

# Long-Range Chromosomal Mapping of the Carcinoembryonic Antigen (CEA) Gene Family Cluster

J. THOMPSON,<sup>\*,1</sup> W. ZIMMERMANN,<sup>\*,2</sup> P. OSTHUS-BUGAT,<sup>\*</sup> C. SCHLEUSSNER,<sup>\*</sup> A.-M. EADES-PERNER,<sup>\*</sup> S. BARNERT,<sup>\*</sup> S. VON KLEIST,<sup>\*</sup> T. WILLCOCKS,<sup>†,1</sup> I. CRAIG,<sup>†</sup> K. TYNAN,<sup>‡,1</sup> A. OLSEN,<sup>‡</sup> AND H. MOHRENWEISER<sup>‡</sup>

<sup>\*</sup>Institute of Immunobiology, University of Freiburg, Stefan-Meier-Strasse 8, D-7800 Freiburg, Germany; <sup>†</sup>Genetics Laboratory, Biochemistry Department, University of Oxford, Oxford, United Kingdom; and <sup>‡</sup>Human Genome Center, Biomedical Sciences Division, Lawrence Livermore National Laboratories, Livermore, California 94550

Received August 2, 1991; revised November 25, 1991

A long-range physical map of the carcinoembryonic antigen (CEA) gene family cluster, which is located on the long arm of chromosome 19, has been constructed. This was achieved by hybridization analysis of large DNA fragments separated by pulse-field gel electrophoresis and of DNA from human/rodent somatic cell hybrids, as well as the assembly of ordered sets of cosmids for this gene region into contigs. The different approaches yielded very similar results and indicate that the entire gene family is contained within a region located at position 19q13.1-q13.2 between the CYP2A and the D19S15/D19S8 markers. The physical linkage of nine genes belonging to the CEA subgroup and their location with respect to the pregnancy-specific glycoprotein (PSG) subgroup genes have been determined, and the latter are located closer to the telomere. From large groups of ordered cosmid clones, the identity of all known CEA subgroup genes has been confirmed either by hybridization using gene-specific probes or by DNA sequencing. These studies have identified a new member of the CEA subgroup (CGM8), which probably represents a pseudogene due to the existence of two stop codons, one in the leader and one in the N-terminal domain exons. The gene order and orientation, which were determined by hybridization with probes from the 5' and 3' regions of the genes, are as follows: cen/3'-CGM7-5'/3'-CGM2-5'/5'-CEA-3'/5'-NCA-3'/5'-CGM1-3'/3'-BGP-5'/3'-CGM9-5'/3'-CGM6-5'/5'-CGM8-3'/PSGcluster/qter. © 1992 Academic Press, Inc.

its relatively high concentrations in the sera of many cancer patients, it is a widely used marker in the postoperative surveillance of such patients (Shively and Beatty, 1985). Indeed, CEA is also gaining in importance for immunolocalization of tumors and their metastases using specific radiolabeled monoclonal antibodies (Bischoff-Delaloye *et al.*, 1989), and studies in rodents indicate that such antibodies may also have a future role in tumor therapy (Sharkey *et al.*, 1991). With the successful molecular cloning of CEA and related genes or cDNA clones, other clinical applications have been reviewed (Thompson *et al.*, 1991). Although the *in vivo* function of this glycoprotein has not been determined, strong evidence unequivocally shows that CEA functions *in vitro* as a Ca<sup>2+</sup>-independent, cellular adhesion molecule (Benchimol *et al.*, 1989; Oikawa *et al.*, 1989). It has also been reported that CEA specifically binds *Escherichia coli* strains of human origin via a lectin/carbohydrate interaction (Leusch *et al.*, 1990), and its possible role as an accessory molecule in binding tumor cells to collagen type I has also been discussed (Pignatelli *et al.*, 1990).

Soon after the discovery of CEA, a number of cross-reacting antigens with variable expression patterns were identified, e.g., the nonspecific cross-reacting antigen (NCA) discovered by von Kleist *et al.* (1972), and independently by Mach and Pusztaszeri (1972), was found in a variety of different normal tissues such as lung, spleen, and liver, as well as in polymorphonuclear granulocytes, macrophages, and monocytes (Bordes *et al.*, 1975). On the other hand, the biliary glycoprotein I (BGP-I), another closely related molecule, was discovered in the epithelium of bile canaliculi (Svenberg, 1976). More recently, through molecular cloning studies, it has been found that these molecules are encoded by a gene family, composed of at least 19 genes, which can be divided into two main subgroups based on sequence comparisons (summarized in Barnett and Zimmermann, 1990). One subgroup contains the genes coding for CEA and the classical CEA cross-reacting antigens, such as NCA and BGP, together with a number of other genes whose products have not been unequivocally determined. These

## INTRODUCTION

The carcinoembryonic antigen (CEA) family consists of a heterogeneous group of glycoproteins whose molecular weights range from 20,000 to 200,000 (Thompson and Zimmermann, 1988). CEA was the first member of this family to be described in colonic tumors as well as in some fetal tissues (Gold and Freedman, 1965a,b). Due to

The DNA sequence for CGM8 has been submitted to GenBank and given Accession No. M80575.

<sup>1</sup> Joint first authors: The data presented were contributed equally from the three laboratories listed.

<sup>2</sup> To whom correspondence should be addressed.

genes have been named CGM1, CGM2 (Thompson *et al.*, 1989), and CGM6 (Arakawa *et al.*, 1990; Berling *et al.*, 1990). The second subgroup consists of the pregnancy-specific glycoprotein (PSG) genes, whose protein products were discovered by Tatarinov and Masyukevich (1970) and independently by Bohn (1971). PSGs are synthesized by the placenta in increasing amounts during pregnancy, and they are secreted into the maternal blood (Lin *et al.*, 1974). Because PSGs are also found in tumors of trophoblastic origin, they are potentially useful as tumor markers (Tatarinov, 1978).

The CEA gene family belongs to the immunoglobulin superfamily (Paxton *et al.*, 1987), whereby all CEA-related proteins consist of one immunoglobulin variable-like (N-terminal) domain and a different number of immunoglobulin constant-like domains, ranging from none (e.g., in CGM7) to six in CEA. These IgC-like regions are designated A and B domains and together they build a highly conserved repeating unit, which is present in three copies in CEA.

A number of groups have shown that the CEA gene family is clustered on the long arm of chromosome 19 in the region 19q13.1–19q13.3 (Kamarck *et al.*, 1987; Zimmermann *et al.*, 1988; Barnett *et al.*, 1989; Niemann *et al.*, 1989; Willcocks *et al.*, 1989; Streydio *et al.*, 1990; Thompson *et al.*, 1990; Brandriff *et al.*, 1992, accompanying manuscript). More detailed analyses by pulse-field gel electrophoresis (PFGE) and chromosome walking have revealed that most members of the PSG subgroup are located within an 800-kb *SacII* restriction endonuclease fragment, on which individual genes are closely linked and show coordinated transcriptional patterns (Thompson *et al.*, 1990). As a basis toward the identification of all members of the CEA gene family and toward studying their evolution, as well as investigations on the regulation of their expression at the transcriptional level, this paper describes fine structural mapping studies of the CEA gene family. The locations of all known members of the CEA subgroup, with respect to each other, have been determined by PFGE analysis of overlapping cosmid clones and somatic cell hybrid studies.

## MATERIALS AND METHODS

**Pulse-field gel electrophoresis.** DNA was isolated from two different primary human fibroblast cultures (line 902 from a healthy individual, and line 1243 from a child with spinal muscular dystrophy). Also studied was DNA from an Epstein-Barr virus-immortalized, lymphoblast cell line (cell line L1164 from a patient with a defect on the X chromosome). All cells were a kind gift from the Institute of Human Genetics (Freiburg University, FRG). Cells were prepared in agarose blocks for DNA extraction according to the European Molecular Biology Laboratory (EMBL, Heidelberg, FRG) protocol. After digestion with rare-cutting restriction endonucleases (5 units/ $\mu$ g DNA for 4 h), fragment separation was carried out on a CHEF-type model electrophoresis unit (EMBL) at 4 V/cm and a pulse time of 90 s for 40–60 h on an 0.8% agarose gel at 15°C, according to the manufacturer's instructions, for separation of fragments up to approximately 1 Mb. Size markers were chromosomes from *Saccharomyces cerevisiae* and either were isolated from the yeast strain 2180 1RA using a modification of the method described by Schwartz and Cantor (1984) or were pur-

chased from Beckman.  $\lambda$  DNA oligomers were prepared from phage strain cI857 Sam7 (Maniatis *et al.*, 1982). DNA was depurinated twice in 0.25 M HCl for 10–15 min and transferred for 48 h according to the alkaline method of Chomczynski and Qasba (1984) onto Pall Biodyne (Pall Biosupport Division) or GeneScreen Plus (New England Nuclear) membranes. After neutralization, the membranes were air-dried for 1 h and baked for 1 h at 80°C in a vacuum oven, followed by a 2-min exposure to uv irradiation (254 nm). Gene-specific DNA fragments were radiolabeled with [ $\alpha$ -<sup>32</sup>P]dATP (Amersham) after Feinberg and Vogelstein (1983). Prehybridization was carried out in 40% deionized formamide, 1% sodium dodecyl sulfate (SDS), 1 M NaCl, 10% dextran sulfate for 30 min at 42°C, followed by hybridization in the same buffer at 42°C for 12–24 h, after addition of 100  $\mu$ g of denatured salmon sperm DNA and approximately  $4 \times 10^6$  dpm/ml radiolabeled probe. Final washes were performed at low stringency in 2 $\times$  SSPE (1 $\times$  SSPE is 0.18 M NaCl, 10 mM sodium phosphate (pH 7.4), and 1 mM EDTA), 1% SDS twice for 15 min at 65°C, followed by high stringency in 0.1 $\times$  SSPE, 0.1% SDS twice for 15 min at 62–65°C.

For distinguishing the CEA and NCA genes, a 408-bp *RsaI/PstI* fragment (CEA 3'a) and a 1.5-kb *EcoRI* fragment (NCA 3') were taken from the 3'-untranslated regions of corresponding cDNA clones. The specificity of these probes has been described elsewhere (Zimmermann *et al.*, 1988; Berling *et al.*, 1990). An additional CEA gene-specific probe (5'CEA) was taken from the 5'-region of a genomic clone (cosCEA1), which we recently isolated (Schrewe *et al.*, 1990). This 526-bp *SstI/SacII* fragment is located approximately 1 kb upstream from the transcriptional start site of the CEA gene. For identification of all PSG subgroup but not CEA subgroup genes, the total coding region minus the 3'-untranslated portion of a PSG1a cDNA clone (Zimmermann *et al.*, 1989) was hybridized as described (Thompson *et al.*, 1990). The specific probe taken for CGM6 was a 297-bp *BglI/EcoRI* restriction fragment from the 3'-untranslated region of an incomplete CGM6 cDNA clone reported elsewhere (Berling *et al.*, 1990). To differentiate the CGM2 gene, a 781-bp *PstI* fragment was taken from a genomic clone ( $\lambda$ hsCGM2-1), which contains the N-terminal domain exon and surrounding intron sequences (Thompson *et al.*, 1989). For recognition of the BGP gene, a 396-bp fragment was used from the 3'-untranslated region of a BGP cDNA clone (4-13), which extends from a *PstI* site located 19 bp downstream from the stop codon to the end of the clone (Hinoda *et al.*, 1989). Finally, a 1.9- and a 1.3-kb *BamHI* fragment containing the N-terminal domain exons of CGM7 and CGM8, respectively (see below), were used.

**Screening the cosmid libraries.** The cosmid library was constructed from human chromosome 19 (de Jong *et al.*, 1989). Clones containing CEA-related genes were isolated after hybridization with the 534-bp *PstI* fragment from the repeat regions of a CEA cDNA clone (pCEA1), described elsewhere (Zimmermann *et al.*, 1987), followed by washing up to 0.1 $\times$  SSC (1 $\times$  SSC is 0.15 M NaCl, 0.015 M sodium citrate, pH 7.0), 0.1% SDS at 55°C. The assembly of cosmid contigs was achieved by employing methods described by Carrano *et al.* (1989) and Branscomb *et al.* (1990). Cosmids within these contigs that contained individual CEA-related genes were identified as described below.

**Identification of CEA-related genes by hybridization of cosmid clones with gene-specific probes.** Cosmid clones shown to contain CEA-related gene sequences were hybridized with specific probes for gene identification. To differentiate clones containing the CGM2 gene, the 781-bp *PstI* genomic fragment described above was used, applying a high-stringency wash. For identification of cosmid clones carrying the CEA and NCA genes, the following oligonucleotides based on complementary CEA- and NCA-specific sequences from the A1 and A domains, respectively, were used as probes:

CEA oligonucleotide: 5'-GTAGCTTGCTGTGTCATTTTC-3'  
NCA oligonucleotide: 5'-GTTCTCTTTGACGCTGAGTA-3'

Both oligonucleotides were hybridized overnight at 45°C in 0.5 M sodium phosphate buffer (pH 7.2), 7% SDS, 1 mM EDTA according to Church and Gilbert (1984). Final washes were twice for 20 min in 6 $\times$  SSPE, 0.1% SDS at 55°C in each case. Cosmid clones bearing the CGM6 gene were identified by hybridization with the amplification

product from a polymerase chain reaction, which covers the last 235 bp of the 3'-untranslated region of an incomplete CGM6 cDNA clone (Berling *et al.*, 1990), followed by a high-stringency wash (0.1× SSPE, 0.1% SDS at 65°C). Clones containing the BGP gene were identified by hybridization with a 245-bp *HindIII/BamHI* fragment corresponding to the Cyt domains of a BGP cDNA clone (TM1-CEA) described by Barnett *et al.* (1989) and washing up to a stringency of 0.1× SSC, 0.1% SDS at 50°C. Recently, we received sequence information (W. N. Khan, L. Frängsmyr, S. Teglund, and S. Hammarström, Umea, Sweden, personal communication) for a new CEA subgroup member (CGM9). Given these data, we synthesized a gene-specific oligonucleotide from the N-terminal domain exon, which was used to probe cosmid clones. The CGM9 oligonucleotide

5'-TCCTGTGTCCTGGGTGACAT-3'

was hybridized and washed as described for the CEA and NCA oligonucleotides with a final wash temperature of 60°C.

*Analysis of cosmid clones by polymerase chain reaction or fragment subcloning and DNA sequencing.* Cosmid clones containing unknown CEA-related genes were digested with restriction enzymes, size-separated on agarose gels, and transferred to charged nylon membranes prior to hybridization with a 220-bp *PstI* probe covering part of the leader and N-terminal domains of an NCA cDNA clone (pNCA1) described elsewhere (Kodelja *et al.*, 1989) under low-stringency conditions (final wash 2× SSPE, 0.1% SDS at 65°C). Some of these clones were further analyzed by the polymerase chain reaction for amplification of regions within the N-terminal domains, using the following two oligonucleotide primers:

PCRCEAall5': 5'-GGAATTCGAGAGGGGAAGGAGGT-3'  
PCRCEAall3': 5'-GGAATTCATCAGCAGGGATGCATTGG-3'

These oligonucleotides were derived from sequences that are conserved for all known members of the CEA subgroup. Only in comparison with CGM2 (Thompson *et al.*, 1989) are one and two mismatches found, respectively. *EcoRI* linkers were included to assist in subcloning the fragments for sequencing (bold print). PCR was performed using 1–10 ng nonlinearized cosmid DNA over 30 cycles using a Bio-Med thermocycler 60 (Bachhofer), basically according to Sambrook *et al.* (1989), with the following modifications: The melting temperature was 93°C, and annealing was for 15 s at 56°C. *Taq* polymerase and the amplification buffer were from Promega, and the dNTPs were supplied by Pharmacia. Amplified fragments were subcloned into a Bluescript phagemid vector (Stratagene), either directly as blunt-ended fragments into the *EcoRV* site or after *EcoRI* digestion into the *EcoRI* site.

For cosmid clones containing the N-terminal domain exons of CGM7, CGM6, and CGM8, a 1.9-kb and two 1.3-kb *BamHI* fragments, respectively, were subcloned into Bluescript vectors. The PCR amplification products from CGM1 and the subcloned CGM6, CGM7, and CGM8 fragments were sequenced by the dideoxy-chain termination method (Sanger *et al.*, 1977) with universal and internal primers using a kit from Pharmacia.

*Determination of gene orientation.* To determine the transcriptional direction of the genes within all overlapping cosmid clones that contain CEA-related gene sequences, the hybridization patterns of gene-specific probes from the 3'-untranslated regions of the different genes were compared with the hybridization signals gained with the 220-bp *PstI* probe from the leader and N-terminal domains of pNCA1 (Kodelja *et al.*, 1989) under low-stringency washing conditions (final wash 2× SSPE, 0.1% SDS at 65°C), with *EcoRI*- and *SstI*-digested cosmid clones. For determining the presence of the 3'-untranslated region of the BGP and CGM6 genes, the same probes described above for hybridization of the fragments separated by PFGE or for identification of the CGM6-containing cosmid clones, respectively, were used at high-stringency washes (0.1× SSPE, 0.1% SDS at 65°C). The probe used to determine the presence of the 3'-untranslated region of CGM7 was an oligonucleotide originally synthesized on the basis of sequence

data from the complementary strand of the 3'-untranslated region of CGM1 (results to be published), with the sequence 5'-CTGTGAGG-TCTCCAGA-3'. In comparison with cDNA sequence data from CGM7 (Kuroki *et al.*, 1991), two mismatches were found at positions 6 and 7, which are reversed in the complementary strand of CGM7 (AG instead of GA). Hybridization was carried out at 37°C in 0.5 M sodium phosphate buffer (pH 7.2), 7% SDS, 1 mM EDTA according to Church and Gilbert (1984). Final washes were twice for 20 min in 6× SSPE, 0.1% SDS at 43°C.

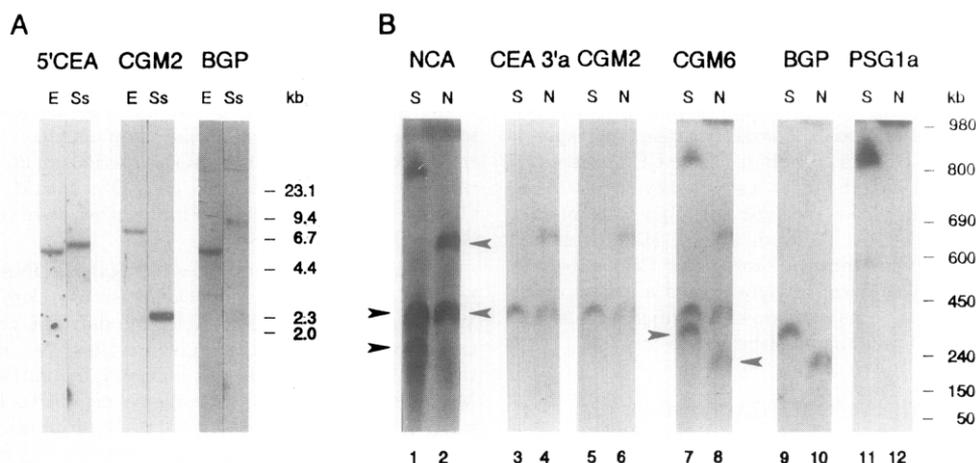
*Analysis of somatic cell hybrids.* Hybrid DNA was kindly provided by Drs. Duncan Shaw and David Brook (Cardiff, UK) and Dr. Bé Wieringa (Nijmegen, The Netherlands). The chromosome compositions of the hybrids used are given in the papers describing their isolation. In brief, 20XP3542-1-4 (hamster/human) contains about 20–30 Mb of chromosome 19q (Stallings *et al.*, 1988). TVBID (hamster/human) contains 1pter–1q31::19q12–19qter plus other human chromosomes. TV7C (hamster/human) contains the reciprocal translation 19pter–19q12::1q31.2–1qter plus other human chromosomes (Brook *et al.*, 1987a). G175B3 and G175C4 are cloned sublines derived from the mouse/human line G175AoXiB, which is heterogeneous for human chromosome 19 (Brook *et al.*, 1987b). WILF M2 and WILF M6 are subclones of the hybrid line WILF1 and contain unique fragments of human chromosome 19 plus further human chromosome material. J64034638 contains a rearranged chromosome 19 translocated onto a Chinese hamster chromosome. These hybrids have been characterized by Brook *et al.* (1991). 908K1B18 contains the translocated chromosome fragment 19cen–19q13.2 as its only human genetic contribution (Schonk *et al.*, 1989).

DNA purification from peripheral blood and from somatic hybrids followed by restriction endonuclease analyses was as described (Sambrook *et al.*, 1989). DNA was transferred to Hybond-N membranes (Amersham International), and hybridization procedures were performed as recommended by the suppliers with the following probes: For determination of the presence of members of the CEA subgroup, a 2.6-kb *XbaI* fragment from the 5'-end of the CEA gene was used (Willcocks *et al.*, 1987). To differentiate PSG subgroup members and the CGM6 gene, the same probes as those for the PFGE analyses were used. All probes were radiolabeled to specific activities of 2–5 × 10<sup>8</sup> dpm/μg DNA. Filters were washed up to 65°C in 0.1× SSC, 1% sodium phosphate, 0.1% SDS for the CEA and PSG probes. The CGM6 probe was washed similarly but in 0.2× SSC.

## RESULTS

### *Determination of Linkage of CEA Gene Family Members by Pulse-Field Gel Electrophoresis*

Prior to analysis of the DNA fragments separated by PFGE, it was necessary to determine the specificity of the probes used to identify individual genes. Probes taken from the 3'-untranslated regions of CEA and NCA cDNA clones (Zimmermann *et al.*, 1988), as well as from the CGM6 cDNA clone (Berling *et al.*, 1990), have been shown elsewhere to be specific under high-stringency washing conditions. Similarly, the PSG1a probe was described elsewhere as having subgroup specificity at high wash stringency (Thompson *et al.*, 1990). A specificity control was carried out for the 5'-untranslated region CEA probe (5'CEA), along with the probes for CGM2 and BGP. As seen in Fig. 1A, these probes only hybridize with one *EcoRI* and one *SstI* DNA restriction fragment each, from total genomic DNA digests, under high-stringency washing conditions, indicating gene specificity. These results, however, do not rule out the existence of very closely related genes with restriction fragments of identical size.



**FIG. 1.** Hybridization of human genomic DNA digests with specific probes from CEA-related genes. (A) Specificity control of *EcoRI*-digested (E) and *SstI*-digested (Ss) human DNA hybridized with specific probes indicated above the lanes and washed at high stringency (see Materials and Methods). (B) Human lymphoblast DNA (cell line L1164) digested with *SacII* (S) and *NruI* (N) followed by PFGE and transfer to charged nylon membranes. The gene-specific probes used for hybridization are indicated above the lanes. Lanes 1, 2, 7, and 8 were washed at low stringency ( $2\times$  SSPE, 1% SDS at  $65^{\circ}\text{C}$ ) and the rest at high stringency. Arrowheads indicate signals remaining after high-stringency wash. The numbers at the right margins show the sizes of  $\lambda$  *HindIII* and yeast chromosome DNA markers in kb.

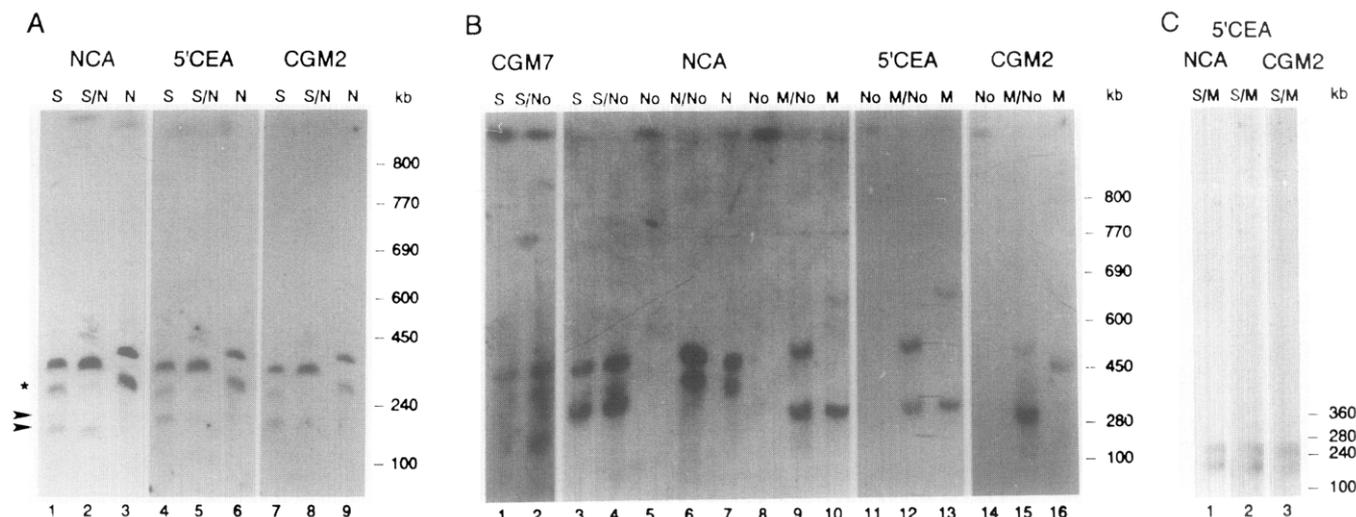
Digestion of cellular DNA with the rare-cutting restriction endonucleases *SacII* and *NruI*, followed by transfer to membranes and sequential hybridization and stripping of the same membrane with the various gene-specific probes, allowed determination of physical linkage of the genes (Fig. 1B). The probes for the NCA, CEA, and CGM2 genes all hybridized with common 250-kb *SacII* and 420-kb *NruI* fragments (lanes 1–6)<sup>3</sup>. The probes for the CGM6 and BGP genes, on the other hand, hybridized with common 360-kb *SacII* and 240-kb *NruI* fragments (lanes 7–10). Most or all of the PSG genes are located on a common 800-kb *SacII* fragment (lane 11). Multiple bands seen in some lanes, even after high-stringency washes, probably represent incompletely digested DNA, possibly due to partial methylation at the restriction sites.

**Mapping the CGM7/CGM2/CEA/NCA cluster.** Closer analyses of the CEA gene cluster were carried out to determine the positions and orientation of these genes relative to each other. Sequence analysis reported elsewhere (Schrewe *et al.*, 1990) has revealed the existence of two *SacII* restriction sites in the 5'-region of the CEA gene. Therefore, using a genomic probe that lies upstream from these sites, the *SacII* fragment that hybridizes should be different from that with the probe from the 3'-untranslated region of the CEA gene, assuming that these two *SacII* sites are not fully methylated. The results of the hybridization with the 526-bp *SstI/SacII* 5'-fragment compared to those with the CGM2 and NCA probes can be seen in Fig. 2A for fibroblast cell line 902 DNA. These results revealed that the 5'CEA and CGM2 probes hybridize with a common 220-kb *SacII* fragment (lanes 4 and 7), whereas the NCA (lane 1) and the CEA 3' probe (results not shown) hybridize with a common

180-kb *SacII* fragment. Neither of these fragments was visible in Fig. 1B, presumably because the DNA of the lymphoblast cells analyzed there was differentially methylated to the fibroblast cell DNA shown in Fig. 2A. These results indicate the following order of these three genes: CGM2–5' CEA 3'–NCA.

To construct a long-range restriction map of this region, both single and double digestions were carried out using combinations of rare-cutting restriction endonucleases. Partial digestions with single enzymes, which are presumably caused by methylation at the restriction sites, assist in mapping. In a comparison of the results shown in Figs. 2A with 2B, differences in the *SacII* fragment patterns are seen between the two fibroblast cell lines analyzed. Cell line 902 (Fig. 2A, lanes 1, 4, and 7) revealed three *SacII* fragments with each gene-specific probe. It would appear that the 420-kb and the 350-kb *SacII* DNA fragments contain all three genes. However, other investigations using the BGP-specific probe revealed that the 350-kb *SacII* fragment seen here does not contain these genes but is due to cross-hybridization with the BGP gene cluster-containing *SacII* fragment (see below). The 420-kb fragment can be split into a 180-kb *SacII* fragment, which contains the NCA gene and 3'-end of the CEA gene, and a 220-kb *SacII* fragment, containing the 5'-end of the CEA gene and the CGM2 gene. In an analysis of the fibroblast cell line 1243 with the NCA probe (Fig. 2B, lane 3), the two larger *SacII* fragments correspond in size to those found with cell line 902 (420 and 350 kb), but the smallest is slightly larger (280 kb in line 1243 compared with 220 and 180 kb in line 902) and contains all three genes. This indicates that the *SacII* sites within the CEA gene must be fully methylated in cell line 1243, but not so in cell line 902. The *SacII* site immediately beyond the CGM2 gene, on the other hand, is fully methylated in cell line 902, but not so in cell line 1243 (see also Fig. 7). After first isolating and identifying CGM7 gene-containing cosmid

<sup>3</sup> Exact fragment size determination is not possible using this method so that size variability for the same fragment between figures is sometimes found.



**FIG. 2.** Mapping the CGM7/CGM2/CEA/NCA cluster by PFGE. (A) Determination of the gene order and orientation of the CEA gene. Digests of human DNA (fibroblast line 902) hybridized with the probes shown above the lanes at high wash stringency. (B) Single and double digests of human DNA (fibroblast line 1243) hybridized with the probes indicated above the lanes. S, *SacII*; N, *NruI*; No, *NotI*; M, *MluI*. (C) Fibroblast DNA (cell line 1243) digested with *SacII/MluI* (S/M) and hybridized with the probes shown above the lanes. The asterisk indicates a cross-hybridizing fragment containing the BGP gene cluster. Arrowheads show the size difference between the smallest NCA-hybridizing *SacII* (180 kb) and the CEA/CGM2-hybridizing *SacII* fragments (220 kb).

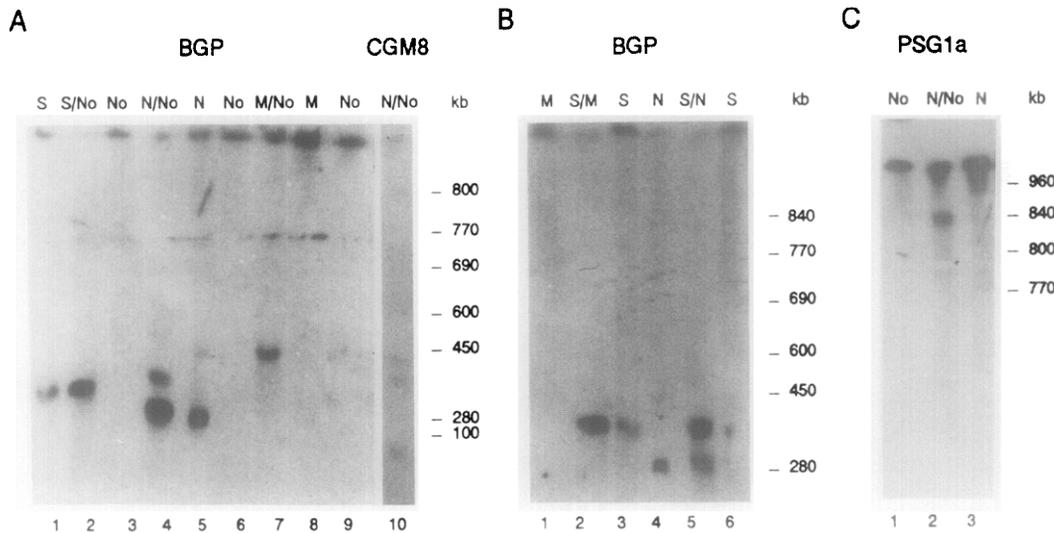
clones from the CGM2/CEA/NCA contig (see below), we hybridized the cell line 1243 genomic DNA *SacII* digests (Fig. 2B, lane 1) and found that this gene must be located beyond the *SacII* restriction site which is digestible only in this cell line. The hybridizing *SacII* fragments are 170 and 420 kb, and the cross-hybridizing BGP gene cluster fragment of 350 kb is also visible. Although the fragments are faint, the pattern is identical to that in Fig. 2B, lane 2, because there are no internal *NotI* sites in this region.

The genomic DNAs of both cell lines have common *NruI* fragments of 420 and 350 kb (Fig. 2A, lanes 3, 6, and 9; and Fig. 2B, lane 7), each of which contains the CEA, NCA, and CGM2 genes. After double digestion with *SacII* and *NruI*, the ensuing pattern is identical to the one obtained by *SacII* single digestion (Fig. 2A, compare lanes 2, 5, and 8 with lanes 1, 4, and 7, respectively). The gene probes all hybridized with a *NotI* fragment that ran in the low mobility range (Fig. 2B, lanes 5, 8, 11, and 14). Further digestion of the *SacII* or *NruI* digests with *NotI* did not change their hybridization patterns and was identical for the CEA, CGM2, and NCA gene probes, but is shown only for the NCA probe (Fig. 2B, compare lane 3 with lane 4, and lane 7 with lane 6), as well as the CGM7 probe (Fig. 2B, compare lanes 1 and 2), indicating the absence of unmethylated *NotI* sites within these DNA fragments.

The 5'CEA and NCA probes hybridize with identical *MluI* fragments of 650 and 280 kb (Fig. 2B, lanes 10 and 13). CGM2, on the other hand, hybridizes weakly with the 650-kb and in addition with a 450-kb *MluI* fragment (Fig. 2B, lane 16). Therefore, the 650-kb fragment contains all three genes and is split into a 280-kb fragment containing the CEA and NCA genes and a 450-kb fragment that contains the CGM2 gene. The 650- and 450-kb *MluI* fragments each become 150 kb shorter after

further digestion with *NotI* (Fig. 2B, lanes 9, 12, and 15). To combine these results, *SacII/MluI* double digestions were carried out (Fig. 2C). All three probes hybridized with fragments of 250 and 200 kb. As we have shown that the CGM2 gene is on a *MluI* separate fragment from the CEA and NCA genes, double digestion with *SacII/MluI* must have yielded different fragments of similar lengths. The only possible order of restriction fragments is shown in a long-range map (Fig. 7).

**Mapping the CGM6/BGP/CGM8 cluster.** The single and double digests for construction of a long-range restriction map of the BGP gene cluster are shown in Figs. 3A and 3B. As probes for the BGP and CGM6 genes yielded identical patterns, only the results with the BGP probe are shown. Both these genes are located on *MluI* and *NotI* fragments that migrate in the low mobility range and whose sizes cannot, therefore, be determined (Fig. 3A, lanes 6 and 8; Fig. 3B, lane 1). However, after double digestion with the same enzymes, a 450-kb fragment is visible (Fig. 3A, lane 7). This indicates the presence of the CGM6 and BGP genes on this smaller fragment. Both probes also hybridized with the same 280-kb and a partial 450-kb *NruI* fragment (Fig. 3A, lane 5; and Fig. 3B, lane 4). The 450-kb *NruI* fragment becomes about 100 kb shorter after *NotI* digestion (Fig. 3A, compare lanes 5 and 4). Therefore, the 280-kb *NruI* fragment must be located within the 450-kb *NotI/MluI* fragment, approximately 100 kb from the *NotI* site. Furthermore, this *NotI/MluI* fragment is adjacent to another *NotI/MluI* fragment of >450 kb that probably contains the CEA gene cluster, indicated by a question mark in Fig. 7. As the faintly visible 450-kb *NruI* fragment is also shortened by approximately 100 kb after *SacII* digestion, we assume that the *SacII* and *NotI* sites must be close together (Fig. 3B, lanes 4 and 5). The hybridizing *SacII* fragment has a length of 350 kb (Fig. 3A, lane 1; and Fig.



**FIG. 3.** Mapping the BGP/CGM6/CGM8 cluster (A and B) and the PSG gene cluster (C) by PFGE. Digested fibroblast DNA (cell line 1243) was hybridized with the BGP gene-specific probe (A and B), the CGM8-specific probe (A), or the PSG1a probe (C). Abbreviations as for Fig. 2.

3B, lanes 3 and 6), so the other *NruI* site must also be close to a *SacII* site. This interpretation is supported by the *NotI/SacII* and *MluI/SacII* double digests, where the 350-kb *SacII* fragment is not further digested (Fig. 3A, lanes 1 and 2; Fig. 3B, lanes 2 and 3). After we isolated and identified the CGM8 gene-containing cosmids from the BGP contig (see below and Fig. 5), we carried out a hybridization using the CGM8 probe and were able to confirm that this gene is located on a *NotI/NruI* fragment of approximately 100 kb (Fig. 3A, lane 10), indicating that this gene lies beyond the *NruI* site adjacent to the BGP and CGM6 genes on the 100-kb *NruI/NotI* fragment. All three genes hybridize with a common 400-kb *NruI/NotI* fragment. These results lead to the long-range restriction map covering the CGM6/BGP cluster shown in Fig. 7.

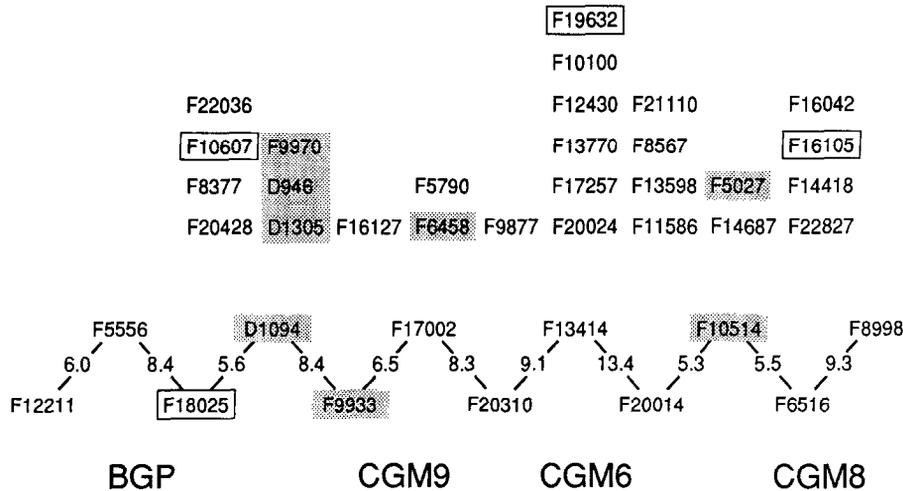
**Mapping the PSG gene cluster.** The hybridization conditions used with the PSG1a probe were shown elsewhere to differentiate between the PSG gene and the CEA gene subgroup members (Thompson *et al.*, 1990). This probe hybridized with a 950-kb and a 1-Mb *NruI* fragment, as well as giving a strong signal in the low mobility range (Fig. 3C, lane 3). The smallest *NruI* fragment was shortened by approximately 110 kb after *NotI* digestion (Fig. 3C, compare lanes 2 and 3). As no 110-kb *NotI/NruI* fragment could be visualized after hybridization, we assume that this does not contain PSG genes. *NotI* digestion yielded a large hybridizing fragment, which migrated in the low mobility range (lane 1). The long-range map for the PSG gene cluster is shown in Fig. 7.

#### CEA Gene Subgroup Mapping to Cosmid Contigs

In parallel to the long-range mapping studies, we have also carried out gene mapping by assembling cosmids into contigs and assigning genes to specific cosmids. For this purpose, contigs of overlapping sets of cosmids were

assembled using a strategy of high-resolution restriction fragment fingerprinting of clones from a human chromosome 19 library. Those cosmids belonging to the CEA family were identified by hybridization with a probe covering the constant domain coding regions of a CEA cDNA clone (Tynan *et al.*, unpublished data). One large contig of 175 kb was shown to contain four distinct regions of hybridization to the CEA IgC probe, suggesting that four genes are localized to this contig. Hybridization with gene-specific probes showed that three of these genes were CGM2, CEA, and NCA and confirmed their order as indicated by PFGE (Fig. 7). The fourth area of hybridization, directly upstream of the CGM2 gene, indicates the presence of a closely linked additional member of this family. Sequence data from a subcloned 1.9-kb *BamHI* fragment containing the first exon and 254 nucleotides from the second exon revealed that this gene corresponds to the recently described CGM7 cDNA clone (Kuroki *et al.*, 1991).

Another contig spanned approximately 160 kb and contained groups of cosmids positive for the BGP and CGM6 probes. DNA sequencing of a 1.3-kb *BamHI* fragment covering part of the first and second exons confirmed the identity of the CGM6 gene by comparison with corresponding cDNA sequence data (Arakawa *et al.*, 1990; Berling *et al.*, 1990). A visual representation of all the cosmids in this contig (Fig. 4) shows the location of the BGP and CGM6 positive clones as well as those clones in the contig that were positive with the CEA repeat region probe. The identification of a third group of cosmids containing a CEA repeat region in this contig indicated the location of another CEA family gene immediately upstream of the CGM6 gene. As can be seen in Fig. 5, sequence data from a subcloned 1.3-kb *BamHI* fragment containing part of the first and second exons revealed that this gene (CGM8) represents a new member of the CEA family. The presence of stop codons in the first and second exons strongly indicates that this is a pseudogene. Hybridization with members of this

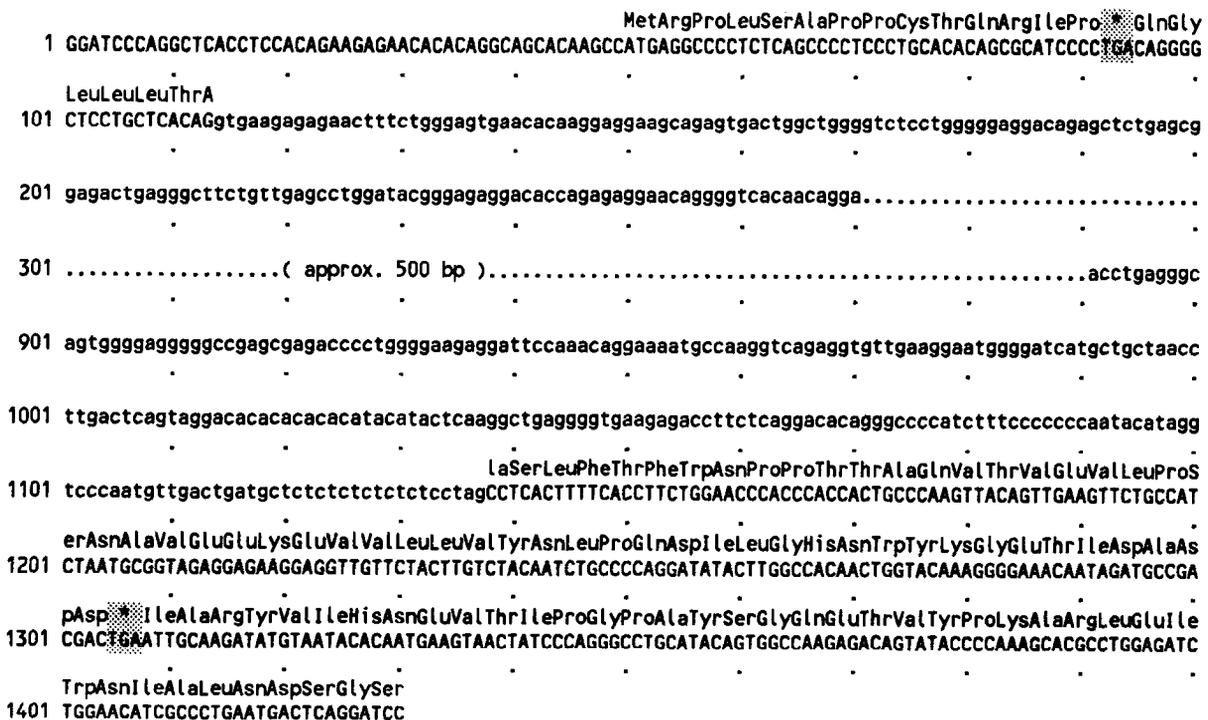


**FIG. 4.** Linkage of BGP, CGM9, CGM6, and CGM8 in contig 29. The screen image is produced using a cosmid/contig database visual access tool constructed at Lawrence Livermore National Laboratory called "Browser" and based on the X11 SUN windowing system (Carrano *et al.*, 1990). Browser generally uses color to represent the estimated confidence ( $\log_{10}$  likelihood ratio, referred to as the L value) of overlap between cosmids. However, in this case the overlap between members of the minimum tiling path is indicated by slanting bars with the strength of the bond displayed. Shaded regions indicate non-CEA IgC-positive or anonymous cosmids. Boxed cosmids have been mapped to bands 19q13.1-13.2 by fluorescence *in situ* hybridization.

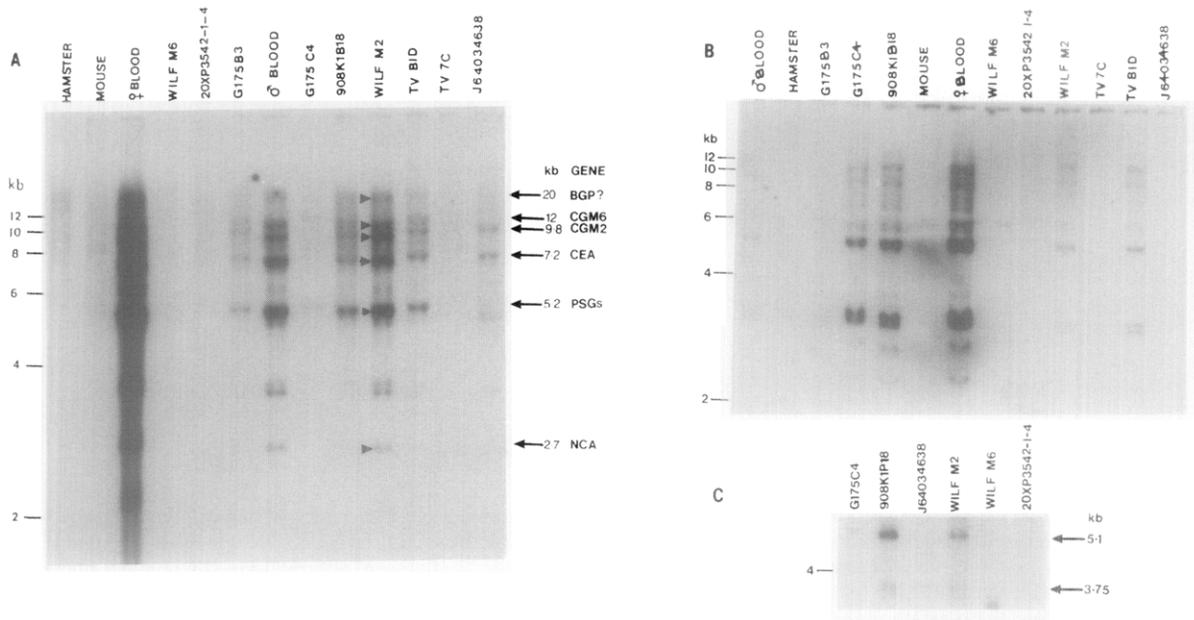
contig using a probe specific for the recently discovered member of the CEA gene subgroup, CGM9 (see Materials and Methods), under high-stringency conditions identified four positive clones (F9933, F17002, F5790, and F20310), which are located between the BGP and the CGM6 gene-containing clones (see Fig. 4).

A third contig spanning approximately 70 kb also contained cosmid clones that hybridized with the CEA repeat region probe. Closer analyses by restriction endonuclease digestion and hybridization with the 220-bp *Pst*I

probe from the N-terminal domain of an NCA cDNA clone (pNCA1) revealed the presence of one CEA-related gene. A DNA fragment was amplified by the polymerase chain reaction using the PCRCEA11 oligonucleotide primers and subcloned. This gene was identified by DNA sequencing to be the CGM1 gene reported elsewhere (Thompson *et al.*, 1989). The identification of cosmid clones linking this contig with the 175-kb contig reveals that the CGM1 gene lies directly downstream of the NCA gene (Tynan *et al.*, unpublished).



**FIG. 5.** DNA sequence data from the first and second exons of CGM8. Stop codons are shaded and indicated by an asterisk. Exon sequences are in capital letters and introns in lowercase. Dots indicate nonsequenced regions.



**FIG. 6.** Hybridization of probes to *Eco*RI digests of human DNA and human/rodent hybrids containing various portions of human chromosome 19. Molecular weight markers are shown on the left and hybridization signals of interest on the right. (A) A 2.6-kb genomic fragment from the 5'-region of the CEA gene was used as a probe. (B) A 1.3-kb *Eco*RI fragment covering the coding region of PSG1a was used as a probe; the order of the hybrid DNA digests is different from that in A. (C) A 0.3-kb *Bgl*I/*Eco*RI fragment from the 3'-untranslated region of CGM6 was used to probe a subset of the hybrid DNA digests used in A.

#### Determination of the Directions of Transcription for CEA-Related Genes

The orientation of the CGM2, CEA, and NCA genes has been determined by restriction mapping (results to be published elsewhere). Cosmid clones containing the N-terminal domain exons of CGM6 (clones F10514, F11586, and F21110) and CGM8 (F8998, F5027, and F14687) are seen to be directly adjacent in Fig. 4 (clone F6516 was not tested), and due to their high overlap, it appears probable but has not yet been confirmed that these two genes are orientated head-to-head. As two of the cosmid clones (clones F17002 and F20310) that hybridize with the N-terminal domain CGM9-specific probe also hybridize with the CGM6-specific 3'-untranslated region probe, we assume the orientation of CGM9 and the other genes to be as shown in Fig. 7. Although the CGM1 gene probe was not used for hybridization in the PFGE studies, because there are no rare-cutting restriction enzyme sites in this region, its position relative to those of the other genes, as determined by contig analyses, has also been included in Fig. 7.

#### Chromosomal Location of CEA Gene Family by Somatic Cell Hybrid Analysis

Under the conditions used, the 2.6-kb *Xba*I fragment probe from the 5'-region of the CEA gene hybridizes to both CEA and PSG family members (Fig. 6A). Although this probe does not contain the N-domain exon, the hybridization patterns were compared and the fragments are similar in size to those obtained using a probe from that region (Thompson *et al.*, 1989). One of the strongest

hybridizing *Eco*RI restriction endonuclease fragments is of 7.2 kb and contains the 5'-region of the CEA gene itself (cf. Schrewe *et al.*, 1990; Willcocks *et al.*, 1990). An *Eco*RI fragment of 5.2 kb also hybridizes strongly because this is common to a number of PSG genes (Thompson *et al.*, 1989). Other CEA family members have N-domain-containing *Eco*RI fragment sizes as follows: NCA, 2.7 kb (Thompson *et al.*, 1987); CGM2, 9.8 kb (Thompson *et al.*, 1989); BGP, >20 kb (Barnett *et al.*, 1989); and CGM6, 12 kb (unpublished results). J64034638 contains a subset of the bands hybridizing to the *Xba*I probe, while hybrid G175C4 apparently contains the remaining bands. The CGM2 and CEA genes (NCA and CGM1 signals are too faint to be seen in Fig. 6A) are contained in J64034638, while the PSGs, BGP, and CGM6 bands are absent and apparently cosegregate with the G175C4 hybrid.

The PSG probe also recognizes a number of *Eco*RI fragments (Fig. 6B) and has been shown elsewhere to be specific for PSG subgroup but not CEA subgroup members under these hybridization conditions (Thompson *et al.*, 1990). A comparison of Figs. 6A and 6B shows that those clones that contained all the CEA subgroup bands also contain all the PSG subgroup bands (908K1B18, TVBID, G175B3, and WILF M2), and those that were negative with the CEA probe were also negative with the PSG probe (WILF M6, TV7C, and 20XP3542-1-4). The J64034638 hybrid, on the other hand, contains no PSG bands, whereas hybrid G175C4 contains them all (Fig. 6B).

The CGM6 probe was hybridized to six of the hybrids used in the previous mapping experiments (Fig. 6C). Under the conditions used, a number of *Eco*RI frag-

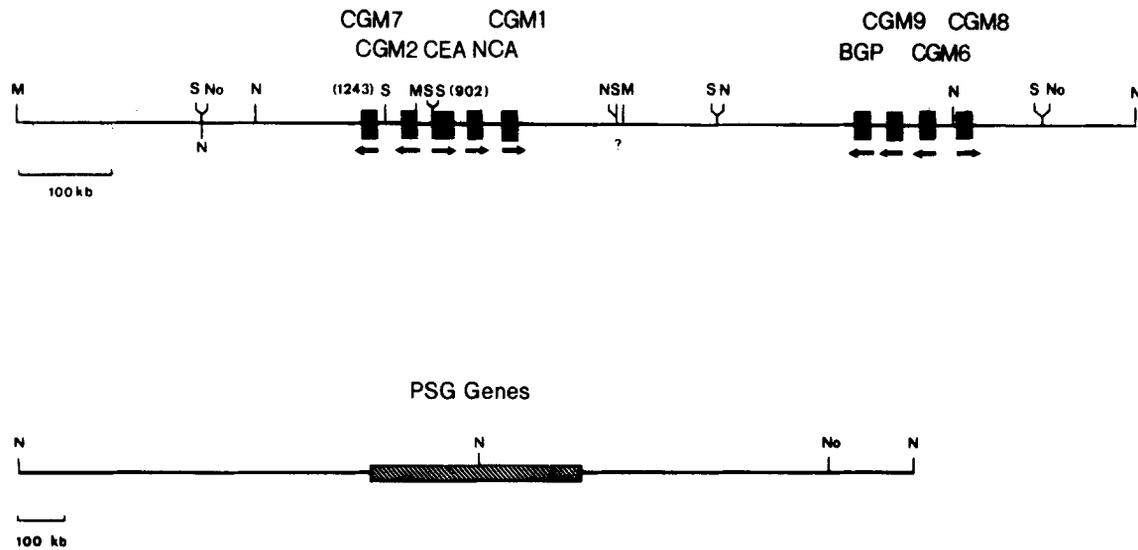


FIG. 7. Long-range restriction map of the CEA gene family. The genes are shown as blocks and named above. The direction of transcription is indicated by arrows. Although linkage of the CGM7/CGM2/CEA/NCA/CGM1 cluster with the BGP/CGM6/CGM8 cluster could not be confirmed, the probable linkage indicated by a question mark is indeed supported by hybridization data in the accompanying paper (Brandriff *et al.*, 1992). For the PSG cluster, the gene order has not been differentiated. Abbreviations for the restriction enzyme sites are shown in Figs. 1–3.

ments hybridized, with the strongest signal at 5.1 kb, which was shown elsewhere to represent the CGM6 gene (Berling *et al.*, 1990). Hybrids WILF M6, 20XP3542-1-4, and J64034638 are negative, whereas hybrids WILF M2, 908K1B18, and G175C4 are positive. An unknown CEA cross-hybridizing band of 3.75 kb is also seen in hybrids 908K1B18, J64034638, and WILF M2. These results confirm the assumption made with the CEA probe that the BGP/CGM6 cluster found using PFGE analyses cosegregates with the PSG gene cluster in hybrid G175C4.

#### DISCUSSION

With the help of pulse-field gel electrophoresis and hybridization using gene-specific probes, we have constructed a long-range physical map of the carcinoembryonic antigen gene family cluster on the long arm of chromosome 19. These results have been confirmed and supplemented by assembling an ordered set of cosmids for this region and DNA sequencing, as well as by hybridization analysis of a number of human/rodent somatic hybrid cells. The data summarized in Fig. 7 indicate that the CEA gene family is contained within a region located at position 19q13.1–q13.2 between the markers for the cytochrome P450 subfamily IIA (CYP2A) and D19S15/D19S8 (Table 1). The PSG subgroup genes were found to reside telomeric of the CEA subgroup genes, which has been confirmed by *in situ* hybridization of metaphase and interphase chromosomes (see Brandriff *et al.*, 1992, accompanying paper). The order of all known members of the CEA subgroup, including a newly described pseudogene (CGM8), has been determined. This order and the relative distances are further supported by the data obtained from the mapping of cosmids from this region by fluorescence *in situ* hybridization to sperm pronuclear and somatic interphase chromosomes (Brandriff *et al.*, 1992, accompanying paper). Differential hy-

bridization using 5' and 3' gene probes together with fine restriction mapping showed the orientation of individual genes. The gene order and their directions of transcription are as follows: cen/3'-CGM7-5'/3'-CGM2-5'/5'-CEA-3'/5'-NCA-3'/5'-CGM1-3'/3'-BGP-5'/3'-CGM9-5'/3'-CGM6-5'/5'-CGM8-3'/PSG cluster/qter.

One general problem in constructing a long-range restriction map is that many rare-cutting restriction endonucleases recognize sequences containing the dinucleotide CpG. This dinucleotide is normally underrepresented in the genome due to methylation at the 5'-position of the cytosine rest followed by nucleotide substitution during evolution (Bird, 1987). However, CpG islands that contain unmethylated CpG dinucleotides exist. These CpG islands generally also contain recognition sites for a number of rare-cutting enzymes (Bird, 1987), which creates mapping problems, as DNA fragment overlaps cannot easily be found in double digests. In mapping the CEA gene family, such problems were encountered, although a map could be constructed. In Fig. 7, CpG islands possibly exist where recognition sites of rare-cutting restriction endonucleases are clustered. It has been reported that CpG islands are often found in the upstream regulatory regions of eukaryotic genes (Bird, 1987). Indeed, in the case of the CEA gene, whose exon structure has been analyzed elsewhere in detail (Schrewe *et al.*, 1990; Willcocks *et al.*, 1990), such a region is found within the CEA promoter (Fig. 7). The existence of additional putative CpG islands within the CEA gene family cluster indicates that other, as yet unidentified genes may also exist close to these sites.

Gene clusters have been found for a number of gene families, e.g., the  $\beta$ -globin family (Efstratiadis *et al.*, 1980; Slightom *et al.*, 1980), the major histocompatibility complex gene family (Steinmetz and Hood, 1983), immunoglobulins (Tonegawa, 1983), and homeotic genes (Holland and Hogan, 1988). Genes within a cluster can

**TABLE 1**  
**Localization of the CEA Gene Family in the Region 19q13.1-q13.2**

Chromosome 19 marker: Hybrid cell line	p		q13.1-q13.2										q13.3					
	C3	D19S11	D19S7	D19S9	GPI	MSK37	TGF $\beta$	CYP2A	CEA	CGM6	PSG	D19S15	D19S8	ATP1A3	APOC2	PVS	CKMM	CGB
G175B3	-	+	+	+	nt	+	+	+	+	nt	+	+	+	+	+	nt	+	-
TV7C	+	+	+	+	-	-	-	-	nt	-	-	-	-	-	-	-	nt	nt
TVBID	-	+	+	+	+	+	+	+	nt	+	+	+	+	+	+	+	+	nt
WILF M6	-	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	nt
WILF M2	+	-	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	-
908K1B18	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-
J64034638	nt	nt	+	+	+	+	+	+	+	-	nt	-	-	-	nt	-	-	nt
G175C4	-	+	+	+	nt	-	-	-	+	+	nt	+	+	+	-	-	-	-
20XP3542-1-4	-	+	+	+	-	-	-	-	-	-	-	-	-	-	+	+	+	-

Note: Human chromosome 19 markers present in the hybrid cell lines used to position the CEA/CGM6/PSG gene clusters on chromosome 19. The markers are ordered according to our most recent analyses and the HGM10.5 chromosome 19 report (Ropers and Pericak-Vance, 1990). The gene families are positioned on the basis of segregation patterns identical to those of flanking markers: the CEA gene cluster is like that of the CYP2A marker; the CGM6 and PSG gene clusters are like those of D19S15/D19S8.

be more easily regulated with regard to the coordination of their expression. For example, the order of the  $\beta$ -globin genes within that cluster represents the order with which they are expressed during ontogenesis (Fritsch *et al.*, 1980). Similarly, the positional expression pattern of Hox genes along the anterior/posterior axis of the mouse has been shown to correlate with the gene order within that cluster (Holland and Hogan, 1988). It has been well documented that CEA-related molecules are expressed at an early stage during fetal development of humans (von Kleist *et al.*, 1986; Nap *et al.*, 1988) and mice (Huang *et al.*, 1990). It would be of interest to see whether variability in the temporal or spatial expression patterns of CEA-related genes exists during development and if this correlates with their genomic organization. Indeed, we have recently found temporal variability in the transcriptional activity of two mouse PSG analogues (results to be published elsewhere). Common regulatory elements that control the expression of more than one gene within a cluster could also exist. Such a situation may be expected for the PSG genes, because most of them are coordinately expressed in the placenta (Thompson *et al.*, 1990).

These studies have provided a basis for future investigations regarding the regulation of transcription of individual or groups of CEA-related genes and the search for *cis*- and *trans*-acting regulatory factors.

#### ACKNOWLEDGMENTS

We thank Dr. Y. Matsuoka and Dr. S. Hammarström for personal communications, Dr. R. Paxton and Dr. T. Barnett for the BGP-specific cDNA probes, and Dr. B. Trask for the *in situ* hybridization data. The assistance of the members of the Human Genome Center staff is gratefully acknowledged. The work at Freiburg was supported by the Deutsche Forschungsgemeinschaft. At Lawrence Livermore National Laboratory, the work was performed under the auspices of the U.S. Department of Energy, Office of Health and Environmental Research under Contract W-7405-ENG-48. At Oxford the work was supported by the Imperial Cancer Research Fund, London.

#### REFERENCES

- Arakawa, F., Kuroki, M., Misumi, Y., Oikawa, S., Nakazato, H., and Matsuoka, Y. (1990). Characterization of a cDNA clone encoding a new species of the nonspecific crossreacting antigen (NCA), a member of the CEA gene family. *Biochem. Biophys. Res. Commun.* **166**: 1063-1071.
- Barnett, T., and Zimmermann, W. (1990). Proposed nomenclature for the carcinoembryonic antigen (CEA) gene family. *Tumor Biol.* **11**: 59-63.
- Barnett, T. R., Kretschmer, A., Austen, D. A., Goebel, S. J., Hart, J. T., Elting, J. J., and Kamarck, M. E. (1989). Carcinoembryonic antigens: Alternative splicing accounts for the multiple mRNAs that code for novel members of the carcinoembryonic antigen family. *J. Cell Biol.* **108**: 267-276.
- Barnett, T. R., Pickle, W., Rae, P. M. M., Hart, J., Kamarck, M., and Elting, J. (1989). Human pregnancy-specific  $\beta$ 1-glycoproteins are coded within chromosome 19. *Am. J. Hum. Genet.* **44**: 890-893.
- Benchimol, S., Fuks, A., Jothy, S., Beauchemin, N., Shiota, K., and Stanners, C. P. (1989). Carcinoembryonic antigen, a human tumor marker, functions as an intercellular adhesion molecule. *Cell* **57**: 327-334.

- Berling, B., Kolbinger, F., Grunert, F., Thompson, J. A., Brombacher, F., Buchegger, F., von Kleist, S., and Zimmermann, W. (1990). Molecular cloning of a carcinoembryonic antigen (CEA) gene family member expressed in leukocytes of chronic myeloid leukaemia patients and bone marrow. *Cancer Res.* **50**: 6534–6539.
- Bird, A. P. (1987). CpG islands as gene markers in the vertebrate nucleus. *Trends Genet.* **3**: 342–347.
- Bischoff-Delaloye, A., Delaloye, B., Buchegger, F., Gilgien, W., Studer, A., Curchod, S., Givel, J.-C., Mosimann, F., Pettavel, J., and Mach, J.-P. (1989). Clinical value of immunoscintigraphy in colorectal carcinoma patients: A prospective study. *J. Nuclear Med.* **30**: 1646–1656.
- Bohn, H. (1971). Nachweis und Charakterisierung von Schwangerschafts-proteinen in der menschlichen Placenta, sowie ihre quantitative Bestimmung im Serum schwangerer Frauen. *Arch. Gynäkol.* **210**: 440–457.
- Bordes, M., Knobel, S., and Martin, F. (1975). Carcinoembryonic antigen (CEA) and related antigens in blood cells and haematopoietic tissues. *Eur. J. Cancer* **11**: 783–786.
- Brandriff, B. F., Gordon, L. A., Tynan, K. T., Olsen, A., Mohrenweiser, H. W., Fertitta, A., Carrano, A. V., and Trask, B. J. (1992). Order and genomic distances among members of the carcinoembryonic antigen (CEA) gene family determined by fluorescence *in situ* hybridization. *Genomics* **12**: 773–779.
- Branscomb, E., Slezak, T., Pae, R., Galas, D., Carrano, A. V., and Waterman, M. (1990). Optimizing restriction fragment fingerprinting methods for ordering large genomic libraries. *Genomics* **8**: 351–366.
- Brook, J. D., Beresford, H. R., Shaw, D. J., Old, L. J., and Rettig, W. J. (1987a). The localisation on human chromosome 19 of three genes for cell surface antigens defined by monoclonal antibodies. *Cytogenet. Cell Genet.* **45**: 156–162.
- Brook, J. D., Knight, S. J. L., Roberts, S. H., Harley, H. G., Walsh, K. V., Rundle, S. A., Freyne, K., Koch, M. C., Epstein, N. D., Wieringa, B., Schonk, D., Smeets, H., Haddingham, K., Siciliano, M. J., Palmer, D. K., Miles, J. S., Wolf, C. R., Fonatsch, C., and Shaw, D. J. (1991). The physical map of chromosome arm 19q: Some new assignments, confirmations and reassessments. *Hum. Genet.* **87**: 65–72.
- Brook, J. D., Skinner, M., Roberts, S. H., Rettig, W. J., Almond, J. W., and Shaw, D. J. (1987b). Further mapping of markers around the centromere of human chromosome 19. *Genomics* **1**: 320–328.
- Carrano, A. V., Branscomb, E., de Jong, P. J., Mohrenweiser, H., Olsen, A., and Slezak, T. (1990). The interaction of high-resolution electrophoresis and computational analysis in genome mapping. In "Electrophoresis, Supercomputing and the Human Genome, Proceedings of the 1st Int. Conf. Tallahassee, FL" (C. R. Cantor and H. A. Lim, Eds.), World Scientific, Cleveland.
- Carrano, A. V., Lamberdin, J., Ashworth, L. K., Watkins, B., Branscomb, E., Slezak, T., Raff, M., de Jong, P. J., Keith, D., McBride, L., Meister, S., and Kronick, M. (1989). A high resolution, fluorescence-based, semi-automated method for DNA fingerprinting. *Genomics* **4**: 129–136.
- Chomczynski, P., and Qasba, P. K. (1984). Alkaline transfer of DNA to GeneScreenPlus membrane. *Biochem. Biophys. Res. Commun.* **122**: 340–344.
- Church, G. M., and Gilbert, W. (1984). Genomic sequencing. *Proc. Natl. Acad. Sci. USA* **88**: 1991–1995.
- de Jong, P. J., Yokabata, K., Chen, C., Lohman, F., Pederson, L., McNinch, J., and Van Dilla, M. (1989). Human chromosome-specific partial digest libraries in lambda and cosmid vectors. *Cytogenet. Cell Genet.* **51**: 985.
- Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., De Riel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. E., Blechl, A. E., Smithies, O., Baralle, F. E., Sholders, C. C., and Proudfoot, N. J. (1980). The evolution of the human  $\beta$ -globin gene family. *Cell* **21**: 653–668.
- Feinberg, A. P., and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**: 6–13.
- Fritsch, E. F., Lawn, R. M., and Maniatis, T. (1980). Molecular cloning and characterization of the human  $\beta$ -globin gene cluster. *Cell* **19**: 959–972.
- Gold, P., and Freedman, S. O. (1965a). Demonstration of tumor-specific antigens in human colonic carcinomata by immunological tolerance and absorption techniques. *J. Exp. Med.* **121**: 439–462.
- Gold, P., and Freedman, S. O. (1965b). Specific carcinoembryonic antigens of the human digestive system. *J. Exp. Med.* **122**: 467–481.
- Hinoda, Y., Neumaier, M., Hefta, S. A., Drzeniek, Z., Wagener, C., Shively, L., Hefta, L. J. F., Shively, J. E., and Paxton, R. J. (1988). Molecular cloning of a cDNA coding biliary glycoprotein I: Primary structure of a glycoprotein immunologically cross-reactive with carcinoembryonic antigen. *Proc. Natl. Acad. Sci. USA* **85**: 6959–6963.
- Hinoda, Y., Neumaier, M., Hefta, S. A., Drzeniek, Z., Wagener, C., Shively, L., Hefta, L. J. F., Shively, J. E., and Paxton, R. J. (1989). Correction to: Molecular cloning of a cDNA coding biliary glycoprotein I. Primary structure of a glycoprotein immunologically cross-reactive with carcinoembryonic antigen. *Proc. Natl. Acad. Sci. USA* **86**: 1668.
- Holland, P. W. H., and Hogan, B. L. M. (1988). Expression of homeobox genes during mouse development: A review. *Genes Dev.* **2**: 773–782.
- Huang, J. Q., Turbide, C., Daniels, E., Jothy, S., and Beauchemin, N. (1990). Spatio-temporal expression of murine carcinoembryonic antigen (CEA) gene family members during mouse embryogenesis. *Development* **10**: 573–588.
- Kamarck, M. E., Elting, J. J., Hart, J. T., Goebel, S. J., Rae, P. M. M., Nothdurft, M. A., Nedwin, J. J., and Barnett, T. R. (1987). Carcinoembryonic antigen family: Expression in a mouse L-cell transfectant and characterization of a partial cDNA in bacteriophage  $\lambda$ gt11. *Proc. Natl. Acad. Sci. USA* **84**: 5350–5354.
- Kodolja, V., Lucas, K., Barnert, S., von Kleist, S., Thompson, J. A., and Zimmermann, W. (1989). Identification of a carcinoembryonic antigen gene family in the rat: Analysis of the N-terminal domains reveals immunoglobulin-like, hypervariable regions. *J. Biol. Chem.* **264**: 6906–6912.
- Kuroki, M., Arakawa, F., Matsuo, Y., Oikawa, S., Misumi, Y., Nakazato, H., and Matsuoka, Y. (1991). Molecular cloning of nonspecific cross-reacting antigens (NCAs) in human granulocytes: Sequences of three new members and distribution of mRNAs for CEA family members. *J. Biol. Chem.* **266**: 11810–11817.
- Leusch, H.-G., Hefta, S. A., Drzeniek, Z., Hummel, K., Markos-Pusztai, Z., and Wagener, C. (1990). *Escherichia coli* of human origin binds to carcinoembryonic antigen (CEA) and non-specific cross-reacting antigen (NCA). *FEBS Lett.* **261**: 405–409.
- Lin, T.-M., Halpert, S. P., and Spellacy, W. N. (1974). Measurement of pregnancy-associated plasma proteins during human gestation. *J. Clin. Invest.* **54**: 576–582.
- Mach, J.-P., and Pusztaszeri, G. (1972). Carcinoembryonic antigen (CEA): Demonstration of a partial identity between CEA and a normal glycoprotein. *Immunochimistry* **9**: 1031–1034.
- Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). "Molecular Cloning: A Laboratory Manual," Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Nap, M., Mollgard, K., Burtin, P., and Fleuren, G. J. (1988). Immunohistochemistry of carcinoembryonic antigen in the embryo, fetus and adult. *Tumor Biol.* **9**: 145–153.
- Niemann, S. C., Schonk, D., van Dijk, P., Wieringa, B., Grzeschik, K. H., and Bartels, I. (1989). Regional localization of the gene encoding pregnancy specific beta-1-glycoprotein 1 (PSBG1) to chromosome 19q13.1. *Cytogenet. Cell Genet.* **51**: 1053.
- Oikawa, S., Inuzuka, C., Kuroki, M., Matsuoka, Y., Kosaki, G., and Nakazato, H. (1989). Cell adhesion activity of non-specific cross-reacting antigen (NCA) and carcinoembryonic antigen (CEA) expression on CHO cell surface: Homophilic and heterophilic adhesion. *Biochem. Biophys. Res. Commun.* **164**: 39–45.

- Paxton, R., Mooser, G., Pande, H., Lee, T. D., and Shively, J. E. (1987). Sequence analysis of carcinoembryonic antigen: Identification of glycosylation sites and homology with the immunoglobulin supergene family. *Proc. Natl. Acad. Sci. USA* **84**: 920-924.
- Pignatelli, M., Durbin, H., and Bodmer, W. F. (1990). Carcinoembryonic antigen functions as an accessory adhesion molecule mediating colon epithelial cell-collagen interactions. *Proc. Natl. Acad. Sci. USA* **87**: 1541-1545.
- Ropers, H. H., and Pericak-Vance, M. A. (1990). Report of the committee on the genetic constitution of chromosome 19. *Cytogenet. Cell Genet.* **55**: 218-228.
- Sambrook, J. E., Fritsch, J., and Maniatis, T. (1989). "Molecular Cloning: A Laboratory Manual," 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**: 5463-5467.
- Schonk, D., Coerwinkel-Driessen, M., van Dalen, I., Oerlemans, F., Smeets, B., Schepens, J., Hulsebos, T., Cockburn, D., Boyd, Y., Davis, M., Rettig, W., Shaw, D., Roses, A., Ropers, H., and Wieringa, B. (1989). Definition of subchromosomal intervals around the myotonic dystrophy gene region at 19q. *Genomics* **4**: 384-396.
- Schrewe, H., Thompson, J., Bona, M., Hefta, L. J. F., Maruya, A., Hassauer, M., Shively, J. E., von Kleist, S., and Zimmermann, W. (1990). Cloning of the complete gene for the carcinoembryonic antigen: Analysis of its promoter indicates a region conveying cell type-specific expression. *Mol. Cell. Biol.* **10**: 2738-2748.
- Schwartz, D. C., and Cantor, C. R. (1984). Separation of yeast chromosome-sized DNAs by pulsed field gel electrophoresis. *Cell* **37**: 67-75.
- Sharkey, R. M., Weadock, K. S., Natale, A., Haywood, L., Aninipot, R., Blumenthal, R. D., and Goldenberg, D. M. (1991). Successful radioimmunotherapy for lung metastasis of human colonic cancer in nude mice. *J. Natl. Cancer Inst.* **83**: 627-632.
- Shively, J. E., and Beatty, J. D. (1985). CEA-related antigens: Molecular biology and clinical significance. *CRC Crit. Rev. Oncol. Hematol.* **2**: 355-399.
- Slightom, J. L., Blechl, A. E., and Smithies, O. (1980). Human fetal A $\gamma$ - and G $\gamma$ -globin genes: Complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell* **21**: 627-638.
- Stallings, R. L., Olson, E., Strauss, A. W., Thompson, L. H., Bachinski, L. L., and Siciliano, M. J. (1988). Human creatine kinase genes on chromosomes 15 and 19, and proximity of the gene for the muscle form to the genes for apolipoprotein C2 and excision repair. *Am. J. Hum. Genet.* **43**: 144-151.
- Steinmetz, M., and Hood, L. (1983). Genes of the major histocompatibility complex in mouse and man. *Science* **222**: 727-733.
- Streydio, C., Lacka, K., Swillens, S., and Vassart, G. (1990). Structure, evolution and chromosomal localization of the human pregnancy-specific  $\beta$ 1 glycoprotein gene family. *Genomics* **6**: 579-592.
- Svenberg, T. (1976). Carcinoembryonic antigen-like substances of human bile: Isolation and partial characterization. *Int. J. Cancer* **17**: 588-596.
- Tatarinov, Y. S. (1978). Trophoblast-specific beta1-glycoprotein as a marker for pregnancy and malignancies. *Gynecol. Obstet. Invest.* **9**: 65-97.
- Tatarinov, Y. S., and Masyukevich, V. N. (1970). Immunological identification of a new  $\beta$ 1-globulin in the blood serum of pregnant women. *Byull. Eksp. Biol. Med.* **69**: 66-68.
- Thompson, J., Barnert, S., Berling, B., von Kleist, S., Kodolja, V., Lucas, K., Mauch, E.-M., Rudert, F., Schrewe, H., Weiss, M., and Zimmermann, W. (1989). Structure, expression and evolution of the human and rat carcinoembryonic antigen (CEA) gene families. In "The Carcinoembryonic Antigen Gene Family" (A. Yachi and J. Shively, Eds.), pp. 65-74, Elsevier, Amsterdam.
- Thompson, J., Grunert, F., and Zimmermann, W. (1991). Carcinoembryonic antigen gene family: Molecular biology and clinical perspectives. *J. Clin. Lab. Anal.* **5**: 344-366.
- 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). The human pregnancy-specific glycoprotein genes are tightly linked on the long arm of chromosome 19 and are coordinately expressed. *Biochem. Biophys. Res. Commun.* **167**: 848-859.
- Thompson, J. A., Mauch, E.-M., Chen, F.-S., Hinoda, Y., Schrewe, H., Berling, B., von Kleist, S., Shively, J. E., and Zimmermann, W. (1989). Analysis of the size of the carcinoembryonic antigen (CEA) gene family: Isolation and sequencing of N-terminal domain exons. *Biochem. Biophys. Res. Commun.* **158**: 996-1004.
- Thompson, J., and Zimmermann, W. (1988). The carcinoembryonic antigen gene family: Structure, expression and evolution. *Tumor Biol.* **9**: 63-83.
- Tonegawa, S. (1983). Somatic generation of antibody diversity. *Nature* **302**: 575-581.
- von Kleist, S., Chavanel, G., and Burtin, P. (1972). Identification of an antigen from normal human tissue that cross-reacts with the carcinoembryonic antigen. *Proc. Natl. Acad. Sci. USA* **69**: 2492-2494.
- von Kleist, S., Winkler, J., Migule, I., and Böhm, N. (1986). Carcinoembryonic antigen (CEA) expression in early embryogenesis: A study of the first trimester of gestation. *Anticancer Res.* **6**: 1265-1272.
- Willcocks, T. C., and Craig, I. W. (1990). Characterization of the genomic organization of human carcinoembryonic antigen (CEA): Comparison with other family members and sequence analysis of the 5' controlling region. *Genomics* **8**: 492-500.
- Willcocks, T. C., Craig, S. P., Coates, D., and Craig, I. W. (1987). Coding sequence for carcinoembryonic antigen (CEA) assigned to human chromosome 19q13. *Cytogenet. Cell Genet.* **46**: 716.
- Willcocks, T. C., Craig, S. P., and Craig, I. W. (1989). Assignment of the coding sequence of carcinoembryonic antigen (CEA) and normal cross-reacting antigen (NCA) to human chromosome 19q13. *Ann. Hum. Genet.* **53**: 141-148.
- Zimmermann, W., Ortlieb, B., Friedrich, R., and von Kleist, S. (1987). Isolation and characterization of cDNA clones encoding the human carcinoembryonic antigen reveal a highly conserved repeating structure. *Proc. Natl. Acad. Sci. USA* **84**: 2960-2964.
- Zimmermann, W., Weber, B., Ortlieb, B., Rudert, F., Schempp, W., Fiebig, H.-H., Shively, J. E., von Kleist, S., and Thompson, J. A. (1988). Chromosomal localization of the carcinoembryonic antigen gene family and differential expression in various tumors. *Cancer Res.* **48**: 2550-2554.
- Zimmermann, W., Weiss, M., and Thompson, J. A. (1989). cDNA cloning demonstrates the expression of pregnancy-specific glycoprotein genes, a subgroup of the carcinoembryonic antigen gene family, in fetal liver. *Biochem. Biophys. Res. Commun.* **163**: 1197-1209.