Cell, Vol. 75, 1215–1225, December 17, 1993, Copyright © 1993 by Cell Press

Mutations of a *mutS* Homolog in Hereditary Nonpolyposis Colorectal Cancer

Fredrick S. Leach,[1*] Nicholas C. Nicolaides,[1*] Nickolas Papadopoulos,[1*] Bo Liu,[1*] Jin Jen,[1] Ramon Parsons,[1] Päivi Peltomäki,[2] Pertti Sistonen,[2] Lauri A. Aaltonen,[2] Minna Nyström-Lahti,[2] X.-Y. Guan,[3] Ji Zhang,[3] Paul S. Meltzer,[3] Jing-Wei Yu,[4] Fa-Ten Kao,[4] David J. Chen,[5] Karen M. Cerosaletti,[6] R. E. Keith Fournier,[6] Sean Todd,[7] Tracey Lewis,[7] Robin J. Leach,[7] Susan L. Naylor,[7] Jean Weissenbach,[8] Jukka-Pekka Mecklin,[2] Heikki Järvinen,[2] Gloria M. Petersen,[9] Stanley R. Hamilton,[10] Jane Green,[11] Jeremy Jass,[12] Patrice Watson,[13] Henry T. Lynch,[13] Jeffrey M. Trent,[3] Albert de la Chapelle,[2] Kenneth W. Kinzler,[1] and Bert Vogelstein[1]

[1]The Johns Hopkins Oncology Center
Baltimore, Maryland 21231
[2]Department of Medical Genetics
University of Helsinki
00290 Helsinki
Finland
[3]National Center for Human Genome Research
Bethesda, Maryland 20892
[4]Eleanor Roosevelt Institute for Cancer Research
Denver, Colorado 80206
[5]Life Sciences Division
Los Alamos National Laboratory
Los Alamos, New Mexico 87545
[6]Molecular Medicine
Fred Hutchinson Cancer Research Center
Seattle, Washington 98104
[7]University of Texas Health Science Center
San Antonio, Texas 78284
[8]Genethon, Evry,
and Unite de Genetique Moleculaire Humaine
Institut Pasteur
Centre National de la Recherche Scientifique URA1445
75015 Paris
France
[9] The Johns Hopkins University School
of Public Health and Hygiene
Baltimore, Maryland 21205
[10]Department of Pathology
The Johns Hopkins University School of Medicine
Baltimore, Maryland 21205
[11]Memorial University of Newfoundland
St. John's, Newfoundland A1B 3V6
Canada
[12]Department of Pathology
University of Auckland School of Medicine
PB 92019 Auckland
New Zealand
[13]Department of Preventive Medicine
and Public Health
Creighton University School of Medicine
Omaha, Nebraska 68178

* These authors contributed equally to this work.
Correspondence should be addressed to A. de la Chapelle or K. W. Kinzler.

Summary

Recent studies have shown that a locus responsible for hereditary nonpolyposis colorectal cancer (HNPCC) is on chromosome 2p and that tumors developing in these patients contain alterations in microsatellite sequences (RER$^+$ phenotype). We have used chromosome microdissection to obtain highly polymorphic markers from chromosome 2p16. These and other markers were ordered in a panel of somatic cell hybrids and used to define a 0.8 Mb interval containing the HNPCC locus. Candidate genes were then mapped, and one was found to lie within the 0.8 Mb interval. We identified this candidate by virtue of its homology to *mutS* mismatch repair genes. cDNA clones were obtained and the sequence used to detect germline mutations, including those producing termination codons, in HNPCC kindreds. Somatic as well as germline mutations of the gene were identified in RER$^+$ tumor cells. This *mutS* homolog is therefore likely to be responsible for HNPCC.

Introduction

HNPCC (Lynch syndrome) is one of the most common cancer predisposition syndromes, affecting as many as 1 in 200 individuals in the western world (Lynch et al., 1993). Affected individuals develop tumors of the colon, endometrium, ovary, and other organs, often before 50 years of age. Although the familial nature of this syndrome was discovered nearly a century ago (Warthin, 1913), the role of heredity in its causation remained difficult to define (Lynch et al., 1966). Recently, however, linkage analysis in two large kindreds demonstrated association with polymorphic markers on chromosome 2 (Peltomäki et al., 1993a). Studies in other families suggested that neoplasia in a major fraction of HNPCC kindreds is linked to this same chromosome 2p locus (Aaltonen et al., 1993).

HNPCC is defined clinically by the occurrence of early-onset colon cancer and other specific cancers in first degree relatives spanning at least two generations (Lynch et al., 1993). The predisposition is inherited in an autosomal dominant fashion. It was initially expected that the gene(s) responsible for HNPCC was a tumor suppressor gene, as other previously characterized cancer predisposition syndromes with this mode of inheritance are caused by suppressor gene mutations (reviewed by Knudson, 1993). But the analysis of tumors from HNPCC patients suggested a different mechanism. Most loci encoding tumor suppressor genes undergo somatic losses during tumorigenesis (Stanbridge, 1990). In contrast, both alleles of

chromosome 2p loci were found to be retained in HNPCC tumors (Aaltonen et al., 1993). During this search for chromosome 2 losses, however, it was noted that HNPCC tumors exhibited somatic alterations of numerous microsatellite sequences.

Widespread subtle alterations of the cancer cell genome were first detected in a subset of sporadic colorectal tumors using the arbitrarily primed polymerase chain reaction (PCR) (Peinado et al., 1992). These alterations were subsequently found to represent deletions of up to 4 nt in genomic poly(A) tracts (Ionov et al., 1993). Other studies showed that a similar distinctive subgroup of sporadic tumors had insertions or deletions in a variety of simple repeated sequences, particularly microsatellite sequences consisting of dinucleotide or trinucleotide repeats (Ionov et al., 1993; Thibodeau et al., 1993; Aaltonen et al., 1993). Interestingly, these sporadic tumors had certain features in common with those developing in HNPCC kindreds, such as a tendency to be located on the right side of the colon and to be near diploid. These and other data suggested that HNPCC and a subset of sporadic tumors were associated with a heritable defect causing replication errors (RER) of microsatellites (Ionov et al., 1993; Aaltonen et al., 1993).

The mechanism underlying the postulated defect could not be determined from the study of tumor DNA, but studies in simpler organisms provided an intriguing possibility (Levinson and Gutman, 1987; Strand et al., 1993). This work showed that bacteria and yeast containing defective mismatch repair genes manifest instability of dinucleotide repeats. The disruption of genes primarily involved in DNA replication or recombination had no apparent effect on the fidelity of microsatellite replication (reviewed by Kunkel, 1993). These pivotal studies suggested that defective mismatch repair might be responsible for the microsatellite alterations in the tumors from HNPCC patients (Strand et al., 1993).

We here describe our efforts to identify a HNPCC gene based on the clues described above. Physical and linkage mapping were used to delineate a 0.8 Mb region of chromosome 2 harboring the HNPCC gene (*COCA1* or *FCC* locus). A new homolog of the *mutS* mismatch repair gene was discovered, and the gene was shown to reside within this small region of chromosome 2. Finally, cDNA clones of the gene were obtained and the sequence used to find mutations in HNPCC patients.

Results

Somatic Cell Hybrids

A panel of human–hamster, human–mouse, and human–rat hybrid cell lines was developed to facilitate HNPCC mapping. Hybrids containing only portions of chromosome 2 were obtained by microcell-mediated chromosome transfer or by standard cell fusions following X irradiation of the chromosome 2 donor (see Experimental Procedures). Additionally, two hybrids were used that contained a (X;2)(q28;p21) translocation derived from human fibroblasts. In previous studies, the HNPCC locus was mapped to the 25 cM region surrounding marker 123 and bordered by markers 119 and 136 (Peltomäki et al., 1993a). Thirty-eight hybrids were screened with these three chromosome 2p markers. Of the hybrids, eight proved useful for mapping the relevant portion of chromosome 2p. For example, hybrids L1 and L2 contained the distal half of the region, including marker 123, while hybrid y3 contained the half proximal to marker 123 (Figure 1).

Polymorphic Markers

To more finely map the HNPCC locus, additional polymorphic markers were obtained in three ways. First, a genomic clone containing 85 kb surrounding the 123 marker was used for fluorescence in situ hybridization (FISH) to localize it to chromosomal band 2p16 (Figures 2A and 2B). The 2p16 band region was then microdissected, and the sequences within this band were amplified using PCR and subcloned into plasmid vectors (see Experimental Procedures). The accuracy of the microdissection was confirmed using dual-color FISH by simultaneously hybridizing microdissected material and a genomic clone containing marker 123 (Figures 2C and 2D). The subclones were screened by hybridization to a $(CA)_{15}$ probe, and hybridizing clones identified and sequenced. These sequences were then used to design oligonucleotide primers for PCR analysis of genomic DNA. Nineteen $(CA)_n$ repeat markers were identified in this way. Of these, four were highly polymorphic and mapped to the region between markers 119 and 136, as assessed by the somatic cell hybrid panel exhibited in Figure 1. Second, eight additional $(CA)_n$ markers, cloned randomly from human genomic DNA using a poly(CA) probe, were found to lie between markers 119 and 136 by linkage analysis in Centre d'Etude du Polymorphisme Humain (CEPH) pedigrees. Five of

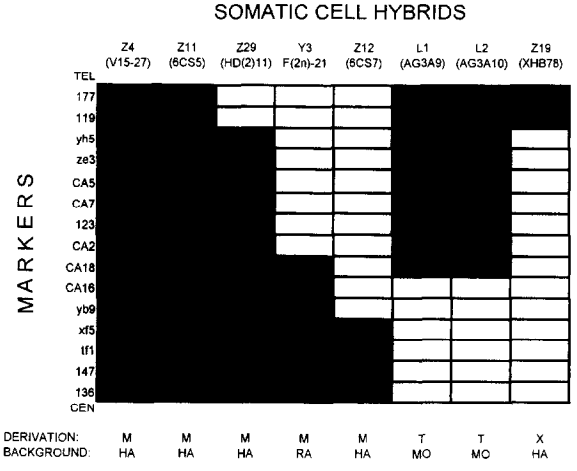

Figure 1. Somatic Cell Hybrids

PCR was used to determine whether each of the listed markers was present (closed box) or absent (open box) in the indicated hybrid. The laboratory name of each hybrid and the formal name (in parentheses) is listed. The hybrid panel was also validated with ten additional polymorphic markers outside of the 136–177 region. M, hybrid derived from microcell-mediated chromosome 2 transfer; T, derived from t(X; 2) translocation; X, derived from X-irradiated chromosome 2 donor. MO, mouse–human hybrid; HA, hamster–human hybrid; RA, rat–human hybrid; TEL, telomere; CEN, centromere.

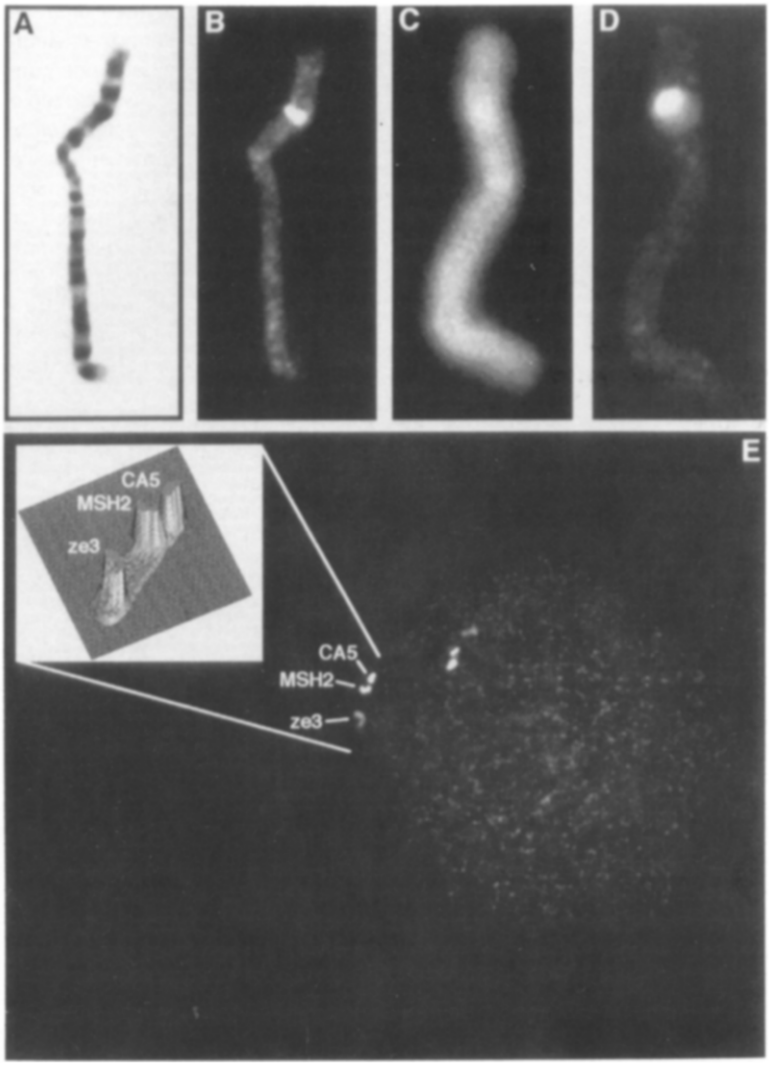

Figure 2. FISH Analysis to Determine the Proximity and Ordering of DNA Sequences within Chromosome Band 2p16

(A, B) FISH mapping of the 123 marker.
(A) G-banded metaphase chromosome 2.
(B) Identical chromosome as in (A) following FISH with a biotin-labeled P1 clone for the 123 marker. Results localize the 123 marker to chromosome band 2p16.
(C, D) Cohybridization documenting the coincident localization of a microdissection (Micro-FISH) probe from chromosome 2p16 and the 123 marker.
(C) DAPI-stained metaphase chromosome 2.
(D) Simultaneous hybridization of the biotin-labeled 123 probe (appearing as an intensely staining smaller circle) and the Spectrum-Orange labeled 2p16 Micro-FISH probe (appearing as a diffusely staining larger circle).
(E) Representative example of an interphase nucleus simultaneously hybridized with P1 clones for CA5, *hMSH2*, and ze3. The results were used to directly measure the distances between markers to establish the order and relative distance between markers (according to Trask et al., 1989).
Inset: The image processing program NIH Image was used to provide an average gray value displayed as a surface plot to support the length measurements and to graphically illustrate the relative order information. The surface plot presented defines the specified interphase chromosome and the relative order CA5–MSH2–ze3.

Table 1. Polymorphic Markers Used for Linkage Analysis

Marker	Derivation	cM	Lod	Heterozygosity	YAC Clones	P1 Clones
177 (AFM267zc9)	T	---		0.84		
119 (AFM077yb7)	T	6.1	5.5	0.77	11E1	406
yh5 (AFM337yh5)	T	6.4	15.4	0.76	4F4, 1E1	838, 839, 840
ze3 (AFM200ze3)	T	0.0	-----	0.61	4F4,1E1, 9H6, 4A10	836, 837
CA5 (CA5)	M	2.1	4.7	0.77	7F10, 4E2, 5A11	820
CA7 (CA7)	M	1.7	3	0.78	6B8	
123 (AFM093xh3)	T	2.4	9.9	0.76	3D11, 8C7	210, 211
CA2 (CA2)	P	0.0	-----	0.75	3D11, 8C7	210, 211
CA18 (CA18)	M	4.3	17.1	0.71	8E5	
CA16 (CA16)	M	0.0	-----	0.69	8E5	
yb9 (AFM320yb9)	T	1.1	3.9	0.80	264	
xf5 (AFM310xf5)	T	2.7	17.6	0.76		
tf1 (AFM348tf1)	T	0.0	-----	0.79		
147 (AFM199vb6)	T	2.6	9.8	0.73		
136 (AFM172xe7)	T	2		0.73		
134 (AFM168xg11)	T	1		0.76		387, 388, 389

Each of the markers listed was found by screening a genomic library with a $(CA)_n$ probe. The libraries were made from total genomic DNA (T), microdissected human chromosome 2p16 (M), or a genomic P1 clone containing marker 123 (P). The laboratory name and the formal name (in parentheses) of each marker is listed. The genetic distance between the indicated marker and the marker listed above it was determined by linkage analysis in CEPH families, as was the heterozygosity. The odds for pairwise inversion of loci is given in all cases in which these odds were greater than 1000:1. Those YAC and P1 clones obtained in this study that contained marker sequences are also listed.

these were particularly informative and were used in our subsequent studies. Finally, one additional marker was identified by screening subclones of a genomic P1 clone containing marker 123 with a $(CA)_{15}$ probe. Through these analyses, thirteen new polymorphic markers were identified in the 25 cM interval between markers 119 and 136, resulting in an average marker spacing of ~2 cM (Table 1; data not shown). These markers were mapped with respect to one another by linkage in CEPH and HNPCC pedigrees, as well as by analysis of somatic cell hybrids. These two mapping techniques provided consistent and complementary information. For example, the relative positions of CA16 and CA18 could not be distinguished through linkage analysis but could be determined with the somatic cell hybrids L1, L2, and Y3. Conversely, the relative position of the ze3 and yh5 markers could not be determined through somatic cell hybrid mapping, but could be discerned by linkage analysis.

Genomic Clones

Many of the polymorphic markers shown in Table 1 were used to derive genomic clones containing 2p16 sequences. Genomic clones were obtained by PCR screening of human P1 and yeast artificial chromosome (YAC) libraries with these polymorphic markers, with ten additional sequence tagged sites (STS) derived from chromosome 2p16 microdissection, or with YAC junctions (see Experimental Procedures). Twenty-three P1 clones, each containing 85–95 kb, were obtained, as well as 35 YAC clones, containing 300–1800 kb. In some cases, the YAC clones confirmed the linkage and somatic cell hybrid maps. For example, markers ze3 and yh5 were both found in YACs 4F4 and 1E1, while CA16 and CA18 were both found in YAC 8E5, documenting their proximity. The highest density of genomic clones (28 YAC and 17 P1 clones) was obtained between markers yb9 and yh5 (Table 1), which became the region most likely to contain the HNPCC gene during the course of these studies. The region between yh5 and yb9 was predicted to contain ~ 9 Mb (assuming 1 Mb/cM). Based on the sizes of the YAC clones, and taking into account their chimerism (see Experimental Procedures), we estimated that they contained over 70% of the sequences between yh5 and yb9.

Analysis of HNPCC Families

The markers described in Table 1 were then used to analyze six large HNPCC kindreds previously linked to chromosome 2p (Peltomäki et al., 1993a; R. P. and M. N.-L., unpublished data). Two hundred thirteen individuals, including 56 members affected with colorectal or endometrial cancer, were examined. Four of the kindreds were from the United States, one was from Newfoundland and one was from New Zealand. To increase the number of affected individuals that could be examined, we obtained formalin-fixed paraffin-embedded sections of normal tissues from deceased individuals and purified DNA from them (Goelz et al., 1985a). A single allele of each of the markers was found to segregate with disease in each of the six families (i.e., the allele was found in over 50% of affected individuals). No allele of any marker was shared among the affected members of more than three kindreds.

Fourteen of the affected members contained only a subset of the expected alleles and therefore had undergone recombination between markers 119 and 136. Eleven of these individuals appeared to have simple single recombination events. The most informative of these was in individual 148 from the J kindred and individual 44 from kindred 621 (Figure 3). Individual 621-44 apparently retained the disease linked allele at markers distal to CA5 while demonstrating multiple recombinants at more proximal loci, thus placing the CA5 marker at the proximal border of the HNPCC locus. Individual J-148 apparently retained the disease-linked allele at all markers proximal to yh5 while exhibiting recombinants at yh5 and 119, thus placing the distal border at yh5. Assuming that the same gene was involved in both the J and the 621 kindreds, the HNPCC gene was predicted to reside between markers CA5 and yh5, an area spanning approximately 2 cM (Table 1).

The DNA of three affected individuals (C-202, 4-156, and 4-92) appeared to have undergone two recombinations in the area. There was probably one recombinant per generation, and this could be demonstrated in C-202 by analysis of DNA from his parents; in the other cases, parental DNA was not available. All three individuals retained disease-linked alleles at CA5 and ze3 but not at more proximal and distal loci (Figure 3). Combined with the data from the patients with single recombinations, the double recombinants suggested that the HNPCC gene most likely resided between CA5 and ze3.

To determine the physical distances separating CA5, ze3, and yh5, metaphase and interphase FISH analysis was carried out. Dual-color FISH with P1 clones containing these markers was performed with P1 clones 820 and 838 (containing the markers CA5 and yh5, respectively),

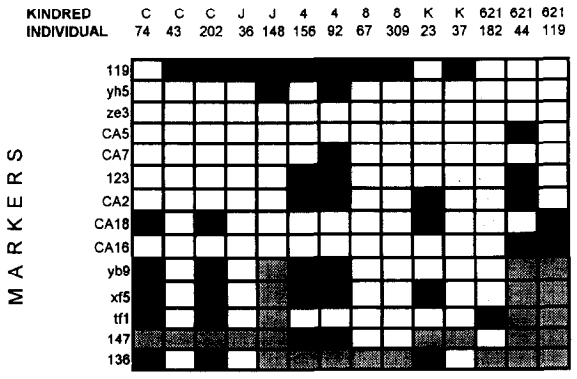

Figure 3. Linkage Analysis of HNPCC Pedigrees

All affected individuals in which meiotic recombination occurred between markers 119 and 136 are included. A closed box indicates that the individual did not contain the allele associated with disease in his or her family. An open box indicates that the individual had an allele that was the same size as the disease-associated allele but does not indicate that the patient necessarily had the disease-associated allele, because phase usually could not be determined. A hatched box indicates that the marker was not studied. All individuals had colon or endometrial cancer at less than 60 years of age, or had progeny with such disease.

labeled with biotin and detected with fluorescein-labeled avidin, and clone 836 (containing marker ze3), labeled with Spectrum–Orange (Meltzer et al., 1992). The hybridization signals of these markers appeared coincident on metaphase chromosomes, confirming that they resided within an interval of <1.0 Mb. When FISH was performed on interphase nuclei, the relative positions of the three markers could be determined and the distances between them estimated (Trask et al., 1989). The results confirmed that the orientation of the markers was telomere–yh5–ze3–CA5–centromere (data not shown). Direct measurement of the distance between yh5 and ze3 was estimated at <0.3 Mb, consistent with the presence of both of these markers on YAC clones 4E4 and 1E1. Measurements of 48 interphase chromosomes provided an estimate of the distance between ze3 and CA5 at <0.8 Mb, independently confirming the linkage data.

Candidate Genes

On the basis of the mapping results described above, we could determine whether a given gene was a candidate for HNPCC by determining its position relative to the CA5–yh5 domain. The first gene considered was a human homolog of the Drosophila *SOS* gene (reviewed by Egan and Weinberg, 1993). This gene transmits signals from membrane-bound receptors to the ras pathway in diverse eukaryotes. It was considered a candidate because another ras-interacting gene, *NF1*, causes a cancer predisposition syndrome (Viskochil et al., 1990; Wallace et al., 1990), and *Sos1* has been localized to chromosome 2p16–21 by in situ hybridization (Webb et al., 1993; P. Chardin, personal communication). Using PCR to amplify *Sos* sequences from the hybrid panel, however, *Sos1* was found to be distal to the CA5–yh5 domain (present in hybrid Z19 but not in Z29).

We next examined the interferon-inducible RNA-activated protein kinase gene *PKR*. This gene has been shown to have tumor suppressor ability (reviewed by Lengyel, 1993) and to map close to 2p16 (Hanash et al., 1993). Initially, we could not exclude *PKR* from the HNPCC domain, and therefore determined the sequence of its coding region in two individuals from HNPCC kindred C. Reverse transcriptase was used to generate cDNA from lymphoblastoid-derived RNA of these two individuals, and PCR performed with primers specific for *PKR*. The *PKR* products were sequenced and no deviations from the published sequence was identified within the coding region (Meurs et al., 1990). Subsequent studies showed that the *PKR* gene was distal to the yh5 marker, and thus could be excluded as a basis for HNPCC.

We then considered human homologs of the *mutL* and *mutS* mismatch repair genes previously shown to produce microsatellite instability in bacteria and yeast when disrupted (Levinson and Gutman, 1987; Strand et al., 1993). A human homolog of the yeast *mutL*-related gene, *PMS1* (Kramer et al., 1989), does not appear to reside on chromosome 2p (M. Liskay, personal communication). To identify homologs of *mutS*, we used degenerate oligonucleotide primers to PCR-amplify cDNA from colon cancer cell lines. The same primers previously had been used to identify the yeast *MSH2* gene on the basis of its *mutS* homology (Reenan and Kolodner, 1992). Under nonstringent conditions of PCR, a fragment of the expected size was obtained and this fragment was cloned into plasmid vectors. Most of the clones contained ribosomal RNA genes, representing abundant transcripts with weak homology to the degenerate primers. A subset of the clones, however, contained sequences similar to that of the yeast *MSH2* gene, and one such clone, pNP-23, was evaluated further. The human gene from which this clone is derived is hereafter referred to as *hMSH2*.

The insert from clone pNP-23 was used as a probe in Southern blots of somatic cell hybrid DNA. This insert hybridized to one or two fragments in human genomic DNA digested with PstI or EcoRI, respectively, and these fragments were present in hybrid Z30, containing most of human chromosome 2p. Analysis of other hybrids showed that the fragment was present in hybrids Z11, Z29, L1, and L2, but not Z12, Y3, or Z19, thereby localizing the

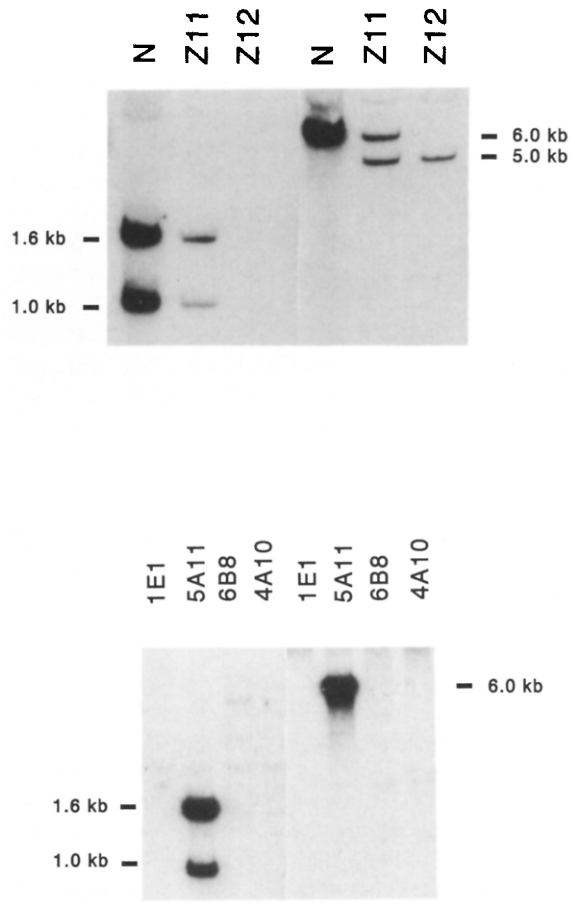

Figure 4. *hMSH2* Gene Localization

Southern blots containing EcoRI- (left three lanes) and PstI-digested (right three lanes) DNA from the indicated somatic cell hybrids (top) or YAC clones (bottom) were hybridized with a radiolabeled insert from cDNA clone pNP-23. Southern blotting and hybridization were performed as described (Vogelstein et al., 1987). Autoradiographs are shown. The 5.0 kb PstI fragment in hybrids Z11 and Z12 is derived from hamster DNA.

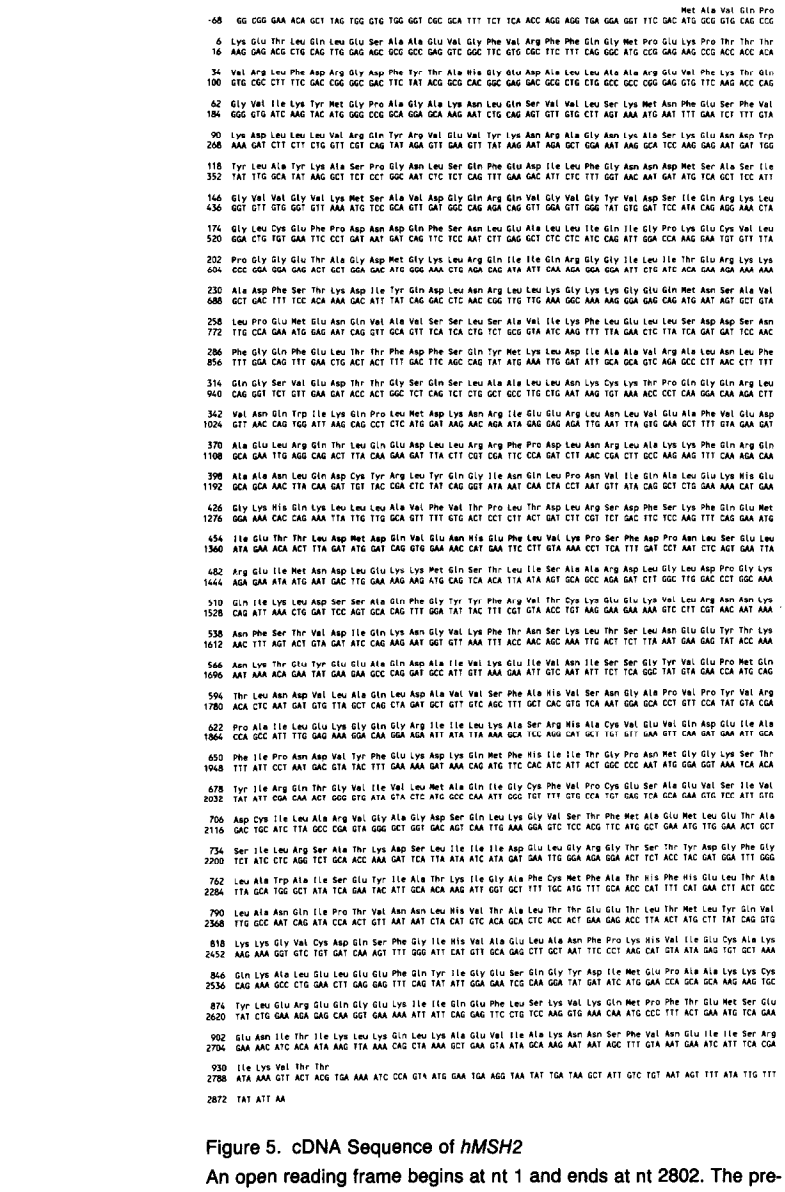

Figure 5. cDNA Sequence of *hMSH2*

An open reading frame begins at nt 1 and ends at nt 2802. The predicted amino acid sequence is shown. The sequence downstream of nt 2879 was not determined.

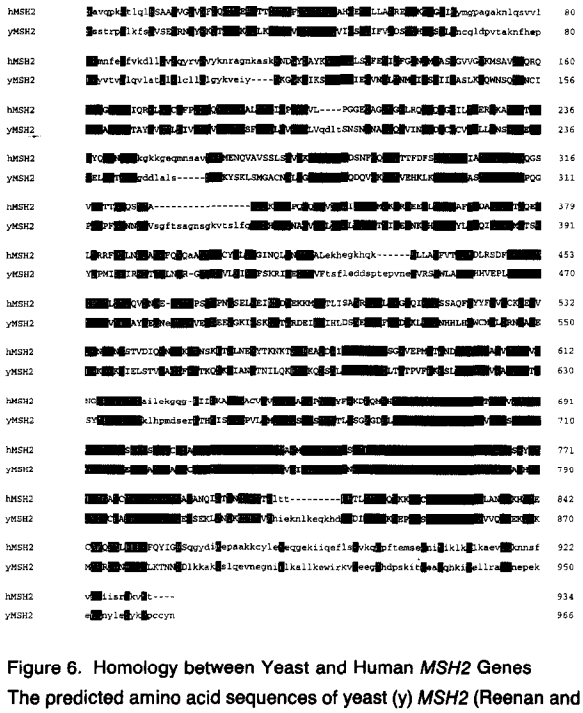

Figure 6. Homology between Yeast and Human *MSH2* Genes

The predicted amino acid sequences of yeast (y) *MSH2* (Reenan and Kolodner, 1992) and human *MSH2* genes were aligned using the MACAW (version 1.0) program. Blocks (10) of significant homology were identified. Amino acids in conserved blocks are capitalized and shaded based on the mean of their pairwise score (Schuler et al., 1991).

hMSH2 gene to a region bordered by markers CA18 and 119 (examples in Figure 4). The YAC clones listed in Table 1 were then analyzed, and EcoRI and PstI fragments of the expected size identified in YAC 5A11, derived from screening the YAC library with the CA5 marker (Figure 4).

To confirm the Southern blots, we designed nondegenerate primers on the basis of the sequence of pNP-23. Several sets of primers were tested so that genomic DNA could be used as a template for PCR; an intervening intron prevented the original primers from being used effectively with templates other than cDNA. PCR with these primers was perfectly consistent with the Southern blot results. The expected 101 bp fragment was present in hybrids Z4, Z11, Z29, L1, and L2, and in YAC 5A11, but not in other hybrids or YAC clones (data not shown).

The localization of *hMSH2* sequences to YAC 5A11 demonstrated the proximity of these sequences to marker CA5. To determine the distance and relative orientation of *hMSH2* with respect to CA5, we performed interphase FISH analysis. P1 clones containing CA5, ze3, and *hMSH2* sequences (clones 820, 836, and M1015, respectively) were simultaneously hybridized to interphase nuclei following fluorescein and Spectrum–Orange labeling (Meltzer et al., 1992). The results demonstrated that *MSH2* resides within the HNPCC locus defined by linkage analysis to lie between CA5 and ze3, and less than 0.3 Mb from marker CA5 (Figure 2E).

cDNA libraries generated from human colon cancer cells or from human fetal brain tissues were then screened with the insert of pNP-23 to obtain additional sequences from this gene (see Experimental Procedures). Initially, 75 cDNA clones were identified and partially sequenced. PCR products representing the ends of the cDNA sequence contig were then used as probes to rescreen the cDNA libraries. This so called cDNA walk was repeated again with the new contig ends. Altogether, 147 cDNA clones were identified. The composite sequence derived from these clones is shown in Figure 5. An open reading frame began 69 nt downstream of the 5′ end of the cDNA contig, and continued for 2802 bp. The methionine initiating this open reading frame was in a sequence context compatible with efficient translation (Kozak, 1986) and was preceded by in-frame termination codons. RNA from placenta and brain were used in a PCR-based procedure (rapid amplification of cDNA ends, Frohman et al., 1988) to independently determine the position of the 5′ end of

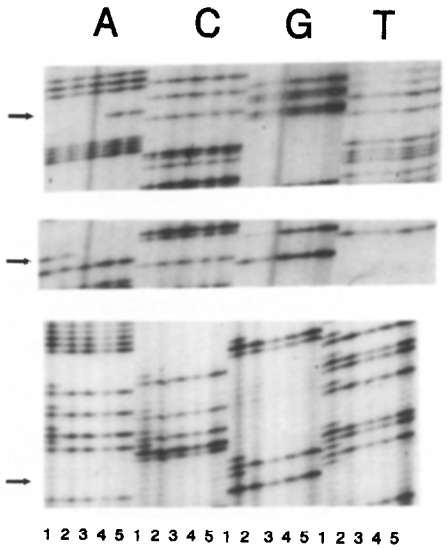

Figure 7. Germline and Somatic Mutations of *hMSH2*

Autoradiographs of polyacrylamide gels containing the sequencing reactions derived from PCR products are shown. The 1.4 kb PCR products containing a conserved region of *hMSH2* were generated from genomic DNA samples as described in Experimental Procedures. Antisense primers were used for sequencing in the top and middle panels, and sense primers were used for reactions in the bottom panel. The ddA mixes from each sequencing reaction were loaded in adjacent lanes to facilitate comparison, as were those for C, G, and T. The DNA samples were derived from the tumor (lane 1) and normal colon (lane 2) of patient Cx10, an RER$^-$ colon tumor cell line (lane 3), and lymphocytes of patients J-42 (lane 4) and J-143 (lane 5).

(Top) A transition (C to T at codon 622) in lymphocyte DNA can be observed in HNPCC patients J-42 and J-143.

(Middle) A transition (C to T at nt codon 639) in tumor (lane 1) and normal colonic mucosa (lane 2) of patient Cx10.

(Bottom) A substitution of a TG dinucleotide for an A at codon 663 can be observed in DNA of the tumor of patient Cx10, (lane 1), but not in DNA from her normal colon (lane 2). Arrows mark the substitutions in (A) and (B) and the TG dinucleotide insertion site in (C).

hMSH2 transcripts. This analysis demonstrated that the 5′ ends of all detectable transcripts were less than 100 bp upstream of the sequence shown in Figure 5, and were heterogeneous upstream of nt −69. The region of highest homology to the yeast *MSH2* gene encompassed the helix–turn–helix domain perhaps responsible for *mutS* binding to DNA (Reenan and Kolodner, 1992). The yeast and human MSH2 proteins were 77% identical between codons 615 and 788. There were ten other blocks of similar amino acids distributed throughout the length of these two proteins (966 and 934 amino acids in yeast and human, respectively), as shown in Figure 6.

Mutations of *hMSH2*

The physical mapping of *hMSH2* to the HNPCC locus was intriguing but could not prove that this gene was responsible for the disease. To obtain more compelling evidence, we determined whether germline mutations of *hMSH2* were present in the two HNPCC kindreds that originally established linkage to chromosome 2 (Peltomäki et al., 1993a). Intron–exon borders within the most conserved region of *hMSH2* (Figure 6) were determined by sequencing genomic PCR fragments containing adjacent exons. Genomic DNA samples from the lymphocytes of affected members of these two kindreds were then used as templates for PCR to determine the sequence of this domain (see Experimental Procedures). The DNA from individual J-42, afflicted with colon and endometrial cancer at ages 42 and 44, respectively, was found to contain one allele with a C to T transition at codon 622 (CCA to CTA), resulting in a substitution of leucine for proline (Figure 7, top). Twenty additional DNA samples from unrelated individuals all encoded proline at this position. Twenty-one members of the J kindred were then analyzed by direct sequencing of PCR products. All eleven affected individuals contained one allele with a C to T transition in codon 622, while all ten unaffected members contained two normal alleles, thus documenting perfect segregation with disease. Importantly, this proline was at a highly conserved position, the identical residue being found in all known *mutS*-related genes from prokaryotes and eukaryotes (Figure 6; Reenan and Kolodner, 1992).

No mutations of the conserved region of *MSH2* were identified in kindred C, so we next examined other parts of the *hMSH2* transcript. RNA was purified from lymphoblastoid cells of patient C-202, a 27 year old male with colon cancer. Reverse transcriptase–coupled PCR (RT–PCR) was used to generate four *hMSH2*-specific products encompassing codons 91–934 from this RNA (see Experimental Procedures). An abnormal smaller RT–PCR product was identified with one of the primer pairs used. Mapping and sequencing studies using various *MSH2* primers showed that the abnormal product was the result of a presumptive splicing defect that removed codons 265–314 from the *hMSH2* transcript. The abnormal transcript was found to segregate with disease in the C kindred, and was not found in 20 unrelated individuals.

We next wished to determine whether *hMSH2* was altered in one of the more recently linked families (R. P. and M. N.-L., unpublished data), and chose kindred 8 for detailed analysis. DNA and RNA were obtained from lymphoblastoid cells of 8-143, a 42 year old male with colon cancer. The conserved region of *hMSH2* was amplified from genomic DNA using PCR and directly sequenced. A substitution of C for T was noted in the polypyrimidine tract 6 bp upstream of the exon beginning at codon 669 (tttgttttgtag/GCCCCAATAT to tttgttctgtag/GCCCCAATAT, with upper case letters indicating exon sequences). However, this substitution was also found in two of twenty unrelated normal individuals, and was therefore a polymorphism apparently unrelated to the disease, with an allele frequency of 0.05. Most of the *hMSH2*-coding region (codons 91–934) was then amplified by RT–PCR, as described above, and no abnormal transcripts were detected. Sequencing of the PCR products, however, revealed a C to T transition at codon 406 (CGA to TGA) causing substitution of a termination codon for an arginine residue. RNA was available from the lymphocytes of a second affected member of kindred 8, and the same stop codon was identified. This alteration was not found in twenty other unrelated individuals.

Finally, we wished to determine whether mutations of

Table 2. Summary of Mutations

Sample	Source	Type	Codon	cDNA Nucleotide Change	Predicted Coding Change
Family J	HNPCC kindred	Germline	622	CCA to CTA	Proline to leucine
Family C	HNPCC kindred	Germline	265–314	793 to 942 deletion	In-frame deletion
Family 8	HNPCC kindred	Germline	406	CGA to TGA	Arginine to stop
Cx10	RER$^+$ tumor	Germline	639	CAT to TAT	Histidine to tyrosine
		Somatic	663	ATG to TGTG	Frameshift

Mutations were detected by sequencing PCR products, and each was confirmed by replicate PCR analyses and by examination of other affected members of the kindred (see text). The indicated germline mutations were each heterozygous; i.e., lymphocyte DNA contained one wild-type allele and one mutant allele. In Cx10, cloning of the PCR products showed that the somatic mutation at codon 663 occurred in the allele not affected by the germline mutation at codon 639.

this gene occurred in RER$^+$ tumors from patients without evident family histories of cancer. The conserved region of *MSH2* was studied in four colorectal tumor cell lines from such patients using genomic DNA as templates for PCR. One tumor (from patient Cx10) was found to contain two *hMSH2* alterations. The first was a C to T transition in codon 639 (CAT to TAT), resulting in a substitution of tyrosine for histidine. This change was not found in any of 20 samples from unrelated individuals, but was present in the DNA from normal colon of this patient, and was therefore likely to represent a germline change (Figure 7, middle). Like the missense mutation in the J kindred, the Cx10 alteration was at a position perfectly conserved in all *mutS* homologs (Figure 6; Reenan and Kolodner, 1992). The second alteration in the tumor from Cx10 was a substitution of a TGT dinucleotide for an A in codon 663 (ATG to TGTG). The resultant 1 bp insertion was predicted to cause a frameshift, producing a termination codon 36 nt downstream. This mutation was demonstrated in both RNA and DNA purified from the Cx10 tumor, but was not present in the patient's normal colon, and so represented a somatic mutation (Figure 7, bottom). The PCR products from Cx10 were cloned and sequenced, and the insertion mutation at codon 663 and the transition at codon 639 were shown to reside on different alleles.

Discussion

Three major conclusions can be drawn from the experiments described here. First, physical mapping and linkage analysis localized the HNPCC locus on chromosome 2 to an 0.8 Mb segment bordered by markers CA5 and ze3. Second, a new human homolog of the yeast *MSH2* gene was identified, and this gene shown to lie in the same 0.8 Mb interval. Third, alterations of the *hMSH2* gene occurred in the germline of patients with RER$^+$ tumors, with or without classical HNPCC, and additional somatic alterations of this gene occurred in tumors (summarized in Table 2). The alterations were at highly conserved regions or significantly altered the expected gene product and likely represented mutations with important functional effects. These results strongly suggest that mutations of *MSH2* are responsible for HNPCC and the RER$^+$ positive phenotype found in tumors.

These data are consistent with studies described in the accompanying paper, which show that RER$^+$ tumors are hypermutable and have a defect in mismatch repair that can be demonstrated biochemically (Parsons et al., 1993 [this issue of *Cell*]). They have substantial implications for understanding the neoplastic disease observed in HNPCC. In particular, they suggest that the microsatellite alterations previously observed in tumors from these patients are not epiphenomena, but are intrinsically related to pathogenesis. Additionally, the mutations observed in yeast and bacteria with defective *mutS*-related genes are not confined to insertions and deletions at simple repeated sequences, though these sequences provide convenient tools for analysis (Modrich, 1991). Similarly, one would expect that many mutations, in addition to microsatellite insertions or deletions, would be found in HNPCC tumors (Parsons et al., 1993).

Colorectal tumors from HNPCC patients are distinguished by their relatively normal cytogenetic composition (Kouri et al., 1990), and sporadic RER$^+$ tumors have been demonstrated to have substantially fewer chromosome losses than those occurring in RER$^-$ cases (Thibodeau et al., 1993; Aaltonen et al., 1993). These data suggest that genetic heterogeneity is critical for colorectal cancer development, but can be generated in two different ways (Thibodeau et al., 1993). Most commonly, it develops through gross alterations resulting in aneuploidy, as suggested nearly eighty years ago (Boveri, 1914). In HNPCC-derived tumors and RER$^+$ sporadic tumors, the diversity is presumably more subtle, consisting of multiple small sequence changes distributed throughout the genome. The latter mechanism of generating diversity may be less dangerous to the host, since HNPCC patients, as well as patients with RER$^+$ sporadic tumors, appear to have a better prognosis than would be expected from histopathologic analysis of their tumors (Ionov et al., 1993; Thibodeau et al., 1993; Lothe et al., 1993; Lynch et al., 1993).

These findings raise a number of more difficult questions. Why would mutations in *hMSH2* result in predisposition to specific tumors (Jass and Stewart, 1992; Watson and Lynch, 1993), rather than tumors of every cell type? The *hMSH2* is expressed in lymphocytes and many other cells and its evolutionary and presumptive function would classify it as a so called housekeeping gene. Additionally, why are mutations in microsatellites only observed in tumors from these patients, rather than in their normal cells? One hypothesis to answer these questions is the following. Cells of HNPCC patients develop normally, as they are

not significantly defective in mismatch repair, their one wild-type allele providing sufficient capacity. Occasionally, however, a tumor arises in such a patient because of a mutation in a growth controlling gene such as *APC* (Powell et al., 1992). Once a small tumor forms, the mismatch repair defect is accentuated somewhat, perhaps because of altered methylation (Baylin et al., 1991), which is known to affect mismatch repair in prokaryotes (Pukkila et al., 1983). This in turn leads to inactivation of the second *mutS* allele through a subtle somatic mutation. In the benign tumors of HNPCC patients, the rate of somatic mutation may be higher than the rate of gross chromosomal changes, unlike the situation in non-HNPCC tumors. Once the second *mutS* allele is inactivated by mutation, tumor progression is accelerated because the cell rapidly develops other mutations (i.e., RAS, p53, DCC; Fearon and Vogelstein, 1990). This hypothesis is consistent with several observations. First, the incidence of adenomas in HNPCC is not particularly high, while the rate of conversion of adenoma to carcinoma may be elevated significantly (Jass and Stewart, 1992; Lynch et al., 1993). This is in contrast with patients with inherited mutations of the tumor suppressor gene *APC*, who have a dramatically increased rate of adenoma formation but a normal rate of conversion of adenoma to carcinoma (Bussey, 1990). Second, colon tumors from HNPCC patients have been shown to contain mutations of *APC*, p53, and RAS at frequencies similar to those found in sporadic colorectal cancers (Aaltonen et al., 1993). Third, normal cells of HNPCC patients are mismatch repair proficient, while RER^+ tumor cells have a profound deficiency (Parsons et al., 1993). Fourth, tumors from HNPCC patients have very few allelic losses compared with other patients. Finally, at least some tumors from patients with a germline mutation of *mutS* acquire a somatic mutation of the wild-type *mutS* allele during their development (Table 2).

Finally, the potential role of *hMSH2* in HNPCC may be the paradigm for a variety of similar pathogenetic mechanisms. Human homologs of mismatch repair genes may cause HNPCC or other cancer predisposition syndromes, in addition to those linked to chromosome 2 (Lindblom et al., 1993). Although characteristic of HNPCC tumors, the RER^+ phenotype is also found in tumors from patients who are unlikely to have HNPCC on the basis of age and family history (Ionov et al., 1993; Thibodeau et al., 1993). These patients include those with colorectal tumors, as well as those with endometrial, gastric, and pancreatic cancers (Risinger et al., 1993; Burks et al., 1993; Han et al., 1993; Peltomäki et al., 1993b). Do tumors arise in these patients as a result of weak germline defects in mismatch repair genes, or through somatic mutations of such genes? These questions should be answerable by sequencing *hMSH2* and related genes in these patients.

Experimental Procedures

Polymorphic Markers

All markers were obtained by screening human genomic libraries with radiolabeled $(CA)_n$ probes (Weber and May, 1989). The T markers (see Table 1) were generated from a library made from total human genomic DNA, as described in Weissenbach et al., 1992. The M markers were made from libraries generated from microdissected chromosome 2p16, as described below. The CA2 marker was generated from a library made from P1 clone 210 digested to completion with Sau3A and cloned into the XhoI site of λYES (Elledge et al., 1991). The sequences of the clones obtained from these libraries were determined, and primers surrounding the CA repeats were chosen. Only primers giving robust amplification and high heterozygosity were used for detailed analysis of HNPCC kindreds. All markers used in this study were shown to be derived from chromosome 2p by both linkage analysis in the CEPH pedigrees and evaluation in the somatic cell hybrid panel shown in Figure 1. The sequences of the primers and other details specific for each marker have been deposited with the Genome Data Base. Linkage analyses to obtain the map of the marker loci in CEPH families 1331, 1332, 1347, 1362, 1413, 1416, 884, and 102 (Weissenbach et al., 1992) were performed using the CLINK program of the LINKAGE program package (Lathrop et al., 1984) with the no sex difference option and 11-point computations. The odds for the best locus order supported by the data were evaluated against pairwise inversions of the loci.

Chromosome Microdissection and FISH

G-banded metaphase chromosomes were microdissected with glass microneedles and amplified by PCR as previously described (Guan et al., 1993 Kao and Yu, 1991). For dual-color FISH, the PCR product was fluorochromosome labeled (Spectrum–Orange, Imagenetics, Naperville, Illinois) or biotinylated in a secondary PCR reaction. P1 clones were labeled by nick translation or by degenerate oligonucleotide primers (Guan et al., 1993). FISH was carried out as previously described (Guan et al., 1993) and visualized with a Zeiss Axiophot equipped with a dual bandpass filter. For analysis of interphase FISH patterns, the distance between hybridization signals was measured in a minimum of 24 nuclei (Trask et al., 1989).

Somatic Cell Hybrids

Methods for the derivation of microcell-mediated chromosome 2 hybrids have been described previously (Chen et al., 1992; Spurr et al., 1993). Some hybrids were generated following fusion of X-irradiated donor cells containing human chromosome 2 to Chinese hamster ovary cells (Chen et al., 1994). Mouse hybrids were derived by fusing hypoxanthine phosphoribosyltransferase-deficient L cells (A9) with human fibroblasts (GM7503) containing a t(X;2)(q28,p21) translocation and selecting in media containing hypoxanthine–aminopterin–thymidine.

Genomic Clones

The markers described in Table 1 were used to screen YAC or P1 libraries by PCR. The CEPH A library was obtained from Research Genetics, Incorporated, and consisted of 21,000 YAC clones, arrayed in a format allowing facile screening and unambiguous identification of positive clones. The sizes of ten of the YAC clones containing markers were determined by transverse alternating pulsed-field gel electrophoresis using a GeneLine II apparatus from Beckman and found to average 0.7 Mb (range 0.2–1.8 Mb). In some cases, inverse PCR was used to determine the YAC junctions (Joslyn et al., 1991), and the derived sequence used for so called chromosome walking with the YAC or P1 libraries. The junctions were also used to design primers to test whether the ends of the YAC clones could be localized to chromosome 2p16 (and therefore presumably be nonchimeric). Three of four YAC clones that were tested in this way had both ends within the expected region of chromosome 2, as judged by analysis with the somatic cell hybrid panel. The human genomic P1 library was also screened by PCR (Genome Systems, Incorporated). P1 clones M1015 and M1016, containing the *hMSH2* gene, were used to determine intron–exon borders using sequencing primers from the exons and SequiTherm polymerase (Epicentre Technologies).

cDNA Clones and PCR Products

cDNA generated from the RNA of colorectal cancer cells with reverse transcriptase was used as template for PCR with the degenerate primers 5′-CTG GAT CCA C(G/A/T/C) G G(G/A/T/C)C C(G/A/T/C)A A(T/C)A TG-3′ and 5′-CTG GAT CC(G/A) TA(G/A) TG(G/A/T/C) GT (G/A/T/C) (G/A)C(G/A) AA-3′. These two primers were used previously to identify the yeast *MSH2* gene and were based on sequences con-

served among related mammalian and bacterial genes (Reenan and Kolodner, 1992). The optimal PCR conditions for detecting the *hMSH2* gene consisted of 35 cycles at 95°C for 30 s, 41°C for 90 s, and 70°C for 90 s in the buffer described previously (Sidransky et al., 1991). PCR products were cloned into T-tailed vectors as described (Holton and Graham, 1991) and sequenced with modified T7 polymerase (United States Biochemical). The insert from one clone (pNP-23) containing human sequences homologous to the yeast *MSH2* gene was then used to screen cDNA libraries generated from RNA of SW480 colon cancer cells (Clontech) or of fetal brain (Stratagene). After two further rounds of screening, positive clones were converted into plasmids and sequenced using modified T7 polymerase (Kinzler et al., 1991). In some cases, the inserts were amplified using one *hMSH2*-specific primer and one vector-specific primer, and then sequenced with SequiTherm Polymerase (Epicentre Technologies). To determine the 5′ end of *MSH2* transcripts, rapid amplification of cDNA ends was performed (Frohman et al., 1988) using brain and placenta cDNA (Clontech).

To detect mutations, PCR products were generated from cDNA and human genomic DNA templates, then sequenced directly using SequiTherm. In some cases, the PCR products were cloned into T-tailed vectors for sequencing to confirm the direct sequencing data. The primers used to amplify the conserved region of *MSH2* from genomic DNA were 5′-CCA CAA TGG ACA CTT CTG C-3′ and 5′-CAC CTG TTC CAT ATG TAC G-3′, resulting in a 1.4 kb fragment containing *hMSH2* codons 616–705; and primers 5′-AAA ATG GGT TGC AAA CAT GC-3′ and 5′-GTG ATA GTA CTC ATG GCC C-3′, resulting in a 2.0 kb fragment containing *MSH2* cDNA codons 684–784. Primers for RT–PCR were 5′-AGA TCT TCT TCT GGT TCG TC-3′ and 5′-GCC AAC AAT AAT TTC TGG TG-3′ for codons 91–433; 5′-TGG ATA AGA ACA GAA TAG AGG-3′ and 5′-CCA CAA TGG ACA CTT CTG C-3′ for codons 352–705; 5′-CAC CTG TTC CAT ATG TAC G-3′ and 5′-AAA ATG GGT TGC AAA CAT GC-3′ for codons 616–784; and 5′-GTG ATA GTA CTC ATG GCC C-3′ and 5′-GAC AAT AGC TTA TCA ATA TTA CC-3′ for codons 684–934.

Acknowledgments

The authors thank S. Booker, S. M. Stewart, J. Cavalieri, S. Slominski, and S. Luscombe for clinical coordination and specimen collection; J. M. Jessup for providing tumor cell lines and xenografts; P. Gold and M. Cunningham for assistance and advice with P1 cloning; M. Castro for oligonucleotide synthesis; J. A. Pietenpol, L.-K. Su, and T. Tokino for critically reading the manuscript; and T. Gwiazda for preparing it. This work was supported by the Clayton Fund, the Folkhälsan Institute of Genetics, the Academy of Finland, the Finnish Cancer Foundation, the Sigrid Juselius Foundation, the Health Research Council and Cancer Society of New Zealand, the Auckland Medical Research Foundation, Nebraska State Health Department of Cancer, the Council for Tobacco Research, American Cancer Society, Department of Energy grants DOE/ERN/F139 and DE-FG0291ER-61139, and grants CA35494, CA09320, CA47527, GM26449, CA09243, CA41183, CA57435, and CA42705 from the National Institutes of Health. B. V. is an American Cancer Society Research Professor.

Received November 29, 1993; revised November 30, 1993.

References

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

Baylin, S. B., Makos, M., Wu, J., Chiu-Yen, R.-W., de Bustros, A., Vertino, P., and Nelkin, B. D. (1991). Abnormal patterns of DNA methylation in human neoplasia: potential consequences for tumor progression. Cancer Cells *3*, 383, 390.

Boveri, T. (1914). Zur Frage der Entstehung maligner Tumoren. Volume 1 (Jena, Germany: Gustave Fischer Verlag).

Bussey, H. J. R. (1990). Historical Developments in Familial Adenomatous Polyposis. In Familial Adenomatous Polyposis, L. Herrera, ed. (New York: Alan R. Liss, Incorporated), pp. 1–7.

Burks, R. T., Kessis, T. D., Cho, K. R., and Hedrick L. (1993). Microsatellite instability in endometrial carcinoma. Oncogene, in press.

Chen, D. J., Park, M. S., Campbell, E., Oshimura, M., Liu, P., Zhao, Y., White, B. F., and Siciliano, M. J. (1992). Assignment of a human DNA double-strand break repair gene (*XRCC5*) to chromosome 2. Genomics *13*, 1088–1094.

Chen, D. J., Marrone, B., Nguyen, T., Stackhouse, M., Zhao, Y., and Siciliano, M. J. (1994). Regional assignment of a human radiation repair gene (*XRCC5*) to 2q35 by X-ray hybrid mapping. Genomics, in press.

Egan, S. E., and Weinberg, R. A. (1993). The pathway to signal achievement. Nature *365*, 781–783.

Elledge, S. J., Mulligan, J. T., Ramer, S. W., Spottswood, M., and Davis, R. W. (1991). Lambda YES: a multifunctional cDNA expression vector for the isolation of genes by complementation of yeast and *Escherichia coli* mutations. Proc. Natl. Acad. Sci. USA *88*, 1731–1735.

Fearon, E. R., and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. Cell *61*, 759–767.

Frohman, M. A, Dush, M. K., and Martin, G. R. (1988). Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA *85*, 8998–9002.

Goelz, S. E., Hamilton, S. R., and Vogelstein, B. (1985a). Purification of DNA from formaldehyde-fixed and paraffin-embedded human tissue. Biochem. Biophys. Res. Commun. *130*, 118–126.

Goelz, S. E., Vogelstein, B., Hamilton, S. R., and Feinberg, A. P. (1985b). Hypomethylation of DNA from benign and malignant human colon neoplasms. Science *228*, 187–190.

Guan, X.-Y., Trent, J. M., and Meltzer, P. S. (1993). Generation of band-specific painting probes from a single microdissected chromosome. Hum. Mol. Genet. *2*, 1117–1121.

Han, H.-J., Yanagisawa, A., Kato, Y., Park, J.-G., and Nakamura, Y. (1993). Genetic instability in pancreatic cancer and poorly differentiated type of gastric cancer. Cancer Res. *53*, 5087–5089.

Hanash, S. M., Beretta, L., Barcroft, C. L., Sheldon, S., Glover, T. W., Ungar, D., and Sonenberg, N. (1993). Mapping of the gene for interferon-inducible dsRNA-dependent protein kinase to chromosome region 2p21–22: a site of rearrangements in myeloproliferative disorders. Genes Chromosomes Cancer *8*, 34–37.

Holton, T. A., and Graham, M. W. (1991). A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. Nucl. Acids Res. *19*, 1156.

Ionov, Y. M., Peinado, A., Malkhosyan, S., Shibata, D., and Perucho M. (1993). Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature *363*, 558–561.

Jass, J. R., and Stewart, S. M. (1992). Evolution of hereditary nonpolyposis colorectal cancer. Gut *33*, 783–786.

Joslyn, G., Carlson, M., Thilveris, A., Albertsen, H., Gelbert, L., Samowitz, W., Groden, J., Stevens, J., Spirio, L., Robertson, M., Sargeant, L., Krapcho, K., Wolff, E., Burt, R., Hughes, J. P., Warrington, J., McPherson, J., Wasmuth, J., Le Paslier, D., Abderrahim, H., Cohen, D., Leppert, M., and White, R. (1991). Identification of deletion mutations and three new genes at the familial polyposis locus. Cell *66*, 601–613.

Kao, F.-T., and Yu, J.-W. (1991). Chromosome microdissection and cloning in human genome and genetic disease analysis. Proc. Natl. Acad. Sci. USA *88*, 1844–1848.

Kinzler, K. W., Nilbert, M. C., Su, L.-K., Vogelstein, B., Bryan, T. M., Levy, D. B., Smith, K.J., Preisinger, A. C., Hedge, P., McKechnie, D., Finniear, R., Markham, A., Groffen, J., Boguski, M. S., Altschul, S. F., Horii, A., Ando, H., Miyoshi, Y., Miki, Y., Nishisho, I., and Nakamura, Y. (1991). Identification of FAP locus genes from chromosome 5q21. Science *253*, 661–665.

Kouri, M., Laasonen, A., Mecklin, J. P., Järvinen, H., Franssila, K., and Pyrhonen, S. (1990). Diploid predominance in hereditary nonpolyposis colorectal carcinoma evaluated by flow cytometry. Cancer *65*, 1825–1829.

Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell *44*, 283–292.

Kramer, W., Kramer, B., Williamson, M. S., and Fogel, S. (1989). Cloning and nucleotide sequence of DNA mismatch repair gene *PMS1* from *Saccharomyces cerevisiae*: homology of *PMS1* to procaryotic *MutL* and *HexB*. J. Bacteriol. *171*, 5339–5346.

Knudson, A. G. (1993). All in the (cancer) family. Nature Genet. *5*, 103–104.

Kunkel, T. A. (1993). Slippery DNA and diseases. Nature *365*, 207–208.

Lathrop, G. M., Lalouel, J. M., Julier, C., and Ott, J. (1984). Strategies for multilocus linkage analysis in humans. Proc. Natl. Acad. Sci. USA *81*, 3443–3446.

Lengyel, P. (1993). Tumor-suppressor genes: news about the interferon connection. Proc. Natl. Acad. Sci. USA *90*, 5893–5895.

Levinson, G., and Gutman, G. A. (1987). High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. Nucl. Acids Res. *15*, 5323–5338.

Lindblom, A., Tånnergard, P., Werelius, B., and Nordenskjöld, M. (1993). Genetic mapping of a second locus predisposing to hereditary nonpolyposis colon cancer. Nature Genet. *5*, 279–282.

Lothe, R. A., Peltomäki, P., Meling, G.-I., Aaltonen, L. A., Nyström-Lahti, M., Pylkkänen, L., Heimdal, K., Andersen, T. I., Møller, P., Rognum, T. O., Fosså, S. D., Haldorsen, T., Langmark, F., Brøgger, A., de la Chapelle, A., and Børresen, A.-L. (1993). Genomic instability in colorectal cancer: relationship to clinicopathological variables and family history. Cancer Res., in press.

Lynch, H. T., Omaha, M. W., and Shaw, M. D. (1966). Hereditary factors in cancer. Arch. Int. Med. *117*, 206–212.

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

Meltzer, P. S., Guan, X.-Y., Burgess, A., and Trent, J. M. (1992). Rapid generation of region specific probes by chromosome microdissection and their application. Nature Genet. *1*, 24–28.

Meurs, E., Chong, K., Galabru, J., Thomas, N. S. B., Kerr, I. M., Williams, B. R. G., and Hovanessian, A. G. (1990). Molecular cloning and characterization of the human double-stranded RNA-activated protein kinase induced by interferon. Cell *62*, 379–390.

Modrich, P. (1991). Mechanisms and biological effects of mismatch repair. Annu. Rev. Genet. *25*, 229–253.

Parsons, R., Li, G.-M., Longley, M. J., Fang, W.-h., Papadopoulos, N., Jen, J., de la Chapelle, A., Kinzler, K. W., Vogelstein, B., and Modrich, P. (1993). Hypermutability and mismatch repair deficiency in RER$^+$ tumor cells. Cell *75*, this issue.

Peinado, M. A., Malkhosyan, S., Velazquez, A., and Perucho, M. (1992). Isolation and characterization of allelic losses and gains in colorectal tumors by arbitrarily primed polymerase chain reaction. Proc. Natl. Acad. Sci. USA *89*, 10065–10069.

Peltomäki, P., Aaltonen, L. A., Sistonen, P., Pylkkänen, L., Mecklin, J.-P., Järvinen, 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. (1993a). Genetic mapping of a locus predisposing to human colorectal cancer. Science *260*, 810–812.

Peltomäki, P., Lothe, R. A., Aaltonen, L. A., Pylkkänen, L., Nystrom-Lahti, M., Seruca, R., David, L., Holm, R., Ryberg, D., Haugen, A., Brøgger, A., Børresen, A.-L., and de la Chapelle, A. (1993b). Microsatellite instability is associated with tumors that characterize the hereditary nonpolyposis colorectal carcinoma syndrome. Cancer Res., in press.

Powell, S. P., Zilz, N., Beazer-Barclay, Y., Bryan, T. M., Hamilton, S. R., Thibodeau, S. N., Vogelstein, B., and Kinzler, K. W. (1992). *APC* mutations occur early during colorectal tumorigenesis. Nature *359*, 235–237.

Pukkila, P. J., Petseon, J., Herman, G., Modrich, P., and Meselson, M. (1983). Effects of high levels of DNA adenine methylation on methyl-directed mismatch repair in *Escherichia coli*. Genetics *104*, 571–582.

Reenan, R. A., and Kolodner, R. D. (1992). Isolation and characterization of two *Saccharomyces cerevisiae* genes encoding homologs of the bacterial HexA and MutS mismatch repair proteins. Genetics *132*, 963–973.

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

Schuler, G. D., Altschul, S. F., and Lipman, D. J. (1991). A workbench for multiple alignment construction and analysis. Proteins Struct. Funct. Genet. *9*, 180–190.

Sidransky, D., von Eschenbach, A., Tsai, Y. C., Jones, P., Summerhayes, I., Marshall, F., Paul, M., Green, P., Hamilton, S. R., Frost, P., Vogelstein, B. (1991). Identification of p53 gene mutations in bladder cancers and urine samples. Science *252*, 706–709.

Spurr, N. K., Cox, S., Naylor, S. (1993). Report and abstracts of the Second International Workshop on human chromosome 2 mapping. Cytogenet. Cell Genet. *64*, 69–92.

Stanbridge, E. J. (1990). Human tumor suppressor genes. Annu. Rev. Genes *24*, 615–657.

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

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

Trask, B., Pinkel, D., and van den Engh, G. (1989). The proximity of DNA sequences in interphase cell nuclei is correlated to genomic distance and permits ordering of cosmids spanning 250 kilobase pairs. Genomics *5*, 710–717.

Viskochil, D., Buchberg, A. M., Xu, G., Cawthon, R. M., Stevens, J., Wolff, R. K., Culver, M., Carey, J. C., Copeland, N. G., Jenkins, N. A., White, R., and O'Connell, P. (1990). Deletions and a translocation interrupt a cloned gene at the neurofibromatosis type 1 locus. Cell *62*, 187–192.

Vogelstein, B., Fearon, E. R., Hamilton, S. R., Preisinger, A. C., Willard, H. F., Michelson, A. M., Riggs, A. D., and Orkin, S. H. (1987). Clonal analysis using recombinant DNA probes from the X chromosome. Cancer Res. *47*, 4806–4813.

Wallace, M. R., Marchuk, D. A., Andersen, L. B., Letcher, R., Odeh, H. M., Saulino, A. M., Fountain, J. W., Brereton, A., Nicholson, J., Mitchell, A. L., Brownstein, B. H., Collins, F. S. (1990). Type 1 neurofibromatosis gene: identification of a large transcript disrupted in three NF1 patients. Science *249*, 181–186.

Warthin, A. S. (1913). Heredity with reference to carcinoma: as shown by the study of the cases examined in the pathological laboratory of the University of Michigan, 1895–1913. Arch. Int. Med. *12*, 546–555.

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

Webb, G. C., Jenkins, N. A., Largaespada, D. A., Copeland, N. G., Fernandez, C. S., and Bowtell, D. D. L. (1993). Mammalian homologues of the Drosophila Son of sevenless gene map to murine chromosomes 17 and 12 and to human chromosomes 2 and 14, respectively. Genomics *18*, 14–19.

Weber, J. L., and May P. E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. *44*, 338–396.

Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morissette, J., Millasseau, P., Vaysseix, G., and Lathrop, M. (1992). A second-generation linkage map of the human genome. Nature *359*, 794–801.

GenBank Accession Number

The accession number for the sequences reported in this paper is U04045.