Available online at www.sciencedirect.com
SCIENCE DIRECT®
Gene 310 (2003) 79–89
GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES
www.elsevier.com/locate/gene

Molecular characterization of a new gene, CEAL1, encoding for a carcinoembryonic antigen-like protein with a highly conserved domain of eukaryotic translation initiation factors

Andreas Scorilas[a,b,1], Pui Ming Chiang[a], Dionyssios Katsaros[c], George M. Yousef[a,b], Eleftherios P. Diamandis[a,b,*]

[a]*Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, 600 University Avenue, Toronto, ON, Canada M5G 1X5*
[b]*Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada M5G 1L5*
[c]*Department of Gynecology, Gynecologic Oncology Unit, University of Turin, 10126 Turin, Italy*

Received 28 October 2002; received in revised form 24 January 2003; accepted 27 February 2003
Received by R. Di Lauro

Abstract

Carcinoembryonic antigen (CEA) is a complex immunoreactive glycoprotein belonging to a large and heterogeneous group of cross-reacting proteins known as the CEA gene family, which contains 29 genes/pseudogenes. CEA is used as a valuable serum tumor marker for monitoring response to therapy in patients with various solid tumors. Through the positional cloning approach we have identified and characterized a CEA-like gene (CEAL1), a novel member of the CEA multigene family. We have characterized the complete genomic structure of CEAL1, as well as one alternative splice variant and determined its chromosomal localization. The new gene is comprised of eight exons, with seven intervening introns and it is localized to chromosome 19q13.2 between the markers D19S574 and D19S219, ~60 kb upstream of the BCL3 gene. The protein-coding region of the gene is formed of 903 bp, encoding for a 300-amino-acid polypeptide with a predicted molecular weight of 32.6 kDa and isoelectric point of 5.74. The CEAL1 protein contains two Immunoglobulin-like (Ig-like) transmembrane domains, which are present in most of the CEA proteins, as well as one highly conserved domain of eukaryotic translation initiation factors. The identified alternative spliced variant has one more exon of 134 bp. This splice variant is expected to encode for a truncated protein of 142 amino acids with the eIF5A domain and without Ig homology domain. CEAL1 mRNA is expressed in a variety of tissues, but highest levels are found in the prostate, uterus, fetal brain, mammary, adrenal gland, skeletal muscle, small intestine and kidney. CEAL1 is highly expressed in BT-474, BT20, T47D and, at much lower levels, in MCF7 breast cancer cell lines. The new gene is also highly expressed in the LNCaP prostate cancer cell line. The CEAL1 gene was found to be down-regulated by dexamethasone in BT-474 breast cancer cell lines. Our data suggest that this gene is overexpressed in a subset of ovarian cancers which are clinically more aggressive.

Keywords: Carcinoembryonic antigen; Ovarian cancer; eIF-5A; Tumor marker; Gene expression; Cancer

Abbreviations: BGP, biliary glycoprotein; CEA, carcinoembryonic antigen; CEAL1, CEA-like gene 1; CGM, CEA gene family member; CEACAM, carcinoembryonic antigen-related cellular adhesion molecule; GPI, glycosyl phosphatidylinositol; IgC, immunoglobulin constant; IgSF, immunoglobulin superfamily; IgV, immunoglobulin variable; ITAM, immunotyrosine activation motif; ITIM, immunotyrosine inhibition motif; NCA, non-specific cross-reactive antigen; PSG, pregnancy specific glycoprotein; RACE, rapid amplification of cDNA ends; RT–PCR, reverse transcription–polymerase chain reaction; eIF-5A, eukaryotic translation initiation factor 5A; SAGE, serial analysis of gene expression; EST, expressed sequence tag.

* Corresponding author. Tel.: +1-416-586-8443; fax +1-416-586-8628.
E-mail addresses: ediamandis@mtsinai.on.ca (E.P. Diamandis), scorilas@netscape.net (A. Scorilas).
[1] Present address: National Center for Scientific Research 'Demokritos', IPC, 153 10 Athens, Greece.

1. Introduction

Carcinoembryonic antigen (CEA/CD66e/CEACAM5) is a complex immunoreactive glycoprotein containing ~60% carbohydrate and has a molecular weight of ~180 kDa (Oikawa et al., 1987). It was initially thought that this antigen is an oncofetal antigen, which was expressed at high levels in the fetal colon and in colon adenocarcinoma but was absent from healthy adult colon (Nap et al., 1988). Today CEA is a widely used test for diagnosis, prognosis and monitoring of colon cancer as well as other malignances (Duffy, 2001).

0378-1119/03/$ - see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0378-1119(03)00499-2

CEA belongs to a large and heterogeneous group of cross-reacting proteins known as the CEA gene family. The human CEA gene family contains 29 genes that are clustered within a region of 1.5 Mb on chromosome 19q13.2 (Olsen et al., 1994). Based on sequence homology, members of the CEA family can be divided into three subgroups: the CEA subgroup containing 12 genes, the PSG (pregnancy specific glycoprotein) subgroup containing 11 genes and a third subgroup containing six pseudogenes (Teglund et al., 1994). The genes of the CEA family are organized into two clusters of 250 and 850 kb, separated by a region of ~700 kb containing several unrelated genes. The PSG and the third family genes are located telomerically to the CEA subgroup genes, where the third subgroup genes are interspersed between the PSG genes (Hammarstrom, 1999).

Amino acid sequence analyses have revealed that the CEA gene family belongs to the immunoglobulin superfamily (IgSF) (Thompson et al., 1991; Zhou et al., 2001). The C-terminal domains of the CEA subgroup members are mostly membrane-bound. Most of the CEA subgroup members contain a hydrophobic transmembrane domain (TM) followed by either a short or a long cytoplasmic domain (CYT), whereas others are attached to the cell membrane via a glycosyl phosphatidylinositol (GPI) linkage. The PSG subgroup genes encode for glycoproteins that contain a short hydrophilic C-terminal tail, of which there are four different types, and are exported as secreted macromolecules (Teglund et al., 1994; Hammarstrom, 1999; Zhou et al., 2001). Several members of the CEA subfamily, have been shown to function in vitro as homophilic and heterophilic intercellular cell-adhesion molecules (Oikawa et al., 1992). It has also been shown that the N-terminal domain is directly involved in cell-adhesion phenomena and the primordial member, biliary glycoprotein, seems to function in signal transduction or regulation of signal transduction, possibly in association with other CEA subfamily members (Hammarstrom, 1999).

Many studies have shown that members of CEA gene family are expressed in a number of tumors of epithelial origin such as colorectal carcinoma, lung adenocarcinoma, mucinous ovarian carcinoma and endometrial adenocarcinoma. These genes were further suggested to play a role in cancer invasion and metastasis, as alterations in cell adhesion are causally involved in these processes (Duffy, 2001). However, great caution must be taken when comparing the levels of expression of CEA family members in normal and cancerous tissues. In many cases, the conclusion that a particular antigen is up or down-regulated in tumors will depend on whether the comparisons were made against poorly or fully differentiated normal cells. Since elevated CEA serum concentrations can be found in colon, breast and lung adenocarcinomas, CEA is one of the most widely used human tumor markers worldwide (Shively and Beatty, 1985).

Eukaryotic translation initiation factor 5A (eIF-5A), formerly known as eIF-4D, is a small, ubiquitous protein with a molecular weight of 18 kDa that is highly conserved among the eukaryotic species (Caraglia et al., 2001). eIF-5A contains the hypusine residue (N^{e}-(4-amino-2-hydroxybutyl)lysine). This unusual basic amino acid is formed by the post-translational transfer of the butylamine group from spermidine to the e-amino group of a specific lysine residue of eIF-5A precursor and the subsequent hydroxylation at carbon 2 of the incoming 4-aminobutyl moiety. In vitro studies have revealed that eIF-5A promotes the formation of the first peptide bond during the final stage of the initiation phase of protein synthesis. eIF-5A metabolic pathway has been found to correlate with the transglutaminase expression and activity. There is evidence that eIF-5A can be modified through a reaction catalyzed by transglutaminase (TGase) in vitro (Beninati et al., 1995; Caraglia et al., 2001). It has been suggested that eIF-5A may play a role in cell growth and differentiation. Agents that prevent the hypusine formation of eIF-5A induce reversible arrest at the G_1-S boundary of the cell cycle, which inhibits the growth of mammalian cells (Caraglia et al., 2001).

During our efforts to identify new carcinoembryonic antigen-like genes that might be involved in malignancy, we cloned a new gene, tentatively named CEAL1 (for CEA like gene 1). Here, we describe the identification and molecular characterization of the new gene as well as its expression pattern in tissues and cancer cell lines. We also investigate its possible hormonal regulation and we describe the identification of one alternatively spliced form. We further describe the expression of the new gene in a subset of ovarian and breast cancers and preliminarily investigate its role in ovarian cancer prognosis.

2. Materials and methods

2.1. Cloning and identification of the new gene

In our search for cancer related genes encoding for new carcinoembryonic antigens, an analysis was performed for different conserved motifs of the immunoglobulin superfamily genes, using the TBLASTN algorithm and the human expressed sequence tag (EST) database. Any matching ESTs, we identified were obtained from the I.M.A.G.E. consortium through Research Genetics Inc. (Hunstville, AL). These clones were then propagated, purified and sequenced from both directions with an automated sequencer, using the flanking T3 and T7 vector primers. A BAC clone (BC94899) with high homology with the examined ESTs was identified. We obtained genomic sequences derived from this clone, covering an area of chromosome 19q13.2, from the Lawrence Livermore National Laboratory (LLNL). The genomic sequences were in the form of different contigs of various lengths. The chromosome 19 *Eco* RI restriction map (Ashworth et al., 1995), long PCR strategies and sequencing, using genomic

DNA, were used to construct a contiguous genomic area of interest. Long PCR products were gel-purified, cloned into the TOPO-XL cloning vector and propagated in LB medium with ampicillin. Plasmids were then purified and sequenced as previously described (Sanger et al., 1977). We utilized the BLAST alignment tool to determine the exact location of any EST identified above within this genomic sequence. Bioinformatic approaches were used, as described elsewhere (Bonet et al., 1998; Yousef et al., 2000a,b, 2001; Scorilas et al., 2001a,b), to predict presence of new genes. A putative new carcinoembryonic antigen-related gene, with a highly conserved eIF domain protein was then identified.

Based on the alignment of the ESTs and the exon prediction results, we proceeded to design sets of primers for reverse transcription–polymerase chain reaction (RT–PCR), in order to determine the exact sequence of the mRNA species. This design allowed for the generation of overlapping RT–PCR fragments, thus enabling determination of the entire mRNA sequence. Based on results from RT–PCR with a panel of human tissues (see below), we utilized colon and prostate tissues for further analysis. Tissue cDNAs were prepared as described below and amplified with different sets of primer combinations (Table 1).

In order to verify the 3′- and 5′-ends of the CEAL1 mRNA, we utilized Marathon-Ready fetal liver cDNA (Clontech, Palo Alto, CA) to perform nested 3′- and 5′-RACE. Four gene-specific reverse primers were used (CEAL1-R1, CEAL1-R′1, CEAL-1-R8 and CEAL1-R′8) along with the forward primers shown in Table 1. Two rounds of RACE reactions were performed with 5 μl of cDNA. The reaction mix and PCR conditions were selected according to the manufacturer's recommendations. All RACE reactions were performed using the Perkin Elmer 9600 Thermocycler (Perkin Elmer, Norwalk, CT). In brief, the initial denaturation was for 5 min at 94 °C, followed by heating at 94 °C for 5 s and 72 °C for 2 min, for five cycles; then, 94 °C for 5 s and 66 °C for 2 min, for five cycles; then, 94 °C for 5 s and 63 °C for 2 min for 30 cycles for the first reaction and 25 cycles for the nested PCR reaction. Positive bands were gel-purified using a gel purification kit cloned and sequenced.

Table 1
Primers used for RT–PCR analysis of the CEAL1 and actin genes

Gene	Primer name	Sequence[a]	Length of PCR product (bp)
CEAL1	CEAL1-F1	TTC CAG AGC AGC CTC AAA AGA	1108 and 974
	CEAL1-R8	TGG GCC TGT TAC AGT CTC CTG	
	CEAL1-F′1	TGA TCT CAG CCC TTC CCT GTA	1643 and 1509
	CEAL1-R′8	GCT GAA TGA ATG CAT AGT TTG	
	CEAL1-R1	TCT TTT GAG GCT GCT CTG GAA	
	CEAL1-R′1	TAC AGG GAA GGG CTG AGA TCA	
Actin	ACTINS	ATC TGG CAC CAC ACC TTC TA	838
	ACTINAS	CGT CAT ACT CCT GCT TGC TG	

[a] All nucleotide sequence are given in the 5′ → 3′ orientation.

2.2. Tissue expression of two CEAL1 splice variants

The tissue expression profile for both CEAL1 alternative splice forms was elucidated by performing RT–PCR using total RNA from 26 normal human tissues (Clontech, Palo Alto, CA). The PCR primers used were CEAL-F1 and CEAL-R8 (Table 1). The PCR conditions were the same as those described below, and reverse transcription was performed using SuperScript II, according to the manufacturer's instructions (Gibco BRL, Gaithersburg, MD). A PCR for actin, as described elsewhere (Yousef et al., 2001a), was included as a control of cDNA quality (for primers see Table 1).

2.3. In silico analysis of CEAL1 expression

Analysis of CEAL1 mRNA expression in a variety of normal and cancer libraries obtained from different tissues were performed using the database of the Cancer Genome Anatomy Project (CGAP) of the National Cancer Institute (NCI). Quantitative CEAL1 levels were performed by 'Virtual Northern' analysis of Serial Analysis of Gene Expression (SAGEmap) database. Comparison between normal and cancerous libraries was done using the Expressed Sequence Tag (EST) database of the CGAP.

2.4. Cancer cell lines and hormonal stimulation experiments

The BT-474, T47D, BT20 and MCF7 breast cancer as well as the LNCaP prostate cancer cell lines were purchased from the American Type Culture Collection (ATCC), Rockville, MD. Cells were cultured in RPMI media (Gibco BRL, Gaithersburg, MD) supplemented with glutamine (200 mmol/l), bovine insulin (10 mg/l), fetal bovine serum (10%), antibiotics and antimycotics, in plastic flasks, to near confluency. The cells were then aliquoted into 24-well tissue culture plates and cultured to 50% confluency. Twenty-four hours before the experiments, the culture media were changed into phenol red-free media containing 10% charcoal-stripped fetal bovine serum. For stimulation experiments of BT-474 cells, various steroid hormones dissolved in 100% ethanol were added into the culture media at a final concentration of 10^{-8} M. Cells stimulated with 100% ethanol were included as controls. The cells were cultured for 24 h, then harvested for mRNA extraction.

2.5. Ovarian and breast tumors

Thirty-seven ovarian tumor tissues were obtained from patients who had undergone surgery for primary ovarian carcinoma at the Department of Obstetrics and Gynecology, Gynecological Oncology Unit, University of Turin, Turin, Italy. Fifteen breast tumor tissues were obtained from female patients at participating hospitals of the Ontario Provincial Steroid Hormone Receptor Program. Ovarian patient age ranged from 33 to 78 years with a median of 60. All patients had a histologically-confirmed diagnosis and received no treatment before surgery. Classification followed the World Health Organization criteria. Clinical staging was performed according to the International Federation of Gynecology and Obstetrics (FIGO) staging system. Histologic grading was performed according to the criteria of Day et al.; highly differentiated tumors were grade 1 and undifferentiated tumors grade 3 (Day et al., 1975). Tissue specimens with different types of ovarian tumors were confirmed histologically, and 50–100 mg of tumor tissue were pulverized on dry ice, followed by total RNA extraction using the Trizol method (Gibco BRL).

2.6. RT–PCR for the CEAL1 gene

Two micrograms of total RNA from tissues and cell lines were reverse-transcribed into first strand cDNA using the SuperScript preamplification system (Gibco BRL). The final volume was 20 μl. Two gene-specific primers were used (CEAL1-F1 and CEAL1-R8) (Table 1) and PCR was carried out in a reaction mixture containing 1 μl of cDNA, 10 mM Tris–HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 200 μM dNTPs (deoxynucleoside triphosphates), 150 ng of primers and 2.5 units of HotStar DNA polymerase (Qiagen Inc., Valencia, CA) on a Mastercycler Gradient Thermocycler (Eppendorf Scientific Inc., Westbury, NY). The cycling conditions were 95 °C for 15 min to activate the Taq DNA polymerase, followed by 35 cycles of 94 °C for 30 s, 62 °C for 30 s, 72 °C for 1 min and a final extension step at 72 °C for 10 min. Equal amounts of PCR products were electrophoresed on 2% agarose gels and visualized by ethidium bromide staining.

2.7. Cloning and sequencing of the PCR products

To verify the identity of the PCR products, they were cloned into the pCR 2.1-TOPO vector (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. The inserts were sequenced from both directions using vector-specific primers, with an automated DNA sequencer.

2.8. Structure analysis of the CEAL1 protein

Multiple alignment was performed using the 'Clustal X' software package and the multiple alignment program available from the Baylor College of Medicine, Houston, TX. Hydrophobicity study was performed using the Baylor College of Medicine search launcher. Protein structure analysis was performed by the SAPS (structural analysis of protein sequence) program (Brendel et al., 1992). Sequence analysis tools were utilized to detect the presence of possible sites of post-translational modification on the CEAL protein. We used the analysis program PROSITE (Bairoch et al., 1997) to detect *N*- and *O*-glycosylation, as well as the presence of kinase phosphorylation motifs.

2.9. Mapping and chromosomal localization of CEAL1

As indicated above, the CEAL1 gene was identified within a genomic sequence from a BAC clone covering chromosome 19q13.2. We subjected the sequence encompassing the CEAL1 gene to the Webcutter restriction analysis tool to determine the size of the resultant *Eco* RI fragments. We then compared these results to the published *Eco* RI map of chromosome 19 (Ashworth et al., 1995), which is available through the LLNL Human Genome Center.

3. Results

3.1. Cloning and identification of the carcinoembryonic antigen-like gene (CEAL1)

Using the BC94899 BAC clone that was obtained and processed as described above, we identified a putative new gene formed of at least eight exons. Screening of the human expressed sequence tag database (dbEST) revealed eight EST clones (Table 2) with $>98\%$ identity with our predicted exons. The full-length sequence of one EST (I.M.A.G.E ID 2132102) showed 100% identity with one more exon as an alternative splicing product.

To identify the full mRNA structure of the gene and to determine the exon/intron boundaries, PCR reactions were performed using primers located in different predicted exons, using a panel of 28 human tissue cDNAs as templates. PCR products were purified and sequenced. Two of these primers (CEAL1-F1 and CEAL1-R8) (Table 1) were able to amplify the full coding region of the gene from different tissues. Comparing the sequenced cDNA with the genomic structure indicated the presence of a gene formed of eight exons with seven intervening introns. 5′- and 3′-RACE reactions and sequencing were performed in order to obtain the 5′- and 3′-ends of the gene. We identified 300 bp of sequence upstream of CEAL1-F1 primer as well as 679 bp downstream to CEAL1-R8 primer.

An in-frame methionine start codon was found in the first exon. Translation of the mRNA sequence in all possible reading frames revealed the presence of only one frame that gives an uninterrupted polypeptide chain containing many amino acid loops which are also present in carcinoembryonic antigen related proteins. The new peptide also

Table 2
EST clones with > 98% identity to exons of the CEAL1 gene

GenBank accession no.	Tissue	I.M.A.G.E. ID	Homologous exons
BF880890	Lung	2132102	1–4
BG675607	Skin	4746917	1, 3–8
BF062458	Colon	3320263	1, 3, 4
AA460340	Soares total fetus	796468	1, 3, 4
N51814	Multiple sclerosis lesions	279125	5–7
BE221800	Kidney	3132797	7, 8
AW082351	Pooled	2581218	8
AA972606	Pooled	1582991	8

contains a highly conserved eIF domain of eukaryotic initiation factors and one immunoglobulin like domain, as discussed below.

3.2. Structural characterization of the CEAL1 gene

As shown in Fig. 1, the CEAL1 gene is formed of 8 coding exons and 7 intervening introns, spanning an area of 12,609 bp of genomic sequence on chromosome 19q13.2. All of the exon/intron splice sites (mGT…AGm) conform to the consensus sequence for eukaryotic splice sites (Iida, 1990). Lengths of the coding exons are 240, 369, 151, 84, 47, 86, 54 and 922 bp, respectively. Nucleotides 15267–15272 (AATAAA) (numbers refer to our GenBank submission accession no. AF406955) are identical with a consensus polyadenylation signal (Proudfoot et al., 1976) and are followed, after 12 nucleotides, by the poly(A) tail not found in the genomic sequence. Another potential polyadenylation signal (AATAAA) was discernible in the 3′-untranslated region, seven bases upstream of the poly(A) tail. The protein-coding region of the gene is formed of 903 bp, encoding a deduced 300-amino-acid polypeptide with a predicted molecular weight of 32.6 kDa and isoelectric point of 5.74.

Comparative analysis revealed that the CEAL1 protein sequence has a relatively low degree of homology with other members of the carcinoembryonic antigen multigene family. CEAL1 shows ~30% protein identity and ~50% protein similarity with the CGM1, CGM2, CGM6, CGM7 and CEA proteins. Fig. 2 shows the alignment of the deduced amino acid sequence of CEAL1 with other members of the carcinoembryonic antigen gene family. Two Ig-like transmembrane domains were identified at amino acid positions 14–36 and 158–182 which are also present in most of the CEA proteins (Fig. 1). The underlined region in Fig. 1 indicates a 9-amino-acid sequence, found in the eIF5A domain of several members of the eukaryotic translation initiation factors (Caraglia et al., 2001). The amino acid loop (SAMGXXXIV) at positions 92–103 was also found in human eIF5A and eIF5A2 as well as in *Mus musculus*, *Oryctolagus cuniculus* and *Gallus gallus* eIF5A.

Hydrophobicity analysis revealed a hydrophobic N-terminal region (Fig. 3), consistent with the possibility that this region has a signal sequence or an N-terminal Ig-like transmembrane domain. The transmembrane domains, predicted by Tmpred and also evident in the Kyte–Doolittle hydrophobicity plot (Fig. 3), is in keeping with observations for other members of CEA subgroup.

Using sequence analysis tools, we were able to identify various putative post-translational modification sites (Table 3). There are numerous potential sites for *O*-glycosylation. Furthermore, several possible sites of phosphorylation have been identified for cAMP-dependent protein kinase, protein kinase C, and casein kinase 2. In addition, several *N*-myristoylation sites have been predicted (Table 3).

3.3. Splice variants of the CEAL1 gene

PCR screening for CEAL1 transcripts using gene-specific primers (CEAL1-F1 and CEAL-R8) (Table 1) revealed the presence of two bands in most of the tissue cDNAs examined (Fig. 4). The two bands were gel-purified, cloned and sequenced. The lower band represents the classical form of the gene, and the upper band is splice variant 1. As shown in Fig. 1, this variant (CEAL1 – splice variant 1) has one more exon of 134 bp (for full sequence, see GenBank accession no. AF406955). This splice variant is expected to encode for a truncated protein of 142 amino acids with the eIF5A domain but without the Ig homology domain.

Table 3
Putative post-translational modification sites in the CEAL1 gene

Modification	Residue	Position[a]
O-Glycosylation	Thr	216, 219, 224, 268
	Ser	210, 214, 218
N-Glycosylation	Asn	104
Protein kinase C phosphorylation	Thr	131, 224
	Ser	190
Casein kinase II phosphorylation	Thr	59
N-Myristoylation	Gly	6, 29, 55, 91, 157, 165

[a] The residues are numbered according to the sequence shown in Fig. 1 and our GenBank submission (accession no. AF406955).

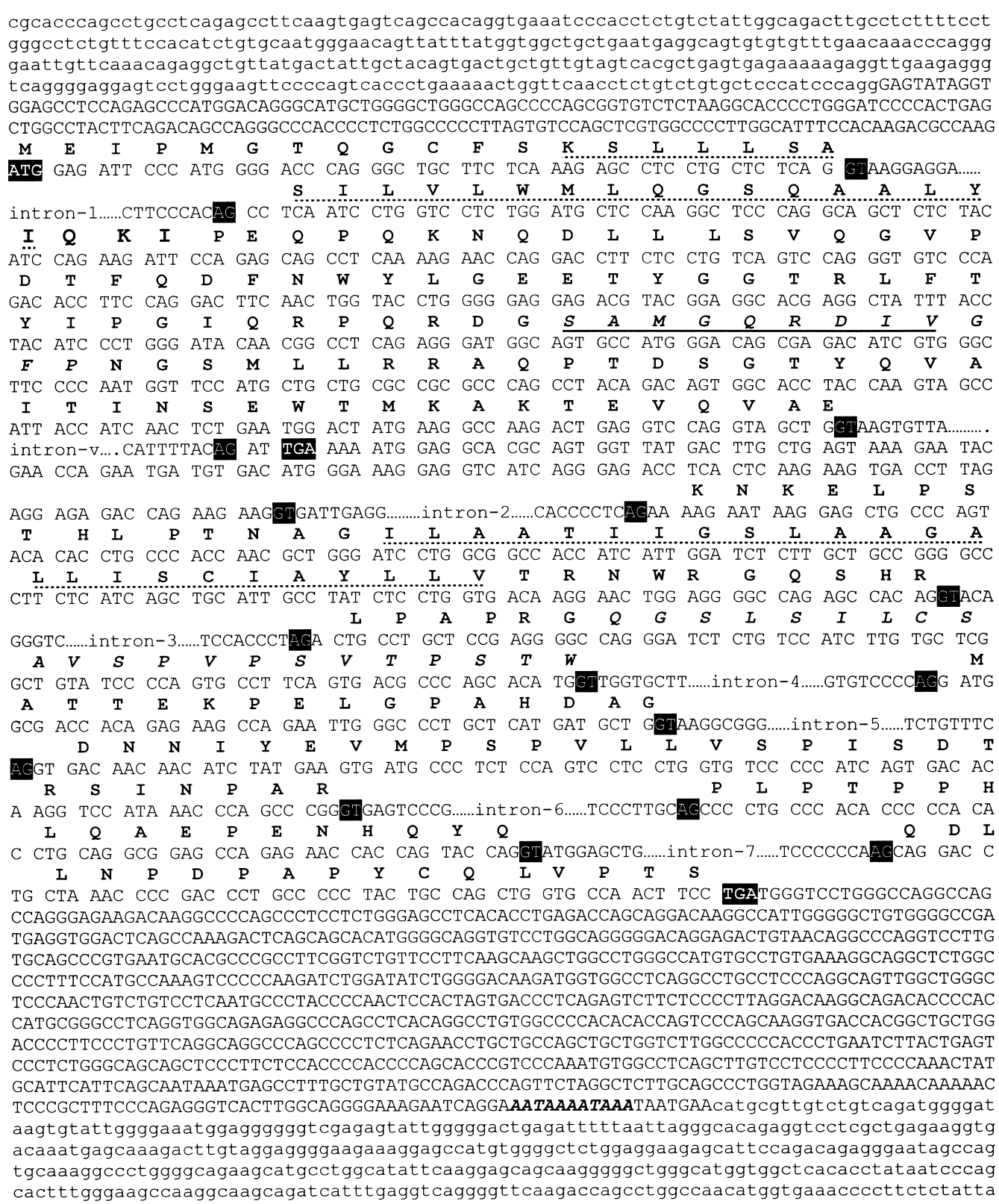

Fig. 1. Genomic organization and partial genomic sequence of the CEAL1 gene. Intronic sequences are not shown except for short sequences around the splice junctions. Exons and partial sequence of intervening introns are shown in capital letters. For full sequence, see GenBank accession no. AF406955. The start, stop codons and the exon–intron junctions are in bold background. Exon 3 is not present in the classical form of the gene. The translated amino acids of the coding region are shown by a single-letter abbreviation. The broken line identifies the Ig-like transmembrane domains. The solid line indicates a 9-amino-acid sequence, found in the eIF5A domain of several members of the eukaryotic translation initiation factors. Putative polyadenylation signals are shown in bold and italic.

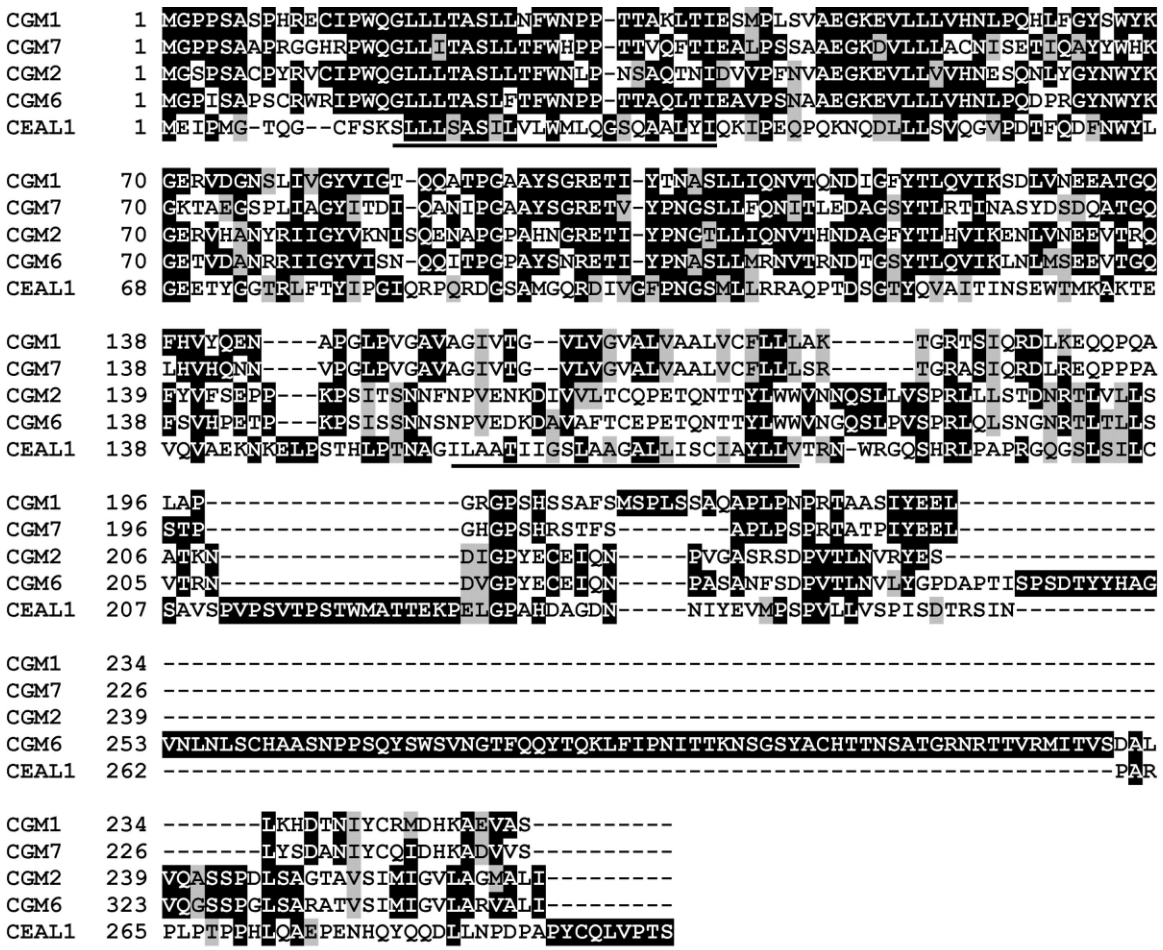

Fig. 2. Alignment of the deduced amino acid sequence of CEAL1 gene with other members of the carcinoembryonic antigen gene family. For full gene names see 'Abbreviations'. Dashes represent gaps to bring the sequences to better alignment. Identical residues are highlighted in black, and similar residues in gray.

3.4. Mapping and chromosomal localization of the CEAL1 gene

Restriction analysis of overlapping BAC clones spanning the chromosomal area of interest and the *Eco* RI restriction map (Ashworth et al., 1995) allowed us to locate the new gene on chromosome 19q13.2 between the markers D19S574 and D19S219, ~60 kb upstream of the BCL3 gene.

3.5. Expression patterns of CEAL1

RT–PCR analysis (with primers CEAL1-F1 and CEAL1-R8) of different human tissues was used to identify the expression pattern of CEAL1. Actin was used as a control gene. In each of the 26 tissues tested, a CEAL1 transcript was identified. As shown in Fig. 4, the classic form of the CEAL1 gene is highly expressed in the prostate, uterus, fetal brain, mammary gland, adrenal gland, skeletal muscle, small intestine and kidney. Lower levels of expression are seen in lung, cerebellum, testis, liver, pancreas, bone marrow and ovary. The CEAL1-splice variant 1 also followed a similar expression. In order to verify the RT–PCR specificity, representative PCR products were cloned and sequenced.

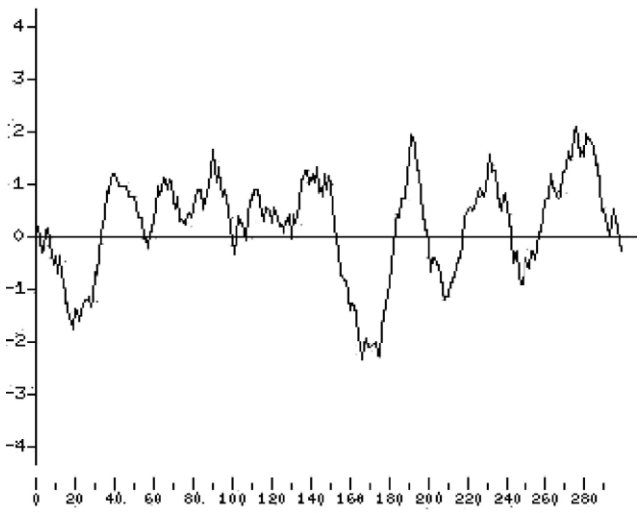

Fig. 3. Plot of hydrophobicity and hydrophilicity of CEAL1 protein. For details see text.

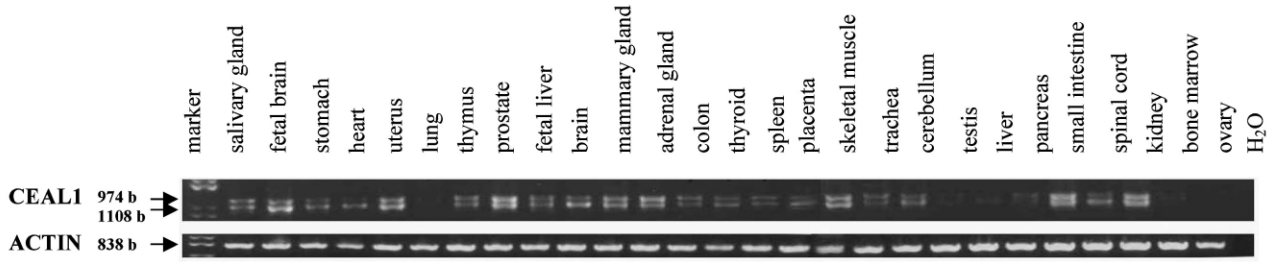

Fig. 4. Tissue expression of the CEAL1 gene as determined by RT–PCR (upper panel). Actin (lower panel) is a control gene.

3.6. Expression and hormonal regulation of CEAL1 in cancer cell lines

As shown in Fig. 5, CEAL1 is highly expressed in BT-474, BT20, T47D, and to a much lower extent, in MCF7 breast cancer cell lines. This gene is also highly expressed in the LNCaP prostate cancer cell line. We used the steroid hormone receptor-positive breast carcinoma cell line BT-474 to evaluate whether *CEAL1*expression is hormonally regulated. It seems that CEAL1 gene expression is down-regulated by dexamethasone. Norgestel, dihydrotestosterone (DHT) and aldosterone do not appear to have any effect to CEAL1 expression (Fig. 5).

3.7. In silico expression analysis of CEAL1 gene

Analysis of CEAL1 mRNA expression in a variety of normal and cancer libraries from different tissues were performed. Two unique CEAL1 SAGEtags (CGGGCCTCAG and GCACTGCCAT) was blasted against the SAGEmap database. CEAL1 was found to be ubiquitously expressed in many tissues. High expression densities were found in colon cancer, ovarian cancer, brain tumors, normal colon, pancreas, prostate and vascular endothelium. Human EST database screening, available through the CGAP and NCBI databases, showed high representation of the gene in the pancreas, kidney, brain, colon and bladder tumors.

3.8. Expression of CEAL1 in ovarian and breast cancer tissues

Examples of CEAL1 gene expression in ovarian and breast tumors are shown in Fig. 6. Tumors from 37 ovarian cancer patients with histological and clinical follow-up information were used. RT–PCR analysis (with primers CEAL1-F1 and CEAL1-R8) of the 37 ovarian cancer tissues was used to identify the expression CEAL1. The PCR cycling conditions were 95 °C for 15 min to activate the Taq DNA polymerase, followed by 35 cycles of 94 °C for 30 s, 62 °C for 30 s, 72 °C for 1 min and a final extension step at 72 °C for 10 min. Actin was used as a control gene. The distributions of CEAL1 qualitative expression status (positive or negative; undetectable) between subgroups of

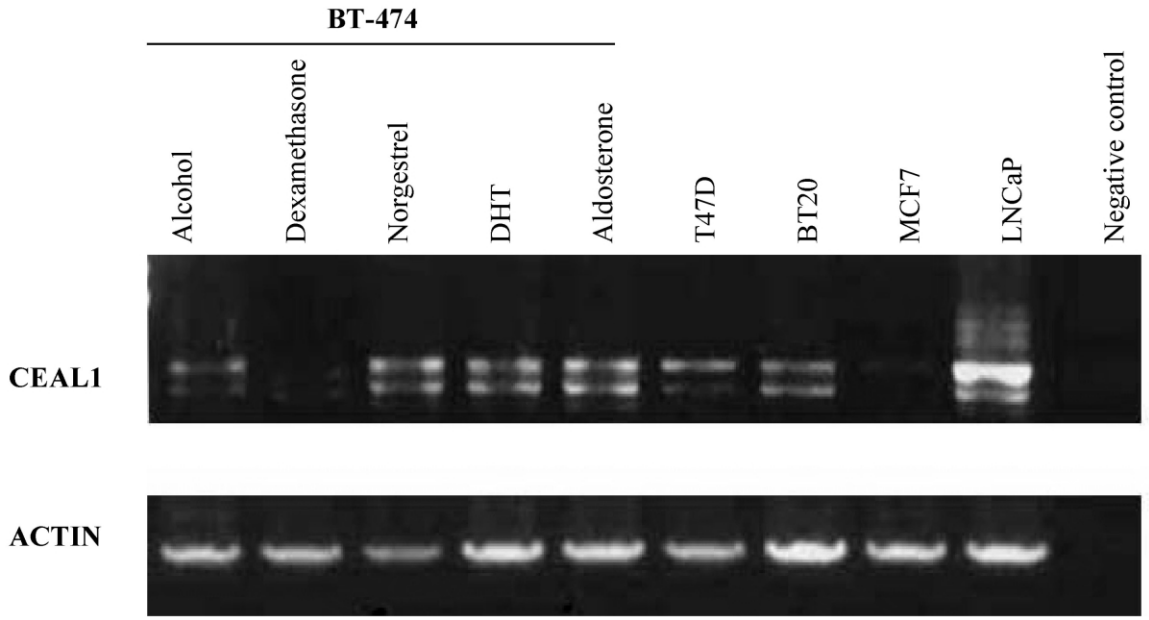

Fig. 5. Expression and hormonal regulation of the of the CEAL1 gene in breast and prostate cancer cell lines (LNCaP). Steroids were at 10^{-8} M final concentration. Actin (not regulated by steroid hormones) was used as a control gene. For interpretations see text. DHT, dihydrotestosterone.

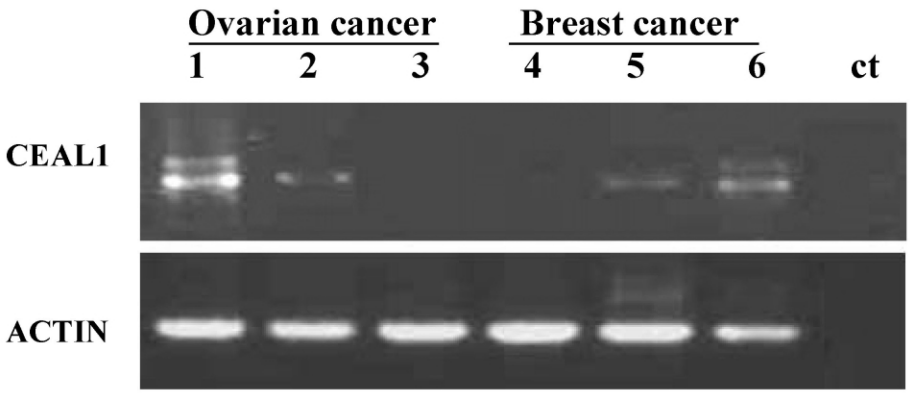

Fig. 6. Expression of the CEAL1 gene in ovarian and breast cancer. 1–3, ovarian tumors with high, low and no expression, respectively, of the CEAL1 gene. 4–6, breast tumors with no expression, low and high expression, respectively, of the CEAL1 gene. Actin is a control gene. ct, negative control.

patients differing by tumor grade, disease stage and success of debulking were examined by the Fisher's exact test (Table 4). Although CEAL1 expression was very low in one normal ovarian cDNA examined (Fig. 4), most ovarian (22/37; 60%) tumors had high levels of CEAL1 expression. High expression of CEAL1 was found more frequently in stage III and suboptimal debulking patients ($P = 0.005$ and $P = 0.023$, respectively). Differences in residual tumor size between the CEAL1-positive and -negative groups were found to be statistically significant by the Mann–Whitney test ($P = 0.024$). Median values of the residual tumor size were 0 and 3.5 cm for CEAL1-negative and CEAL1-positive tumors, respectively. Statistically significant association between CEAL1 status and patient age was not observed. Whereas CEAL1 expression was easily detectable in one normal breast cDNA examined (Fig. 4), most breast tumors (12/15; 80%) were negative. Moreover, the expression of CEAL1 in normal breast was found to be lower to cancer breast tissues. Due to the small number of samples, further statistical analysis of breast cancer patients was not attempted.

Table 4
Relationship between CEAL1 expression (negative or positive) status and other clinicopathological variables in ovarian cancer patients

Variable	Patients	No. of patients (%)		P value[a]
		CEAL1 Negative	CEAL1 Positive	
Stage[b]				
I/II	12	9 (75.0%)	3 (25.0%)	0.005
III	25	6 (24.0%)	19 (76.0%)	
Grade[c]				
I/II	16	7 (43.8%)	9 (56.3%)	0.47
II/III	18	5 (27.8%)	13 (72.2%)	
x	3			
Debulking[d]				
OD	16	10 (62.5%)	6 (37.5%)	0.023
SD	21	5 (23.8%)	16 (76.2%)	

[a] Fisher's exact test.
[b] International Federation of Gynaecology and Obstetrics (FIGO) staging system.
[c] I, well differentiated; II, moderately differentiated; III, poorly differentiated; x, status unknown.
[d] OD, optimal (0–1 cm); SD, suboptimal (>1 cm).

4. Discussion

We have cloned a novel human gene, CEAL1, which encodes for a carcinoembryonic antigen-like protein with a highly conserved domain of eukaryotic translation initiation factors. Bioinformatic approaches and EST analysis were first used to delineate the genomic organization of the gene and predict the putative mRNA coding region. RT–PCR and sequencing were used to verify the exons and the splice junctions. The 3′- and 5′-end of the gene were verified using RACE technology. The exon–intron splice junctions were further confirmed by comparing the genomic sequence with the EST or cDNA sequence obtained by RT–PCR and were found to be identical with the consensus splice sequences (Iida, 1990). The CEAL1 gene is formed of eight coding exons and seven intervening introns. One alternative splice variant, having one more exon was also identified.

The CEAL1 protein consists of 300 amino acids and shows ~30% homology with most other members of the human carcinoembryonic antigen family and ~60% homology with the recently identified *Mus musculus* similar to pregnancy-specific glycoprotein 23 (GenBank accession nos. XM_145421 and BAC33195). CEAL1 has a week homology at the gene/nucleotide level with other members of the human CEA family. Nevertheless a significant homology ~82% of the CEAL1, at the gene/nucleotide level, with the recently identified Mus musculus similar to pregnancy-specific glycoprotein 23 gene was found. Although the degree of homology was relatively low, this is not unusual since other members of the CEA family also share low homology (Zhou et al., 2001). In addition, the CEAL1 protein contains two Ig-like transmembrane domains which are present in most of the CEA proteins. Amino acid sequence analyses have revealed that the CEA gene family belongs to the immunoglobulin superfamily (IgSF) (Zhou et al., 2001). Among the CEA subgroup

members, BGP, CGM1 and CGM7 contain a hydrophobic transmembrane domain (TM) followed by either a short or long cytoplasmic domain (CYT), whereas CEA, NCA, CGM2 and CGM6 are attached to the cell membrane via a glycosyl phosphatidylinositol (GPI) linkage (Hammarstrom, 1999). The PSG subgroup of genes encode for glycoproteins that contain a short hydrophilic C-terminal tail, of which there are four different types, and are exported as secreted macromolecules (Teglund et al., 1994; Hammarstrom, 1999). For BGP and CGM1, splice variants were found to differ in the length of their cytoplasmic domains. The longer cytoplasmic domain contains two tyrosine residues which may be phosphorylated. It has been shown that the phosphorylated tyrosines are part of the modified immunoreceptor tyrosine based activation/inhibition motifs (ITAM/ITIM motifs) that might be involved in signal transduction (Hammarstrom, 1999).

CEAL1 protein contains one highly conserved domain of eukaryotic translation initiation factors (eIF-5A). Recently, the eIF-5A metabolic pathway has been found to correlate with transglutaminase expression and activity and it has been suggested that eIF-5A may play a role in cell growth and differentiation (Caraglia et al., 2001). eIF-5A was also found to be a cofactor for the functioning of the two viral transactivator proteins, Rex and Rev, which are required for the expression of the structural proteins of their respective viruses, T-cell leukemia virus and HIV. The interaction of eIF-5A with Rex and Rev promotes the association of specific viral mRNA with ribosomes, thereby increasing the stability or the translational rate of these mRNAs (Katahira et al., 1995).

There is evidence that eIF-5A can be modified through a reaction catalyzed by transglutaminases (TGase) in vitro. It has been reported that the GDP-bound form of tissue TGase (tTGase) associates with eIF-5A in HeLa cells and that the interaction is promoted by Ca^{2+}, Mg^{2+} and retinoic acid. In normal cells, tTGase gene expression is regulated by retinoids while the tTGase gene expression in cancer cells is regulated by growth factors (i.e. epidermal growth factor and transforming growth factor-β) and cytokines (i.e. interleukin-6) (Caraglia et al., 2001). Previous studies have demonstrated that tTGase activity is directly involved in some cellular activities, such as receptor mediated endocytosis, programmed cell death and tumor cell proliferation (Fesus and Thomazy, 1988; Caraglia et al., 2001). Beninati et al. have also revealed that tTGases could induce cell death and apoptosis through the inactivation of eIF-5A (Beninati et al., 1998).

The CEAL1 gene is constitutively expressed in many tissues suggesting that the encoded protein serves an important function. However, the amount of the CEAL1 transcripts seem to vary between tissues. CEA-related genes differ considerably in their expression patterns. There are basically four different expression patterns that can be recognized in normal adult tissues: (1) selective epithelial (CEA and CGM2); (2) granulocytic (CGM1 and CGM6); (3) selective syncytiotrophoblastic (PSGs); and (4) broad (BGP and NCA). CEA and CGM2 are mainly expressed in epithelial cells, especially those lining the gastrointestinal tract (i.e. colon, stomach, tongue and esophagus). CEA can also be found in squamous epithelial cells of the cervix, in secretory epithelia and duct cells of sweat glands and in epithelial cells of the prostate. The expression of CEA in these tissues does not generally begin until the early fetal period (weeks 9–14) and it seems to continue throughout life (Hammarstrom, 1999). Low levels of PSG expression have also been detected in a number of non-placental tissues, including the fetal liver, salivary gland, testis and myeloid cells. BGP and NCA, on the other hand, show a broader pattern of expression in normal adult tissues. Besides being expressed in intestinal epithelia, their transcripts are found in granulocytes, lymphocytes (BGP) and monocytes (NCA). In the case of BGP, their cDNA clones have also been isolated in the liver, pancreas, gallbladder, kidney, urinary bladder, prostate, cervix, endometrium, and probably in the endothelial cells of some organs (Thompson et al., 1991).

We here show preliminarily that CEAL1 transcripts are differentially regulated in ovarian and breast tumors. Furthermore, our data suggest that CEAL1 overexpression seems to be associated with clinically more aggressive ovarian tumors. Many studies have already shown that CEA, NCA and BGP are expressed in a number of tumors of epithelial origin such as colorectal carcinoma, lung adenocarcinoma and mucinous ovarian carcinoma and endometrial adenocarcinoma. CEA, however, seems to display a wider expression pattern compared to NCA and BGP; the antigen is also present in breast, pancreatic, gallbladder, urinary and small cell lung carcinoma, but not in acute lymphoblastic leukemia and hepatocellular carcinoma. CGM2 is limited to being expressed in some epithelial cancers, notably colorectal carcinoma, gastric carcinoma and mucinous ovarian carcinoma, while the granulocyte-associated molecules CGM6 and CGM1 are not detected in any of the investigated epithelial tumors. PSGs, on the other hand, are only found in hydatidform mole and choriocarcinoma. BGP has been observed to be down-regulated in colorectal and in hepatocellular carcinoma but up-regulated in gastric and squamous cell carcinoma of the lung. Likewise, for CGM2, down-regulation in colorectal carcinoma and up-regulation in gastric carcinoma have also been observed. In many cases the conclusion that a particular antigen is up- or down-regulated in tumors will depend on whether the comparisons were made against poorly or fully differentiated normal cells (Hammarstrom, 1999). Tumor CEA and normal CEA do not appear to have any differences at the genetic level. In the study by Fritsche and Mach, normal colon CEA was demonstrated to be indistinguishable from tumor CEA by several immunological, physicochemical and chemical criteria (Fritsche and Mach, 1977). The elevated CEA serum concentrations found in colon, breast and lung

adenocarcinomas have made CEA one of the most widely used human tumor markers (Shively and Beatty, 1985; Duffy, 2001).

The development of new prognostic biomarkers for ovarian cancer may help to improve the diagnostic/prognostic power of CA125. A recent study suggested that CA125 could be used for prediction of optimal primary tumor cytoreduction in stage III tumors (Chi et al., 2000). Kallikreins and SR-A1, located in the same chromosomal area as CEAL1, were also recently found to have significant prognostic/diagnostic value in ovarian cancer (Yousef et al., 2001). Since CEAL1 expression levels are significantly different in patients with optimal and suboptimal cytoreduction, and in patients with early and late stages of the disease (Table 3), CEAL1 might be useful as a new differentiation marker.

References

Ashworth, L.K., Batzer, M.A., Brandriff, B., Branscomb, E., de Jong, P., Garcia, E., Garnes, J.A., Gordon, L.A., Lamerdin, J.E., Lennon, G., 1995. An integrated metric physical map of human chromosome 19. Nat. Genet. 11, 422–427.

Bairoch, A., Bucher, P., Hofmann, K., 1997. The PROSITE database, its status in 1997. Nucleic Acids Res. 25, 217–221.

Beninati, S., Nicolini, L., Jakus, J., Passeggio, A., Abbruzzese, A., 1995. Identification of a substrate site for transglutaminases on the human protein synthesis initiation factor 5A. Biochem. J. 305, 725–728.

Beninati, S., Gentile, V., Caraglia, M., Lentini, A., Tagliaferri, P., Abbruzzese, A., 1998. Tissue transglutaminase expression affects hypusine metabolism in BALB/c 3T3 cells. FEBS Lett. 437, 34–38.

Bonet, M., Steel, M., Warnow, T., Yooseph, S., 1998. Better methods for solving parsimony and compatibility. J. Comput. Biol. 5, 391–407.

Brendel, V., Bucher, P., Nourbakhsh, I.R., Blaisdell, B.E., Karlin, S., 1992. Methods and algorithms for statistical analysis of protein sequences. Proc. Natl. Acad. Sci. USA 89, 2002–2006.

Caraglia, M., Marra, M., Giuberti, G., D'Alessandro, A.M., Budillon, A., del Prete, S., Lentini, A., Beninati, S., Abbruzzese, A., 2001. The role of eukaryotic initiation factor 5A in the control of cell proliferation and apoptosis. Amino Acids 20, 91–104.

Chi, D.S., Venkatraman, E.S., Masson, V., Hoskins, W.J., 2000. The ability of preoperative serum CA-125 to predict optimal primary tumor cytoreduction in stage III epithelial ovarian carcinoma. Gynecol. Oncol. 77, 227–231.

Day, T.G. Jr, Gallager, H.S., Rutledge, F.N., 1975. Epithelial carcinoma of the ovary: prognostic importance of histologic grade. Natl. Cancer Inst. Monogr. 42, 15–21.

Duffy, M.J., 2001. Carcinoembryonic antigen as a marker for colorectal cancer: is it clinically useful? Clin. Chem. 47, 624–630.

Fesus, L., Thomazy, V., 1988. Searching for the function of tissue transglutaminase: its possible involvement in the biochemical pathway of programmed cell death. Adv. Exp. Med. Biol. 231, 119–134.

Fritsche, R., Mach, J.P., 1977. Isolation and characterization of carcinoembryonic antigen (CEA) extracted from normal human colon mucosa. Immunochemistry 14, 119–127.

Hammarstrom, S., 1999. The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues. Semin. Cancer Biol. 9, 67–81.

Iida, Y., 1990. Quantification analysis of 5′-splice signal sequences in mRNA precursors. Mutations in 5′-splice signal sequence of human beta-globin gene and beta-thalassemia. J. Theor. Biol. 145, 523–533.

Katahira, J., Ishizaki, T., Sakai, H., Adachi, A., Yamamoto, K., Shida, H., 1995. Effects of translation initiation factor eIF-5A on the functioning of human T-cell leukemia virus type I Rex and human immunodeficiency virus Rev inhibited trans dominantly by a Rex mutant deficient in RNA binding. J. Virol. 69, 3125–3133.

Nap, M., Mollgard, K., Burtin, P., Fleuren, G.J., 1988. Immunohistochemistry of carcino-embryonic antigen in the embryo, fetus and adult. Tumour Biol. 9, 145–153.

Oikawa, S., Nakazato, H., Kosaki, G., 1987. Primary structure of human carcinoembryonic antigen (CEA) deduced from cDNA sequence. Biochem. Biophys. Res. Commun. 142, 511–518.

Oikawa, S., Kuroki, M., Matsuoka, Y., Kosaki, G., Nakazato, H., 1992. Homotypic and heterotypic Ca(++)-independent cell adhesion activities of biliary glycoprotein, a member of carcinoembryonic antigen family, expressed on CHO cell surface. Biochem. Biophys. Res. Commun. 186, 881–887.

Olsen, A., Teglund, S., Nelson, D., Gordon, L., Copeland, A., Georgescu, A., Carrano, A., Hammarstrom, S., 1994. Gene organization of the pregnancy-specific glycoprotein region on human chromosome 19: assembly and analysis of a 700-kb cosmid contig spanning the region. Genomics 23, 659–668.

Proudfoot, N.J., Cheng, C.C., Brownlee, G.G., 1976. Sequence analysis of eukaryotic mRNA. Prog. Nucleic Acid Res. Mol. Biol. 19, 123–134.

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463–5467.

Scorilas, A., Kyriakopoulou, L., Yousef, G.M., Ashworth, L.K., Kwamie, A., Diamandis, E.P., 2001a. Molecular cloning, physical mapping, and expression analysis of a novel gene, BCL2L12, encoding a proline-rich protein with a highly conserved BH2 domain of the Bcl-2 family. Genomics 72, 217–221.

Scorilas, A., Yousef, G.M., Jung, K., Rajpert-De Meyts, E., Carsten, S., Diamandis, E.P., 2001b. Identification and characterization of a novel human testis-specific kinase substrate gene which is downregulated in testicular tumors. Biochem. Biophys. Res. Commun. 285, 400–408.

Shively, J.E., Beatty, J.D., 1985. CEA-related antigens: molecular biology and clinical significance. Crit. Rev. Oncol. Hematol. 2, 355–399.

Teglund, S., Olsen, A., Khan, W.N., Frangsmyr, L., Hammarstrom, S., 1994. The pregnancy-specific glycoprotein (PSG) gene cluster on human chromosome 19: fine structure of the 11 PSG genes and identification of 6 new genes forming a third subgroup within the carcinoembryonic antigen (CEA) family. Genomics 23, 669–684.

Thompson, J.A., Grunert, F., Zimmermann, W., 1991. Carcinoembryonic antigen gene family: molecular biology and clinical perspectives. J. Clin. Lab. Anal. 5, 344–366.

Yousef, G., Scorilas, A., Magklara, A., Soosaipillai, A., Diamandis, E.P., 2000a. The PRSS6 gene, encoding for the stratum corneum chymotryptic enzyme is a new member of the human kallikrein gene family-Genomic characterization, mapping, tissue expression and hormonal regulation. Gene 258, 24–31.

Yousef, G.M., Scorilas, A., Diamandis, E.P., 2000b. Genomic organization, mapping, tissue expression, and hormonal regulation of trypsin-like serine protease (TLSP PRSS20), a new member of the human kallikrein gene family. Genomics 63, 88–96.

Yousef, G.M., Scorilas, A., Jung, K., Ashworth, L.K., Diamandis, E.P., 2001. Molecular cloning of the human kallikrein 15 gene (KLK15). Up-regulation in prostate cancer. J. Biol. Chem. 276, 53–61.

Zhou, G.Q., Zhang, Y., Hammarstrom, S., 2001. The carcinoembryonic antigen (CEA) gene family in non-human primates. Gene 264, 105–112.