Genetic Mapping of a Locus Predisposing to Human Colorectal Cancer

Päivi Peltomäki,* Lauri A. Aaltonen,* Pertti Sistonen, Lea Pylkkänen, Jukka-Pekka Mecklin, Heikki Järvinen, Jane S. Green, Jeremy R. Jass, James L. Weber, Fredrick S. Leach, Gloria M. Petersen, Stanley R. Hamilton, Albert de la Chapelle,† Bert Vogelstein†

Genetic linkage analysis was used to determine whether a specific chromosomal locus could be implicated in families with a history of early onset cancer but with no other unique features. Close linkage of disease to anonymous microsatellite markers on chromosome 2 was demonstrated in two large kindreds. The pairwise lod scores for linkage to marker D2S123 in these kindreds were 6.39 and 1.45 at zero recombination, and multipoint linkage with flanking markers resulted in lod scores of 6.47 and 6.01. These results prove the existence of a genetically determined predisposition to colorectal cancer that has important ramifications for understanding and preventing this disease.

The role of hereditary factors in the etiology of common human cancers is of great scientific and public importance. For colorectal cancer (CRC), in particular, there has been much debate about the relative contributions of heredity and environment in tumorigenesis. It is ironic that one of the earliest descriptions of an "inherited cancer" involved hereditary adenocarcinoma of the colon in a large family (*1*). Yet 80 years later, there is still no proof that the cancers in these families are due primarily to hereditary factors. One problem in establishing such proof is that colon cancers are so common in the general population that it is difficult to rule out chance clustering and other nonhereditary factors. Moreover, the environment, notably diet, has been shown to play a substantial role in CRC risk (*2*). Members of an individual family are likely to share similar environments, further complicating definitive analysis.

We searched for evidence of a genetic component through linkage analysis. There are two major forms of CRC predisposition. The first, familial adenomatous polyposis (FAP), is recognizable because affected patients have several unusual phenotypic features, most notably the presence of thousands of benign tumors lining the entire large intestine. FAP accounts for ~1% of CRC cases in the Western world (*3*) and the *APC* gene responsible for FAP has recently been identified (*4*). A second form of CRC that shows familial aggregation is hereditary nonpolyposis colorectal cancer (HNPCC). It is more common than FAP, accounting for 4 to 13% of all CRC in industrial nations (*5*). It is impossible to reliably distinguish patients with this form of CRC from "sporadic" cases on physical examination, as neither have diffuse polyposis or other unusual stigmata. HNPCC kindreds are commonly defined as those in which at least three relatives in two generations have CRC, with one of the relatives having been diagnosed at less than 50 years of age (*6*). In addition to the colon, other organs affected with cancer in HNPCC patients include the endometrium, stomach, biliopancreatic system, and urinary tract (*7*).

To prove the existence of a familial colon cancer gene, we studied two large kindreds that met the HNPCC criteria described above. Because we were uncertain that a single gene (rather than a compendium of genes) might be responsible for CRC in any individual family, or that the same gene or genes would be involved in different families, we hypothesized that families sufficiently large for individual linkage analysis might prove critical in gaining statistically significant results. We chose two large kindreds with HNPCC, families C and J, for detailed analysis (see Fig. 1 for pedigrees). The C kindred originated from North America and the J kindred from New Zealand. The mean age of onset of CRC was 41.1 years (SD = 13.1) and 44.4 years (SD = 15.5), respectively. Extracolonic cancers were observed in both families.

Two candidate gene regions on chromosomes 18 and 5 had already been excluded as contributing to HNPCC in our previous studies (*8*). We therefore began a systematic search through the whole genome, with highly informative microsatellite markers (*9–11*). Altogether, 345 microsatellite markers were analyzed before convincing evidence of linkage was obtained. Marker D2S123 [AFM093xh3; (*10*)] showed a highly significant two-point lod score of 6.39 at a recombination fraction (θ) of zero in family C (Table 1; low-stringency criteria). The maximum pairwise lod score for the J family was 1.45 at $\theta = 0.0$. There was no recombination between HNPCC and

P. Peltomäki, L. A. Aaltonen, L. Pylkkänen, A. de la Chapelle, Department of Medical Genetics, P.O. Box 21, University of Helsinki, SF-00014 Helsinki, Finland.
P. Sistonen, Department of Medical Sciences, University of Helsinki, and Finnish Red Cross Blood Transfusion Service, SF-00310 Helsinki, Finland.
J.-P. Mecklin, Department of Surgery, Jyvaskylä Central Hospital, SF-40620 Jyväskylä, Finland.
H. Järvinen, Second Department of Surgery, Helsinki University Central Hospital, SF-00290 Helsinki, Finland.
J. S. Green, Faculty of Medicine, Memorial University of Newfoundland, St. John's, Newfoundland, Canada A1B 3V6.
J. R. Jass, Department of Pathology, University of Auckland School of Medicine, Auckland, New Zealand.
J. L. Weber, Marshfield Medical Research Foundation, Marshfield, WI 54449.
F. S. Leach, G. M. Petersen, S. R. Hamilton, B. Vogelstein, Departments of Oncology, Pathology, and Epidemiology, Johns Hopkins University School of Medicine and School of Public Health and Hygiene, and Johns Hopkins Hospital, Baltimore, MD 21231.

*These authors contributed equally to this work.
†To whom correspondence should be addressed.

Table 1. Pairwise analysis of linkage between chromosome 2 markers and HNPCC by the MLINK program of the LINKAGE program package (*13*). Sex-averaged lod scores are shown separately for families C and J under two different phenotypic criteria. Individuals with adenoma or a single carcinoma other than that of colon or endometrium were considered to be affected (low stringency) or to be of unknown status (high stringency) [see text and (*17*)]. Allele frequencies for the markers were estimated from individuals married into HNPCC families.

Marker	Family	Stringency criteria	Lod score (Z) at recombination fraction (θ) of:								Z_{max}	θ
			0	0.001	0.01	0.05	0.1	0.2	0.3	0.4		
D2S119	C	Low	−1.08	−1.02	−0.63	0.11	0.42	0.48	0.30	0.11	0.50	0.16
D2S119	C	High	−0.37	−0.32	−0.04	0.47	0.63	0.55	0.32	0.12	0.64	0.12
D2S119	J	Low	1.81	2.07	2.81	3.43	3.47	2.96	2.10	1.00	3.49	0.08
D2S119	J	High	2.21	2.45	3.07	3.43	3.32	2.71	1.86	0.86	3.43	0.05
D2S123	C	Low	6.39	6.38	6.28	5.82	5.21	3.91	2.49	1.04	6.39	0.00
D2S123	C	High	5.24	5.23	5.14	4.75	4.25	3.15	1.97	0.79	5.24	0.00
D2S123	J	Low	1.45	1.44	1.42	1.30	1.15	0.83	0.53	0.24	1.45	0.00
D2S123	J	High	0.85	0.85	0.83	0.75	0.66	0.47	0.30	0.14	0.83	0.00
D2S136	C	Low	−1.47	−1.42	−1.02	−0.06	0.39	0.57	0.39	0.13	0.58	0.18
D2S136	C	High	−1.03	−0.99	−0.64	0.13	0.45	0.50	0.30	0.09	0.53	0.16
D2S136	J	Low	1.15	1.15	1.22	1.53	1.66	1.47	1.00	0.40	1.66	0.11
D2S136	J	High	2.38	2.39	2.43	2.46	2.33	1.84	1.18	0.46	2.47	0.04

D2S123 in either family, suggesting very close linkage. With two flanking markers, D2S119 (distal to D2S123) and D2S136 (proximal to D2S123), recombination was observed in both families, and the maximum pairwise lod scores shown by D2S119 and D2S136, respectively, were 0.50 at θ = 0.16 and 0.58 at θ = 0.18 for family C and 3.49 at θ = 0.08 and 1.66 at θ = 0.11 for family J (Table 1; low-stringency criteria).

In linkage analysis, individuals with different neoplasms were classified as follows. Colorectal or endometrial carcinomas [the two most common cancers in HNPCC family members (*7*, *12*)] were considered definitive manifestations of the disorder. Individuals with a colorectal adenoma or with a single carcinoma of the ovary, stomach, hepatobiliary system, small intestine, kidney, or ureter were classified as being of unknown status (high-stringency analysis) or affected (low-stringency analysis). Multipoint linkage analysis was carried out separately in families C and J by the program LINKMAP of the LINKAGE program package with the no sex difference option (*13*). A fixed map of the markers used was constructed from the data available in the CEPH database (version 6.0). Flanking marker distances to D2S123 were 11 centimorgans (cM) and 14 cM for D2S119 and D2S136, respectively. The HNPCC gene maps with D2S123, giving a four-point lod score of 6.47 in family C and 6.01 in family J at zero recombination.

The physical location of marker D2S123 has not yet been directly determined. The approximate locations for the markers used in multipoint linkage analyses were obtained from a genome mapping effort including a total of 96 markers from chromosome 2. Our analysis placed locus D2S123 at a position 5 cM distal to D2S5, which has been mapped to 2p15-16 by in situ hybridization, linkage, and somatic cell hybrid analysis (*14*). Thus, the most likely physical location of the HNPCC gene is 2p15-16.

The linkage identified is unequivocal. Changes in different parameters had little effect on the results. We varied the criteria used for classifying individuals as affected or unaffected (high versus low stringency) and the HNPCC gene frequency (0.001 versus 0.01 or 0.0005). In the J family, low-stringency criteria resulted in an increase in the lod scores for D2S123 and a decrease in those for D2S136; in family C, the change in stringency did not affect the results (Table 1). Variation of the HNPCC gene frequency had almost no influence on the lod scores (for example, when we used a frequency of 0.01, the maximum lod scores for D2S123 were 6.18 in family C and 1.38 in family J, both at zero recombination.

The localization of an HNPCC gene to chromosome 2 in kindreds from two differ-

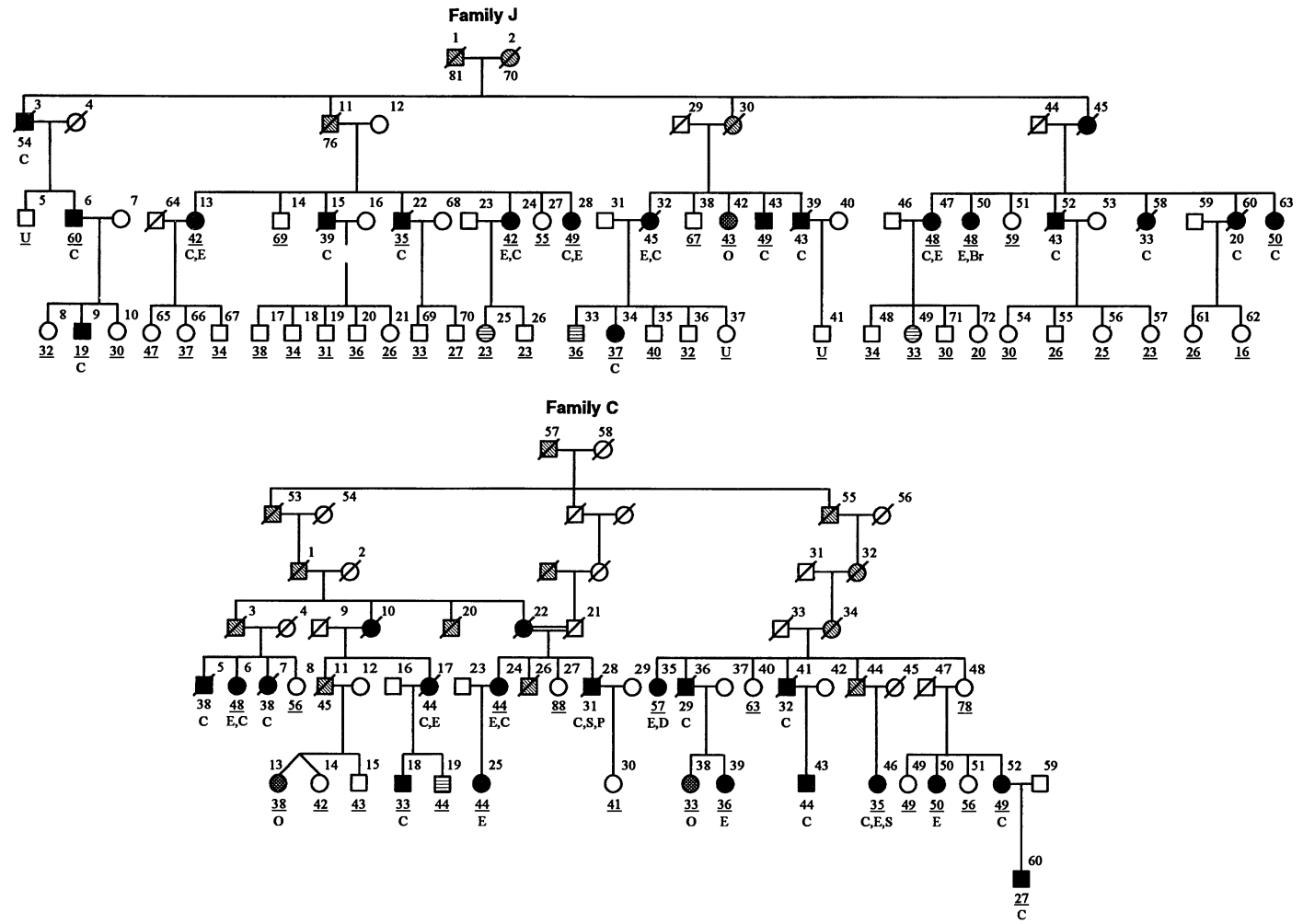

Fig. 1. Pedigrees of families J and C. Numbers above the symbols are patient identifiers; numbers below the symbols indicate age at tumor diagnosis or age at last observation if unaffected; "U" indicates age unknown. If a blood sample was available, the age is underlined. DNA samples from most of the spouses were also available but are not indicated on the pedigrees. Letters signify the site of the tumor: C, colon or rectum; E, endometrium; S, stomach; D, duodenum; O, ovary; P, pancreas; and Br, breast. Marker alleles were omitted to protect confidentiality of family members. General symbols: squares, males; circles, females; all symbols with a diagonal, deceased. Open symbols, no neoplasm detected; solid, colorectal or endometrial carcinoma; dotted, other tumor of HNPCC spectrum; horizontal stripes, colorectal adenoma; and diagonal stripes, diagnosis uncertain.

ent continents proves the existence of a genetically determined form of this disease. Many precedents suggest that any gene that plays a role in a hereditary form of cancer is likely to be involved in a significant number of "sporadic" cases of the same tumor type, by virtue of somatic mutations, germ line alterations, or inherited alterations that are poorly penetrant (*15*). The current study lays the groundwork for further mapping, isolation, and characterization of the HNPCC gene on chromosome 2, and the examination of its role in colorectal and other tumors. If, as expected, this gene is involved in a significant fraction of familial CRC cases, then it is responsible for one of the most common forms of heritable disease yet identified in humans (*5*, *16*). The study of such a gene has obvious implications for public health. Most immediately, these results make presymptomatic diagnosis of susceptibility possible in HNPCC family members and will enable more effective surveillance programs for the early detection and treatment of cancer.

REFERENCES AND NOTES

1. A. S. Warthin, *Arch. Intern. Med.* **12**, 546 (1913).
2. B. Armstrong and R. Doll, *Int. J. Cancer* **15**, 617 (1975); W. C. Willett and B. MacMahon, *N. Engl. J. Med.* **310**, 697 (1984); W. Willett, *Nature* **338**, 389 (1989).
3. J. J. Mulvihill, in *Prevention of Hereditary Large Bowel Cancer*, J. R. Ingall and A. J. Mastromarino, Eds. (Liss, New York, 1983), pp. 61–75; H. J. Järvinen, *Gut* **33**, 357 (1992).
4. K. W. Kinzler *et al.*, *Science* **253**, 661 (1991); I. Nishisho *et al.*, *ibid.*, p. 665; J. Groden *et al.*, *Cell* **66**, 589 (1991); G. Joslyn *et al.*, *ibid.*, p. 601.
5. H. T. Lynch *et al.*, *Cancer* **56**, 939 (1985); J.-P. Mecklin, *Gastroenterology* **93**, 1021 (1987); F. Kee and B. J. Collins, *Gut* **32**, 509 (1991); J. R. Jass and S. M. Stewart, *ibid.* **33**, 783 (1992); R. S. Houlston, A. Collins, J. Slack, N. E. Morton, *Ann. Hum. Genet.* **56**, 99 (1992).
6. H. F. A. Vasen, J.-P. Mecklin, P. Meera Khan, H. T. Lynch, *Dis. Colon Rectum* **34**, 424 (1991).
7. H. T. Lynch *et al.*, *Cancer Genet. Cytogenet.* **53**, 143 (1991); J. P. Mecklin and H. J. Järvinen, *Cancer* **68**, 1109 (1991).
8. P. Peltomaki *et al.*, *Cancer Res.* **51**, 4135 (1991); P. Peltomaki *et al.*, *ibid.* **52**, 4530 (1992).
9. J. L. Weber and P. E. May, *Am. J. Hum. Genet.* **44**, 388 (1989).
10. J. Weissenbach *et al.*, *Nature* **359**, 794 (1992).
11. For most microsatellite markers, the polymerase chain reaction (PCR) was done in the following final reaction conditions: 1 × PCR buffer [10 mM tris (pH 8.3), 1.5 mM $MgCl_2$, 50 mM KCl, 0.01% bovine serum albumin]; 200 μM deoxyguanosine 5′-triphosphate, deoxyadenosine 5′-triphosphate, and deoxythymidine triphosphate; 2 μM deoxycytidine 5′-triphosphate (dCTP); 0.7 μCi of α-[^{32}P]dCTP (3000 Ci/mmol); 10 ng of each primer (1 to 5 primer pairs at a time); 30 ng of genomic DNA template; and 0.3 units of AmpliTaq polymerase (Perkin-Elmer–Cetus) in a volume of 10 μl. Twenty-seven cycles were performed at 94°C for 30 s, 55°C for 75 s, and 72°C for 15 s, and the samples were incubated at 72°C for 6 min after the last cycle. The PCR products were subjected to electrophoresis in 6% polyacrylamide gels containing 7.7 M urea. After electrophoresis, the gels were fixed in 10% acetic acid for 15 min, dried, and exposed to x-ray film.
12. T Hakala, J.-P. Mecklin, M. Forss, H. Järvinen, P. Lehtovirta, *Cancer* **68**, 1656 (1991).
13. G. M. Lathrop, J.-M. Lalouel, C. Julier, J. Ott, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3443 (1984).
14. Y. Shiloh *et al.*, *Nucleic Acids Res.* **13**, 5403 (1985); H. Donis-Keller *et al.*, *Cell* **51**, 319 (1987); A. J. Pakstis, C. M. Castiglione, J. R. Kidd, *Cytogenet. Cell Genet.* **51**, 1057 (1989).
15. E. J. Stanbridge, *Annu. Rev. Genet.* **24**, 615 (1990); R. A. Weinberg, *Science* **254**, 1138 (1991); A. G. Knudson, *Cancer Res.* **45**, 1437 (1985); E. R. Fearon and B. Vogelstein, *Cell* **61**, 759 (1990).
16. L. A. Aaltonen *et al.*, *Science* **260**, 812 (1993).
17. Phenotypes were coded as affected with an autosomal dominant mode of inheritance and four liability classes according to age at the time of observation (healthy) or at diagnosis. Penetrances for heterozygotes were set at 0.15 at age 30 years or below (liability class 1), 0.40 at age 31 to 45 years (liability class 2), 0.70 at age 46 to 60 years (liability class 3), and 0.90 at age 61 or over (liability class 4). Phenocopies were introduced, with frequencies set at 1, 3, and 5% in liability classes 2, 3, and 4, respectively. A model without phenocopies was also tested and was found to have only a minor effect on the results. The frequency of the HNPCC gene was set at 0.001 (but values of 0.01 and 0.0005 were also tested; see text). The allele frequencies of markers D2S119, D2S123, and D2S136 were calculated from 53 to 58 spouses of HNPCC family members.

18. We thank the patients and their families who provided specimens for this study with informed consent, S. Booker and S. Stewart for organizing patient accrual and clinical datas, and C. Wadelius for providing primers for microsatellite markers. Supported by grants from the Sigrid Juselius Foundation, the Academy of Finland, the Finnish Cancer Society, the Duodecim Foundation, the Ida Montin Foundation, the Finnish-Norwegian Medical Foundation, Suomen Astra Oy, the Finnish Foundation for Gastroenterological Research, the Auckland Medical Research Foundation, the Cancer Society of New Zealand, the Health Research Council of New Zealand, the Clayton Fund, the McAshan Fund, and NIH grants HG 00248, CA 47527, and CA 35494. B.V. is an American Cancer Society research professor. Part of this study was carried out at the Folkhälsan Institute of Genetics.

8 April 1993; accepted 15 April 1993

Clues to the Pathogenesis of Familial Colorectal Cancer

Lauri A. Aaltonen,* Päivi Peltomäki,* Fredrick S. Leach,* Pertti Sistonen, Lea Pylkkänen, Jukka-Pekka Mecklin, Heikki Järvinen, Steven M. Powell, Jin Jen, Stanley R. Hamilton, Gloria M. Petersen, Kenneth W. Kinzler, Bert Vogelstein,† Albert de la Chapelle†

A predisposition to colorectal cancer is shown to be linked to markers on chromosome 2 in some families. Molecular features of "familial" cancers were compared with those of sporadic colon cancers. Neither the familial nor sporadic cancers showed loss of heterozygosity for chromosome 2 markers, and the incidence of mutations in *KRAS, P53*, and *APC* was similar in the two groups of tumors. Most of the familial cancers, however, had widespread alterations in short repeated DNA sequences, suggesting that numerous replication errors had occurred during tumor development. Thirteen percent of sporadic cancers had identical abnormalities and these cancers shared biologic properties with the familial cases. These data suggest a mechanism for familial tumorigenesis different from that mediated by classic tumor suppressor genes.

Colorectal cancer (CRC) is one of the most common forms of neoplasia in industrial countries and its incidence is increasing (*1*). Familial forms of CRC account for up to 13% of all CRC cases (*2*). Except for the rare familial adenomatous polyposis (FAP) syndrome, the molecular basis of the cancers that arise in these families has not been previously studied but is of great interest for both scientific and practical reasons. Inherited forms of cancer can provide unique insights into the pathogenesis of specific tumor types that often extend to nonfamilial cases (*3*). In the studies reported here, we evaluated molecular aspects of tumorigenesis in families with hereditary nonpolyposis colorectal cancer (HNPCC). This study revealed surprising findings that are relevant not only to this familial form of cancer but also to individuals with no strong family history of the disease.

HNPCC is defined by the presence of three or more family members with CRC in at least two successive generations, with at

L. A. Aaltonen, P. Peltomaki, L. Pylkkanen, A. de la Chapelle, Department of Medical Genetics, P.O. Box 21, University of Helsinki, SF-00014 Helsinki, Finland.
F. S. Leach, S. M. Powell, J. Jen, S. R. Hamilton, G. M. Petersen, K. W. Kinzler, B Vogelstein, Departments of Oncology, Pathology, Surgery, and Epidemiology, Johns Hopkins University School of Medicine and School of Public Health and Hygiene, and Johns Hopkins Hospital, Baltimore, MD 21231.
P. Sistonen, Department of Medical Sciences, University of Helsinki, and Finnish Red Cross Blood Transfusion Service, SF-00310 Helsinki, Finland.
J.-P. Mecklin, Department of Surgery, Jyvaskyla Central Hospital, SF-40620 Jyvaskylä, Finland.
H. Jarvinen, Second Department of Surgery, Helsinki University Central Hospital, SF-00290 Helsinki, Finland.

*These authors contributed equally to this work.
†To whom correspondence should be addressed.