Assembly and analysis of cosmid contigs in the CEA-gene family region of human chromosome 19

Katherine Tynan[+], Anne Olsen, Barbara Trask, Pieter de Jong, John Thompson[1], Wolfgang Zimmermann[1], Anthony Carrano and Harvey Mohrenweiser[*]
Human Genome Center, Biomedical Sciences Division, L-452, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA and [1]Institute of Immunobiology, University of Freiburg, Stefan-Meier-Strasse 8, D-7800 Freiburg, Germany

ABSTRACT

The carcinoembryonic antigen (CEA)-like genes are members of a large gene family which is part of the immunoglobulin superfamily. The CEA family is divided into two major subgroups, the CEA-subgroup and the pregnancy-specific glycoprotein (PSG)-subgroup. In the course of an effort to develop a set of overlapping cosmids spanning human chromosome 19, we identified 245 cosmids in a human chromosome 19 cosmid library (6 – 7X redundant) by hybridization with an IgC-like domain fragment of the CEA gene. A fluorescence-based restriction enzyme digest fingerprinting strategy was used to assemble 212 probe-positive cosmids, along with 115 additional cosmids from a collection of ~ 8,000 randomly selected cosmids, into five contigs. Two of the contigs contain CEA-subgroup genes while the remaining three contigs contain PSG-subgroup genes. These five contigs range in size from 100 kb to over 300 kb and span an estimated 1 Mb. The CEA-like gene family was determined by fluorescence in situ hybridization to map in the q13.1-q13.2 region of human chromosome 19. Analysis of the two CEA-subgroup contigs provided verification of the contig assembly strategy and insight into the organization of 9 CEA-subgroup genes.

INTRODUCTION

The carcinoembryonic antigen (CEA)-gene family is a large gene family mapping to chromosome 19 [1]. Two major subgroups of the CEA-gene family have been identified. The CEA-subgroup currently consists of nine genes, including CEA, NCA (non-specific cross-reacting antigen), BGP (biliary glycoprotein), and six genes referred to as CEA-Gene family Members (CGM1, 2, 6, 7, 8 and 9) [2,3,4,]. The pregnancy specific glycoprotein (PSG)-subgroup currently consists of eleven identified genes [5]. However, the actual number of genes in the PSG-subgroup is

difficult to ascertain, due to the sequence similarities among the members of the family [6,7], and additional genes may exist.

The CEA family of glycoproteins have a domain organization that has homology to the immunoglobulin superfamily [8, 9]. The proteins consist of a leader peptide, an NH_2-terminal or IgV-like domain, a varying number of IgC-like domains (six copies are present in CEA, two in NCA, and two to three in the PSGs), and a short COOH-terminal domain. The proteins of the CEA-subgroup appear to be membrane-bound while most, if not all of the PSG proteins appear to be secreted [reviewed in 10]. Further heterogeneity among the proteins of some gene family members results from alternative splicing of transcripts and post-translational modifications (glycosylation) of the expressed proteins [reviewed in 10].

CEA is a glycoprotein expressed during tumorigenesis and has importance as a tumor marker in assessing the recurrence of colorectal, breast and lung cancers. Additional members of the CEA family of glycoproteins include molecules that are differentially expressed during fetal development or abundant in some cancers. Tissue distribution and specific gene expression patterns have been documented for a number of these glycoproteins [11,12,13]. CEA and NCA appear to function as homotypic intercellular adhesion molecules [14,15]. The PSG proteins are abundantly expressed during pregnancy and antisera against them induced abortion in pregnant monkeys [16]; however, their exact function(s) remains unknown.

Significant efforts are currently focussed on the construction of physical maps for a number of regions of the human genome. We report here on the verification of a strategy to develop an overlapping cosmid contig map in the CEA gene family region of human chromosome 19. We identified 245 cosmids hybridizing with the IgC-like domain probe from the CEA cDNA. Using a semi-automated fluorescence restriction enzyme digest fingerprinting strategy, and a statistically based assembly algorithm, these cosmids were assembled into 5 contigs covering an ~ 1 Mb region on chromosome 19. We present results

confirming the validity of the contig assembly strategy based on restriction enzyme mapping of individual cosmids within two of these contigs. In addition we describe the organization of the 9 CEA subgroup genes in the 19q13.1-q13.2 region.

MATERIALS AND METHODS

Chromosome 19 cosmid library

The cosmid library was generated from flow-sorted human chromosome 19 DNA [17]. A Chinese hamster-human hybrid cell line (UV5HL9−5B) that contains a single chromosome 19 as the only detectable human material [18] was the source of the chromosome. The library was constructed in the cosmid vector Lawrist 5 [17], which is a modification of the LORIST series of vectors originally constructed by Gibson et al., [19]. Lawrist 5 contains a double cos site, two unique *Sfi* I sites and promoters (T7 and SP6) flanking the insert. The library has been propagated in two bacterial hosts, ED8767 and DH5αMCR. The stability of some clones was appreciably improved in the latter host [20].

Individual clones from the cosmid library were arrayed in microtiter trays to generate a set of master plates. These master plates were then used to inoculate new replicate microtiter arrays and to prepare colony arrays on hybridization membranes (Hybond N−Amersham). A total of 20,000 cosmids have been arrayed (50% in each host). Of these, 50−60% contained human DNA inserts, with an average insert size of 35 kb. An approximate 7×coverage of chromosome 19 was contained in the 12,000 cosmids with human DNA inserts. For initial screening, the contents of all 96 wells of each of the 214 microtiter arrays were pooled, the DNA digested with *Eco*RI, the DNA separated on agarose gels and Southern blotted onto nylon membranes (Gene Screen Plus—DuPont). The first round of screening using these pools identified colony arrays for further analysis. A second round of screening against arrayed cosmids from each of these positive pools revealed the specific probe-positive clone(s).

Preparation of cosmid DNA

Cosmid DNA samples for restriction enzyme digestion, fingerprinting, and *in situ* hybridization were isolated from 4 ml cultures by alkaline lysis and purification on Qiagen tip-20 columns, following the procedure recommended by the manufacturer (Qiagen, Inc. Chatsworth, CA). Undigested DNA samples were analyzed by electrophoresis on agarose gels to determine the integrity of the cosmid DNA. *Eco*RI digests of all DNA samples were analyzed on agarose gels to estimate the insert size and the DNA concentration. Optical density readings at 260 nm were used to determine DNA concentrations on samples used for *in situ* hybridization experiments.

Probes

The CEA cDNA probe (pCEA1) [21] used in these studies was obtained from the American Type Culture Collection. A 534-bp *Pst*I sub-fragment of pCEA1, containing part of repeats A2, B2 and A3 (Fig. 1), was used as a probe for screening the cosmid library. Assignment of CEA-subgroup genes to cosmids previously identified as containing the CEA constant domain was based on hybridization with synthetic oligonucleotides for CEA and NCA (20-mers located in the respective A1 and A domain exons) and CGM9 (20-mer located in the N-terminal domain exon) [4]. Specific gene probes were as follows: CGM2—781bp *Pst*I fragment from the N-terminal domain [22], CGM6—297bp

*Bgl*I/*Eco*RI fragment from the 3' untranslated region) [23], CGM7 and CGM8—1.9 kb and a 1.3 kb *Bam*HI fragments containing the respective N-terminal domains were subcloned and partially sequenced, CGM1, a 595bp fragment from the 3'UTR from CGM1a, and BGP—245bp *Hind*III/*Bam*HI fragment from the Cyt domain [24] (kindly provided by T.R. Barnett, West Haven, CT). Identification of cosmids containing PSG-subgroup genes was based on hybridization to synthetic oligonucleotides corresponding to segments of the carboxyl terminal domains derived from the sequences determined by Streydio et al. [7].

Probes were labeled with [α-^{32}P] dCTP using the random primer labeling kit from Pharmacia [25]. The oligonucleotides were end-labeled with [γ-^{32}P] ATP using T4 polynucleotide kinase [26].

Filter hybridizations

Conditions for hybridization with genomic probes were 42°C in 40% formamide (v/v), 1 M NaCl, 100 mM Tris, 1 mM EDTA, 0.1% SDS (w/v), 5×Denhart solution [26] and 100 μg/ml of heat-denatured sheared herring sperm DNA for approximately 16 h. Filters were washed to a stringency of 0.1×SSC, 0.1% SDS at 50−60°C (1×SSC: 0.15 M NaCl, 0.015 M sodium citrate, pH 7.0). Hybridization conditions for oligonucleotide probes were 42°C in 0.5 M sodium phosphate buffer (pH7.2), 7% SDS, 1 mM EDTA [27] for approximately 16 h. Filters probed with oligonucleotide probes were washed twice for 10 min each in 2×SSC, 0.2% SDS at room temperature, twice for 20 min each in 6×SSC, 0.2% SDS at 55°C, and twice for 5 min each in 0.1×SSC, 0.1% SDS at room temperature or for CGM9 as described by Thompson et. al. [4]. Washed filters were exposed overnight to Kodak XAR 5 X-ray film at −70°C without intensifying screens.

Contig assembly

Cosmid DNAs were analyzed using the fluorescence-based, semi-automated fingerprinting method previously described by Carrano et al. [28,29,30]. This strategy involves digestion of cosmid DNA with a six cutter restriction enzyme, labeling the fragments with a fluorescent dye and then digesting with a four base recognition restriction enzyme, before separating the fragments under denaturing conditions on polyacrylamide gels. The potential overlap of cosmids was estimated by computing an overlap log likelihood score (L value) for all pairs of cosmids based on shared restriction digest fragment sizes [31]. Based on these pairwise comparisons, contigs were assembled using a computerized contig assembly algorithm.

Analysis of overlapping cosmids

Assessment of contig integrity, contig size, and the location of the CEA conserved sequence motif was carried out by restriction enzyme digestion and Southern blot analysis of selected cosmids from a contig. Restriction enzyme digestions were carried out according to manufacturers' specifications, and resulting DNA fragments were resolved by electrophoresis in 0.7−1.0% agarose gels in TAE buffer [26]. The gels were acid-depurinated for 15 min in 0.25 N HCl [32] and alkali blotted in 0.5 N NaOH and 0.5 M NaCl onto nylon membrane (Gene Screen Plus, DuPont). Following transfer, membranes were neutralized in 0.2 M Tris-HCl (pH 7.0) and 2×SSC.

Hybridization with end-specific probes was also employed to confirm the overlap of cosmids in previously assembled contigs. These end-specific probes were generated from selected cosmids

using the T7 and SP6 promoters on either side of the cloning site in the Lawrist 5 vector. Approximately 1−2 µg of *Eco*RI digested cosmid DNA was used as a template in a 20 µλ reaction consisting of 40 mM Tris-HCl, pH7.9, 6 mM MgCl$_2$, 2 mM spermidine, 5 mM NaCl, 10 mM dithiothreitol, 1 unit/µl RNasin (Promega), 500 µM each ATP, GTP, UTP, 12 µM unlabeled CTP, 30−100µCi [α-^{32}P] CTP (Amersham, 800 Ci/mM), and 10 to 50 units of SP6 or T7 RNA polymerase. Reactions were incubated for 90 min at 37°C, then treated at 37°C with 20 units of RNase-free DNaseI (Pharmacia), and ethanol precipitated from 0.25 M ammonium acetate. Colony blots or Southern blots of cosmid DNA were hybridized with labeled transcripts overnight in 5×SSC, 1% SDS, 0.1 mg/ml denatured herring sperm DNA and then washed to a stringency of 0.1×SSC, 1% SDS at 75°C.

Fluorescence *in situ* hybridization to metaphase chromosomes

Fluorescence *in situ* hybridization (FISH) was performed as described previously [33] to localize single cosmids to metaphase bands. Cosmid DNA was labeled with biotin-dUTP via nick translation and detected, after hybridization, with fluoresceinated avidin. The QFH banding pattern produced by incubation in DAPI and actinomycin [34]

RESULTS

Library screening

Screening of the cosmids from the flow-sorted library by hybridization with a 534-bp probe for the shared conserved IgC-like domain region of the CEA gene (Fig. 1) resulted in the identification of 245 CEA IgC domain probe positive cosmids. The library contains ∼ 12,000 cosmids with human DNA inserts, which represents approximately a 7×coverage of chromosome 19. Assuming that the cloning efficiency for this gene region was similar to the average for the entire chromosome, 245 positive cosmids would be consistent with the existence of ∼30 CEA-like genes on this chromosome, although currently only 20 CEA-like genes have been identified. This small difference could mean that additional CEA-like genes remain to be identified or that the cloning efficiency of this region was higher than the average for the remainder of the chromosome.

Contig assembly

The CEA probe positive cosmids, along with 115 cosmids (from a pool of ∼8,000 additional randomly fingerprinted cosmids,

or ∼5×coverage of the chromosome), were assembled into five CEA gene family contigs (Table 1). Two hundred and twelve of the 245 CEA-positive cosmids were included in these five contigs (boxed cosmids Fig. 2A,3A). Thirty three cosmids identified as CEA probe positive by hybridization are not included as members of these five contigs for several reasons. Approximately 50% of the 33 cosmids failed to generate the minimum of 50 fragments needed to have an adequate fingerprint for analysis. The failure to generate >49 fragments was often the result of a small insert, usually associated with instability of the insert DNA. Infrequently, cosmids did not generate usable fingerprints because of technical difficulties such as incomplete restriction enzyme digestion. These 33 cosmids have not been reanalyzed.

In addition to the 212 CEA IgC domain probe positive cosmids, 115 additional CEA probe negative cosmids were assembled into these five contigs (Table 1). The minimum tiling path (a near minimal number of overlapping cosmids required to cover a portion of the genome) ranges from 6 to 21 cosmids and totaled 64 cosmids for the 5 contigs. The cosmids on the tiling path generally have the least overlap, thereby extending the contig the most. Excess cosmids in a contig, (those not needed to form the minimum tiling set) are placed in the 'stack' above the cosmids/bond with which they have most in common (Figs 2A, 3A). Thus, the average coverage (327 cosmids divided by 64 cosmids on the tiling path) is ∼5, which was consistent with the estimated depth of the library.

Confirmation of contig assembly
Restriction mapping

In order to validate the organization of the contig as assembled by the fingerprinting strategy, subsets of cosmids from the two CEA subgroup contigs were digested with *Eco*RI and analyzed on agarose gels to identify common restriction fragments (Figs. 2B, 3B). Twenty one overlapping cosmids were selected for analysis from contig 76. The resultant *Eco*RI restriction map is shown in Fig. 2B. The cosmids used in this analysis were positioned either within the stacks or on the minimum tiling path in Fig. 2A. Additional resolution of the restriction fragment map of contig 76 was obtained by a partial EcoRI digest strategy, which allowed the ordering of the restriction fragments (Fig. 2B). With only two exceptions, cosmids 26017 and 22584 and cosmids 15709 and 8486, it was possible to confirm the overlap of neighboring cosmids as predicted by the contig assembly algorithm. The overlap of these two pairs of cosmids was confirmed by hybridization experiments as described below.

Eleven individual cosmid members from the CEA subgroup contig 31 were also digested with *Eco*RI and analyzed on agarose gels to detect common restriction fragments (Fig. 3B). Again these results confirmed the computerized contig assembly.
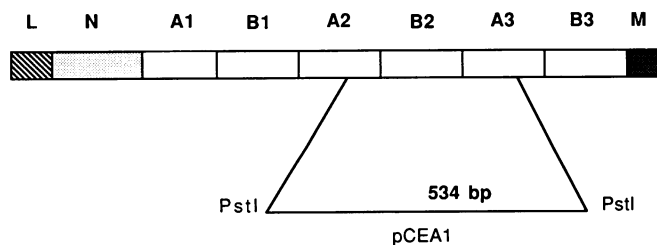

Figure 1. Schematic representation of the domain structure of CEA and the location of the 534-bp PstI fragment used as a probe in these studies. Four major domains are indicated by differential shading and letters above the blocks: a leader sequence (L), an IgV-like N-terminal domain (N), two different types of IgC-like domains (A,B), and a hydrophobic carboxyl-terminal region (M, for membrane-associated).

Table 1. Summary of CEA cosmid contigs

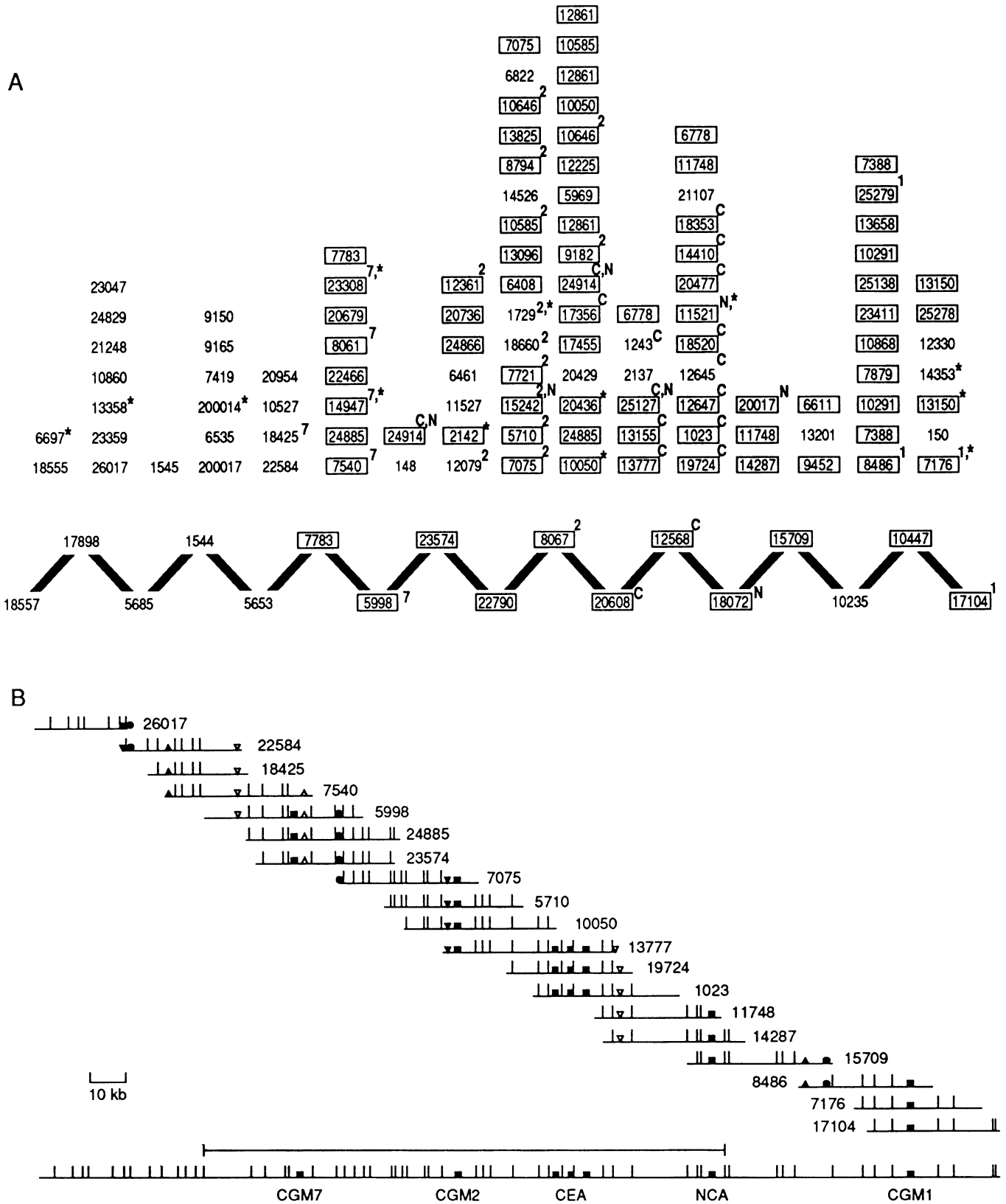| LLNL Contig #	No. of Cosmids in Contig	Cosmids on Tiling Path	CEA Positive Cosmids	Estimated Length (kb)	Sub-Group Assignment
76	118	17	74	280	CEA
31	41	12	30	160	CEA
670	97	21	62	310	PSG
81	34	6	15	100	PSG
234	41	8	36	130	PSG

Figure 2. (A) Display of contig 76. **Boxed** cosmids indicate positive hybridization to the CEA IgC-like domain region probe. **Starred** cosmids have been mapped to cytogenetic bands by fluorescence *in situ* hybridization. Gene specific cosmids are indicated as follows 2-CGM2, C-CEA, N-NCA, 7-CGM7, 1-CGM1. **(B)** EcoRI restriction map of contig 76. The **solid boxes** designate the five gene regions identified by hybridization with the CEA IgC-like domain probe. A partial *EcoRI* digest strategy (Olsen, unpublished data) was used to order the EcoRI fragments in the region indicated by the overbar in the composite map at the bottom. EcoRI sites are indicated by vertical lines to the nearest kb. Some fragments of less than 1 kb may not be included. Specific gene regions are also indicated on the composite map. Positive hybridizations of clones to SP6 endprobes (**filled symbols**) and T7 endprobes (**open symbols**) are indicated.

Hybridization data

Hybridization of cosmid end probes in a number of regions of contig 76 also served to confirm the computerized contig assembly. For example, an SP6 end-probe from 7075 (filled circle symbol) in contig 76 hybridized to a 2.2 kb *Eco* RI fragment in cosmids 23574, 24886 and 5998 (Fig. 2B). End probes from six additional cosmids (26017, 22584, 7540, 13777, 15709 and 8486) in this contig were also used to verify overlaps, and cosmids hybridizing positively to the respective probes are indicated (Fig. 2B). In the early stages of this study, directed walking via cosmid end probes identified anonymous clones which provided links between previously established contigs. Clone 15709 (tiling path contig 76 Fig. 2A) was identified by two independent SP6 end probes from cosmids 11748 and 8486 respectively. Subsequent restriction fingerprinting of this clone, resulted in the assembly of the present contig 76 (Fig. 2A). These probing data also allowed the confirmation of overlap of the two sets of

cosmids, where overlap could not be confirmed from the restriction enzyme digest data.

Contig size estimation

The lengths of the tiling path of the two CEA-subgroup contigs were directly estimated by the addition of restriction fragment lengths. The *Eco*RI restriction maps indicated that contig 76 spans ~280 kb (Fig. 2B) and contig 31 spans ~160 kb (Fig. 3B). In general the first cosmid of the tiling path will contribute 35 kb to the total path length. In contig 76, the addition of each cosmid to the tiling path extended the coverage of the region by 16 kb, while in contig 31 the extension with each new cosmid is 11.5 kb. A 13 kb average extension for each new cosmid added to the tiling path has been measured for a series of contigs in other regions of the chromosome (Carrano, unpublished results). Sizes of the PSG-subgroup contigs listed in Table 1 were calculated assuming a 14 kb extension of the minimum tiling path with the
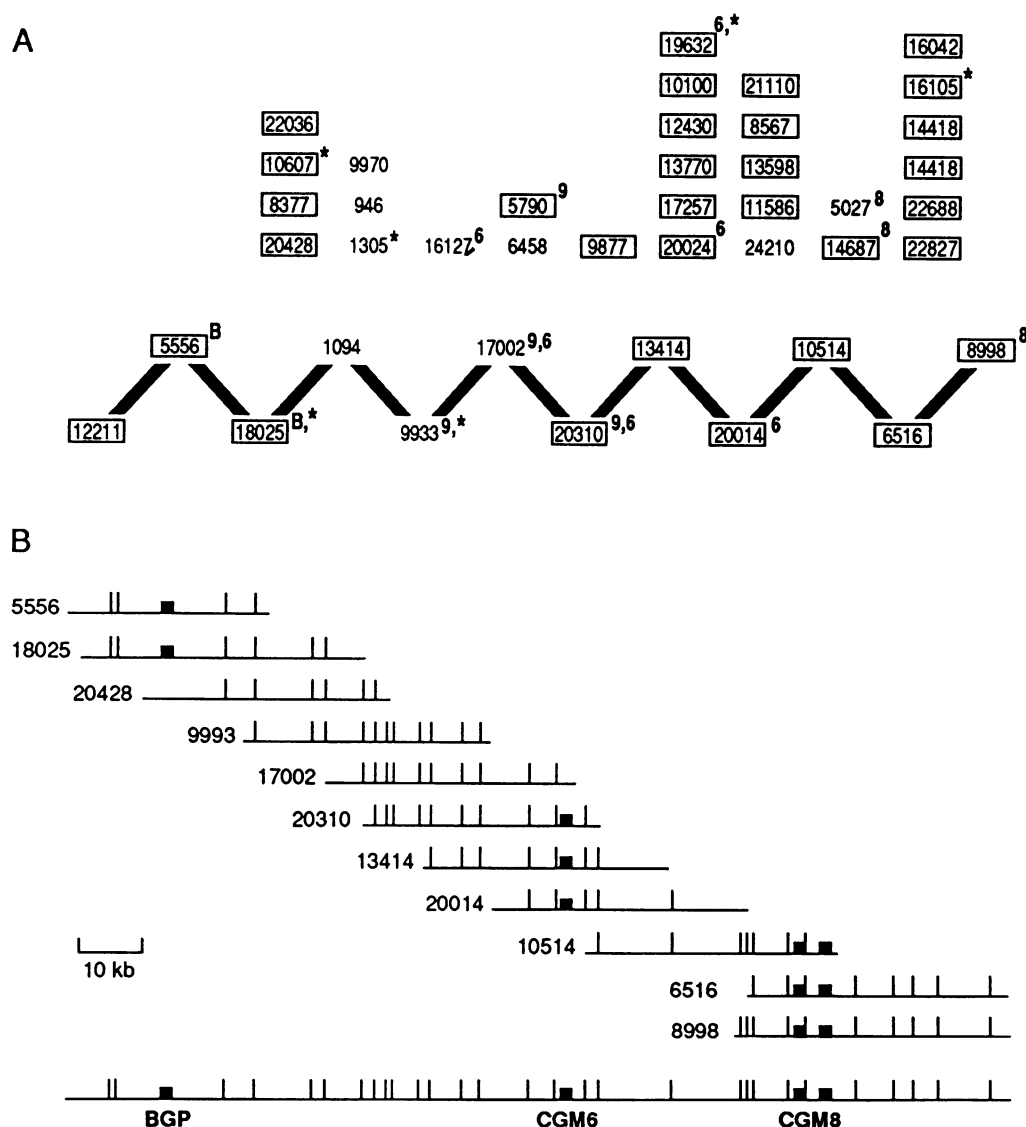

Figure 3. (A) Display of contig 31. **Starred** cosmids have been mapped to cytogenetic bands by fluorescence *in situ* hybridization. **Boxed** cosmids indicate positive hybridization to the CEA IgC-like domain region probe. Gene specific cosmids are indicated as follows B-BGP, 9-CGM9, 6-CGM6, 8-CGM8. (B) EcoRI restriction map of contig 76. The **solid boxes** designate the three gene regions identified by hybridization with the CEA IgC-like domain probe, and the associated gene is indicated.

addition of each new cosmid. The lengths of the tiling paths in the five CEA-positive contigs established by fingerprinting range from 100 kb to over 300 kb. These results suggest that the 5 CEA contigs encompass a minimal region of ~ 1 Mb on human chromosome 19.

Fluorescence *in situ* hybridization

Thirty eight cosmids from the five contigs were mapped by fluorescence in situ hybridization to cytogenetic bands. Twenty six were mapped to 19q13.2, 11 to the border of 19q13.1-q13.2 and one to 19q13.1. One of the 38 cosmids (14353) produced a secondary site at 19p13.1-p13.2. Cosmid 14353 is not positive for the IgC-like probe or the CGM1 probe. It is possible that this sample has DNA from two cosmids or is a cosmid with homology to two segments of chromosome 19. A total of 384 metaphase cells were analyzed to map the 38 cosmids (average of 10.1 cells/cosmid). A total of 1440 hybridization sites were recorded: 1245 (86%) in q13.2, 164 (11%) in q13.1 and 31(2%) in p13.1−13.2. Hybridization efficiency averaged 92% ((1409 sites in q13.1-q13.2)/(384 cells×4 expected hybridiation sites/cell)). The cosmids that have been hybridized from contigs 76 and 31 are starred in Figs. 2A and 3A.

Gene assignment

Using a series of gene- or gene family-specific probes it was possible to identify contigs 76 and 31 as CEA-subgroup contigs while the PSG genes were localized in contigs 81, 234, 670 (Table 1). Southern blots of EcoRI digests were hybridized with the CEA constant domain region probe to identify gene regions. CEA probe positive and presumed gene regions were identified and suggested the presence of at least five tightly linked CEA-family genes in contig 76 (Fig. 2B). The three tightly clustered fragments identified in cosmids 13777, 19724, 1023 probably represent the six IgC-like domains of the CEA gene (Fig. 1). These findings were confirmed by the gene-specific probings and suggested that the left end of contig 76 extends beyond the CEA gene family cluster (Fig. 2A). The specific genes identified in contig 76 are CGM7, CGM2, CEA, NCA, and CGM1 (Fig. 2B). Southern blots of DNA from members of contig 31 were also probed with the CEA IgC-like domain probe identifying three contiguous gene regions (Fig. 3B). However, four specific genes BGP, CGM9, CGM6 and CGM8 have been localized to contig 31 (Fig. 3A). Restriction digestion and probing analysis of cosmids 10514, 6516 and 8998 indicate that the IgC-like probe positive fragments in these cosmids belong to CGM8, a new gene identified in the course of these studies [4]. The remaining two IgC-like regions

in contig 31 are associated with BGP and CGM6, suggesting that CGM9 does not contain IgC-like domains. Contig 31 is an updated version of contig 29 reported by Thompson *et al.* [4]. No cross-hybridization of gene specific probes was observed and and with one exception involving an inconsistency between replicate samples, probe positive cosmids aligned themselves in discrete regions of the contigs (Figs 2B,3B). Thirteen cosmids were fingerprinted more than once, twelve from independent DNA preparations, and as expected for replicate analysis, the results, with one exception, indicated a very high degree of overlap. The results from the independent analysis of cosmid 24914 indicated two nonoverlapping positions (Fig 2A), a discrepancy that could be the result of data inadequacies or sample mishandling.

Identification of specific genes in the PSG-subgroup has been more problematic because of the extremely high sequence similarity among the genes in this subgroup. However, cross-reactivity between PSGs allowed us to assign PSG-subgroup status to three contigs (81,234 and 670) based on positive probings with one or more of three PSG-'specific' oligonucleotides (Table 1). None of the cosmids in these contigs was positive for any of the specific probes for the CEA-subgroup genes.

DISCUSSION

Ordered sets of cloned DNA fragments provide a useful resource for analysis of known genes, studies of gene organization and identification of currently unknown genes. Here we describe the utilization of restriction enzyme digest fingerprint data and subsequent overlap detection by a likelihood statistical approach to assemble five cosmid contigs spanning ~ 1-Mb that includes the CEA-gene family. Two hundred and twelve of 245 cosmids identified as containing the constant domain of the CEA gene and 115 additional chromosome 19 cosmids were included in these contigs. The tiling path for these contigs consists of 64 cosmids. Thus, the redundancy of the cosmid library estimated from the depth of the assembled contigs in the CEA gene region is ~5, which is similar to the estimated redundancy for the entire chromosome 19 library.

A number of lines of evidence confirm the validity of the assembly strategy, and support the computerized assembly of the five CEA gene family contigs. These include the co-localization of 212 CEA IgC-like domain probe positive cosmids into five contigs, out of a present total of ~700 contigs for chromosome 19 and the focussing of gene and family subgroup-specific probes to unique/overlapping cosmids and particular contigs respectively.
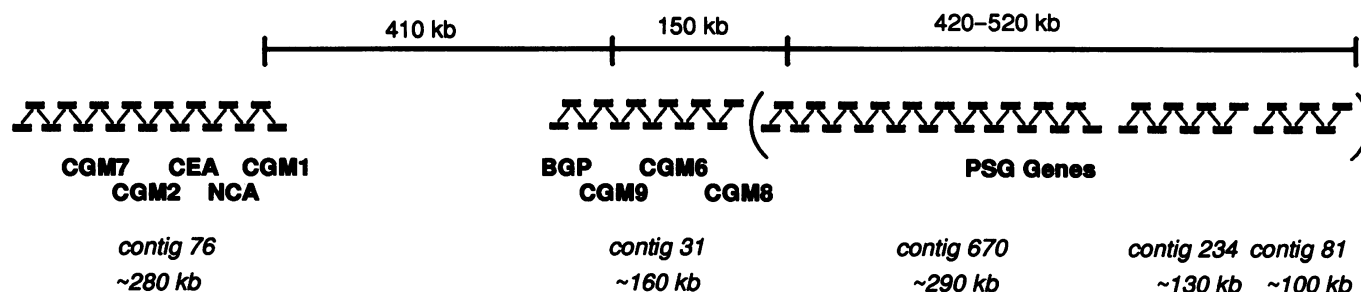

| 410 kb | 150 kb | 420–520 kb |

CGM7   CEA   CGM1   BGP   CGM6   PSG Genes
CGM2   NCA   CGM9   CGM8

contig 76   *contig 31*   *contig 670*   *contig 234*   *contig 81*
~280 kb   *~160 kb*   *~290 kb*   *~130 kb*   *~100 kb*

Figure 4. Summary of the long range physical map of the CEA gene family region located on chromosome 19q13.2. Data was accumulated from the assembly of cosmid contigs, (this paper) pulse field gel electrophoresis [4] and *in situ* hybridization to somatic interphase and sperm pronuclear [35].

The mapping to 19q13.1-q13.2 of 38 cosmids from these contigs by *in situ* hybridization provides additional support for the assembly. Independent experimental confirmation was obtained from complete digest data for both contig 76 and 31. These were entirely consistent with the tiling paths being a set of ordered, overlapping cosmids spanning 280 and 160 kb respectively. The correctness of the order of the cosmids in contig 76 was further substantiated by the partial EcoRI restriction digest map and also by selected cosmid end-probe hybridizations. These data indicate that the fluorescence-based restriction enzyme digest fingerprinting technique [28] and subsequent computerized assembly [31] correctly identify overlapping cosmids.

Assignment of constant domain regions within the CEA-subgroup contigs to particular members of the CEA family [2,3,7] suggested the existence of a previously unidentified CEA-like gene. Subsequent analyses of cosmids 8998 and 14687 led to the identification of a new gene CGM8 in contig 31 [4]. Similar efforts to verify the assembly of the PSG-subgroup contigs and assign specific genes to cosmids have been initiated. The results from preliminary analysis of the PSG cosmids with several gene-specific probes indicate a series of closely spaced genes and also suggest that the contigs are assembled correctly (pers. commun. Sten Hammarstrom, Umea, Sweden).

The overlapping cosmid contig map for the CEA-gene family derived in this study is an initial comprehensive view of the organization of these genes. Brandriff *et al.* [35] have mapped the PSG subgroup distal to the CEA subgroup by two color fluorescence *in situ* hybridization of representative cosmids to metaphase chromatin. Additional fluorescence *in situ* hybridization experiments using sperm pronuclear and somatic interphase chromatin [35] combined with the cosmid contig data; suggest the following order; Cen-//-CGM7-CGM2-CEA-NCA-CGM1-//-BGP-CGM9-CGM6-CGM8-//-PSG-subgroup-//-Tel. This gene order is consistent with pulse field and somatic cell hybrid mapping reported by Thompson *et al.*[4]. Additionally, pulsed field gel electrophoresis analyses [36] indicate that the PSG-subgroup genes are closely linked within an 800 kb *Sac* II restriction enzyme digest fragment. This finding is consistent with our estimate of 520 kb as the sum of the minimal tiling path lengths for the 3 PSG contigs established in this study. This estimate is also consistent with the span of this subgroup as estimated by using fluorescence *in situ* hybridization to sperm pronuclei [35]. A summary of the long-range physical map of the CEA gene family cluster is presented in Fig. 4 (accumulation of data by the present authors and those cited). Thus, the contig assembly strategy positioned the cosmids in an order consistent with other data.

The generation of a cosmid contig based map of the CEA region of chromosome 19 that is consistent with independent results confirms the validity of the semiautomated fluorescence-based fingerprinting technique and the use of likelihood ratios to detect overlap between cosmids. The availability of the cosmid contig map has refined the physical organization of individual members of the family. The cosmids, in turn, are providing material for detailed analyses of gene organization and structure [7], the regulation of gene expression, and the physiological role for these genes.

ACKNOWLEDGEMENTS

REFERENCES

1. Ropers, H. H., Pericak-Vance, M.A. (1991) Cytogenet. Cell Genet. in press.
2. Barnett, T., and Zimmermann, W. (1990) Tumor Biol. **11**: 59–63.
3. Kuroki, M., Arakawa, F., Matsuo, Y., Oikawa, S., Misumi, Y., Nakazato, H., Matsuoka, Y. (1991) J. Biol. Chem. **266**:11810–11817.
4. 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 H., Mohrenweiser, H. (1991) Genomics in press.
5. Thompson, J.A., Grunert, F. and Zimmermann, W. (1991) J. Clin. Lab. Anal. **5**:344–366.
6. Rudert, F., Zimmermann, W., and Thompson, J.A. (1989) J. Mol. Evol. **29**: 126–134.
7. Streydio, C., Swillens, S., Georges, M., Szpirer, C., and Vassart, G. (1990) Genomics **6**: 579–592.
8. Paxton, R.J., Mooser, G., Pande, H., Lee, T.D., and Shively, J.E. (1987) Proc., Natl. Acad. Sci. USA **84**: 920–924.
9. Beauchemin, N., Benchimol, S., Cournoyer, D., Fuks, A., and Stanners, C. (1987) Mol. Cell. Biol. **7**: 3221–3230.
10. Thomas, P., Toth, C.-A., Kulvinder, S.S., Jessup, J.S.,and Steele, JR, G. (1990) Biochimica et Biophysica Acta **1032**:177–189.
11. Huang, J.Q., Turbide, C., Daniels, E., Jothy, S., and Beauchemin, N. (1990) Development **110**: 573–588.
12. 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.
13. Zimmermann, W., Weber, B., Ortlieb, B., Rudert, F., Schempp, W, Fiebig, H.-H., Shively, J.E., von Kleist, S., and Thompson, J.A. (1988) Cancer Res. **48**: 2550–2554.
14. Benchimol, S., Fuks, A., Jothy, S., Beauchemin, N., Shirota, K., and Stanners, C.P. (1989) Cell **57**: 327–334.
15. Oikawa, S., Inuzuka, C., Kuroki, M., Matsuoka, J.,Kosaki, G., and Nakazato, H. (1989) Biochem. Biophys. Res. Commun. **158**: 39–45.
16. Bohn, H. (1986).Pregnancy proteins in animals Hau U. (ed), de Gruyter, Berlin, New York **166**: 247–269.
17. De Jong, P.J., Yokabata, K., Chen, C., Lohman, F., Pederson, L., McNinch, J., and Van Dilla, M. (1989) Cytogen. Cell Genet. **51**: 985.
18. Siciliano, M.J., Carrano, A.V., and Thompson, L.H. (1986) Mutat. Res. **174**: 303–308.
19. Gibson, T.J., Coulson, A.R., Sulston, J.E., and Little, P. F.R. (1987) Gene **53**: 275–281.
20. Yokobata, K., Trenchak, B., and De Jong, P.J. (1991) Nucleic Acids Res. **19**: 403–404.
21. Zimmermann, W., Ortlieb, B., Friedrich, R., and von Kleist, S. (1987) Proc. Natl. Acad. Sci. USA **84**: 2960–2964.
22. 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. (1989) Biochem. Biophys. Res. Commun. **158**: 996–1004.
23. 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.
24. Barnett, T.R., Kretschmer, A., Austen, D.A., Goebe, S.J., Hart, J.T., Elting, J.J.,and Kamarck, M.E. (1989) J. Cell Biol. **108**: 267–276.
25. Feinberg, A.P., and Vogelstein B. (1983) Anal. Biochem.**132**: 6–13.
26. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). Molecular cloning: A Laboratory Manual. Cold Spring Harbor Press, NY.
27. Church, G.,M. and Gilbert, W. (1984) Proc. Natl. Acad. Sci., USA **88**: 1991–1995.
28. 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) Genomics **4**: 129–136.
29. Carrano, A.V., Alleman, J., Amemiya, C., Ashworth, L.K., Aslanidis, C., Branscombe, E.W., Combs, J., Chen, C., Christensen, M., Copeland, A., Lamberdin, J., Mohrenweiser, H., Olsen, A., Slezak, T., Trask, B., and Tynan, K. (1991) Miami Bio/Technology Winter Symposia 1991, Ahmad, F (ed) Advance in Gene Technology: The Molecular Biology of Human Genetic Disease Vol 1, pg 60.

30. Carrano, A.V., Branscomb, E.W., De Jong, P.J., Mohrenweiser, H., Olsen, A., and Slezak, T. (1991) Electrophoresis, Supercomputing and the Human Genome, Proceedings of the 1st Int. Conf. Tallahassee, Fl. Cantor, C.R., and Lim H.A., (ed) World Scientific pg 25–34.
31. Branscomb, E., Slezak, T., Pae, R., Galas, D., Carrano, A.V., and Waterman, M. (1990) Genomics **8**: 351–366.
32. Wahl, G.M., Stern., M., Stark, G.H. (1979) Proc. Natl. Acad. Sci., USA **79**: 3683–3687.
33. Mohrenweiser, H.W., Carrano,A.V.,Fertitta, A., Perry, B., Thompson, L.H., Tucker, J.D., Weber, C.A. (1989) Cytogenet. Cell Genet. **52**:11–14.
34. Schweizer, D. (1976) Chromosoma **58**:307–324.
35. Brandriff, B.F., Gordon, L.A., Tynan, K., Olsen, A., Mohrenweiser, H., Carrano, A.V., and Trask, B.J. (1991) Genomics in press.
36. Thompson, J., Koumari, R., Wagner, K., Barnert, S., Schleussner, C., Schrewe, H., Zimmermann, W., Muller, G., Schempp, W., Zaninetta, D., Ammaturo, D., and Hardman, N. (1990) Biochem. Biophys. Res. Commun. **2**: 848–859.