The Gene for Familial Polyposis Coli Maps to the Long Arm of Chromosome 5

M. LEPPERT, M. DOBBS, P. SCAMBLER, P. O'CONNELL, Y. NAKAMURA, D. STAUFFER, S. WOODWARD, R. BURT, J. HUGHES, E. GARDNER, M. LATHROP, J. WASMUTH, J.-M. LALOUEL, R. WHITE*

The inherited genetic defect in adenomatous polyposis has been localized to a small region on the long arm of chromosome 5. Sixteen DNA marker loci were used to construct a linkage map of the chromosome. When five kindreds segregating a gene for adenomatous polyposis coli were characterized with a number of the markers, significant linkage was found between one marker and the disease gene. Linkage analysis determined the location of the defective gene within a primary genetic map of chromosome 5.

INHERITED CANCER SYNDROMES OFFER an opportunity to identify genetic mechanisms of carcinogenesis. The chromosomal locations of the germline mutations involved can be identified by genetic linkage studies in families segregating rare cancer syndromes.

Colon cancer is one of the most common cancers in the North American population; the lifetime risk of occurrence approaches 5%. The natural history of this tumor includes a relatively well defined intermediate stage, the adenomatous polyp, that defines a discrete step in the pathway of carcinogenesis. Carcinomas are believed to arise within the adenomatous polyp. Several distinct, inherited predispositions to colon cancer have been identified, providing an opportunity to map, and then clone, these genes. One of the most widely recognized predispositions imposes a striking phenotype of multiple, adenomatous, colonic polyps. This condition is observed in two syndromes, familial polyposis coli (FPC) and Gardner syndrome (GS) (*1*). While GS and FPC are similar with respect to colonic polyposis and colorectal cancer, individuals with GS are distinguished by a number of benign extraintestinal growths. These include multiple osteomas, particularly of the mandible and skull, epidermoid cysts, desmoid tumors (*2*), and congenital hypertrophy of the retinal pigment epithelium (*3*). However, the expression of extraintestinal manifestations is variable, and it has become increasingly difficult to clearly separate GS from FPC on clinical grounds, partly because subtle extracolonic lesions in FPC patients are recognized more often now than in the past. Together, the syndromes have an incidence in European and North American populations of 1 in 12,000 (*4*). The syndromes show an autosomal dominant mode of inheritance and are characterized by the appearance of hundreds of adenomatous polyps in the colon, often by the third decade of life. Surgically untreated individuals are at high risk for developing colorectal adenocarcinoma by age 40 (*5*). Because progression from the adenomatous polyp to carcinoma appears similar in both inherited and sporadic colon cancer, the occurrence of numerous adenomatous polyps in young patients with FPC may account for their increased risk of carcinoma (*6, 7*).

After the discovery by Herrera *et al.* (*8*) of a constitutional interstitial deletion on chromosome 5q of the region q13–15 or q15–22 in a mentally retarded patient who exhibited FPC as well as a large desmoid tumor, we focused our attention on DNA probes from the long arm of chromosome 5. We now have evidence supporting the localization of the FPC-GS locus to the long arm of chromosome 5, probably in the region of 5q22.

Families for our study were ascertained at the University of Utah Medical Center from probands seeking medical care for polyposis. Medical status was determined from records documenting colonoscopic and surgical pathology; individuals lacking medical records were classified as unknown. Affected individuals had developed hundreds of colonic polyps during their second or third decade, and untreated family members from earlier generations often had died of documented colon cancer in their forties. Affected individuals had lesions typical of GS in pedi-

Table 1. Lod scores for linkage of three chromosome 5 markers to FPC-GS. A two-allele system was detected for pC11p11 by the enzyme Taq I, giving allelic frequencies of 0.17 and 0.83 for the 4.2- and 2.7-kb fragments, respectively, in our collection of 59 normal linkage families. pTP5E also defines a two-allele system with Taq I, with frequencies of 0.77 and 0.23 for the 11.0- and 5.2-kb alleles. pJO205E-C defines a three-allele system (Msp I) with frequencies of 0.43, 0.35, and 0.22 for the 7.0-, 3.9-, and 3.7-kb fragments. An autosomal dominant mode of inheritance with incomplete penetrance was assumed. Although a frequency of 0.0001 was assumed for the disease-predisposing allele, our results remained constant over a large range of allele frequencies. Penetrance was assigned at 0.5 for ages 0 to 20, 0.75 for ages 21 to 40, and 1.0 for ages 41 and over. These values were estimated from the observed distribution of age at diagnosis among affected members of our pedigrees. No provision was made to allow for sporadic cases.

Pedigree	Lod score at recombination fraction of:					
	0.00	0.05	0.10	0.20	0.30	0.40
			C11p11			
K415	0.00	0.00	0.00	0.00	0.00	0.00
K109	0.04	0.04	0.03	0.02	0.01	0.00
K1452	2.34	2.17	1.96	1.48	0.91	0.32
K1441	0.98	0.84	0.70	0.44	0.22	0.07
K1498	0.00	0.00	0.00	0.00	0.00	0.00
Total	3.37	3.04	2.69	1.94	1.14	0.39
			TP5E			
K415	0.72	0.65	0.51	0.43	0.27	0.12
K109	−2.36	0.69	0.95	0.86	0.52	0.15
K1452	−0.31	−0.24	−0.10	0.09	0.11	0.04
K1441	−2.56	−0.77	−0.29	−0.20	−0.09	−0.04
K1498	0.00	0.00	0.00	0.00	0.00	0.00
Total	−4.51	0.34	0.98	−1.19	−0.81	−0.27
			JO205E-C			
K415	0.43	0.36	0.30	0.15	0.03	−0.03
K109	−6.15	−1.04	−0.26	0.27	0.30	0.16
K1452	−13.3	−3.28	−1.74	−0.95	−0.06	0.00
K1441	−10.1	−1.52	−0.20	0.72	0.82	0.51
K1498	−2.19	−0.78	−0.50	−0.22	−0.09	−0.02
Total	−31.3	−6.26	−2.40	0.11	0.86	0.59

M. Leppert, P. O'Connell, Y. Nakamura, D. Stauffer, S. Woodward, M. Lathrop, J.-M. Lalouel, R. White, Howard Hughes Medical Institute and Department of Human Genetics, University of Utah Medical Center, Salt Lake City, UT 84132.
M. Dobbs and J. Wasmuth, Department of Biological Chemistry, California College of Medicine, Irvine, CA 92717.
P. Scambler, St. Mary's Hospital Medical School, Department of Biochemistry, University of London, London W2 1PG, U.K.
R. Burt, Division of Gastroenterology, University of Utah Health Sciences Center, Salt Lake City, UT 84132.
J. Hughes, St. Mark's Hospital, Salt Lake City, UT 84117.
E. Gardner, Department of Biology, Utah State University, Logan, UT 84321.

*To whom correspondence should be addressed.

grees K109 (16 of 41 typed individuals) and in K1498 (5 of 7). The presentation was consistent with familial polyposis coli in pedigrees K1452 (15 of 68 typed individuals), K1441 (22 of 58), and K415 (3 of 16).

A primary genetic map of chromosome 5 (Fig. 1) was constructed with 16 DNA markers, through the GENE MAPPING system (*9*). This computer program derives the most likely order of the loci by an automated strategy that identifies and examines critical alternate orders. For each order examined, linkage analysis was performed with the LINKAGE programs (*10*), and the likelihood of the order was contrasted with competing alternatives. A test for sex-specific differences in recombination frequencies, allowing for a constant female:male ratio along the map, was significant (χ_1^2 = 12.06), leading to an estimate of 1.6 for the female:male ratio of genetic distance. When the effect of sex was allowed to vary along the map, no further significance was obtained. The linked loci formed a continuous map of the chromosome, spanning 196 centimorgans (cM) in males and 324 cM in females. Each marker locus showed strong evidence of linkage to its nearest neighbor [all lod (logarithm of the odds) scores >5.0].

Several of the markers were tested for linkage in the FPC-GS pedigrees. Three showed evidence of linkage to the disease gene; pairwise lod scores between each of the three markers and FPC-GS are reported in Table 1. The maximum lod score of 3.37 at a recombination fraction of 0, observed with the marker pC11p11, provided significant support for genetic linkage. The two other markers, pJO205E-C and pTP5E, yielded positive but nonsignificant lod scores when tested independently.

Evidence for the location of the FPC-GS gene on the genetic map of the long arm of chromosome 5 was provided by the LINKMAP routine (*11*), which calculated the likelihood of the position of the FPC-GS locus at each of a number of locations along the established genetic map of chromosome 5. The most likely location for the disease gene was very close to C11p11; the 1–lod unit limit of support ranged from 9 cM proximal to 36 cM distal to the C11p11 locus on the female map (Fig. 1, top). The relative likelihood of this location as compared to a distant location is 10^5:1; thus, the multilocus analysis adds confidence to the estimate of map location for the disease gene. Preliminary results of physical mapping of the C11p11 locus with somatic cell hybrids (*12*) have been consistent with the placement of the FPC-GS locus in the region implicated by the 5q deletion.

Localization of the polyposis gene to a specific chromosomal region is a major step toward identifying and isolating the gene. It should now be possible to identify close, flanking DNA markers that will specify the physical region containing the gene, and to clone segments from the region that may identify genes expressed as messenger RNA in normal mucosa, polyps, or carcinoma. Such genes become strong candidates for the polyposis gene.

Multiple genetic events must occur in the pathway to colon cancer. For example, mutation at the K-*ras*2 locus on chromosome 12 is observed more than a third of the time in either pure colonic carcinomas or adenomatous regions of cancerous polyps (*13*). Other genes affecting carcinogenesis may be uncovered by the approach described here. Of particular interest are the genes involved in hereditary nonpolyposis colorectal cancers, Lynch syndromes I and II (*14*), and in several possibly distinct genetic forms of polyposis disease such as Turcot syndrome (*15*). These inherited predispositions to colon and other forms of cancer could reflect mutations in different genes along the same genetic pathway. Moreover, large pedigrees that may be segregating a common gene predisposing to solitary polyps of the colon have been identified (*16, 17*). If such a gene exists, its identification would be relevant for medical screening because of the frequent occurrence of this clinical condition.

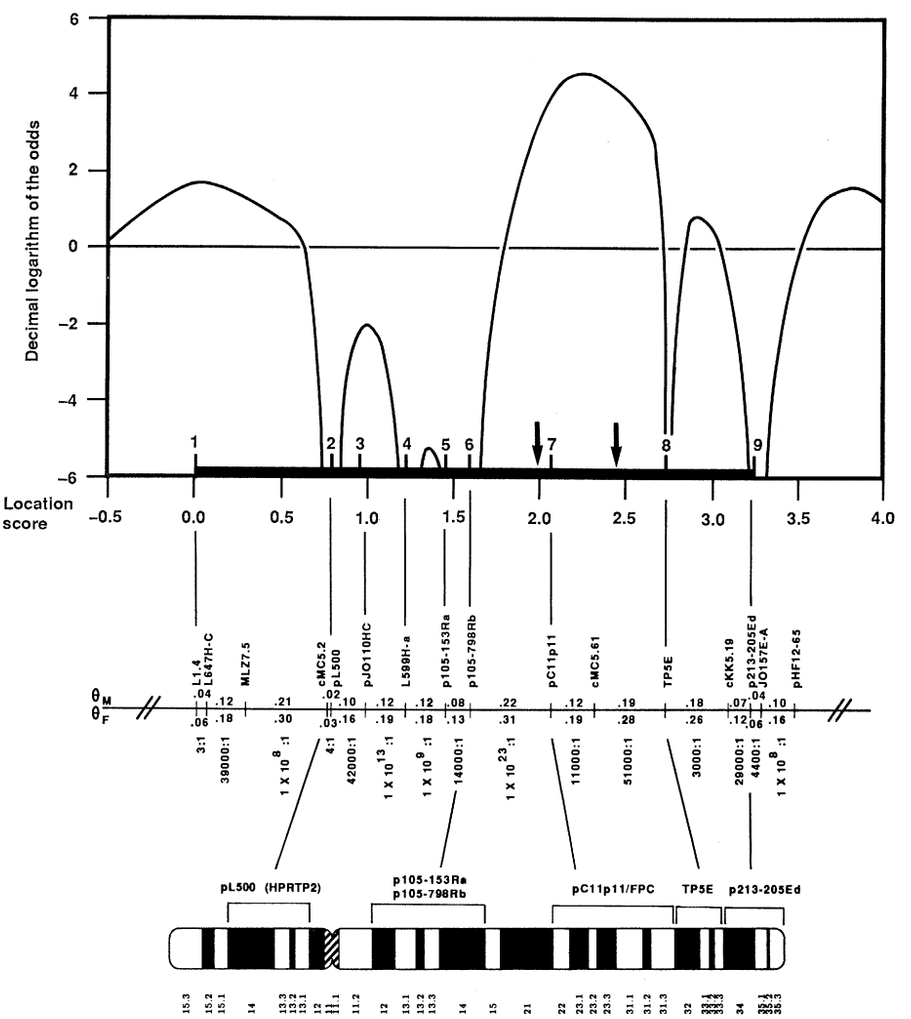

Fig. 1. Genetic linkage map of 16 DNA markers on chromosome 5, with location scores (top) for the FPC-GS gene on the female map. Two arrows on the abscissa indicate the 95% confidence boundaries for the location of the FPC-GS locus. The genetic map (center), also scaled to female genetic disease (in Morgans), indicates estimates of recombination distance (θ) between markers for males (above line) and females (below line). Genotypes were obtained as described (*19, 20*) in a panel of 59 normal reference families (*21, 22*). The most likely gene order was determined with the GENE MAPPING system (*9*), under the assumption of a constant ratio of male:female map distance along the entire chromosome. The physical localizations of selected probes (*12*), carried out on cell lines described (*23, 24*) are indicated on the ideogram of chromosome 5. Descriptions of the polymorphisms and sources for the DNA probes are available from the RFLP and PROBE databases at the Yale Human Gene Mapping Library, 25 Science Park, New Haven, Connecticut 06511.

Resolution of whether the GS and FPC syndromes are due to lesions in the same gene should now be possible with linked genetic markers. Although our studies have not yet permitted a rigorous conclusion, very little of our information was derived from the two families segregating Gardner's syndrome, K109 and K1498. Moreover, clinical studies in extended pedigrees (*5*), have suggested that FPC is a single genetic disease with a wide spectrum of phenotypic expression. An accurate gene marker for FPC-GS would be a useful diagnostic tool for directing clinical screening efforts, such as colonoscopy, to the individuals who are at risk.

Note added in proof: Since submission of this paper, a similar finding has been reported by Bodmer *et al.* (*18*).

REFERENCES AND NOTES

1. E. J. Gardner, *Am. J. Hum. Genet.* **3**, 167 (1951).
2. ———, *ibid.* **14**, 376 (1962).
3. R. A. Lewis, W. E. Crowder, L. A. Eierman, R. L. Nussbaum, R. E. Ferrell, *Ophthalmology* **91**, 916 (1984).
4. A. L. Watne, *Curr. Probl. Cancer* 7, 3 (1982).
5. H. J. R. Bussey, *Familial Polyposis Coli* (Johns Hopkins Univ. Press, Baltimore, 1975), p. 50.
6. C. E. Dukes, *Ann. Eugen. London* **17**, 1 (1952).
7. A. M. O. Veale, *Intestinal Polyposis* (Monograph, 101 pp.) Eugenics Laboratory Memoirs XL, London (Cambridge Univ. Press, New York, 1965).
8. L. Herrera, S. Kakati, L. Gibas, E. Pietrzak, A. A. Sandberg, *Am. J. Med. Genet.* **25**, 473 (1986).
9. G. M. Lathrop, P. Cartwright, J. M. Lalouel, *Cytogenet. Cell Genet.*, in press. [abstract from Human Gene Mapping 9 Workshop, Paris (September 1987)].
10. G. M. Lathrop, J. M. Lalouel, C. Julier, J. Ott, *Am. J. Hum. Genet.* **37**, 482 (1985).
11. ———, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3443 (1984).
12. J. Wasmuth, unpublished data.
13. J. L. Bos *et al.*, *Nature (London)* **327**, 293 (1987).
14. H. T. Lynch, *Cancer* **56**, 934 (1985).
15. J. Turcot, J. P. Despres, F. Pierre, *Dis. Colon Rectum* **2**, 465 (1959).
16. S. R. Woodward, E. J. Gardner, R. W. Burt, L. K. Neff, *Encyclia* **58**, 21 (1983).
17. R. Burt *et al.*, *N. Engl. J. Med.* **312**, 1540 (1985).
18. W. F. Bodmer *et al.*, *Nature (London)* **328**, 614 (1987).
19. W. Cavenee, T. Mohandas, P. Pearson, R. White, *Am. J. Hum. Genet.* **36**, 10 (1984).
20. K. C. Reed and D. A. Mann, *Nucleic Acids Res.* **13**, 7207 (1985).
21. J. Dausset, *La Presse Medicale* **15**, 1801 (1986).
22. R. White *et al.*, *Nature (London)* **313**, 101 (1985).
23. L. R. Carlock, D. Skarecky, S. L. Dana, J. J. Wasmuth, *Am. J. Hum. Genet.* **37**, 839 (1985).
24. J. Overhauser, A. Beaudet, J. J. Wasmuth, *ibid.* **39**, 562 (1986).

25. We thank L. Ballard, M. Culver, J. R. Eldridge, J. Gill, S. Gillilan, M. Hadley, L. Jerominski, C. Martin, M. Mitchell, L. Nelson, B. Ogden, L. Rowe, L. Sargeant, T. Sears, K. Smith, and J. Stevens for technical assistance in this project. R. Foltz assisted in the editing and preparation of the manuscript. Supported in part by NIH grant CA40641 (R.B.). R.W. and J.-M.L. are investigators of the Howard Hughes Medical Institute.

23 July 1987, accepted 14 October 1987

Bacteriophage M13 Procoat Protein Inserts into the Plasma Membrane as a Loop Structure

ANDREAS KUHN

The major coat protein of bacteriophage M13 is synthesized as a precursor, the procoat, with a typical leader (signal) sequence of 23 residues at its NH_2-terminus. A fusion protein that contains the NH_2-terminal 141 residues of cytoplasmic ribulokinase and all but the first ten residues of M13 procoat was made. The fusion protein inserts into the plasma membrane of *Escherichia coli* and is processed by leader peptidase to give rise to a leader peptide of 155 residues and the mature coat protein of 50 residues. The NH_2-terminus of the leader peptide remains in the cytoplasm and is protected from protease added to the medium outside of the cell. This indicates that M13 procoat inserts into the membrane as a loop structure and that the NH_2-terminus of a leader peptide remains within the cytoplasm during membrane insertion.

INSERTION OF PROTEINS INTO THE membrane is mediated either by a cleavable leader (signal) sequence or by an uncleaved signal or insertion sequence (*1*). Cleavable leader sequences of bacterial proteins and proteins inserted across the endoplasmic reticulum are located at the NH_2-termini of pre-proteins, are structurally similar, and share a similar length (13 to 26 residues). Each one consists of a polar, basic region at the NH_2-terminus followed by a stretch of hydrophobic residues and a region coding for the recognition site for the processing enzyme called leader peptidase (*2*). Since the enzymatically active site of the leader peptidase is located on the outer surface of the inner membrane (*3*), only those precursor proteins that are translocated across the membrane are processed to their mature form (*4, 5*). Mutations that alter the hydrophobicity of the leader region inhibit membrane insertion and consequently inhibit processing (*5, 6*). Although the involvement of leader sequences in initiating membrane insertion is now well established, the molecular mechanism of this process is unknown.

The insertion of the coat protein of bacteriophage M13 into the inner membrane of *Escherichia coli* has been extensively studied as a model system (*7, 8*). The coat protein is synthesized in the cytoplasm as a precursor protein, the procoat, with a typical 23-residue cleavable leader sequence at its NH_2-terminus. Mutations that alter the hydrophobic region of the procoat leader strongly inhibit its membrane insertion (*5*). In addition, it has recently been shown that the procoat leader can substitute for the leader sequence of the outer membrane protein A (OmpA) without affecting its membrane insertion (*9*). We have proposed that the hydrophobic regions in the leader and in the mature (membrane anchor) regions of M13 procoat directly partition into the hydrophobic core of the membrane bilayer (*4, 5*). Transmembrane electrochemical potential is then required to translocate the central, acidic region of the procoat across the membrane, leaving the NH_2- and COOH-termini at the inner face of the membrane (Fig. 1A). If this model is correct, fusion of other polar peptides to the NH_2- or COOH-terminus of procoat should not interfere with its insertion into the membrane. In contrast, if this protein inserts linearly across the plasma membrane with its leader peptide first, fusion of extra, polar residues on the NH_2-terminus would block membrane assembly.

To test if the NH_2-terminus of procoat remains in the cytoplasm, I have constructed a fusion protein of 205 amino acids (M_r 21,955) that encompasses the first 141 amino acids of ribulokinase (*ara*B), 63 amino acids of procoat, and a glycyl residue at the fusion point (Fig. 1B). In this construction, the first 10 amino acids of the procoat leader were deleted.

Expression of this plasmid-encoded fusion protein was induced by addition of L-arabinose into the growth medium, and the cells were pulse labeled with [^{35}S]methionine. The fusion protein was immunoprecipitated by antibodies to ribulokinase (Fig. 2, lane 1) and was of higher molecular weight than the nonfused ribulokinase fragment of 162 amino acids (Fig. 2, lane 4). In addition, a small amount of pulse-labeled ribulokinase was seen at a lower molecular weight; its electrophoretic mobility suggests that it is the result of cleavage of the fusion protein. When the same sample was immunoprecipitated with antibody to M13 coat protein (Fig. 2, lane 3), both the full-length fusion protein and a protein which comigrated with M13 coat protein (see markers in Fig. 2, lane 2) were observed. As expected, the ribulokinase fragment itself was not immu-

Microbiology Department, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland.