Identification of a Chromosome 18q Gene That Is Altered in Colorectal Cancers

ERIC R. FEARON, KATHLEEN R. CHO, JANICE M. NIGRO, SCOTT E. KERN, JONATHAN W. SIMONS, J. MICHAEL RUPPERT, STANLEY R. HAMILTON, ANTONETTE C. PREISINGER, GILES THOMAS, KENNETH W. KINZLER, BERT VOGELSTEIN*

Allelic deletions involving chromosome 18q occur in more than 70 percent of colorectal cancers. Such deletions are thought to signal the existence of a tumor suppressor gene in the affected region, but until now a candidate suppressor gene on this chromosomal arm had not been identified. A contiguous stretch of DNA comprising 370 kilobase pairs (kb) has now been cloned from a region of chromosome 18q suspected to reside near this gene. Potential exons in the 370-kb region were defined by human-rodent sequence identities, and the expression of potential exons was assessed by an "exon-connection" strategy based on the polymerase chain reaction. Expressed exons were used as probes for cDNA screening to obtain clones that encoded a portion of a gene termed DCC; this cDNA was encoded by at least eight exons within the 370-kb genomic region. The predicted amino acid sequence of the cDNA specified a protein with sequence similarity to neural cell adhesion molecules and other related cell surface glycoproteins. While the DCC gene was expressed in most normal tissues, including colonic mucosa, its expression was greatly reduced or absent in most colorectal carcinomas tested. Somatic mutations within the DCC gene observed in colorectal cancers included a homozygous deletion of the 5′ end of the gene, a point mutation within one of the introns, and ten examples of DNA insertions within a 0.17-kb fragment immediately downstream of one of the exons. The DCC gene may play a role in the pathogenesis of human colorectal neoplasia, perhaps through alteration of the normal cell-cell interactions controlling growth.

CURRENT THEORIES ON THE ORIGIN OF HUMAN CANCER suggest that it arises, at least in part, from the accumulation of genetic changes in a single cell (*1*). These genetic alterations include the somatic activation of cellular oncogenes through point mutation, rearrangement, or amplification (*2*), and the germline or somatic inactivation of tumor suppressor genes through point mutation or deletion (or both) (*3*). This model is supported by studies of human colorectal tumors demonstrating that the progressive accumulation of genetic changes in both oncogenes and tumor suppressor genes parallels the clinical progression of colorectal tumors from normal epithelium to benign tumors and further to the malignant stage of the disease (*4*). Specifically, mutations in the Ki-*ras* gene are present in approximately 50 percent of large colorectal adenomas and a similar percentage of colorectal carcinomas (*4, 5*). In addition, the loss of sequences on several chromosomes commonly occurs in colorectal cancers; the most frequently affected regions are chromosomes 17p and 18q, each of which is lost in more than 70 percent of carcinomas (*4, 6–8*).

Chromosomal losses in human tumors have been regarded as evidence that the affected regions contain tumor suppressor genes. The tumor suppressor gene represents the critical gene (target) of the allelic loss event (*3*). Such genes have been postulated to encode proteins that regulate normal growth, and thus indirectly suppress neoplastic development. In cells, these genes may act recessively, so that both maternal and paternal copies of the gene product must be inactivated in order for the suppressor function to be eliminated. This model for tumor suppressor genes, originally postulated by Knudson (*3*), has gained support from the study and molecular cloning of the retinoblastoma (RB) gene on chromosome 13q (*9*). Recent evidence suggests that the target of chromosome 17p loss in colorectal tumors is the p53 gene (*10*). The frequent loss of chromosome 18q sequences in colorectal carcinomas would then imply that this region might also contain a tumor suppressor gene, which we sought to identify.

Somatic genetic changes detected on chromosome 18q. We have previously localized the common region of deletion on chromosome 18 to 18q21-qter (*4*). Several genes have been mapped to this region, including the *bcl*-2 gene (*11*), the gastrin-releasing peptide gene (*12*), and the cellular homologue of the *yes*-1 oncogene (*13*). Using probes for these genes and several others from the region, we did not detect any abnormalities in DNA (Southern) blots in any colorectal carcinoma sample tested. However, we found two tumors with interesting hybridization patterns when DNA from these cases was probed with p15-65, a DNA fragment from the D18S8 locus previously mapped to 18q21.3 (*14*). One carcinoma (S115) showed a complete (homozygous) deletion of the sequences detected by this probe. Both the 10.5- and 7.8-kb alleles

E. R. Fearon, J. M. Nigro, J. W. Simons, J. M. Ruppert, A. C. Preisinger, and B. Vogelstein are at the Oncology Center, The Johns Hopkins University School of Medicine, Baltimore, MD 21231. K. R. Cho, S. E. Kern, and S. R. Hamilton are at the Oncology Center and the Department of Pathology, The Johns Hopkins University School of Medicine and Hospital, Baltimore, MD 21205. G. Thomas is at The Institut de Curie, 75231 Paris CX5, France.

*To whom correspondence should be addressed.

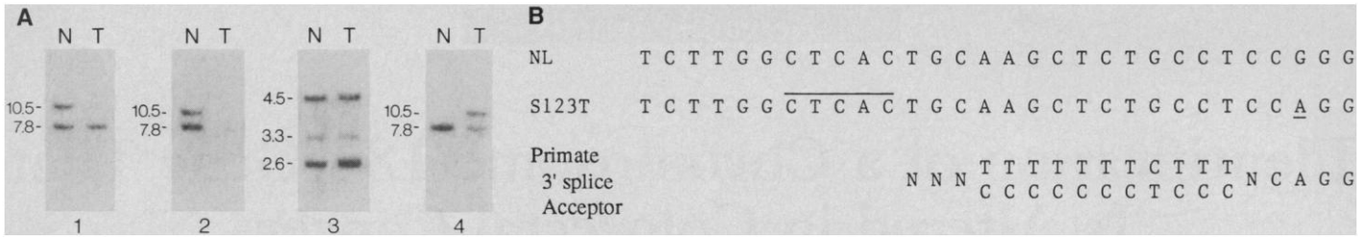

Fig. 1. Alterations in colorectal tumors at a chromosome 18q locus. (**A**) DNA was isolated from cryostat sections (*6*) of tumor tissue (T) or adjacent normal colonic mucosa (N) and Southern blots were prepared as described (*44*). Panel 1 shows an example of allelic loss in a colorectal tumor at the locus detected by p15-65 (*14*). Two alleles were present in the normal DNA, but only the 7.8-kb allele was present in the DNA from the tumor. Panel 2 shows a homozygous loss of sequences detected by p15-65 in tumor S115. In panel 3, rehybridization of this blot to a probe from another chromosome (*45*) demonstrated that approximately equal amounts of DNA were placed in lanes containing either normal or tumor DNA's. Panel 4 shows the gain of heterozygosity at the same locus in the S123 tumor; a 10.5-kb fragment was present in tumor DNA that was not present in DNA from the normal tissue of the same patient. The sizes of the DNA fragments in kilobases are indicated to the left of the panels. (**B**) Sequences at the Msp I site altered in the S123 tumor. The sequence of a normal allele and the allele from the S123 tumor that could not be cleaved by Msp I are shown. The single difference noted was a substitution of A for G in the Msp I cleavage site of the tumor allele; this mutation is underlined. The consensus sequence of the 3′ splice acceptor site for primate genes (*18*) is also shown. The abbreviation N indicates any nucleotide, and where two nucleotides are specified for a position, the one above is more common than the one below. A potential lariat site (CTCAC) (*19*) is indicated by a line over the sequence. The normal allele and the S123 tumor allele were cloned and sequenced as described (*46*).

present in the non-neoplastic tissue of the patient were lost in the tumor (Fig. 1A, panel 2). Rehybridization of this filter to a polymorphic probe derived from another chromosome served to control for equal "loading" of normal and tumor DNA (Fig. 1A, panel 3). Previously, the somatic loss of both maternal and paternal copies of a gene was observed only at the RB locus (*9*), the p53 locus (*15*), a locus on chromosome 11p (*16*), and a locus on 9p (*17*). At least the first three of these loci appear to encode tumor suppressor genes, the identification of a homozygous deletion with the p15-65 probe suggested that the sequences detected might be near a putative tumor suppressor gene on chromosome 18q.

Allelic losses of the sequences detected by p15-65 were observed in 71 percent of colorectal carcinomas studied (for example, Fig. 1A, panel 1). In addition to these allelic losses and the homozygous loss in S115, one other tumor (S123) with abnormal hybridization to the p15-65 probe was identified among 120 colorectal tumors analyzed. DNA from normal colonic mucosa of patient S123 was homozygous for the 7.8-kb Msp I allele detected by p15-65 (Fig. 1A, panel 4). However, DNA from the tumor of this patient was apparently heterozygous at this locus, with alleles of 7.8 and 10.5 kb (Fig. 1A, panel 4). The tumor from this patient had "gained" rather than lost heterozygosity.

To determine the molecular basis for this gain in heterozygosity, genomic DNA clones containing the affected Msp I site were isolated from the carcinoma of patient S123 and the sequences were compared to those of normal DNA. Except for Msp I, the restriction maps of this region were identical in the tumor and normal tissue of patient S123 indicating no gross DNA additions or deletions in the tumor. The affected Msp I site was within an Alu-type repeated element; approximately 300 base pairs (bp) surrounding this Msp I site were 75 to 80 percent identical to recorded human Alu sequences. The sequence of the cloned DNA fragment from the S123 tumor DNA differed from the normal allele at a single base pair, resulting in the replacement of the internal G residue within the Msp I recognition sequence 5′-CCGG-3′ with an A residue (Fig. 1B). This mutation created a potential 3′ splice acceptor site (Fig. 1B), identical to the consensus sequence for intron-exon junctions of primate genes (*18, 19*). Mutations creating splice acceptor sites associated with abnormal RNA processing have been previously noted in inherited diseases such as thalassemias (*20*).

Identification of expressed sequences. The data described above demonstrated that colorectal carcinomas from two different patients had acquired somatic mutations at the same locus on chromosome 18q, and that this locus was within a region of the chromosome in which allelic deletions were frequent. They also suggested that this locus might contain sequences encoding a tumor suppressor gene, and further studies were undertaken to identify expressed sequences from this genomic region. First, phage clones that encompassed a 35-kb region surrounding the mutated Msp I site of tumor S123 were isolated. All Eco RI fragments from the phage clones were subcloned and used in hybridization experiments (Northern blots) with RNA of normal colonic mucosa and cell lines derived from tumors of the colon and several other organs. No expression was detected in these experiments, nor was expression detected in ribonuclease (RNase) protection studies of these RNA samples using selected subfragments from the phage clones. A bidirectional chromosomal walk from this region of chromosome 18q was then achieved with bacteriophage vectors (Fig. 2). More than 140 different clones, spanning approximately 370 kb, were isolated in 30 rounds of walking (Fig. 2). To identify potential exons on the basis

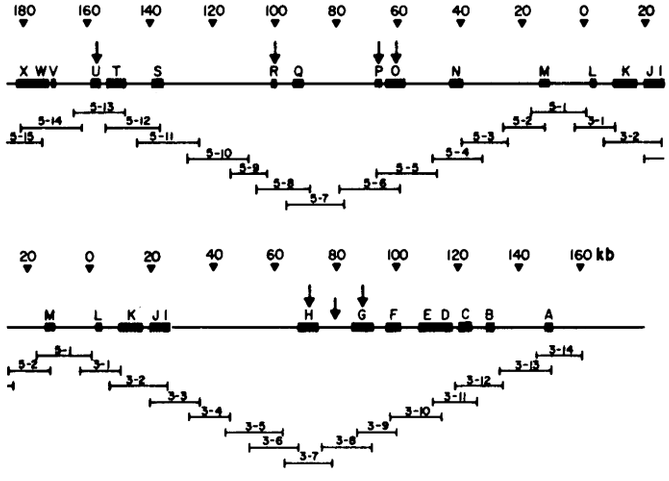

Fig. 2. Map of the chromosomal walk and cross-hybridizing fragments showing the DNA region of approximately 370 kb, cloned in 30 rounds of walking; only the maximal walk for each of the rounds is shown (*47*). The map position at "0" marks the location of p15-65. An Eco RI map for the region was constructed and the Eco RI fragments that hybridized at reduced stringency (*48*) to rodent, chicken, or *Xenopus* DNA samples are indicated by solid boxes and alphabetical letters (A to X). Human fragments G, I, J, K, M, O, and P were used to isolate rat clones as described (*50*). The minimal region of cross-hybridization was identified and sequenced for both human and rat fragments. The locations of the Eco RI fragments that hybridized to cDNA clones (Fig. 5) are indicated by arrows.

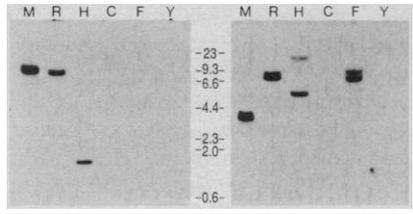

Fig. 3. Cross-species hybridization of human fragments O and P. Autoradiographs of Southern blots of DNA from mouse (M), rat (R), hamster (H), chicken (C), *Xenopus laevis* (F), and *Saccharomyces cerviseae* (Y) hybridized to human fragment O (left) and human fragment P (right) are shown. The size of the corresponding molecular markers in kilobases are indicated between the two panels. DNA (5 μg) was placed in each lane, except for lane Y, which contained 0.5 μg. Southern blotting and hybridization were performed as described (*48*).

of their homology to other species, we isolated every Eco RI fragment from the region and used each as a hybridization probe at reduced stringency to screen DNA samples from various species (mouse, rat, hamster, chicken, and *Xenopus*) (*21*). Twenty-four of the 117 Eco RI fragments hybridized to discrete DNA fragments of at least one of the species tested (Fig. 2). The patterns observed with two of the fragments producing strong cross-species hybridization are illustrated in Fig. 3. Fragment P (subcloned from phage clone W5-6) hybridized to mouse, rat, hamster, and *Xenopus* DNA's, and fragment O (subcloned from phage clone W5-5) hybridized to mouse, rat, and hamster DNA's (Fig. 3). Most of the cross-species hybridizing fragments were then used as probes to screen RNA blots prepared with RNA of various normal tissues or tumor cell lines, and they were also used to screen cDNA libraries from normal colonic mucosa specimens, a colorectal adenoma cell line, a brain tumor, a fibrosarcoma line, and an embryonal carcinoma line. No evidence for expression was obtained from these RNA blots, nor were any hybridizing clones identified in any of the cDNA libraries when the cross-hybridizing fragments were used as probes.

In order to determine whether any of the cross-hybridizing fragments had exon-like structural features, we cloned seven of the homologous fragments from the rat DNA, and determined the regions of cross-hybridization for each fragment for both human and rat DNA's. These regions were then subcloned and sequenced from both species (Fig. 4). The sizes of the regions of sequence similarity ranged from 128 to over 534 bp, and the similarity ranged from 75 to 89 percent. The sequences were examined for open reading frames (ORF's), conservation in the predicted amino acid sequence of the ORF's, consensus mammalian splice acceptor and lariat sequences at the 5′ end of the ORF's, and consensus splice donor sequences at the 3′ region of the ORF's (*18, 19*). Several of these features were found in most of the fragments sequenced. Three of the human-rat fragment pairs (Figs. 2 and 4, fragments G, O, and P) had more exon-like features than the other fragments. The region of cross-species hybridization from human fragment G and its homologous rat fragment predicted ORF's that differed at two amino acid positions (Fig. 4). Similarly, the ORF's predicted for the other two sets of fragments were highly conserved with a single amino acid substitution distinguishing human fragments O and P from their respective rat homologues (Fig. 4). For all three sets of fragments, nucleotide substitutions were predominantly at the third position of codons in the exon-like regions, and sequence homology decreased to 75 percent or less outside these regions.

The striking sequence similarity between the human and rat sequences and their exon-like structural features suggested that the fragments might contain exons. However, as noted above, no expression was detected when these three fragments were used as probes in Northern blots of various RNA's or used to screen cDNA libraries. In addition, RNase protection experiments (*22*), with the use of these three fragments to generate antisense transcripts, did not conclusively demonstrate evidence of expression in any RNA samples tested.

Fig. 4. Sequence of three homologous pairs of fragments from the human and rat. The nucleotide sequence, predicted amino acid (AA) (*49*) sequence and splice acceptor and donor features of the human fragments and their corresponding rat sequences. The nucleotide sequences of the rat fragments were identical to those of the human except where indicated. When the predicted amino acid sequence of the rat differed from that of the human, the human sequence is given on the left of the slash and the rat sequence on the right. The regions in human fragment O and P where no corresponding sequence was available for the rat are indicated by the dashed lines. The predicted intron-exon boundaries are indicated by solid arrowheads, and the potential lariat signal preceding the splice acceptor sequence is overlined. Rat clones were obtained as described (*50*) and sequenced as described (*27*).

```
FRAGMENT G

HUMAN   ATTTGGAAGACTTATTCTTCCTTCTTTGTTTTTCTCCTAG GA CCA CTG AGG TTC CTT TCA CAG ACA GAA TCT GTC ACA GCC
RAT     T GAAAT CT A CC T CT G TG CT  C C  T                      C  C              A
AA                                                          P  L  R  F  L  S  Q  T  E  S  I  T  A

HUMAN   TTC ATG GGA GAC ACA GTG CTA CTC AAG TGT GAA GTC ATT GGG GAG CCC ATG CCA ACA ATC CAC TGG CAG AAG
RAT             T       T   T                          C                A
AA       F  M  G  D  T  V  L  L  K  C  E  V  I  G E/D P  M  P  T  I  H  W  Q  K

HUMAN   AAC CAA CAA GAC CTG ACT CCA ATC CCA GGT GAC TCC CGA GTG GTG GTC
RAT      T           C  C     AC  G  A     G  T     C
AA       N  Q  Q  D  L  T  P I/N P  G  D  S  R  V  V  V

FRAGMENT O

HUMAN   CTCACTCACTTTTTTTTTTCTGTCTTTGCAG TT CCG CCA TGG TTT TTA AAT CAT CCT TCC AAC CTG TAT GCC TAT GAA AGC
RAT        T C G                                          C           C
AA                                          P  P  W  F  L  N  H  P  S  N  L  Y  A  Y  E  S

HUMAN   ATG GAT ATT GAG TTT GAA TGT ACA GTC TCT GGA AAG CCT GTG CCC ACT GTG AAT TGG ATG AAG AAT GGA GAT GTG
RAT        C                 G   G  C  G
AA       M  D  I  E  F  E  C T/A V  S  G  K  P  V  P  T  V  N  W  M  K  N  G  D  V

HUMAN   GTC ATT CCT AGT GAT TAT TTT CAG ATA GTG GTAATT
RAT      G --- --- --- --- --- --- --- --- --- ---
AA       V  I  P  S  D  Y  F  Q  I  V

FRAGMENT P

HUMAN   TACTTTACCTGCTTTTTCTTTTCCCTAG GGA GGA AGC AAC TTA CGG ATA CTT GGG GTG GTG AAG TCA GAT GAA GGC TTT
RAT     ---------------------------- --- --- --- --- --  C               A
AA                                   G  G  S  N  L  R  I  L  G  V  V  K  S  D  E  G  F

HUMAN   TAT CAA TGT GTG GCT GAA AAT GAG GCT GGA AAT GCC CAG ACC AGT GCA CAG CTC ATT GTC CCT AAG CCT GGTAAGACAAT
RAT                                  C        A GT     C     G        C     C       C
AA       Y  Q  C  V  A  E  N  E  A  G  N  A  Q T/S S  A  Q  L  I  V  P  K  P
```

To increase the sensitivity of expression assays, we used the polymerase chain reaction (PCR) (*23*) in an "exon-connection" strategy. Complementary DNA was prepared from total RNA of various cell lines and tissues (see below), and reverse transcriptase was used to prime synthesis with random hexamers (*24*). This cDNA was then used in PCR experiments with oligonucleotide pairs derived from the sequence of two of the potential exons noted above. If the two potential exons were present in the same RNA transcript, then, with the use of the appropriate oligonucleotides, it would be possible to amplify a cDNA product linking these two regions. Trace amounts of DNA, which often contaminate RNA preparations, would not give rise to the same size PCR product in this assay because the exons were separated by intron sequences in DNA. Most pairs of oligonucleotides derived from the seven regions of human-rat sequence similarity described above did not generate discrete PCR fragments when tested by this strategy. However, an oligonucleotide pair derived from the sequence of fragments O and P (Fig. 4) generated a discrete 233-bp PCR product from cDNA of several of the mRNA samples studied, including that derived from a small cell carcinoma of the lung (H82) and a colorectal carcinoma (HCT116) (*25*). The PCR product was cloned (*26*) and sequenced (*27*) from both cell lines, and was found to be the product of splicing the predicted exon of fragment O directly to the predicted exon of fragment P.

Complementary DNA isolation. To confirm and extend these PCR experiments, we constructed a cDNA library from RNA of the H82 cell line (*28*). Approximately 3.0×10^6 recombinant cDNA clones were screened with genomic DNA subclones containing regions of fragments G, O, and P of Fig. 4. Four hybridizing clones were isolated and mapped with respect to one another and to the

```
TGA AAT ATG GAG AAT AGT CTT AGA TGT GTT TGG GTA CCC AAG CTG GCT TTT GTA CTC TTC GGA GCT TCC TTG CTC AGC GCG CAT CTT CAA   bp84
 *      M  E  N  S  L  R  C  V  W  V  P  K  L  A  F  V  L  F  G  A  S  L  L  S  A  H  L  Q   aa28

GTA ACC GGT TTT CAA ATT AAA GCT TTC ACA GCA CTG CGC TTC CTC TCA GAA CCT TCT GAT GCC GTC ACA ATG CGG GGA GGA AAT GTC CTC   bp174
 V  T  G  F  Q  I  K  A  F  T  A  L  R  F  L  S  E  P  S  D  A  V  T  M  R  G  G  N  V  L   aa58

CTC GAC TGC TCC GCG GAG TCC GAC CGA GGA GTT CCA GTG ATC AAG TGG AAG AAA GAT GGC ATT CAT CTG GCC TTG GGA ATG GAT GAA AGG   bp264
 L  D  C  S  A  E  S  D  R  G  V  P  V  I  K  W  K  K  D  G  I  H  L  A  L  G  M  D  E  R   aa88

AAG CAG CAA CTT TCA AAT GGG TCT CTG CTG ATA CAA AAC ATA CTT CAT TCC AGA CAC CAC AAG CCA GAT GAG GGA CTT TAC CAA TGT GAG   bp354
 K  Q  Q  L  S  N  G  S  L  L  I  Q  N  I  L  H  S  R  H  H  K  P  D  E  G  L  Y  Q  C  E   aa118

GCA TCT TTA GGA GAT TCT GGC TCA ATT ATT AGT CGG ACA GCA AAA GTT GCA GTA GCA GGA CCA CTG AGG TTC CTT TCA CAG ACA GAA TCT   bp444
 A  S  L  G  D  S  G  S  I  I  S  R  T  A  K  V  A  V  A  G  P  L  R  F  L  S  Q  T  E  S   aa148

GTC ACA GCC TTC ATG GGA GAC ACA GTG CTA CTC AAG TGT GAA GTC ATT GGG GAG CCC ATG CCA ACA ATC CAC TGG CAG AAG AAC CAA CAA   bp534
 V  T  A  F  M  G  D  T  V  L  L  K  C  E  V  I  G  E  P  M  P  T  I  H  W  Q  K  N  Q  Q   aa178

GAC CTG ACT CCA ATC CCA GGT GAC TCC CGA GTG GTG GTC TTG CCC TCT GGA GCA TTG CAG ATC AGC CGA CTC CAA CCG GGG GAC ATT GGA   bp624
 D  L  T  P  I  P  G  D  S  R  V  V  V  L  P  S  G  A  L  Q  I  S  R  L  Q  P  G  D  I  G   aa208

ATT TAC CGA TGC TCA GCT CGA AAT CCA GCC AGC TCA AGA ACA GGA AAT GAA GCA GAA GTC AGA ATT TTA TCA GAT CCA GGA CTG CAT AGA   bp714
 I  Y  R  C  S  A  R  N  P  A  S  S  R  T  G  N  E  A  E  V  R  I  L  S  D  P  G  L  H  R   aa238

CAG CTG TAT TTT CTG CAA AGA CCA TCC AAT GTA GTA GCC ATT GAA GGA AAA GAT GCT GTC CTG GAA TGT TGT GTT TCT GGC TAT CCT CCA   bp804
 Q  L  Y  F  L  Q  R  P  S  N  V  V  A  I  E  G  K  D  A  V  L  E  C  C  V  S  G  Y  P  P   aa268

CCA AGT TTT ACC TGG TTA CGA GGC GAG GAA GTC ATC CAA CTC AGG TCT AAA AAG TAT TCT TTA TTG GGT GGA AGC AAC TTG CTT ATC TCC   bp894
 P  S  F  T  W  L  R  G  E  E  V  I  Q  L  R  S  K  K  Y  S  L  L  G  G  S  N  L  L  I  S   aa298

AAT GTG ACA GAT GAT GAC AGT GGA ATG TAT ACC TGT GTT GTC ACA TAT AAA AAT GAG AAT ATT AGT GCC TCT GCA GAG CTC ACA GTC TTG   bp984
 N  V  T  D  D  D  S  G  M  Y  T  C  V  V  T  Y  K  N  E  N  I  S  A  S  A  E  L  T  V  L   aa328

GTT CCG CCA TGG TTT TTA AAT CAT CCT TCC AAC CTG TAT GCC TAT GAA AGC ATG GAT ATT GAG TTT GAA TGT ACA GTC TCT GGA AAG CCT   bp1074
 V  P  P  W  F  L  N  H  P  S  N  L  Y  A  Y  E  S  M  D  I  E  F  E  C  T  V  S  G  K  P   aa358

GTG CCC ACT GTG AAT TGG ATG AAG AAT GGA GAT GTG GTC ATT CCT AGT GAT TAT TTT CAG ATA GTG GGA GGA AGC AAC TTA CGG ATA CTT   bp1164
 V  P  T  V  N  W  M  K  N  G  D  V  V  I  P  S  D  Y  F  Q  I  V  G  G  S  N  L  R  I  L   aa388

GGG GTG GTG AAG TCA GAT GAA GGC TTT TAT CAA TGT GTG GCT GAA AAT GAG GCT GGA AAT GCC CAG ACC AGT GCA CAG CTC ATT GTC CCT   bp1254
 G  V  V  K  S  D  E  G  F  Y  Q  C  V  A  E  N  E  A  G  N  A  Q  T  S  A  Q  L  I  V  P   aa418

AAG CCT GCA ATC CCA AGC TCC AGT GTC CTC CCT TCG GCT CCC AGA GAT GTG GTC CCT GTC TTG GTT TCC AGC CGA TTT GTC CGT CTC AGC   bp1344
 K  P  A  I  P  S  S  S  V  L  P  S  A  P  R  D  V  V  P  V  L  V  S  S  R  F  V  R  L  S   aa448

TGG CGC CCA CCT GCA GAA GCG AAA GGG AAC ATT CAA ACT TTC ACG GTC TTT TTC TCC AGA GAA GGT GAC AAC AGG GAA CGA GCA TTG AAT   bp1434
 W  R  P  P  A  E  A  K  G  N  I  Q  T  F  T  V  F  F  S  R  E  G  D  N  R  E  R  A  L  N   aa478

ACA ACA CAG CCT GGG TCC CTT CAG CTC ACT GTG GGA AAC CTG AAG CCA GAA GCC ATG TAC ACC TTT CGA GTT GTG GCT TAC AAT GAA TGG   bp1524
 T  T  Q  P  G  S  L  Q  L  T  V  G  N  L  K  P  E  A  M  Y  T  F  R  V  V  A  Y  N  E  W   aa508

GGA CCG GGA GAG AGT TCT CAA CCC ATC AAG GTG GCC ACA CAG CCT GAG TTG CAA GTT CCA GGG CCA GTA GAA AAC CTG CAA GCT GTA TCT   bp1614
 G  P  G  E  S  S  Q  P  I  K  V  A  T  Q  P  E  L  Q  V  P  G  P  V  E  N  L  Q  A  V  S   aa538

ACC TCA CCT ACC TCA ATT CTT ATT ACC TGG GAA CCC CCT GCC TAT GCA AAC GGT CCA GTC CAA GGT TAC AGA TTG TTC TGC ACT GAG GTG   bp1704
 T  S  P  T  S  I  L  I  T  W  E  P  P  A  Y  A  N  G  P  V  Q  G  Y  R  L  F  C  T  E  V   aa568

TCC ACA GGA AAA GAA CAG AAT ATA GAG GTT GAT GGA CTA TCT TAT AAA CTG GAA GGC CTG AAA AAA TTC ACC GAA TAT AGT CTT CGA TTC   bp1794
 S  T  G  K  E  Q  N  I  E  V  D  G  L  S  Y  K  L  E  G  L  K  K  F  T  E  Y  S  L  R  F   aa598

TTA GCT TAT AAT CGC TAT GGT CCG GGC GTC TCT ACT GAT GAT ATA ACA GTG GTT ACA CTT TCT GAC GTG CCA AGT GCC CCG CCT CAG AAC   bp1884
 L  A  Y  N  R  Y  G  P  G  V  S  T  D  D  I  T  V  V  T  L  S  D  V  P  S  A  P  P  Q  N   aa628

GTC TCC CTG GAA GTG GTC AAT TCA AGA AGT ATC AAA GTT AGC TGG CTG CCT CCT CCA TCA GGA ACA CAA AAT GGA TTT ATT ACC GGC TAT   bp1974
 V  S  L  E  V  V  N  S  R  S  I  K  V  S  W  L  P  P  P  S  G  T  Q  N  G  F  I  T  G  Y   aa658

AAA ATT CGA CAC AGA AAG ACG ACC CGC AGG GGT GAG ATG GAA ACA CTG GAG CCA AAC AAC CTC TGG TAC CTA TTC ACA GGA CTG GAG AAA   bp2064
 K  I  R  H  R  K  T  T  R  R  G  E  M  E  T  L  E  P  N  N  L  W  Y  L  F  T  G  L  E  K   aa688

GGA AGT CAG TAC AGT TTC CAG GTG TCA GCC ATG ACA GTC AAT GGT ACT GGA CCA CCT TCC AAC TGG TAT ACT GCA GAG ACT CCA GAG AAT   bp2154
 G  S  Q  Y  S  F  Q  V  S  A  M  T  V  N  G  T  G  P  P  S  N  W  Y  T  A  E  T  P  E  N   aa718

GAT CTA GAT GAA TCT CAA GTT CCT GAT CAA CCA AGC TCT CTT CAT GTG AGG CCC CAG ACT AAC TGC ATC ATC ATG AGT TGG ACT CCT CCC   bp2244
 D  L  D  E  S  Q  V  P  D  Q  P  S  S  L  H  V  R  P  Q  T  N  C  I  I  M  S  W  T  P  P   aa748

TTG AAC                                                                                                                  bp2250
 L  N                                                                                                                    aa750
```

Fig. 5. Nucleotide sequence and predicted amino acid sequence of the open reading frame (ORF) found in cDNA clones. The sequence shown was derived from 16 overlapping cDNA clones, prepared from mRNA of either the H82 cell line or from normal human brain (*28*). The methionine codon initiating the ORF is designated as amino acid 1. This methionine is in a reasonably favorable context for translation initiation according to the paradigms defined by Kozak (*51*). A stop codon was found six nucleotides upstream of amino acid 1. No stop codon was identified downstream of amino acid 1, however, so the carboxyl terminus of the DCC gene product remains undefined.

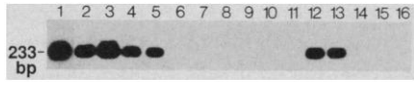

Fig. 6. Expression of DCC in human tumors and colorectal tumor cell lines. RNA was isolated from normal human brain (lane 1), four different normal colonic mucosa specimens (lanes 2 to 5), or colorectal carcinoma cell lines (lanes 6 to 16), and first strand cDNA was prepared as described in (*24*). The cDNA samples were used for polymerase chain reaction (PCR) analysis, according to the conditions specified in (*25*). The PCR products were separated by electrophoresis through an agarose gel, and after Southern transfer, hybridized to a ^{32}P-labeled subclone of fragment P (Fig. 2 and 4). The 233-bp fragment detected represents the exons connecting fragments O and P in the cDNA, and corresponds to nucleotides 986 to 1218 in the cDNA sequence shown in Fig. 5. The human colorectal carcinoma cell lines used were: lane 6, SW948; lane 7, SW1417; lane 8, SW1116; lane 9, SW403; lane 10, SW1463; lane 11, SW48; lane 12, HCT116; lane 13, RKO; lane 14, RCA; lane 15, "C"; and lane 16, MOSER (*52*).

genomic clones shown in Fig. 2. The longest clone was 1.65 kb in length and hybridized to at least 11 different Eco RI fragments in human genomic DNA (Fig. 7A), eight of which were present in the clones from the chromosomal walk shown in Fig. 2. The four cDNA clones isolated were sequenced and subsequently used as probes of cDNA libraries from H82 cells or from normal human brain to obtain additional cDNA clones extending over 2854 base pairs. Sequence analysis indicated that all clones encoded overlapping portions of a transcript in which a single long ORF of 2250 bp extended to the end of the sequenced region. The ORF began with a methionine codon surrounded by nucleotides favorable for translation initiation (nucleotide 1) (see Fig. 5). The methionine was followed by a relatively hydrophobic sequence of 25 amino acids that resembled previously described signal sequences associated with membrane-bound proteins (*29*). The signal sequence was immediately followed by 725 amino acids with significant sequence similarity to the neural cell adhesion molecules and other related cell surface glycoproteins as discussed below. We refer to the gene encoding this transcript as the DCC (deleted in colorectal carcinomas) gene.

Expression pattern. To identify the tissues in which the DCC gene was expressed, cDNA was prepared from several rat organs. Because of the high degree of conservation of the DCC gene (Fig. 4), the same oligonucleotide primers used to demonstrate expression in human cells could also be used in the rat. To assess expression, oligonucleotide pairs from fragments O and P of Fig. 4 were used in a PCR expression assay (*25, 30*). Seventeen of the 18 rat tissues tested appeared to produce the transcript at low levels, with greatest abundance observed in brain (*30*). Similar analysis of human tissues and cell lines revealed that the transcript was present in highest concentration in brain, and was also expressed in normal colonic mucosa and in several tumor cell lines, including those derived from tumors of the lung, brain, and mesenchyme (Fig. 6). In most colorectal carcinomas, however, expression was reduced or absent; of 17 colorectal tumor cell lines studied, only two expressed DCC mRNA amounts in excess of 5 percent of that produced in normal colonic mucosa (examples in Fig. 6).

To determine the size of the transcript produced from this gene, we hybridized (RNA blots, Northern) RNA from normal colonic mucosa or brain with ^{32}P-labeled cDNA clones. A major band of 10 to 12 kb was observed in normal brain RNA, but no bands were seen in the RNA from colonic mucosa, a result consistent with the greater expression observed in brain by PCR analysis.

Genetic alterations detected in colorectal tumors with the cDNA probes. In an attempt to establish the boundaries of the homozygous loss in the S115 tumor with respect to the DCC gene, the cDNA clones were used to probe Southern blots containing S115 DNA. A 430-bp subclone (pKC430, representing nucleotides 1760 to 2205 of the cDNA) detected three Eco RI fragments of 20, 10, and 1.8 kb in DNA from non-neoplastic colonic mucosa of patient S115 (Fig. 7B). However, in DNA from the S115 tumor, the 20-kb fragment was not detected and the 10-kb and 1.8-kb fragments were present at approximately half the intensity observed in normal DNA. In addition, a new fragment of 5 kb was observed only in DNA from the tumor (Fig. 7B). Probes further downstream of the 3′ end of pKC430 also detected fragments in tumor DNA that were present at half the intensity of those in normal DNA, while probes 5′ of pKC430 detected fragments that were homozygously deleted in the tumor. Thus, the DCC gene appeared to be broken by the deletion event on one copy of chromosome 18, and the breakpoint established one boundary of the homozygous loss.

To search for other genetic alterations, we used the cDNA probes in Southern blot analysis of colorectal tumor DNA samples. Three of 51 primary tumors and 2 of 21 tumor xenografts had new fragments not present in normal DNA of the same patient. In addition, 5 of 22 colorectal tumor cell lines had altered fragments not present in 44 DNA samples from normal individuals nor in any of 45 DNA samples from tumor cell lines derived from tissues other than that of the colon or rectum.

In all cases, detailed mapping experiments showed that the new fragments detected by the cDNA probe resulted from insertions in an approximately 170-bp Xba I–Eco 0109 fragment located 165 bp downstream of the exon in fragment P of Fig. 4. The size of the insertion in the tumors varied from 120 to 300 bp (Fig. 8A). Some variation in the size of the Xba I–Eco 0109 fragment in alleles from normal individuals was seen; however, the greatest difference between the size of the smallest and largest of the 88 normal alleles studied was approximately 50 bp, and the largest of the normal alleles was approximately 120 bp smaller than any of the altered alleles in the tumors. The 1.4-kb Eco RI fragment (fragment P) containing the insertion site was isolated from one of the genomic clones of the chromosome walk (Fig. 2) and sequenced; the sequence of the 170-bp Xba I–Eco 0109 fragment (Fig. 8B) showed two regions of TA repeats; one of the regions had eight repeats and the other had 26. Both TA repeat regions were contained within a 130-bp region of alternating purine-pyrimidine base pairs which

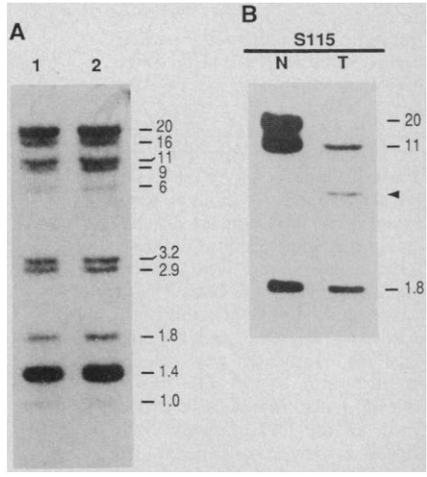

Fig. 7. (**A**) An autoradiograph of a Southern blot prepared from DNA of peripheral blood lymphocytes of two normal individuals, hybridized to a ^{32}P-labeled cDNA probe (pDCC 1.65). The pDCC 1.65 contained nucleotides 591 to 2250 of the cDNA shown in Fig. 5. Eleven Eco RI fragments were detected in normal DNA samples, eight of which were within the 370-kb region cloned by chromosome walking (see Fig. 2 for the location of the eight fragments). The sizes of the fragments in kilobases are indicated to the right of the blot. A 0.45-kb Eco RI fragment detected by pDCC 1.65 migrated off the gel. (**B**) Hybridization of pKC 430 (from the 3′ region of the 1.65-kb cDNA) to normal (N) and cryostat sections of tumor (T) DNA from S115. This probe contained nucleotides 1760 to 2205 of the cDNA shown in Fig. 5. It detected the 10-kb and 1.8-kb fragments in S115 tumor DNA at approximately half the intensity observed in normal DNA, but the 20-kb fragment was not detected. A new fragment of approximately 5 kb (arrowhead) was detected only in the S115 tumor DNA. Southern blotting and hybridization were performed as described in the legend to Fig. 1.

could potentially form Z-DNA (*31*). We were unable to clone alleles with intact copies of the insertion from any of three tumors tested. Indeed, in the clones obtained from alleles containing the insertions, deletions arose during the cloning process in both bacteriophage and plasmid vectors, regardless of the bacterial host strain used (*32*). In addition, we were unable to use PCR to amplify alleles with insertions from any of five cases tested, although we could easily amplify normal alleles lacking the somatically acquired insertions. The fact that the DNA fragments containing the insertions could not be synthesized with Taq polymerase suggested that they had an unusual structure that might similarly interfere with transcription through the insertion in vivo. Experiments to test this hypothesis must await cloning of the fragments containing the tumor-specific insertions.

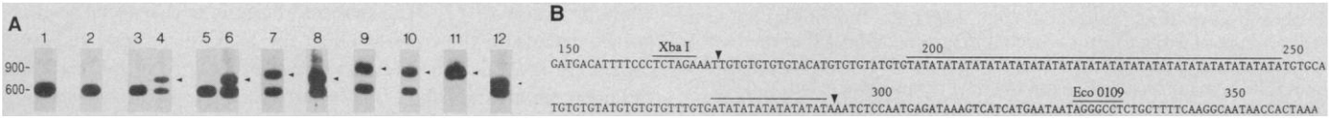

Fig. 8. Analysis of insertions in colorectal tumors. (**A**) DNA from normal and tumor DNA samples were digested with Eco RI and Eco 0109, and Southern blots were prepared as described (Fig. 1). The DNA was then hybridized to a 0.4-kb genomic fragment that contained the exon from fragment P (Fig. 2). Molecular sizes in kilobases are indicated at the left; arrowheads indicate the tumor DNA fragments with insertions. (Lanes 1 and 2) DNA's from peripheral blood lymphocytes from two normal individuals; (lanes 3 and 4) DNA's from non-neoplastic colon and colorectal carcinoma xenograft, respectively, from patient Cx7; (lanes 5 and 6) DNA's from non-neoplastic colon and cryostat sections of colorectal carcinoma, respectively, from patient S175; (lane 7) DNA from a carcinoma xenograft from patient Cx10; (lane 8) DNA from colorectal carcinoma cell line RKO; (lane 9) DNA from colorectal tumor cell line VACO6; (lane 10) DNA from colorectal carcinoma cell line NCI-H630; (lane 11) DNA from colorectal carcinoma cell line SW48; (lane 12) DNA from colorectal carcinoma cell line HCT116. (**B**) Sequence of the 170-bp Xba I–Eco 0109 fragment to which the insertions were localized. The numbers above the sequence indicate the distance in base pairs from the 3′ end of the exon contained in fragment P (Fig. 4). The Xba I and Eco 0109 restriction sites are indicated. The two regions of TA repeats are overlined, and the 130-bp region of alternating purine-pyrimidine sequence is contained between the arrowheads. The sequence was obtained from a plasmid subclone of phage clone W5-6 (Fig. 2) as described (*27*). The insertions in tumor DNA samples were mapped to the Xba I and Eco 0109 fragment by comparison of Southern blot patterns produced by digestion of tumor DNA samples with a combination of Xba I and Eco 0109, Eco RI, and Hind III.

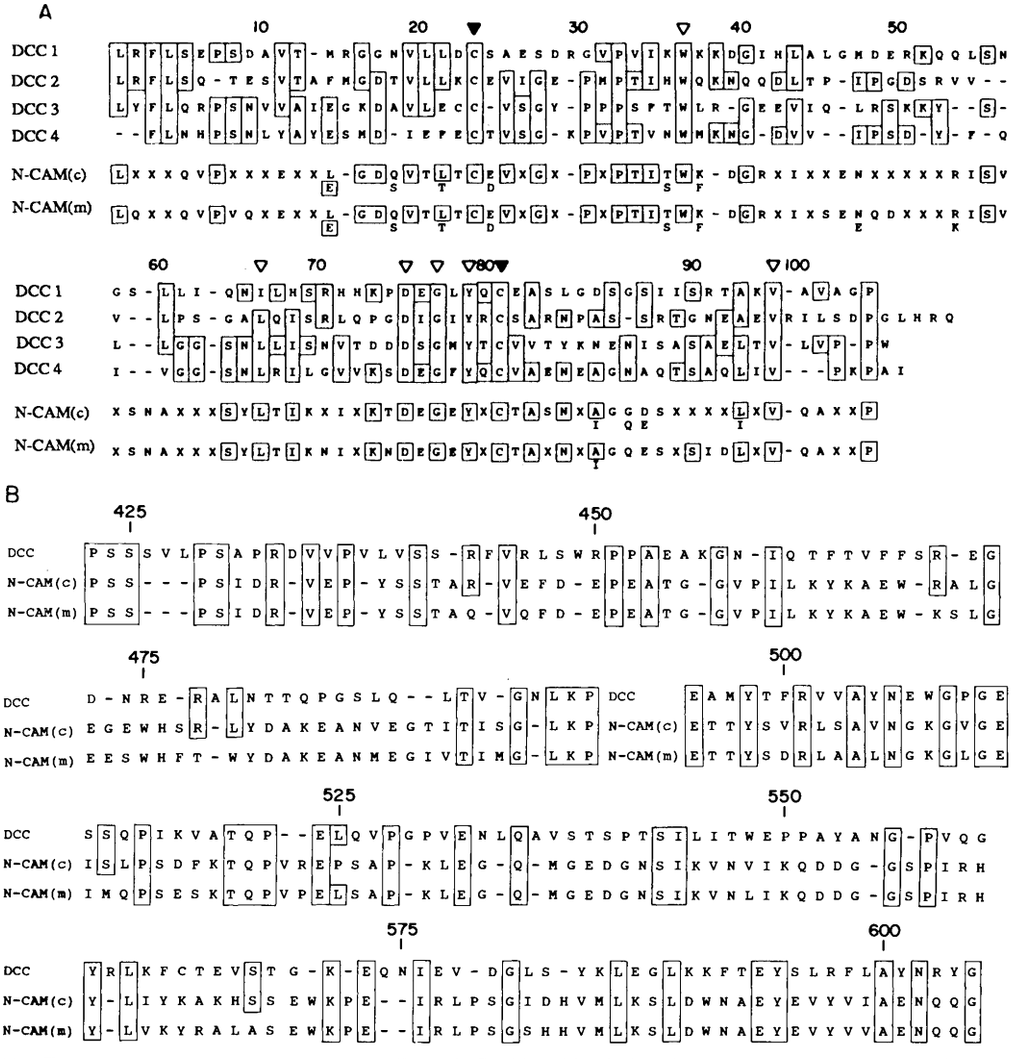

Fig. 9. Homology between DCC and N-CAM. (**A**) Comparison of the sequence homology of the four immunoglobulin-like domains with one another and with chicken N-CAM [N-CAM(c)] and mouse N-CAM [N-CAM(m)]. Each Ig-like domain is approximately 100 amino acids in length. DCC domain 1 includes amino acids 40 to 139, domain 2 includes amino acids 140 to 239, domain 3 includes amino acids 240 to 332, and domain 4 includes amino acids 333 to 422. The N-CAM(c) and N-CAM(m) sequences shown represent the consensus of the five Ig-like domains present in each protein (*39*); if no consensus was present at a particular position of N-CAM(c) or N-CAM(m) (that is, if no two domains contained the identical residue), then the position is indicated by an X. Spaces inserted for alignment are indicated by a dash. The conserved cysteines thought to be involved in intra-domain disulfide pairing are indicated by solid triangles; other amino acid residues highly conserved in N-CAM and other similar Ig-like domains of the C2 class (*37*) are noted by open triangles. Sequences were aligned by inspection to give the greatest overall match. Residues in two or more of the DCC domains were boxed if they were identical. The N-CAM(c) and N-CAM(m) consensus sequences were boxed if they matched the DCC consensus. (**B**) Comparison of the sequence homology between DCC and chicken and mouse N-CAM in the fibronectin type III–related regions. DCC amino acid positions 423 to 605 were compared to amino acids 481 to 662 of the two N-CAM proteins (*39*). The sequences were aligned by inspection, and spaces, indicated by dashs, were inserted to give the greatest overall match. Residues identical in the DCC sequence and either of the N-CAM sequences were boxed.

Potential role of the DCC gene in tumorigenesis. Through a strategy combining chromosome walking, isolation and sequencing of evolutionarily conserved DNA regions, a PCR-based assay for expression, and cDNA cloning, we have identified a highly conserved gene on chromosome 18q. Several observations suggest that this gene should be considered as a candidate colorectal tumor suppressor. First, one allele of the DCC gene was deleted in 29 of 41 (71 percent) colorectal carcinomas, and, as noted previously, allelic losses are thought to indicate that the affected region contains a tumor suppressor gene. Second, the gene was expressed in almost all normal tissues tested, including colonic mucosa, but DCC expression was greatly reduced or absent in 15 of 17 (88 percent) colorectal carcinoma cell lines. Such reduced expression in tumors is consistent with a suppressive function. Third, somatic mutations of the DCC gene detectable by DNA blotting (Southern) were observed in 12 of 94 (13 percent) different colorectal carcinomas. In one tumor (S115), a homozygous deletion was associated with a break in one copy of the DCC gene, eliminating its 5′ half. In addition, a point mutation was observed in tumor S123, and insertions within a 170-bp fragment directly downstream of an exon were found in ten other colorectal tumors. The effects of these mutations on DCC gene expression are not known. Moreover, demonstration of somatic mutations within the DCC gene cannot formally establish that this gene is a suppressor of colorectal tumorigenesis. Validation of this hypothesis will require cloning of a full-length DCC gene product, and demonstration of a biological effect after its introduction into neoplastic colorectal epithelial cells.

The studies described above raise two other issues. First, Lynch *et al.* have described a syndrome with autosomal dominant inheritance in which affected individuals are predisposed to carcinomas of the colon and other organs (*33*). This syndrome is the most common of the various inherited diseases predisposing to colorectal neoplasia, and is estimated to account for at least 5 to 6 percent of colorectal cancers (*34*). The gene causing this syndrome has been linked to the Kidd blood group on chromosome 18q11.1-q21.1 (*35*), near the region of chromosome 18 that contains the DCC gene. Whether the DCC gene is involved in this predisposition syndrome or other syndromes not yet mapped to a particular chromosomal region remains to be determined.

Second, the predicted amino acid sequence of DCC is highly similar to the neural cell adhesion molecules and other related cell surface glycoproteins (*36*). Two areas of high sequence similarity were noted. (i) The DCC gene contained four immunoglobulin (Ig)-like domains of the C2 class, defined by pairs of cysteines separated by 50 to 56 amino acids and other highly conserved residues surrounding the first and second cysteine of each pair (Fig. 9A) (*37*). Potential sites of N-linked glycosylation were found at several positions within this region (*38*). The four Ig-like domains of DCC showed more sequence similarity to one another than to N-CAM, L1, or other members of the Ig super-family. A consensus sequence for the four Ig-like domains of DCC could be derived for 67 percent of the positions; the DCC consensus sequence matched the N-CAM consensus sequence at 42 percent of these positions (Fig. 9A). In addition, the ORF identified within the DCC gene also contained a fibronectin type III–related domain (or domains), similar to the fibronectin-like domains present in N-CAM, L1, leukocyte common antigen-related gene 1 (LAR1), fasciclin II, and other members of this cell adhesion molecule family (*39*). These fibronectin-related domains are carboxyl to the Ig-like domains in all of these proteins, including DCC. Of 195 positions within the fibronectin-like region of DCC, 31 percent were identical in DCC and N-CAM, and several conservative substitutions were also present (Fig. 9B). The extensive sequence similarities in both Ig-like and fibronectin-like domains between DCC and the other members of this family suggest that these proteins may all have been derived from a common precursor that included both these regions (*37*).

That a gene altered in colorectal tumors is similar in sequence to genes involved in cell-surface interactions is provocative. Numerous studies have suggested that neoplasia is associated with alterations of such interactions. For example, a common marker of in vitro transformation is a loss of growth inhibition mediated by cell-cell contact (*40*). In vivo, a hallmark of malignancy is the disorganization of the normal topography of cell-cell and cell-basement membrane interactions (*41*), and the process of metastasis involves the disruption of normal cell contacts and the formation of new ones in foreign environments (*42*). Intercellular adhesion mediated by CAM's has been shown to directly influence cellular differentiation (*36*), and the progressive loss of differentiated features is another common characteristic of neoplastic cells (*43*). Further study of the DCC gene may provide new insights concerning the molecular mechanisms governing colorectal neoplasia, particularly regarding the roles that cell-surface interactions play in controlling abnormal cellular proliferation.

REFERENCES AND NOTES

1. P. C. Nowell, *Science* **194**, 23 (1976); L. Foulds, *J. Chronic Dis.* **8**, 2 (1958); R. A. Weinberg, *Cancer Res.* **49**, 3713 (1989).
2. J. M. Bishop, *Science* **235**, 305 (1987); C. M. Croce, *Cancer Res.* **46**, 6019 (1986).
3. A. G. Knudson, *Cancer Res.* **45**, 1437 (1985); A. Murphree and W. Benedict, *Science* **223**, 1028 (1984); M. F. Hansen and W. K. Cavenee, *Cancer Res.* **47**, 5518 (1987).
4. B. Vogelstein *et al.*, *N. Engl. J. Med.* **319**, 525 (1988).
5. J. Bos *et al.*, *Nature* **327**, 293 (1987); K. Forrester *et al.*, *ibid.*, p. 298.
6. B. Vogelstein *et al.*, *Science* **244**, 207 (1989).
7. E. R. Fearon, S. R. Hamilton, B. Vogelstein, *ibid.* **238**, 193 (1987).
8. M. Muleris *et al.*, *Ann. Genet. (Paris)* **28**, 206 (1985); J. P. Monpezat *et al.*, *Int. J. Cancer* **41**, 404 (1988).
9. S. H. Friend *et al.*, *Nature* **323**, 643 (1986); W.-H. Lee *et al.*, *Science* **235**, 1394 (1987); Y.-K. T. Fung *et al.*, *ibid.* **236**, 1657 (1987).
10. Alterations of the p53 gene in colorectal carcinomas are described by S. J. Baker *et al.* [*Science* **244**, 217 (1989)]. Evidence that p53 is a suppressor gene is summarized by C. A. Finlay, P. W. Hinds, and A. J. Levine [*Cell* **57**, 1083 (1989)] and D. Eliyahu *et al.* [*Proc. Natl. Acad. Sci. U.S.A.* **86**, 8763 (1989)].
11. Y. Tsujimoto, L. R. Finger, J. Yunis, P. C. Nowell, C. M. Croce, *Science* **226**, 1097 (1984); M. L. Cleary, S. D. Smith, J. Sklar, *Cell* **47**, 19 (1986).
12. E. R. Spindel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5699 (1984).
13. K. Semba *et al.*, *Science* **227**, 1038 (1985); M. C. Yoshida *et al.*, *Cytogenet. Cell Genet.* **40**, 786 (1985).
14. The plasmid p15-65 contains a 2.7-kb Sal I–Eco RI fragment derived from sequences adjacent to and including those contained within the anonymous DNA segment OLVII E10, which is an 0.8-kb Hind III–Eco RI fragment that marks the D18S8 locus at 18q21.3 and detects Msp I polymorphism in normal individuals [F. Marthens *et al.*, *Nucleic Acids Res.* **15**, 1348 (1987)]. The p15-65 plasmid was derived from a human genomic DNA library prepared from human DNA partially digested with Mbo I and cloned in λFIX (Stratagene). Two phage clones hybridizing to OLVII E10 were isolated by hybridization selection. Among several tested subfragments from the two phage clones identified, p15-65 produced the highest signal-to-noise ratio when used for DNA (Southern) blot experiments. The p15-65 plasmid (and OLVII E10) detect two polymorphic Msp I sites, and thus give rise to four alleles of 10.5, 9.7, 7.8, and 7.0 kb at frequencies of 17, 4, 49, and 30 percent, respectively ($N = 232$).
15. D. Wolf and V. Rotter, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 790 (1985).
16. W. H. Lewis *et al.*, *Genomics* **3**, 25 (1989).
17. M. O. Diaz *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 5259 (1988).
18. M. B. Shapiro and P. Senapathy, *Nucleic Acids Res.* **15**, 7155 (1987).
19. B. Ruskin, A. R. Krainer, T. Maniatis, M. R. Green, *Cell* **38**, 317 (1984); Y. Zhuang, A. M. Goldstein, A. M. Weiner, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2752 (1989).
20. T. J. Ley, N. P. Anagnou, G. Pepe, A. W. Nienhuis, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4775 (1982); P. A. Sharp, *Annu. Rev. Biochem.* **55**, 1119 (1986).
21. The identification of exons in the Duchenne muscular dystrophy gene and cystic fibrosis genes was facilitated by studies with cross-species hybridization [A. P. Monaco *et al.*, *Nature* **323**, 646 (1986); J. M. Rommens *et al.*, *Science* **245**, 1059 (1989)].
22. E. Winter, F. Yamamoto, C. Almoguera, M. Perucho, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7575 (1985).
23. R. K. Saiki *et al.*, *Science* **239**, 487 (1988).
24. E. Noonan and I. B. Roninson, *Nucleic Acids Res.* **162**, 10366 (1988).
25. The oligonucleotide derived from fragment O was 5′-TTCCGCCATGGTTTT-TAAATCA-3′ and the oligonucleotide derived from fragment P was 5′-AGCCT-CATTTTCAGCCACACA-3′. Cycle times were 1 minute at 95°C, 1 minute at 58°C, and 2 minutes at 70°C (25 cycles were performed).
26. The PCR products were phosphorylated with T4 DNA kinase (Bethesda Research

Laboratories), ligated to Eco RI linkers (New England Biolabs), inserted into λgt10 phage arms (Stratagene), and cloned in *Escherichia coli* C600 cells. Insert-containing clones were identified through hybridization with ^{32}P-labeled probes from fragments O and P (Fig. 4).
27. Eco RI inserts from the phage clones were subcloned in Bluescript SK (Stratagene) and sequenced as described [S. Tabor and C. C. Richardson, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4767 (1987); R. Kraft, J. Tardiff, K. S. Krauter, L. A. Leinwand, *Biotechniques* **6**, 544 (1988)].
28. Double-stranded cDNA was synthesized as described [U. Gubler and B. J. Hoffman, *Gene* **25**, 263 (1983)] except that random hexamers, rather than oligo-(dT), were used to prime first strand synthesis with reverse transcriptase. Both λgt10 and λZAP (Stratagene) were used as vectors in *E. coli* C600 bacterial hosts.
29. M. E. E. Watson, *Nucleic Acids Res.* **12**, 5145 (1984).
30. RNA was isolated as described [P. Chomczynski and N. Sacchi, *Anal. Biochem.* **162**, 156 (1987)] and first strand cDNA was prepared as described (*24*) from the following organs of an adult female Sprague-Dawley rat: liver, kidney, adrenal, heart, lung, stomach, esophagus, spleen, small bowel, breast, bladder, uterus, aorta, psoas, brain, colon, tongue, and skin. Expression was detected in all organs except the liver.
31. A. Rich, A. Nordheim, A. H.-J. Wang, *Annu. Rev. Biochem.* **53**, 791 (1984); R. D. Wells, *J. Biol. Chem.* **263**, 1095 (1988). Both Z-DNA and short repetitive elements have been hypothesized to be prone to recombination [R. Holliday, *Trends Genet.* **5**, 355 (1989); A. P. Jarman and R. A. Wells, *ibid.*, p. 367].
32. In addition to C600, the *E. coli* strain DL538 was unsuccessfully used to clone the tumor-specific alleles. The DL538 strain allows cloning of repeated or methylated sequences (or both) that are difficult to clone in more commonly used hosts [D. A. Whitaker, A. J. D. Campbell, E. M. Southern, N. E. Murray, *Nucleic Acids Res.* **16**, 6725 (1988)].
33. H. T. Lynch *et al.*, *Cancer* **56**, 934 (1985).
34. J.-P. Mecklin, *Gastroenterology* **93**, 1021 (1987); H. T. Lynch *et al.*, *Dis. Colon Rectum* **31**, 439 (1988).
35. H. T. Lynch *et al.*, *Cancer* **56**, 939 (1985); G. Geitvik *et al.*, *Hum. Genet.* **77**, 205 (1987); P. O'Connell *et al.*, *Genomics* **3**, 367 (1988).
36. G. M. Edelman, *Biochemistry* **27**, 3533 (1988).
37. A. F. Williams and A. N. Barclay, *Annu. Rev. Immunol.* **6**, 381 (1988).
38. Six potential sites of N-linked glycosylation [P. V. Wagh and O. P. Bahl, *CRC Crit. Rev. Biochem.* **10**, 307 (1981)] were found at codons 94, 299, 318, 478, 628, and 702.
39. Members of the cell adhesion molecule family were aligned with DCC according to the FASTP algorithm of D. J. Lipman and W. R. Pearson [*Science* **227**, 1435 (1985)]. The identities were: 25 percent over 564 amino acids for chicken N-CAM [B. A. Cunningham *et al.*, *Science* **236**, 799 (1987)]; 24 percent over 585 amino acids for mouse N-CAM [D. Barthels *et al.*, *EMBO J.* **6**, 907 (1987)]; 25 percent over 660 amino acids for mouse L1 [M. Moos *et al.*, *Nature* **334**, 701 (1988)]; 22 percent over 726 amino acids for leukocyte common antigen-related gene 1 (LAR1) [M. Streuli *et al.*, *J. Exp. Med.* **168**, 1523 (1989)]; 21 percent over 372 amino acids for fasciclin II [A. L. Harrelson and C. S. Goodman, *Science* **242**, 700 (1988)]; 28 percent over 198 amino acids for myelin-associated glycoprotein (MAG) [M. Arquint *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 600 (1987); C. Lai *et al.*, *ibid.* p. 4337]; 25 percent over 238 amino acids for amalgam [M. A. Seeger, L. Haffley, T. C. Kaufman, *Cell* **55**, 589 (1988)]. The N-CAM and MAG sequence similarities were found from a search of the NBRF protein [version 20 (7-3-89)] data base. The relationship of other Ig superfamily members not present in the gene bank was determined after we found the sequence similarity to N-CAM. The fibronectin type III–related domain has been described [R. O. Hynes, *Annu. Rev. Cell Biol.* **1**, 67 (1986)].
40. G. J. Todaro and H. Green, *J. Cell Biol.* **17**, 299 (1963).
41. S. L. Robbins, *Pathology* (Saunders, New York, ed. 3, 1967), p. 93; K. S. Pienta, A. W. Partin, D. S. Coffey, *Cancer Res.* **49**, 2925 (1989).
42. I. J. Fidler and I. R. Hart, *Science* **217**, 998 (1982); G. L. Nicholson and G. Poste, *Int. Rev. Exp. Pathol.* **25**, 77 (1983); M. E. Greenberg, R. Brackenberry, G. M. Edelman, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 969 (1984).
43. M. Pignatelli and W. F. Bodmer, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 5561 (1988); P. A. Marks *et al.*, *Can. Res.* **47**, 659 (1987); L. Sachs, *ibid.*, p. 1981.
44. The conditions used were the same as described [B. Vogelstein *et al.*, *Cancer Res.* **47**, 4806 (1987)], except that the hybridization buffer contained 0.5 percent nonfat dried milk and the filter was washed in 44.5 m*M* sodium chloride, 1.8 m*M* sodium citrate, 0.3m*M* tris, *p*H 7.5 at 65°C for 45 minutes.
45. The probe used was pTHH 55 [T. Holm *et al.*, *Nucleic Acids Res.* **16**, 3117 (1988)], generously provided by Y. Nakamura and R. White. Patient S115 was heterozygous at the locus detected by this probe, with one allele of 4.5 kb and a second allele of 3.3 kb; both alleles were retained in the tumor tissue. A constant fragment of 2.6 kb was also detected in both normal and tumor.
46. The affected Msp I site was found within a 1.8-kb Eco RI fragment that was 5 kb from p15-65. Eco RI digests of DNA from the S123 tumor or from normal tissue were fractionated by electrophoresis, and DNA fragments (approximately 1.8 kb in size) were eluted and cloned with the use of a λgt10 vector and [illegible] C600 cells as a host. Phages containing the fragments of interest were identified by hybridization with ^{32}P-labeled inserts from a phage clone containing p15-65 sequences (*14*). The sequences of subclones of the phages containing the relevant Msp I sites were obtained as described (*27*). Because of the heterozygosity found at this Msp I site in normal individuals, we also cloned a normal allele without the Msp I site. The sequence of the normal allele with the Msp I site (the allele present in the normal tissue of patient S123) is shown in Fig. 1B; the sequence of the normal allele without the Msp I site (from a different individual) differed only by a single base pair in this region, with the sequence 5′-CTGG-3′ instead of 5′-CCGG-3′ at the Msp I recognition site. Thus, neither of the alleles found in normal individuals contained a potential 3′ splice acceptor site (5′-CAGG-3′), such as that seen in the S123 tumor (Fig. 1B).
47. Human genomic DNA was partially digested with Mbo I and fragments of 12 to 18 kb were cloned in the λFIX vector (Stratagene) as recommended by the manufacturer. Clones were propagated in *E. coli* C600 or TAP 90 cells [T. A. Patterson and M. Dean, *Nucleic Acids Res.* **15**, 6298 (1987)]. For each round of walking, Eco RI maps were constructed from comparison of digests of overlapping phage clones (starting with phage clones containing p15-65). Eco RI fragments mapping furthest from previously obtained phage clones were used to re-screen the library. Approximately 1×10^6 phage clones were screened for each round of walking; three to seven new clones were generally obtained in each walk, and the clones were purified through three rounds of hybridization selection.
48. Southern blot hybridization was performed as described in (*44*), except that the hybridization and wash temperatures were both reduced to 55°C.
49. A, alanine; R, arginine; N, asparagine; D, aspartic acid; C, cysteine; E, glutamic acid; Q, glutamine; G, glycine; H, histidine; I, isoleucine; L, leucine; K, lysine; M, methionine; F, phenylalanine; P, proline; S, serine; T, threonine; W, tryptophan; Y, tyrosine; and V, valine.
50. Rat genomic clones were obtained in two ways. In some cases, fragments identified on Southern blots by cross-species hybridization (Fig. 3, for examples) were eluted from agarose gels and cloned as described (*46*), with the use of cross-hybridizing, ^{32}P-labeled human clones as probes. In other cases, Mbo I partial digests of rat genomic DNA were cloned in λDASH (Stratagene) as in (*47*), and the library was screened with corresponding human clones of interest.
51. M. Kozak, *Nucleic Acids Res.* **15**, 8125 (1987).
52. Cell lines were obtained from the American Type Culture Collection (Rockville, MD) or from M. Brattain.
53. We thank M. and D. Brattain, A. Gazdar, J. Willson, and J. M. Jessup for providing colorectal tumor cell lines and xenografts and T. Gwiazda for preparation of the manuscript. Supported by the Clayton Fund, the McAshan Fund, NIH training grants GM07184, GM07309, and CA 09243, and NIH grant CA35494.

18 October 1989; accepted 22 November 1989