CGM2, a Member of the Carcinoembryonic Antigen Gene Family Is Down-regulated in Colorectal Carcinomas*

John Thompson‡, Wolfgang Zimmermann, Peter Nollau§, Michael Neumaier§, Julia Weber-Arden, Heinrich Schrewe¶, Ian Craig‖, and Teresa Willcocks**

*From the Institute of Immunobiology, University of Freiburg, D-79104 Freiburg, Federal Republic of Germany, the §Department of Clinical Chemistry, University Hospital Eppendorf, D-20251 Hamburg, Federal Republic of Germany, the ‖Genetics Laboratory, Biochemistry Department, University of Oxford, Oxford OX1 3QU, United Kingdom, and the **Biotechnology Center, University College, Dublin 4, Ireland*

We have determined the precise chromosomal location, the exon structure, and the expression pattern of *CGM2*, a member of the carcinoembryonic antigen (CEA) gene family. CGM2 cDNA was amplified by reverse transcription-polymerase chain reaction (RT/PCR) from the colon adenocarcinoma cell line, LS174T. A defective exon is missing from this cDNA clone, leading to a novel domain organization for the human CEA family with two immunoglobulin-like domains. The derived C-terminal domain predicts that the CGM2 protein is membrane-bound through a glycosyl phosphatidylinositol anchor. RT/PCR analyses identified CGM2 transcripts in mucinous ovarian and colonic adenocarcinomas as well as in adjacent colonic tissue, but not in other tumors including leukocytes from six chronic myeloid leukemia patients. Thus, unlike several other family members, *CGM2* is not expressed in granulocytes but reveals a more CEA-like expression pattern. Northern blot analyses identified a 2.5-kilobase CGM2 mRNA that is strongly down-regulated in colonic adenocarcinomas compared with adjacent colonic mucosa, suggesting a possible tumor suppressor function. In addition, a 3.2-kilobase transcript was observed in a number of colon tumors that is not detectable in normal colonic tissue. This mRNA species could represent a tumor-specific CGM2 splice variant.

The carcinoembryonic antigen (CEA)[1] gene family contains 29 genes (1), which are densely packed within a 1.2-megabase region on the long arm of chromosome 19 (2). The CEA family belongs to the immunoglobulin superfamily (3). The derived protein structures show that CEA family members consist of one N-terminal, immunoglobulin variable (IgV)-like domain (N domain), followed by zero, two, three, four, or six immunoglobulin constant (IgC)-like domains of subtype A or B (reviewed in Ref. 4). Sequence comparisons allow the division of CEA-related genes into two main subgroups, one that encodes CEA and its classical cross-reacting antigens and the second containing 11 genes that encode the pregnancy-specific glycoproteins (PSG). CEA subgroup proteins are mostly membrane-bound, either through a glycosyl phosphatidylinositol (GPI) tail or as integral membrane proteins with transmembrane and cytoplasmic domains. PSG molecules are directly secreted from the cell and possess only very short C-terminal domains. The CEA subgroup contains 12 genes, of which five apparently represent pseudogenes (1, 5). As far as they have been analyzed, the other seven potentially active genes of this subgroup have open reading frames, and transcripts have been identified for six members, *i.e.* for CEA, the nonspecific cross-reacting antigen (NCA), biliary glycoprotein (BGP), and CEA gene family members 1, 6, and 7 (CGM1, CGM6, and CGM7 (summarized in Ref. 4). The remaining gene (CGM2) has so far only been partially characterized at the genomic level (6).

Although CEA-related genes are very similar at the sequence level, they show variable expression patterns. CEA itself is mainly expressed in different epithelial cells, especially those lining the gastrointestinal tract (7). Much higher CEA concentrations are often found in a variety of carcinomas, *e.g.* colonic, lung, and breast adenocarcinomas, where CEA is of great clinical value as a tumor marker (8). The closely related NCA-50/90 is often coexpressed with CEA in a number of tumors (4, 9). The larger glycosylation variant NCA-90 is also present in granulocytes (10) where no CEA is found. A number of other CEA-related molecules are coexpressed in granulocytes, *e.g.* the product of CGM6, NCA-95 (11), BGP, and CGM1 (12–15). BGP is also expressed in epithelia of the bile canaliculi (16). Interestingly, it was recently shown that the expression of the human BGP gene, as well as of its murine homologue Bgp1, is down-regulated in colonic tumors as compared with the adjacent colonic mucosa, and the levels of rat BGP homologues (cell-CAM105) are reduced in hepatocellular carcinomas (17–19). On the other hand, the NCA gene is up-regulated whereas the CEA mRNA steady state levels stay more or less the same in colorectal cancers (17). It could be shown *in vitro* that CEA, NCA, and BGP are capable of mediating cell adhesion (20–22) so that their deregulation, as found for other cell adhesion molecules (23, 24), could play a causal role during tumor development, *e.g.* BGP down-regulation could facilitate tumor cell release from their normal environment.

As a prerequisite for the cDNA cloning strategy that we adopted, we describe here the genomic organization and exon sequencing of the CGM2 gene. We have shown that this last known, potentially expressible member of the CEA subgroup is in fact expressed. Internal primers specific for this gene have
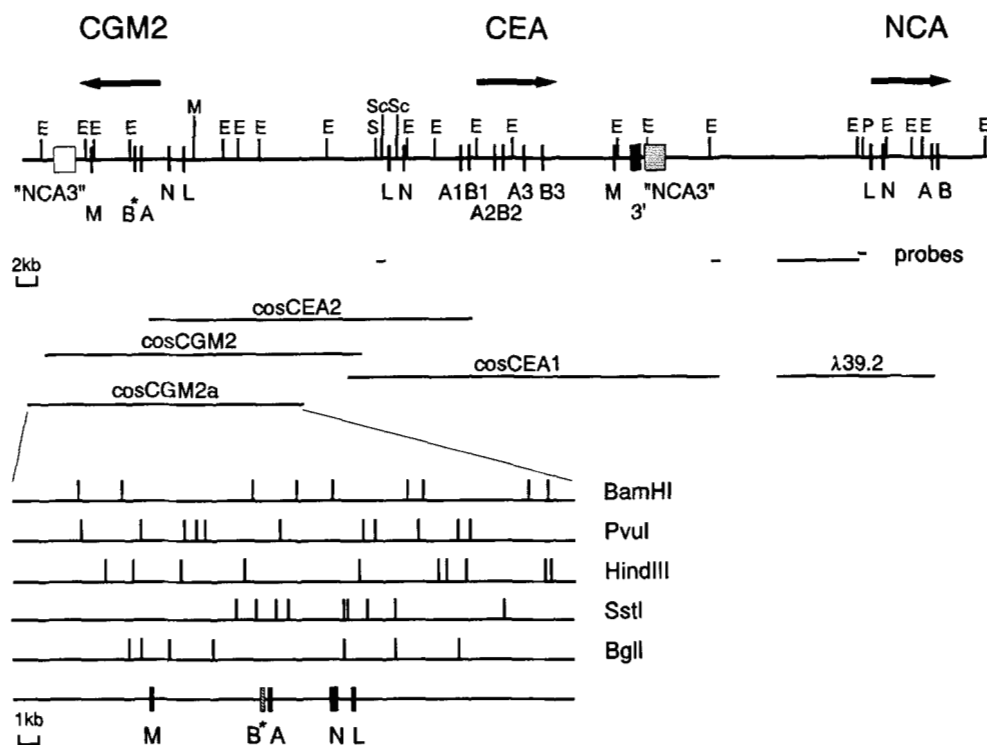
FIG. 1. **Fine mapping of the *CGM2/CEA/NCA* gene region.** Only the *CGM2, CEA,* and *NCA* genes are shown. Their relative positions within the CEA gene cluster have been described elsewhere (5). The *upper part* reveals the *Eco*RI (*E*) restriction enzyme sites over this gene region. Rare-cutting restriction enzyme sites for *Mlu*I (*M*) and *Sac*II (*Sc*) as well as the *Pst*I (*P*) and *Sst*I (*S*) sites relevant for probe preparation are also included. The gene names are written *above* together with *arrows* that indicate directions of transcription. The exons are depicted as *blocks*. *Filled-in blocks* with *letters below* indicate exons encoding the 5′-untranslated region and most of the leader sequence (*L*), the N-terminal IgV-like domain (*N*), IgC-like domains (*A* and *B*), the hydrophobic membrane domain (*M*), and a 3′-untranslated region (*3′*). The *striped box* indicates the pseudo-B exon (*B**) of *CGM2*; the *open* and *stippled boxes* show putative NCA-like 3′-untranslated region exons ("*NCA3*") in the *CGM2* and *CEA* genes, respectively, whereby the former has only been mapped through hybridization to the indicated *Eco*RI DNA fragment. The position of terminal DNA probes from cosmid clone cosCEA1 (25) and from λ39.2 (37) that co-hybridize with the same 13 kb *Eco*RI DNA fragment in digested genomic DNA are shown *below* this map. Similarly, the 527-bp *Sst*I/*Sac*II fragment from cosCEA1 and the 667-bp *Eco*RI/*Pst*I fragment from the NCA gene, which were used to isolate cosCEA2 and cosCGM2, respectively, are shown (*probes*). The *central part* shows the positions and sizes of four overlapping cosmid clones (cosCEA1, cosCEA2, cosCGM2, and cosCGM2a) and of the *NCA* gene-containing λ clone λ39.2 with respect to this region. The *lower part* reveals individual restriction enzyme maps of clone cosCGM2a with the exons from the *CGM2* gene shown *below*.

been used to assay for CGM2 transcripts in RNAs from a large number and variety of tumors and some normal tissues by reverse transcription followed by polymerase chain reaction (RT/PCR). With the knowledge that other members of the CEA subgroup are deregulated during tumorigenesis, we have determined the steady state levels of CGM2 transcripts in colonic tumors in comparison with adjacent mucosa by Northern blot analyses.

EXPERIMENTAL PROCEDURES

Cosmid Isolation and Analysis—The cosmid library used for screening was obtained from W. Lindenmaier (Gesellschaft für Biotechnologische Forschung, Braunschweig, Germany) and has been described elsewhere (25). Specific probes (see "Results") were hybridized at high stringency washing conditions, *i.e.* 0.1 × SSPE, 0.1% SDS at 65 °C (1 × SSPE is 0.18 M NaCl, 10 mM sodium phosphate, pH 7.4, 1 mM EDTA), to isolate clones cosCEA2 and cosCGM2. Clone cosCGM2a was isolated employing two degenerate oligonucleotides as described before (26). Genes contained in cosmid clones overlapping the CEA gene locus were identified by PCR amplification of regions within the N-terminal domains using oligonucleotides PCRCEAall5′(5′-**GGAATTC**GCAGAG-GGGAAGGAGGT-3′) and PCRCEAall3′ (5′-**GGAATTC**ATCAGCA-GGGATGCATTGG-3′) as primers. These oligonucleotides were derived from sequences that are conserved for all known members of the CEA subgroup. In comparison with *CGM2* (6) one and two mismatches are found, respectively. *Eco*RI linkers were included (bold print). PCR was performed using 1–10 ng of nonlinearized cosmid DNA over 30 cycles using a bio-med Thermocycler 60 (Bachhofer, Reutlingen, Germany), basically according to Sambrook *et al.* (27), with the following modifications: the melting temperature was 93 °C, and annealing was for 15 s at 56 °C. Amplified fragments were subcloned into a Bluescript pha-

gemid vector (Stratagene, Heidelberg, Germany) as blunt-ended fragments into the *Eco*RV site. DNA from these double-stranded templates was sequenced by the dideoxy chain termination method (28) with universal primers using a kit from Pharmacia (Freiburg, Germany). A series of four oligonucleotides, corresponding to different exons of *CEA* were used to identify homologous regions within the *CGM2* gene by hybridization, following multiple restriction enzyme digestions of cosCGM2a, as recently described (26). Those regions exhibiting homology were subcloned into pUC18 for double-stranded sequencing in both directions. Restriction enzyme mapping was achieved after complete digestion of the cosmids with either a single enzyme or double digestions in different combinations.

cDNA Cloning—Based on sequence data of *CGM2*, two oligonucleotide primers were synthesized from the leader (CGM2-L, 5′-**AAGGTT**TGTC-CATACAGAGTGTGCATT-3′) and the 3′-untranslated regions (CGM2–3′UT, 5′-AGATCTAGAAATYRAGAAACTACACCAA-3′) to clone CGM2 transcripts from the LS174T colon adenocarcinoma cell line (American Type Culture Collection, Rockville, MD) by RT/PCR. Total RNA was extracted using the guanidinium thiocyanate method, followed by pelleting through a cesium chloride cushion (29). For reverse transcription, 5 µg of total RNA was used as a template using random hexamer primers at 42 °C in 15 µl with a kit and the protocol from Pharmacia. Following this step, 3 µl of the mixture were taken, and the *CGM2* primers were added (50 pmol each) along with 0.5 unit of *Taq* polymerase (Promega, Heidelberg, Germany), and the volume was made up to 10 µl with water. Thirty-five amplification cycles were carried out at an annealing temperature of 55 °C (72 s/step). The obtained product (5 µl) was used as a template for a second amplification step using an extended 5′-oligonucleotide (CGM2-PCR5′, 5′-CCATGGGGTTCCCCTTCAGCCTGTCCATA-CAGAGTGTGCA-3′) to encompass the translational start codon, together with the CGM2–3′UT primer, at an annealing temperature of 52 °C over 30 cycles. After phosphorylation, the product (5 µl) was blunt-end

A

Exon			Intron			Exon	
name	size (bp)	splice donor		size (bp)	splice acceptor		size (bp)	name
L	nd	...CTGCTCACAG LeuLeuThrA	gtgagtggaggacttcctgggagt...	880	...actgatgttctctcccctcctag	CCTCGCTTTTT.. laSerLeuLeu	363	N
N	363	...TACGTATTCT TyrValPheS	gtgagtgatacctccatgacttct...	3000	...cttttctctctcttatctgcacag	CGGAGCCACCC.. erGluProPro	276	A
A	276	...AATGTCCGCT AsnValArgT	gtgagtatcttctgttcctctgtg...	390	...acggctttattctctttgctccgg	GTGGCTTGGAC..	254	ψB
ψB	254	...ACAGTCCGTG	gtacatggatccctggaccgttgg...	3700	...acttctcttttctttccatggcag	ATGAGTCAGTA.. yrGluSerVal	127	M
M	127	...GGGAAGAGTG	gtaggtattatgacctttcctctt...	nd				

B

```
                                                        ▼
gaccaggagacttccgcttccctctgaggacatcacttacggctttattctctttgctccgg
↦ ψB
GTGGCTTGGACACCCCTTCTA---CCCCAGTTAT---ATCTTACGAGTTAGGAGCAAACC

TTAGACTCTTCATCCATGTGGCCTCTAACCCACCCTCTCCGTATTTTTGGCTGAGTAATG

GAAACCTTCTGCAACACATGCAAGAGCTCTTTATCCCCAA-AGCACTGCAAAGAATAGTG

TATCCTATACCTGCTATGCCCACAAATTGGGCCACTGACCGCAATAGTACCACAGTCAAGA

TGATCACAGTCCGTGgtacatggatccctggaccgttggcatcatatttttgagatggggt
```

FIG. 2. **Nucleotide sequences of exon/intron borders of the *CGM2* gene (*A*) and the nonspliceable B exon (*B*).** *Capital letters* indicate coding regions with the amino acid sequences *below*. Intron sequences are shown in *lower case letters*. Splice donor and acceptor sites are *underlined*, and the degenerate splice acceptor for the nonspliceable B exon (ψB) is *shaded*. Abbreviations for the exons are as in the legend to Fig. 1. In (*B*), the mutated base (A → G) at the splice acceptor site of the pseudo-B exon (ψB) is additionally indicated with an *arrowhead*. *Dashes* reveal deleted nucleotides after alignment with the B1 intron/exon region of the *CEA* gene (25).

ligated into pUC18 using the Pharmacia Sureclone kit, following the manufacturer's instructions, and sequenced in both orientations. DNA sequence comparisons were carried out using the program "Clustal" (PC-gene, IntelliGenetics, Inc., Mountain View, CA), which is based on average linkage cluster analysis (30) and was used to construct dendrograms.

RNA Isolation, RT/PCR Assay, and Northern Blot Analyses—Total RNA was isolated either from deep frozen normal tissues, primary brain, breast, ovarian, and colonic tumors as well as white blood cells from chronic myeloid leukemia patients or from human tumors propagated in nude mice as described previously (9). Its integrity was first tested in an RT/PCR assay using primers to identify β-actin and/or c-*abl* transcripts. CGM2 transcripts were then assayed, also using an RT/PCR system. The primers used for amplification had at least five mismatches to all other members of the CEA gene family. The 5'-primer was the CGM2-L oligonucleotide (see above), and the 3'-primer was from the M domain (CGM2-M, 5'-**CTGCAG**CGGTTCCAGCTGAGAG-GTCAGGT-3' (see Fig. 3)). Amplification was carried out over 30 or 40 cycles with an annealing temperature of 56 °C (15 s) and denaturation at 94 °C (15 s), using 50 pmol of each each primer. Plasmid pCGM2 containing the full length CGM2 cDNA was used as a positive control, and the specificity was confirmed using 5 ng of plasmid DNA containing cDNA of other CEA subgroup members as templates. Water replaced RNA in the negative control. For Northern blot analyses, fresh tissue samples were obtained from colon cancer patients during surgery. From each patient, both tumor and adjacent colonic mucosa tissue samples were obtained and processed separately, essentially as described recently (17). From each sample, 15 μg of total RNA were analyzed by hybridization with the complete CGM2 cDNA as a probe according to standard procedures. Final stringency washings were carried out twice at 65 °C for 30 min in 0.1 × SSPE, 0.1% SDS.

RESULTS

Isolation of a Cosmid Clone and Determination of the Exon Structure of the CGM2 Gene—In order to gain sequence information from the complete coding region of the last known member of the CEA subgroup that is potentially transcriptionally active, a cosmid clone containing the *CGM2* gene (cosCGM2a) was isolated from the same human genomic library using the same method of isolation as that described for the *CEA* gene (31). In addition to identifying a CEA-containing cosmid, the oligonucleotide probes hybridized less intensely to a cosmid containing the entire coding and flanking regions of the *CGM2* gene as determined through partial, double-stranded DNA sequencing of restriction endonuclease digested fragments and comparison with the partial *CGM2* gene sequence published elsewhere (6). The organization of the *CGM2* gene was determined by digestion of the cosmid with a series of restriction enzymes and hybridization with oligonucleotides from different domains of the *CEA* gene (26). The results suggest an exon arrangement analogous to that of the *NCA* gene (Figs. 1 and 2*A*). Sequencing of each homologous region revealed that the first exon codes for 21 amino acids of the leader sequence (*L*) and the second encodes the remaining 13 plus the entire N-terminal domain (*N*). The third exon encodes an IgC-like domain of subtype A (*A*); the fourth exon reveals homology to the subtype B IgC-like domains (*B*) of the *NCA* gene but has a degenerate splice acceptor site, where the requisite A is substituted by a G (Fig. 2). Within the B exon two deletions of three nucleotides each are followed by a single deletion (Fig. 2*B*) when compared with the *CEA* gene B1 exon (25). This single deletion (nucleotide position 409 in Ref. 32) leads to a frameshift that is quickly followed by a stop codon. Taken together, these data suggest that this is a pseudo-B exon. The final exon encodes a 30 amino acid, hydrophobic membrane (*M*) domain whose sequence similarity to corresponding regions of CEA, NCA-50/90, and NCA-95 (4) suggest its posttranslational substitution by a glycosyl phosphatidylinositol tail. In analogy to the corresponding exon of the *CEA* gene, 39 nucleotides of the
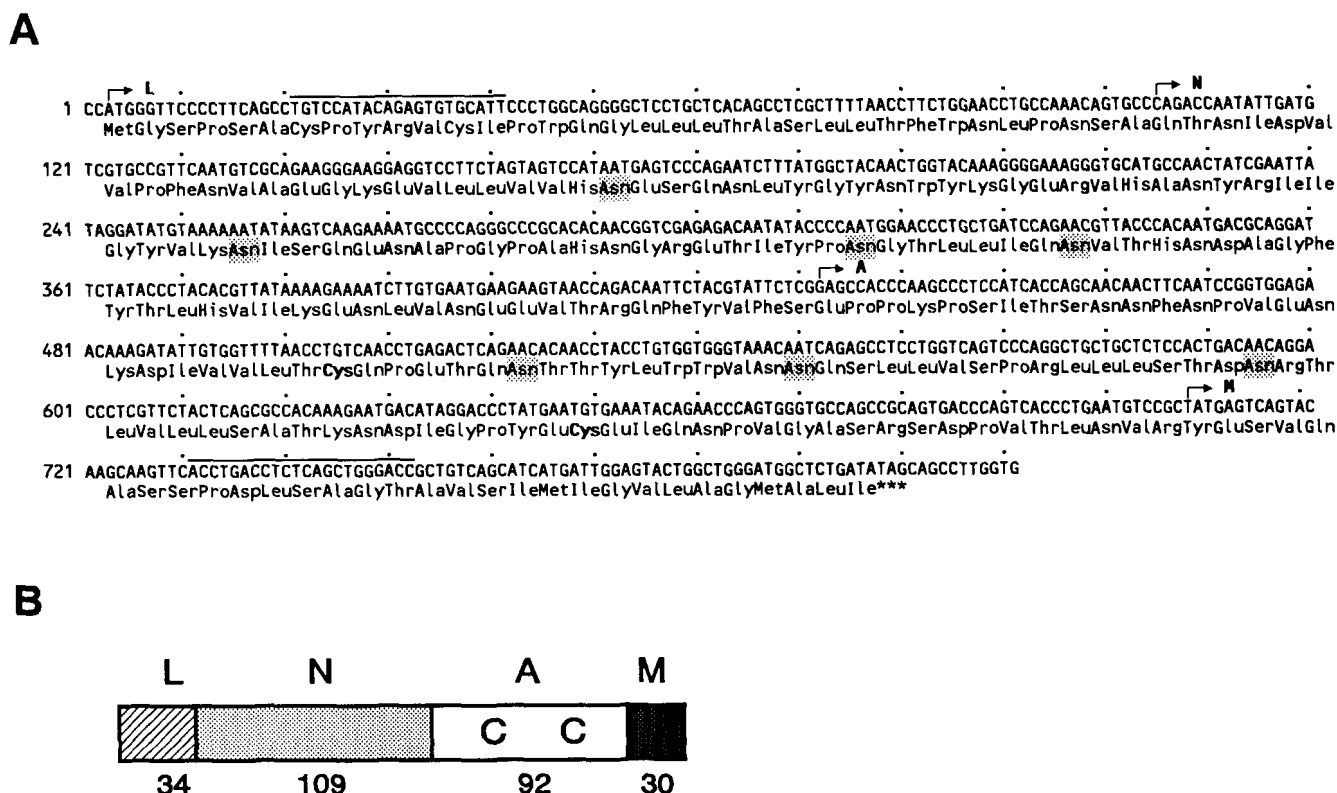
A

```
         ┌→ L                          .                    .                    .                    .                    .                 ┌→ N
   1 CCATGGGTTCCCCTTCAGCCTGTCCATACAGAGTGTGCATTCCCTGGCAGGGGGCTCCTGCTCACAGCCTCGCTTTTAACCTTCTGGAACCTGCCAAACAGTGCCCAGACCAATATTGATG
     MetGlySerProSerAlaCysProTyrArgValCysIleProTrpGlnGlyLeuLeuLeuThrAlaSerLeuLeuThrPheTrpAsnLeuProAsnSerAlaGlnThrAsnIleAspVal

 121 TCGTGCCGTTCAATGTCGCAGAAGGGAAGGAGGTCCTTCTAGTAGTCCATAATGAGTCCCAGAATCTTTATGGCTACAACTGGTACAAAGGGGAAAGGGTGCATGCCAACTATCGAATTA
     ValProPheAsnValAlaGluGlyLysGluValLeuLeuValValHisAsnGluSerGlnAsnLeuTyrGlyTyrAsnTrpTyrLysGlyGluArgValHisAlaAsnTyrArgIleIle

 241 TAGGATATGTAAAAAATATAAGTCAAGAAAATGCCCCAGGGCCCGCACACAACGGTCGAGAGACAATATACCCCAATGGAACCCTGCTGATCCAGAACGTTACCCACAATGACGCAGGAT
     GlyTyrValLysAsnIleSerGlnGluAsnAlaProGlyProAlaHisAsnGlyArgGluThrIleTyrProAsnGlyThrLeuLeuIleGlnAsnValThrHisAsnAspAlaGlyPhe

 361 TCTATACCCTCACACGTTATAAAAGAAAATCTTGTGAATGAAGAAGTAACCAGACAATTCTACGTATTCTCGGAGCCACCCAAGCCCTCCATCACCAGCAACAACTTCAATCCGGTGGAGA
     TyrThrLeuHisValIleLysGluAsnLeuValAsnGluGluValThrArgGlnPheTyrValPheSerGluProProLysProSerIleThrSerAsnAsnPheAsnProValGluAsn

 481 ACAAAGATATTGTGGTTTTAACCTGTCAACCTGAGACTCAGAACACAACCTACCTGTGGTGGGTAAACAATCAGAGCCTCCTGGTCAGTCCCAGGCTGCTGCTCTCCACTGACAACAGGA
     LysAspIleValValLeuThrCysGlnProGluThrGlnAsnThrThrTyrLeuTrpTrpValAsnAsnGlnSerLeuLeuValSerProArgLeuLeuLeuSerThrAspAsnArgThr

 601 CCCTCGTTCTACTCAGCGCCACAAAGAATGACATAGGACCCTATGAATGTGAAATACAGAACCCAGTGGGTGCCAGCCGCAGTGACCCAGTCACCCTGAATGTCCGCTATGAGTCAGTAC
     LeuValLeuLeuSerAlaThrLysAsnAspIleGlyProTyrGluCysGluIleGlnAsnProValGlyAlaSerArgSerAspProValThrLeuAsnValArgTyrGluSerValGln

 721 AAGCAAGTTCACCTGACCTCTCAGCTGGGACCGCTGTCAGCATCATGATTGGAGTACTGGCTGGGATGGCTCTGATATAGCAGCCTTGGTG
     AlaSerSerProAspLeuSerAlaGlyThrAlaValSerIleMetIleGlyValLeuAlaGlyMetAlaLeuIle***
```

B

Fig. 3. **Nucleotide and deduced amino acid sequence (A) and deduced domain organization (B) of the CGM2 cDNA.** *A*, nucleotides are *numbered* at the *left*. Domain borders were deduced from the exon borders and are indicated by *arrows*. Domain abbreviations are as in the legend to Fig. 1. The stop codon is depicted by three *asterisks*. The positions of the primers used for RT/PCR screening of CGM2 transcripts are *overlined* in the sequence. Putative N-glycosylation sites are *shaded*. *B*, the deduced domain organization of the CGM2 preprotein is depicted with differentially *shaded boxes*. The number of amino acids is listed *below* each domain. Cystein (*C*) positions are indicated within the A domain. Abbreviations for the domains are as in the legend to Fig. 1.

3'-untranslated region and an intact splice donor site follow the M-domain.

Cloning of CGM2 cDNA—In the search for putative CGM2 transcripts, RT/PCR was performed on total RNA from the colon adenocarcinoma cell line LS174T using *CGM2*-specific oligonucleotides based on genomic sequences from the leader (CGM2-L) and from the 3'-untranslated regions (CGM2–3'UT), yielding an 819-bp amplification product. This cell line was originally chosen because it is known to express various members of the CEA subgroup (33). A second amplification step was then carried out using the extended 5'-primer, CGM2-PCR5' with CGM2–3'UT to generate a full length CGM2 cDNA clone with respect to the coding region. Direct amplification from RNA with this primer combination proved to be unsuccessful. The resulting fragment had a size of 833 bp and was cloned for sequencing. The DNA and derived amino acid sequence, along with the domain structure, are shown in Fig. 3. Compared with the genomic sequence and the published N-domain exon sequence, two nucleotide substitutions are found at positions 341 (C → T) and 749 (A → G) of the cDNA sequence, which do not lead to amino acid changes. The translated region consists of 798 nucleotides including the stop codon and encodes a putative protein with 265 amino acids that would yield a polypeptide with a relative molecular mass of 25,689 after removal of the 34-amino acid leader sequence. Seven potential *N*-glycosylation sites are present, and assuming a carbohydrate mass of 3000/site the expected size of the mature CGM2 protein would be approximately 47,000 with the M-domain or 44,000 after its removal. The domain structure, derived after sequence comparisons with other CEA-related molecules, reveals that the CGM2 protein has only one immunoglobulin C2 set domain (34) of subtype A. As expected from the size of the PCR product, no B domain is present. Amino acid sequence comparisons were carried out with other members of the CEA family for the N-terminal domain and A domain (data not shown). In both cases, CGM2 is more closely related to the CEA than to the PSG subgroup. Within the CEA subgroup, however, it is the least similar member apart from the CEA subgroup members CGM7 (N domain sequence comparisons) and the BGP A2 domain (A domain sequence comparisons), which segregate alone.

Search for CGM2 Transcripts in Various Normal Tissues and Tumors—A wide range search was carried out to determine the expression pattern of the *CGM2* gene in a large number of tumors and a small collection of normal tissues. To achieve this, the specificity of the PCR-primers was first tested using cDNAs for various CEA family members as templates. All cDNAs tested (Fig. 4*A*) were negative, apart from CGM2 cDNA that yielded a 732-bp amplification product (Fig. 4*C*). 118 human tumor samples of 25 different tumor types, 4 tumor cell lines, 5 normal colonic mucosae (adjacent to tumors), 1 placenta, and 1 amnion tissue sample have been analyzed for the expression of *CGM2* at the transcriptional level (Fig. 4; data not shown). 8 of 12 ovarian mucinous adenocarcinomas, 5 of 5 normal colonic mucosae, 5 of 7 colonic adenocarcinomas, and the colon adenocarcinoma cell line LS174T revealed amplification products of the expected size. The following other tumors were all negative for CGM2 transcripts: 55 ductal and 6 lobular breast carcinomas; 16 other breast tumors (including comedo, intraductal, mucinous, medullar, and anaplastic carcinomas as well as sarcomas); 2 meningiomas, 4 mucinous, 3 serous, and 1 endometrioid ovarian carcinoma; 2 colorectal adenocarcinomas; the leukocytes from 6 chronic myeloid leukemia patients; 5 lung tumors (2 small cell, 1 large cell, 1 adenocarcinoma, and 1 squamous epithelial carcinoma); single gall bladder, stomach,
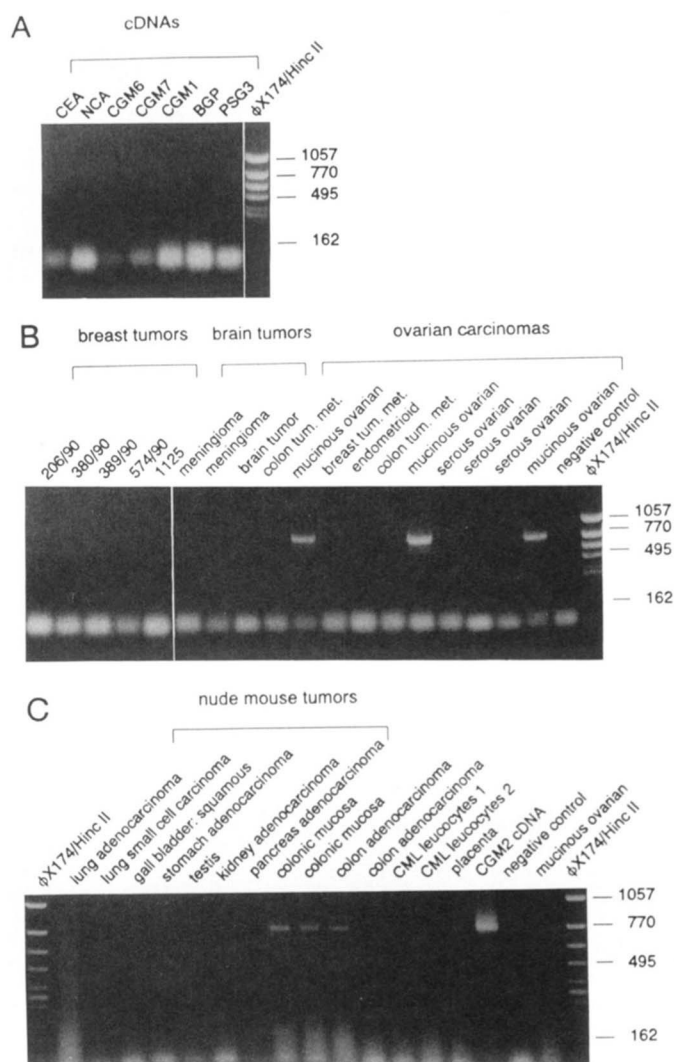
FIG. 4. **Analysis of CGM2 expression in various tissues and tumors using RT/PCR.** RNAs extracted from tumors, tumor metastases (*tum. met.*), and normal tissues (*B* and *C*) and to test specificity, cDNAs from different CEA family members (*A*) were used as templates. An amplification product with the expected size of 732 bp is clearly visible in some lanes. Amplification was carried out over 30 (*A* and *C*) or 40 cycles (*B*).

FIG. 5. **Northern blot analyses to determine the steady state levels of CGM2 mRNA in colonic adenocarcinomas and adjacent colonic mucosae.** *A*, RNA was isolated from colonic adenocarcinomas (*T*) or adjacent mucosa (*N*) of 11 patients. After electrophoretic separation and transfer, the membrane was hybridized with a CGM2-specific and a cytokeratin 18-specific (*Cyt 18*) probe and washed at high stringency conditions. Sizes of the hybridizing signals are indicated at the *left* in kb. *B*, histograms showing the relative steady state levels of the 2.5-kb transcript in tumors and adjacent mucosae (*normal*). The relative CGM2 mRNA steady state levels are expressed as percentages, with the normal colonic mucosa set at 100%. Standard deviations are indicated by *bars*.

kidney, testis, and pancreas tumors; and 1 colonic (SW403) and 1 pancreatic (ASPC-1) adenocarcinoma cell line. The placenta and amnion tissues were also negative.

CGM2 Gene Expression in Colonic Adenocarcinomas and Adjacent Colonic Mucosa—As the expression of some members of the CEA subgroup is known to be deregulated in colonic adenocarcinomas, we decided to determine quantitatively CGM2 expression at the transcriptional level from colon tumors with respect to the adjacent mucosa. Northern blot hybridization with the CGM2 cDNA probe revealed an mRNA of 2.5 kb in normal mucosae after high stringency washing steps. Cross-hybridization to other members of the CEA gene family was not observed, since no additional bands were detected (Fig. 5A). The mRNA species hybridizing with the CGM2 probe is down-regulated in all 11 tumors analyzed compared with the adjacent colon mucosae (Fig. 5B). This finding excludes cross-hybridization of the probe used with NCA transcripts, which are also 2.5 kb in size (35), as the latter are up-regulated in colonic tumors (17). Since carcinomas are known to contain variable proportions of tumor stromal tissue, we used a cytokeratin 18 probe to correct for the fraction of epithelial mRNA
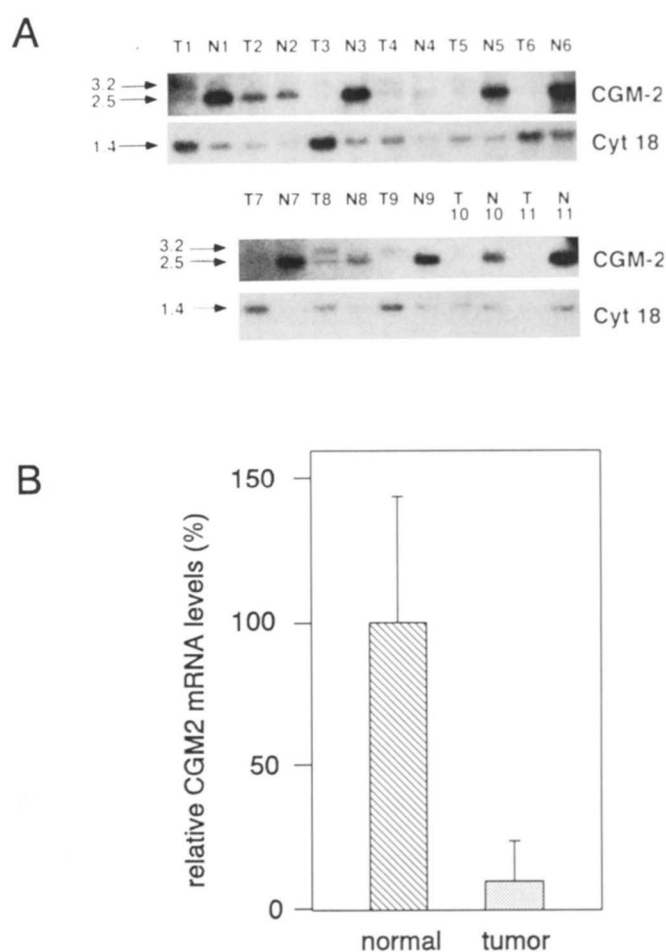
in these samples by calculation of tumor-to-normal ratios. For this purpose, the Northern blot signals were quantified by densitometry from the integer over the mRNA bands in the autoradiograms for both CGM2 and cytokeratin 18. The relative expression (RE) of CGM2 in the tumors (RE(CGM2-tu)) with respect to normal expression was then calculated using the formula, RE(CGM2-tu) = (cyt(n) ÷ cyt(tu)) × (CGM2(tu) ÷ CGM2(n)), where cyt and CGM2 are the expression levels of cytokeratin 18 and CGM2 mRNA, in normal (n) or tumor (tu) tissues, respectively. From the eleven patients investigated, RNA samples of one did not give a positive cytokeratin 18 signal in the Northern blot analysis and had to be excluded from the evaluation. For the remaining ten patients, the relative cytokeratin 18 corrected expression of CGM2 was 5 ± 2.2 in normal colonic mucosa and only 0.5 ± 0.7 in the tumors (Fig. 5B).

Besides finding a strong down-regulation of CGM2 mRNA steady state levels, the majority of tumors that still exhibited a residual expression also showed an additional mRNA species of 3.2 kb of approximately the same intensity as the 2.5-kb mRNA (Fig. 5A). This larger mRNA species was found exclusively in

tumors, although the general CGM2 mRNA steady state level was low. Cross-hybridization to CEA or related mRNA species of a similar size can be disregarded, since the probe did not detect CEA mRNA in the normal tissue that is known to express substantial levels of CEA transcripts (32).

Isolation of CGM2 Cosmid Clones That Overlap the CEA Gene Locus—RT/PCR analyses revealed that the CGM2 gene is often coexpressed with the CEA and NCA genes (data not shown). As it has been reported that these three genes are directly adjacent to each other on the long arm of chromosome 19 (36), we decided to extend the fine restriction enzyme map encompassing these genes as a basis for future investigations to locate and identify putative common, cis-regulatory elements. In order to isolate clones overlapping the CEA (25, 26) and NCA (37) genes, we used a CEA gene-specific probe (5'-CEA) from the 5'-region of a genomic cosmid clone cosCEA1 (25). This 527-bp *Sst*I/*Sac*II fragment is located at nucleotides −1769 to −1242 upstream from the translational start site of the CEA gene and was used to screen a cosmid library under high stringency washing conditions. Its specificity under these conditions has been described elsewhere (5). Apart from the original CEA clone (cosCEA1), a second clone was also isolated (cosCEA2). Hybridization with a 220-bp *Pst*I/*Pst*I probe from the 5'-untranslated region and part of the leader of a NCA cDNA clone, pNCA1 (38) under low stringency washing conditions (2 × SSPE, 1% SDS at 65 °C), revealed two hybridizing *Sst*I fragments of 7.7 and 2.7 kb in cosCEA2, indicating the presence of an N-terminal domain exon from a second CEA-related gene within this clone. The 2.7-kb fragment is comparable in size to that determined elsewhere for the CGM2 gene (39), and indeed the presence of CGM2 close to the CEA gene is supported by a long range restriction enzyme map and analyses of groups of overlapping clones (5, 36). This was further supported by identifying the two *Sac*II as well as the *Mlu*I restriction enzyme sites predicted to exist in between these two genes within clone cosCEA2 (Fig. 1). The final proof that this clone also contains the N-terminal domain exon of the CGM2 gene was provided by PCR analyses using the PCRCEAall oligonucleotide primers (see "Experimental Procedures"). The sequence of a subcloned PCR amplification product from this clone was identical to the CGM2 sequence published elsewhere (6).

Another cosmid clone (cosCGM2) was also shown by PCR and DNA sequencing to contain the N-terminal domain exon of the CGM2 gene. This clone was originally isolated from the same cosmid library with a 667-bp *Eco*RI/*Pst*I fragment, which is located at nucleotides −1247 to −580 upstream of the NCA gene translational start site (37). Restriction endonuclease digestion revealed a number of fragments that are common to cosCEA2, and cosCGM2 is also digestible with *Mlu*I. However, cosCGM2 was found not to contain the *Sac*II sites present in cosCEA1 and cosCEA2. Restriction enzyme mapping of these clones confirmed the degrees of overlap for these three clones (Fig. 1).

In a previous publication we were able to show that the gene encoding NCA is located immediately downstream of the CEA gene (5). As we have genomic clones containing the total CEA gene (25, 26) as well as a clone containing the start of the NCA gene (37) at our disposal, we attempted to confirm linkage at the fine structural level. Based on the assumption that these two genes are in a head-to-tail orientation, DNA fragments from the upstream region of the NCA clone λ39.2 (7.6-kb terminal *Eco*RI fragment (see Ref. 37)) and the downstream region of the CEA clone, cosCEA1 (0.65-kb terminal *Eco*RI fragment (see Ref. 25), hybridized with the same 13-kb *Eco*RI genomic DNA fragment in the presence of unlabeled human genomic DNA to saturate repetitive sequences in the probes (data not shown). The complete *Eco*RI restriction enzyme map covering the CGM2/CEA/NCA genes is shown in Fig. 1 together with other restriction enzyme sites for the CGM2 locus.

The restriction enzyme map including the CEA and NCA gene loci show that the CGM2 gene has an inverted orientation on the chromosome. It is situated approximately 20 kb upstream of the CEA gene, the same distance that separates the CEA and NCA genes. Hybridization analyses with a 1.5-kb *Eco*RI probe from the 3'-untranslated region of the NCA cDNA clone 9 (35) reveal similar sequences in the 3'-region of the CGM2 gene. This indicates that the whole 3'-region is similar in its basic structure to that of the CEA gene (25).

DISCUSSION

Through chromosome walking, we have isolated cosmid clones and carried out detailed restriction endonuclease mapping and sequence analyses to determine the structure of the CEA-related CGM2 gene. These investigations verify its close linkage to the CEA and NCA genes, as also shown elsewhere by analyses of groups of overlapping clones (36) and allow exact determination of the distances between the three genes. In the long range map, these three genes have been located in between the CGM7 and CGM1 genes, within the CEA subgroup cluster closest to the centromere, as reported elsewhere (5). CGM2 contains five exons covering the coding region, whereby exon four represents a nonspliceable B-domain exon, with a degenerate splice acceptor site. A single nucleotide deletion within this exon causes a frameshift and soon leads to a stop codon. Indeed, this nonspliceable B-domain exon is missing in the CGM2 cDNA clone that was amplified by RT/PCR from LS174T cell RNA. Products of an identical size from other RNA sources (Fig. 4) indicate the same domain organization. The predicted structure of the CGM2 protein is unique within the human CEA family (4), with the omnipresent leader (L) and N-terminal, IgV-like domain (N domain), followed by just one IgC-like domain of subtype A. Other members have either no IgC-like domains (CGM1 and CGM7), or at least two (e.g. NCA, CGM6 and BGPb, d, h, and l splice variants (reviewed in Ref. 4)). It remains to be tested as to whether this difference has any functional implications or not. The short hydrophobic domain (M) shows homology to corresponding regions of CEA, NCA-50/90, and NCA-95 that are posttranslationally replaced by a glycosyl phosphatidylinositol moiety, which anchors the molecules into the membrane. Therefore, a similar modification may be expected for the CGM2 product. Although distantly related to the CEA subgroup members, the derived CGM2 protein can be placed within this subgroup. This affiliation was expected from the position of the gene within the CEA subgroup gene cluster on chromosome 19 (5). The close linkage of CEA gene family members, together with their highly conserved DNA sequences and basic domain structures (4), suggests that they have evolved recently by unequal crossing over. Interestingly, inversions must also have occurred to explain the head-to-head genomic arrangement of CGM2 and CEA.

In order to study the expression pattern of CGM2 at a broader level, the RT/PCR method was utilized because it allows unequivocal differentiation of CGM2 transcripts. The expression studies were mainly carried out on tumors because of limited availability of normal human tissues. Although CGM2 expression is down-regulated in colonic tumors (Fig. 5), the high sensitivity of the RT/PCR method allowed discernment of transcripts there, and a comparison with the expression pattern of other CEA family members could still be made. The CGM2 expression pattern found is most similar to that of CEA, *i.e.* expression in normal human colon as well as colonic and mucinous ovarian adenocarcinomas (9, 32). No expression was

found in leukocytes of chronic myeloid leukemia patients, although it is known that cells of the granulocyte lineage express most of the NCA species but not CEA (11–15), or in placenta that expresses most PSGs (40). However, the expression of CGM2 is more limited than that of CEA because a number of tumors known to express CEA transcripts, *e.g.* 60% of 71 breast tumors examined,[2] were all negative for CGM2 transcripts. In the RNAs investigated, when active, the *CGM2* gene is always coexpressed with the *CEA* and *NCA* genes, which together with their close genomic locations indicates that common or closely related regulatory elements may be responsible for coordinating this. Indeed, such an element for regulating the *CGM2* and *CEA* genes could be located in their common 20- kb upstream region, as they are in a head-to-head orientation to each other (Fig. 1).

A regulated coexpression of different members of the CEA family may have functional implications. Coexpression within the same cell and their presence on the cell membrane are reminiscent of other members of the immunoglobulin superfamily that functionally interact forming, *e.g.* the T-cell and B-cell receptor complexes (41). In a similar manner, CEA family members could build complexes on the cell surface. Preliminary data do indeed indicate that CEA family members form complexes on the surface of neutrophils.[3] CEA, NCA-50/90, NCA-95, and possibly CGM2 are GPI-linked to the cell membrane (11, 42, 43, and results presented here). GPI-linked proteins are capable of signal transduction as has been shown for several proteins including NCA-95 (44, 45). Other CEA family members that are also often coexpressed with CEA, NCA, and CGM2, *i.e.* splice variants of BGP (46 and data not shown), represent integral membrane proteins. That BGP may be involved in signal transduction has been shown for the rat BGP homologue, which becomes tyrosine phosphorylated in the cytoplasmic domain through the insulin receptor tyrosine kinase (47). It has also been demonstrated that human BGP is tyrosine phosphorylated (48), a characteristic that suggests its role in signal transduction. The existence of CEA family member complexes could combine the adhesive properties already reported for various family members (20–22, 49) with signal transduction, either through the GPI-linked members, through the integral membrane proteins, or through a combination of both.

The down-regulation of the *CGM2* gene expression at the transcriptional level in colonic adenocarcinomas as compared with the adjacent mucosa is similar to the down-regulation of the *BGP* gene (17, 18), which indicates a potential tumor suppressor function of these genes. Indeed, it has recently been shown through transfection of cDNAs for the rat BGP counterpart cellCAM105, into tumor cell lines, that the tumor phenotype is suppressed.[4] On the other hand, some members of the CEA-family are up-regulated in tumors, *e.g.* NCA at the transcriptional level (17, 50) and CEA at the protein level (51). This differential expression of closely related cell adhesion molecules could play a direct role in either suppression or stimulation of tumorigenesis and/or progression. Down-regulation of other adhesion molecules has been implicated in tumor progression, *e.g.* E-cadherin (23) or the *DCC* gene product, a putative adhesion molecule (24). Alternatively, overexpression of CEA in rodent cells has been shown to enhance liver metastases formation in athymic nude mice (52).

There are two possible explanations for the identity of the 3.2-kb mRNA species that is only seen in tumors (Fig. 5A). First, a *de novo* low level expression of a new CEA-like gene

could occur, as detected through cross-hybridization using the CGM2 probe. Due to the lack of cross-hybridization with other family members known to be coexpressed in colonic mucosa, it would follow that this mRNA species must have a higher sequence similarity to *CGM2* than to other family members. Second, as down-regulation of *CGM2* expression occurs in colorectal cancers, an alternatively spliced CGM2 mRNA species could be expressed. Although alternative splicing of mRNAs coding for members of the CEA gene family has been found (4), to date it has not been reported to occur as a cancer-specific phenomenon. In the case of the cell adhesion molecule CD44, tumor-specific splice variants have been demonstrated to promote metastasis (53).

In the future, it will be of interest to study the putative role of the CGM2 protein in mediating cell adhesion and signal transduction or its ability to build molecular complexes with other CEA family members. Experiments are now underway to see whether overexpression of the CGM2 protein can suppress the malignant phenotype and to test for loss of heterozygosity and mutations in the *CGM2* gene locus of cancer patients.

REFERENCES

1. Teglund, S., Olsen, A., Khan, W. N., Frängsmyr, L., and Hammarström, S. (1994) *Genomics* **23**, 669–684
2. Brandriff, B. F., Gordon, L. A., Tynan, K. T., Olsen, A. S., Mohrenweiser, H. W., Fertitta, A., Carrano, A. V., and Trask, B. J. (1992) *Genomics* **12**, 773–779
3. Paxton, R., Mooser, G., Pande, H., Lee, T. D., and Shively, J. E. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84**, 920–924
4. Thompson, J. A., Grunert, F., and Zimmermann, W. (1991) *J. Clin. Lab. Analysis* **5**, 344–366
5. Thompson, J., Zimmermann, W., Osthus-Bugat, P., Schleussner, C., Eades-Perner, A.-M., Barnert, S., von Kleist, S., Willcocks, T., Craig, I., Tynan, K., Olsen, A., and Mohrenweiser, H. (1992) *Genomics* **12**, 761–772
6. Thompson, J. A., Mauch, E.-M., Chen, F.-S., Hinoda, Y., Schrewe, H., Berling, B., Barnert, S., von Kleist, S., Shively, J. E., and Zimmermann, W. (1989a) *Biochem. Biophys. Res. Commun.* **158**, 996–1004
7. Nap, M., Mollgard, K., Burtin, P., and Fleuren, G. J. (1988) *Tumor Biol.* **9**, 145–153
8. Shively, J. E., and Beatty, J. D. (1985) *CRC Crit. Rev. Oncol. Hematol.* **2**, 355–399
9. Thompson, J., Mössinger, S., Reichardt, V., Engels, U., Beauchemin, N., Kommoss, F., von Kleist, S., and Zimmermann, W. (1993) *Int. J. Cancer* **55**, 311–319
10. Daniel, S., Nagel, G., Johnson, J. P., Lobo, F. M., Hirn, M., Jantscheff, P., Kuroki, M., von Kleist, S., and Grunert, F. (1993) *Int. J. Cancer* **55**, 303–310
11. Berling, B., Kolbinger, F., Grunert, F., Thompson, J. A., Brombacher, F., Buchegger, F., von Kleist, S., and Zimmermann, W. (1990) *Cancer Res.* **50**, 6534–6539
12. Kuroki, M., Arakawa, F., Matsuo, Y., Oikawa, S., Misumi, Y., Nakazato, H., and Matsuoka, Y. (1991) *J. Biol. Chem.* **266**, 11810–11817
13. Nagel, G., Grunert, F., Kuijpers, T. W., Watt, S. M., Thompson, J., and Zimmermann, W. (1993) *Eur. J. Biochem.* **214**, 27–35
14. Stoffel, A., Neumaier, M., Gaida, F. J., Fenger, U., Drzeniek, Z., Haubeck, H. D., and Wagener, C. (1993) *J. Immunol* **150**, 4978–4984
15. Mahrenholz, A. M., Yeh, C. H., Shively, J. E., and Hefta, S. A. (1993) *J. Biol. Chem.* **268**, 13015–13018
16. Svenberg, T. (1976) *Int. J. Cancer* **17**, 588–596
17. Neumaier, M., Paululat, S., Chan, A., Matthaes, P., and Wagener, C. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 10744–10748
18. Rosenberg, M., Nédellec, P., Jothy, S., Fleiszer, D., Turbide, C., and Beauchemin, N. (1993) *Cancer Res.* **53**, 4938–4945
19. Hixson, D. C., and McEntire, K. D. (1989) *Cancer Res.* **49**, 6788–6794
20. Benchimol, S., Fuks, A., Jothy, S., Beauchemin, N., Shirota, K., and Stanners, C. P. (1989) *Cell* **57**, 327–334
21. Oikawa, S., Inuzuka, C., Kuroki, M., Matsuoka, Y., Kosaki, G., and Nakazato, H. (1989) *Biochem. Biophys. Res. Commun.* **164**, 39–45
22. Rojas, M., Fuks, A., and Stanners, C. P. (1990) *Cell Growth & Differ.* **1**, 527–533
23. Risinger, J. I., Berchuck, A., Kohler, M. F., and Boyd, J. (1994) *Nat. Genet.* **7**, 98–102
24. Fearon, E. R., Cho, K. R., Nigro, J. M., Kern, S. E., Simons, J. W., Ruppert, J. M., Hamilton, S. R., Preisinger, A. C., Thomas, G., Kinzler, K. W., and Vogelstein, B. (1990) *Science* **247**, 49–56
25. Schrewe, H., Thompson, J., Bona, M., Hefta, L. J. F., Maruya, A., Hassauer, M., Shively, J. E., von Kleist, S., and Zimmermann, W. (1990) *Mol. Cell. Biol.* **10**, 2738–2748
26. Willcocks, T. C., and Craig, I. W. (1990) *Genomics* **8**, 492–500

---

[2] J. Thompson, H.-J. Weig, and W. Zimmerman, manuscript in preparation.

[3] F. Grunert, personal communication.

[4] S.-H. Lin, personal communication.

27. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
28. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
29. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299
30. Higgins, D. G., and Sharp, P. M. (1988) *Gene (Amst.)* **73,** 237–244
31. Willcocks, T. C., Craig, S. P., and Craig, I. W. (1989) *Ann. Hum. Genet.* **53,** 141–148
32. Zimmermann, W., Ortlieb, B., Friedrich, R., and von Kleist, S. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 2960–2964
33. Toribara, N. W., Sack, T. L., Gum, J. R., Ho, S. B., Shively, J. E., Willson, J. K. V., and Kim, Y. S. (1989) *Cancer Res.* **49,** 3321–3327
34. Williams, A. F., and Barclay, A. N. (1988) *Annu. Rev. Immunol.* **6,** 381–405
35. Neumaier, M., Zimmermann, W., Shively, L., Hinoda, Y., Riggs, A. D., and Shively, J. E. (1988) *J. Biol. Chem.* **263,** 3202–3207
36. Tynan, K., Olsen, A., Trask, B., de Jong, P., Thompson, J., Zimmermann, W., Carrano, A., and Mohrenweiser, H. (1992) *Nucleic Acids Res.* **20,** 1629–1636
37. Thompson, J. A., Pande, H., Paxton, R. J., Shively, L., Padma, A., Simmer, R. L., Todd, C. W., Riggs, A. D., and Shively, J. E. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 2965–2969
38. Kodelja, V., Lucas, K., Barnert, S., von Kleist, S., Thompson, J. A., and Zimmermann, W. (1989) *J. Biol. Chem.* **264,** 6906–6912
39. Thompson, J. A., Barnert, S., Berling, B., von Kleist, S., Kodelja, V., Lucas, K., Mauch, E.-M., Rudert, F., Schrewe, H., Weiss, M., and Zimmermann, W. (1989) in *The Carcinoembryonic Antigen Gene Family* (Yachi, A., and Shively, J., eds) pp. 65–74, Elsevier Science Publishers B.V., Amsterdam
40. Thompson, J., Koumari, R., Wagner, K., Barnert, S., Schleussner, C., Schrewe, H., Zimmermann, W., Müller, G., Schempp, W., Zaninetta, D., Ammaturo, D., and Hardman, N. (1990) *Biochem. Biophys. Res. Commun.* **167,** 848–859
41. Weiss, A., and Littman, D. R. (1994) *Cell* **76,** 263–274
42. Hefta, L. J. F., Schrewe, H., Thompson, J. A., Oikawa, S., Nakazato, H., and Shively, J. E. (1990) *Cancer Res.* **50,** 2397–2403
43. Kolbinger, F., Schwarz, K., Brombacher, F., von Kleist, S., and Grunert, F. (1989) *Biochem. Biophys. Res. Commun.* **161,** 1126–1134
44. Lund-Johansen, F., Olweus, J., Symington, F. W., Arli, A., Thompson, J. S., Vilella, R., Skubitz, K., and Horejsi, V. (1993) *Eur. J. Immunol.* **23,** 2782–2791
45. Shenoy-Scaria, A. M., Gauen, L. K., Kwong, J., Shaw, A. S., and Lublin, D. M. (1993) *J. Cell Biol.* **13,** 6385–6392
46. Barnett, T. R., Drake, L., and Pickle, W., II (1993) *Mol. Cell. Biol.* **13,** 1273–1282
47. Margolis, R. N., Schell, M. J., Taylor, S. I., and Hubbard, A. L. (1990) *Biochem. Biophys. Res. Commun.* **166,** 562–566
48. Afar, D. E. H., Stanners, C. P., and Bell, J. C. (1992) *Biochim. Biophys. Acta* **1134,** 46–52
49. Kuijpers, T. W., Hoogerwerf, M., van der Laan, L., Nagel, G., and van der Schoot, E. (1992) *J. Cell Biol.* **118,** 457–466
50. Chi, K., Jessup, J. M., and Frazier, M. L. (1991) *Tumor Biol.* **12,** 298–308
51. Boucher, D., Cournoyer, D., Stanners, C. P., and Fuks, A. (1989) *Cancer Res.* **49,** 847–852
52. Hashino, J., Fukuda, Y., Oikawa, S., Nakazato, H., and Nakanishi, T. (1994) *Biochem. Biohys. Res. Commun.* **200,** 1748–1753
53. Günthert, U., Hofmann, M., Rudy, W., Reber, S., Zöller, M., Haußmann, I., Matzku, S., Wenzel, A., Ponta, H., and Herrlich, P. (1991) *Cell* **65,** 13–24