal cell carcinoma; *Br*, breast cancer; *CRC*, colorectal carcinoma; *Cx*, cervical carcinoma; *L*, lung cancer; *Ov*, ovarian cancer; *RP*, carccinoma of the renal pelvis; *SB*, carcinoma of the small bowel.

we have found that the *hMLH1* gene covers ~58 kilobases (not including the promoter region) and contains 19 exons. The organization of this gene and the genomic clones obtained are summarized in Fig. 1. The sequences of the intron-exon junctions are presented in Fig. 2 and the sequences of primers that are useful for amplifying individual exons are presented in Table 1.

Linkage Analysis in a Large HNPCC Kindred. We have been studying a very large HNPCC family from northeast England. It was one of the earliest colorectal cancer families documented, published

over 20 years ago by Dunstone and Knaggs (39). In their original report there were 36 members with cancer, 26 of whom had colorectal cancer. This pedigree now extends over six generations with at least 53 members having a confirmed diagnosis of cancer. Haplotype analysis using two markers (D2S391 and D2S123) tightly linked to *hMSH2* (7, 30) had previously excluded linkage from this region (40), and so it was a candidate for linkage to the chromosome 3p locus. The segregation of chromosome 3p markers within part of this large pedigree is shown in Fig. 3. Whole or part of a common haplotype

Wild Type

ATTAATGAGCAGGGACATGAGGgtacgtaaacg
I   N   E   Q   G   H   E   intron
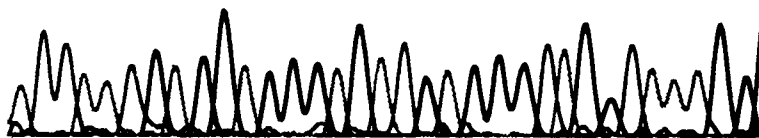
Fig. 4. Analysis of the mutation present in the family illustrated in Fig. 3. A portion of the sequence chromatograms generated by sequencing exon 13 from one unaffected (*top*) and one affected (*bottom*) individual from this family using Sequenase and dye primers. The DNA sequence and predicted protein sequence of the wild-type and mutant alleles deduced from this analysis is illustrated above each sequencing chromatogram. *Underlined*, position that is heterozygous for the extra T due to the mutation.

Mutant
+1
ATTAATGAGCAGGGACATTGAGGgtacgtaaacg
                     GAGGgtacgtaaacg
I   N   E   Q   G   H   *

(Fig. 3, *shaded*) has been inherited by many of the family members sampled, most of whom have developed a malignancy. Linkage results for unaffected members at risk of inheriting the disease gene have been omitted from Fig. 3. Various recombination events narrow the region of interest to a segment bordered by D3S1561 and D3S1298. Thus visual analysis of haplotype segregation confirms linkage to a region of chromosome 3p which includes the *hMLH1* locus. These results are confirmed by formal linkage analysis which gives a maximum lod score of 3.12 for the marker D3S1561.

Analysis in Kindred Members for *hMLH1* Mutations. For primary mutational analysis, all 19 exons of *hMLH1* were sequenced with Sequenase and dye primers using template DNA from one affected individual. Fig. 4 shows sequencing chromatograms generated by sequencing exon 13 from one affected and one unaffected family member. The sequencing chromatogram from the affected family member shows the presence of multiple heterozygous peaks consistent with the presence of a +1 T frame shift mutation in one copy of *hMLH1*. This frame shift mutation creates an in-frame stop codon at codon 519 which is predicted to lead to the synthesis of a truncated MLH1 protein missing the last 238 amino acids of MLH1. It also alters an *Nla*III recognition sequence which may be useful for diagnostic purposes within this family. No other significant sequence change was observed in this individual. Nineteen total members of this family were then examined by DNA sequencing with *Taq* polymerase and dye terminators and nine individuals were found to be heterozygous for this +1 T frame shift mutation. All of the individuals who had this frame shift mutation are predicted to be gene carriers by linkage analysis using markers flanking *hMLH1* and many of them have developed cancer. The remaining members of this family who were examined had the normal sequence and none of these individuals were predicted to be gene carriers by linkage analysis. Linkage analysis within this family using the frame shift mutation, D3S1561 and D3S1611 as markers, and assuming all individuals with colorectal or endometrial cancer are affected gave a multipoint lod score of 5.70.

Discussion

HNPCC is a common cancer predisposition syndrome which appears to be responsible for a small but significant proportion of some of the most prevalent types of cancer in the Western world. The recent demonstration that HNPCC is caused by inherited defects in DNA mismatch repair has made it possible to directly begin to address a number of important questions about HNPCC (5–9, 25–28). These questions include the relative importance of the different HNPCC genes to cancer prevalence, defining the cancer spectrum caused by HNPCC, and determining the relationship between specific mutations and the severity or age of onset of disease. In addition, it should be possible to resolve the issue of the frequency of germline mutations in the mismatch repair genes and determine the relationship between sporadic cancers and HNPCC. Because of the implications of these questions, there is considerable interest in establishing molecular diagnostic methods to address them. In this article we have described the intron-exon structure of the *hMLH1* gene, which is presently thought to be the second most common HNPCC-causing gene (29, 33). This information will make possible DNA-based diagnostic methods for detecting mlh1 mutations using a variety of clinical samples.

In the studies described here, we have used direct DNA sequencing on an Applied Biosystems DNA sequencer to detect an mlh1 frame shift mutation and follow its segregation in an HNPCC kindred. This method of mutational analysis is particularly applicable to situations like that described here where archival samples must be analyzed and RNA samples are not available. The HNPCC kindred we have analyzed (39) is a particularly large kindred which has been studied for a number of years and for which a considerable amount of clinical information has been accumulated. The observation that HNPCC in this family is due to inheritance of an *mlh1* mutation provides an opportunity to begin to address issues of penetrance and expressivity and markers of preclinical disease.

In our initial analysis of expressivity, it is notable that all family members with colorectal cancer have inherited the same *mlh1* mutation. This has been confirmed, if a DNA sample was available, either by direct sequence analysis or by haplotype analysis or in the absence of a sample by showing that a descendent has the mutation and hence inferring its transmission. We have estimated the risk of colorectal cancer in gene carriers to be 35% by age 50 years and 80% by age 70 years.[4] Analysis of the mutation carrier status of family members who have developed extracolonic cancers will in the long-term determine the tumor spectrum of HNPCC. It is evident that one family member with cancer of the renal pelvis and another with cancer of the small bowel are mutation carriers while a woman with premenopausal breast cancer is not. These results would be consistent with the view that transitional cell carcinomas of the upper renal tract and small bowel cancers are an integral part of the HNPCC phenotype, but that breast cancer is not (41). The status of ovarian cancer in the family is a little unclear: one woman who died at 70 of disseminated intraabdominal adenocarcinoma was presumed to have ovarian cancer (despite normal-sized ovaries) because of a very high CA-125 and a normal barium enema, and she had not inherited the disease haplotype (Fig. 3). However, a second woman with histologically proven ovarian cancer does have the disease haplotype, suggesting that her cancer arose as a result of inheriting a susceptibility. Extension of this type of analysis will, in time, allow an accurate elaboration of the sites at increased risk of cancer, the definitive estimate of the age-specific risks of each cancer, and molecular insights into the mechanisms by which cells which are heterozygous for a mutant *hMLH1* allele develop into tumors.

Acknowledgments

References

1. Fitzgibbons, R. J., Lynch, H. T., Stanislav, G. V., Watson, P. A., Lanspa, S. J., Marcus, J. N., Smyrk, T., Kriegler, M. D., and Lynch, J. F. Recognition and treatment of patients with hereditary nonpolyposis colon cancer (Lynch syndromes I and II). Ann. Surg., *206:* 289–295, 1987.

2. Lynch, H. T., Smyrk, T., Watson, P., Lanspa, S. J., Boman, B. M., Lynch, P. M., Lynch, J. F., and Cavalieri, J. Hereditary colorectal cancer. Semin. Oncol., *18:* 337–366, 1991.

3. Lynch, H. T., Smyrk, T. C., Watson, P., Lanspa, S. J., Lynch, J. F., Lynch, P. M., Cavalieri, R. J., and Boland, C. R. Genetics, natural history, tumor spectrum, and pathology of hereditary nonpolyposis colorectal cancer: an updated review. Gastroenterology, *104:* 1535–1549, 1993.

4. Mecklin, J-P., and Jarvinen, H. J. Tumor spectrum in cancer family syndrome (hereditary nonpolyposis colorectal cancer). Cancer (Phila.), *68:* 1109–1112, 1991.

5. Bronner, C. E., Baker, S. M., Morrison, P. T., Warren, G., Smith, L. G., Lescoe, M. K., Kane, M., Earabino, C., Lipford, J., Lindblom, A., Tannergard, P., Bollag, R. J., Godwin, A. R., Ward, D. C., Nordenskjold, M., Fishel, R., Kolodner, R., and Liskay, R. M. Mutation in the DNA mismatch repair gene homologue *hMLH1* is associated with hereditary nonpolyposis colon cancer. Nature (Lond.), *368:* 258–261, 1994.

6. Fishel, R. A., Lescoe, M. K., Rao, M. R. S., Copland, N., Jenkins, N., Garber, J., Kane, M., and Kolodner, R. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell, *75:* 1027–1038, 1993.

7. Leach, F. S., Nicolaides, N. C., Papadopoulos, N., Liu, B., Jen, J., Parsons, R., Peltomaki, P., Sistonen, P., Aaltonen, L. A., Nystrom-Lahti, M., Guan, X-Y., Zhang, J., Meltzer, P. S., Yu, J-W., Kao, F-T., Chen, D. J., Cerosaletti, K. M., Fournier, R. E. K., Todd, S., Lewis, T., Leach, R. J., Naylor, S. L., Weissenbach, J., Mecklin, J-P., Jarvinen, H., Petersen, G. M., Hamilton, S. R., Green, J., Jass, J., Watson, P., Lynch, H. T., Trent, J. M., de la Chapelle, A., Kinsler, K. W., and Vogelstein, B. Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. Cell, *75:* 1215–1225, 1993.

8. Nicolaides, N. C., Papadopoulos, N., Liu, B., Wei, Y., Carter, K. C., Ruben, S. M., Rosen, C. A., Haseltine, W. A., Fleischmann, R. D., Fraser, C. M., Adams, M. D.,

Venter, J. C., Dunlop, M. G., Hamilton, S. R., Petersen, G. M., de la Chapelle, A., Vogelstein, B., and Kinzler, K. Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. Nature (Lond.), *371:* 75–80, 1994.

9. Papadopoulos, N., Nicolaides, N. C., Wei, Y-F., Ruben, S. M., Carter, K. C., Rosen, C. A., Haseltine, W. A., Fleischmann, R. D., Fraser, C. M., Adams, M. D., Venter, J. C., Hamilton, S. R., Petersen, G. M., Watson, P., Lynch, H. T., Peltomaki, P., Mecklin, J-P., de la Chapelle, A., Kinzler, K. W., and Vogelstein, B. Mutation of a *mutL* homolog in hereditary colon cancer. Science (Washington DC), *263:* 1625–1629, 1994.

10. Aaltonen, L. A., Peltomaki, P., Leach, F., Sistonen, P., Pylkkanen, S. M., Mecklin, J.-P., Jarvinen, H., Powell, S., Jen, J., Hamilton, S. R., Petersen, G. M., Kinzler, K. W., Vogelstein, B., and de la Chapelle, A. Clues to the pathogenesis of familial colorectal cancer. Science (Washington DC), *260:* 812–816, 1993.

11. Aaltonen, L. A., Peltomaki, P., Mecklin, J-P., Jarvinen, H., Jass, J. R., Green, J. S., Lynch, H. T., Watson, P., Tallqvist, G., Juhola, M., Sistonen, P., Hamilton, S. R., Kinzzler, K. W., Vogelstein, B., and de la Chapelle, A. Replication errors in benign and malignant tumors from hereditary nonpolyposis colorectal cancer patients. Cancer Res., *54:* 1645–1648, 1994.

12. Honchel, R., Halling, K. C., Schaid, D. J., Pittelkow, M., and Thibodeau, S. N. Microsatellite instability in Muir-Torre syndrome. Cancer Res., *54:* 1159–1163, 1994.

13. Risinger, J. I., Berchuck, A., Kohler, M. F., Watson, P., Lynch, H. T., and Boyd, J. Genetic instability of microsatellites in endometrial carcinoma. Cancer Res., *53:* 5100–5103, 1993.

14. Strand, M., Prolla, T. A., Liskay, R. M., and Petes, T. D. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature (Lond.), *365:* 274–276, 1993.

15. Levinson, G., and Gutman, G. A. High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. Nucleic Acids Res., *15:* 5313–5338, 1987.

16. Jiricny, J. Colon cancer and DNA repair: have mismatches met their match? Trends Genet., *10:* 164–168, 1994.

17. Chong, J., Fukayama, M., Hayashi, Y., Takizawa, T., Koike, M., Konishi, M., Kikuchi-Yanoshita, R., and Miyaki, M. Microsatellite instability in the progression of gastric carcinoma. Cancer Res., *54:* 4595–4597, 1994.

18. Gonzalez-Zulueta, M., Ruppert, J. M., Tokino, K., Tsai, Y. C., Spruck, I. I. I., C. H., Miyao, N., Nichols, P. W., Hermann, G. G., Horn, T., Steven, K., Summerhayes, I. C., Sidransky, D., and Jones, P. A. Microsatellite instability in bladder cancer. Cancer Res., *53:* 5620–5623, 1993.

19. Ionov, Y., Peinado, M. A., Malkbosyan, S., Shibata, D., and Perucho, M. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature (Lond.), *260:* 558–561, 1993.

20. Peltomaki, P., Lothe, R. A., Aaltonen, L. A., Pylkkanen, L., Nystrom-Lahti, M., Seruca, R., David, L., Holm, R., Ryberg, D., Haugen, A., Brogger, A., Borresen, A-L., and de la Chapelle, A. Microsatellite instability is associated with tumours that characterize the hereditary non-polyposis colorectal carcinoma syndrome. Cancer Res., *53:* 5853–5855, 1993.

21. Shibata, D., Peinado, M. A., Ionov, S., Malkhosyan, S., and Perucho, M. Genomic instability in repeated sequences in an early somatic event in colorectal tumorigenesis that persists after transformation. Nature Genet., *6:* 273–281, 1994.

22. Shridhar, V., Siegfried, J., Hunt, J., del Mar Alonso, M., and Smith, D. I. Genetic instability of microsatellite sequences in many non-small cell lung carcinomas. Cancer Res., *54:* 2084–2087, 1994.

23. Thibodeau, S. N., Bren, G., and Schaid, D. Microsatellite instability in cancer of the proximal colon. Science (Washington DC), *260:* 816–819, 1993.

24. Wada, C., Shionoya, S., Fyjino, Y., Tokuhiro, H., Akahoshi, T., Uchida, T., and Ohtani, H. Genomic instability of microsatellite repeats and its association with the evolution of chronic myelogenous leukemia. Blood, *83:* 3449–3456, 1994.

25. Liu, B., Parsons, R. E., Hamilton, S. R., Petersen, G. M., Lynch, H. T., Watson, P., Markowitz, S., Willson, J. K. V., Green, J., de la Chapelle, A., Kinzler, K. W., and Vogelstein, B. *hMSH2* mutations in hereditary nonpolyposis colorectal cancer kindreds. Cancer Res., *54:* 4590–4594, 1994.

26. Kolodner, R. D., Hall, N. R., Lipford, J., Kane, M. F., Rao, M. R. S., Morrison, P., Wirth, L., Finan, P. J., Burn, J., Chapman, P., Earabino, C., Merchant, E., and Bishop, D. T. Structure of the human *MSH2* locus and analysis of two Muir-Torre kindreds for *msh2* mutations. Genomics, *24:* 516–526, 1994.

27. Mary, J-L., Bishop, D. T., Kolodner, R., Lipford, J. R., Kane, M., Torhorst, J., Muller, H., Spycher, M., and Scott, R. J. Mutational analysis of the *hMSH2* gene reveals a three base pair deletion in a family predisposed to colorectal cancer development. Hum. Mol. Genet., *3:* 2067–2069, 1994.

28. Lynch, H. T., Drouhard, T., Lanspa, S., Smyrk, T., Lynch, P., Lynch, J., Vogelstein, B., Nystrom, L. M., Sistonen, P., Peltomaki, P., and de la Chapelle, A. Mutation of an mutL homologue in a Navajo family with hereditary nonpolyposis colorectal cancer. J. Natl. Cancer Inst., *86:* 1417–9, 1994.

29. Nystrom-Lahti, M., Parsons, R., Sistonen, P., Pylkkanen, L., Aaltonen, L. A., Leach, F. S., Hamilton, S. R., Watson, P., Bronson, E., Fusaro, R., Cavalieri, J., Lynch, J., Lanspa, S., Smyrk, T., Lynch, P., Drouhard, T., Kinzler, K. W., Vogelstein, B., Lynch, H. T., de la Chapelle, A., and Peltomaki, P. Mismatch repair genes on chromosomes 2p and 3p account for a major share of hereditary nonpolyposis colorectal cancer families evaluable by linkage. Am. J. Hum. Genet., *55:* 659–665, 1994.

30. Peltomaki, P., Aaltonen, L. A., Sistonen, P., Pylkkanen, L., Mecklin, J.-P., Jarvinen, H., Green, J. S., Jass, J. R., Weber, J. L., Leach, F. S., Petersen, G. M., Hamilton, S. R., de la Chapelle, A., and Vogelstein, B. Genetic mapping of a locus presisposing to human colorectal cancer. Science (Washington DC), *260:* 810–812, 1993.

31. Sandkuijl, L. A., and Bishop, D. T. Results of the joint analysis of the EUROFAP

---

[4] N. R. Hall and D. T. Bishop, unpublished data.

linkage data: summary. *In:* J. Mohr (ed.), Genetics and Clinics of HNPCC: Proceedings of the Fourth Workshop, pp. 26–27. Copenhagen: EUROFAP, 1993.

32. Prolla, T. A., Pang, Q., Alani, E., Kolodner, R. D., and Liskay, R. M. Interactions between the MSH2, MLH1 and PMS1 proteins during the initiation of DNA mismatch repair. Science (Washington DC), *265:* 1091–1093, 1994.

33. Lindblom, A., Tannergard, P., Nordenskjold, W., and Nordenskjold, M. Genetic mapping of a second locus predisposing to hereditary non-polyposis colon cancer. Nature Genet., *5:* 279–282, 1993.

34. Hultman, T., Stahl, S., Hornes, E., and Uhlen, M. Direct solid-phase sequencing of genomic and plasmid DNA using magnetic beads as solid support. Nucleic Acids Res., *17:* 4937–4946, 1989.

35. Jones, M. H., Yamakawa, Y., and Nakamura, Y. Isolation and characterization of 19 dinucleotide repeat polymorphisms on chromosome 3p. Hum. Mol. Genet., *1:* 131–133, 1992.

36. Nystrom-Lahti, M., Sistonen, P., Mecklin, J-P., Pylkkanen, L., Aaltonen, L. A., Jarvinen, H., Weissenbach, J., de la Chapelle, A., and Peltomaki, P. Close linkage to chromosome 3p and conservation of ancestral founding haplotype in hereditary nonpolyposis colorectal cancer families. Proc. Natl. Acad. Sci. USA, *9:* 6054–6058, 1994.

37. Tannergard, P., Zabarovsky, E., Stanbridge, E., Nordeskjold, M., and Lindblom, A. Sublocalization of a locus at 3p21.3–23 predisposing to hereditary nonpolyposis colorectal cancer. Hum. Genet., *94:* 210–214, 1994.

38. Lathrop, G. M., and Lalouel, J. M. Easy calculation of LOD scores and genetic risks on small computer. Am. J. Hum. Genet., *36:* 460–465, 1984.

39. Dunstone, G. H., and Knaggs, T. W. L. Familial cancer of the colon and rectum. J. Med. Genet., *9:* 451–455, 1972.

40. Hall, N. R., Bishop, D. T., Burn, J., Chapman, P., Knaggs, T. W. L., Williams, M. A. T., and Finan, P. J. Genetic linkage to chromosome 2p markers in hereditary nonpolyposis colorectal cancer. Br. J. Surg., *81:* 754, 1994.

41. Watson, P., and Lynch, H. T. Extracolonic cancer in hereditary nonpolyposis colorectal cancer. Cancer (Phila.), *71:* 677–685, 1993.

248