High expression of a cytokeratin-associated protein in many cancers

Kristi A. Egland*, Xiu Fen Liu†, Stephen Squires†, Satoshi Nagata†, Yan-Gao Man‡, Tapan K. Bera†, Masanori Onda†, James J. Vincent†, Robert L. Strausberg§, Byungkook Lee†, and Ira Pastan†¶

*South Dakota Health Research Foundation–Signal Transduction Institute, University of South Dakota School of Medicine and Sioux Valley Hospitals and Health System, Sioux Falls, SD 57105; †Laboratory of Molecular Biology, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892; ‡Department of Gynecologic and Breast Cancer Pathology, Armed Forces Institute of Pathology, Washington, DC 20306; and §J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850

Contributed by Ira Pastan, February 15, 2006

We have described previously a cDNA library made from membrane-bound polysomal mRNA prepared from breast and prostate cancer cell lines. The library is highly enriched for cDNAs encoding membrane proteins, secreted proteins, and cytokeratins. To characterize this library, 25,277 cDNA clones were sequenced and aligned with various databases; 1,439 clones did not align with known genes. From this set of clones we identified a previously uncharacterized gene encoding a 334-aa protein. Although protein structural motif prediction programs indicate that the gene encodes a membrane protein comprising a signal sequence, a series of leucine-rich repeats, and a single transmembrane domain with a cytoplasmic tail, confocal microscopy of MCF7 breast cancer cells demonstrates that the protein is not directly associated with the plasma membrane or intracellular membranes but instead colocalizes with intermediate filaments and cytokeratins within the cell. Immunofluorescence studies also show that protein expression is increased greatly in mitotic MCF7 cells, and immunohistochemistry demonstrates its expression in human breast cancer cells. Analysis of mRNA levels in 25 different normal tissues by RT-PCR shows that this gene is expressed highly in normal prostate and salivary gland, very weakly in colon, pancreas, and intestine, and not at all in other tissues. RT-PCR studies on human cancer samples show that the RNA is expressed highly in many cancer cell lines and cancer specimens, including 26 of 33 human breast cancers, 3 of 3 prostate cancers, 3 of 3 colon cancers, and 3 of 3 pancreatic cancers. We name the protein CAPC, cytokeratin-associated protein in cancer.

breast cancer | colon cancer | intermediate filaments | mitosis | prostate cancer

Several different experimental and *in silico* approaches have been developed to identify new genes and proteins that could be useful in the diagnosis or treatment of cancer (1–3). To identify genes that encode membrane or membrane-associated proteins that are present in breast and prostate cancers but have limited expression in normal and essential organs, we developed a molecular approach in which a cDNA library is generated from membrane-associated polyribosomal RNA that encodes membrane and secreted proteins and cytokeratins. The mRNA was isolated from four breast cancer cell lines, one normal breast cell line, and one prostate cancer cell line. To remove housekeeping genes and genes expressed in vital organs, the membrane-associated polyribosomal cDNA library (MAPcL) was subtracted with RNA from normal brain, liver, lung, kidney, and muscle. To determine which genes are represented in the subtracted MAPcL, one sequencing reaction was performed on the 5′ end of 15,581 clones (1). From this initial set of sequences, we identified a breast cancer gene designated BASE (breast cancer and salivary gland expression) (1). In the present work we extended this analysis by performing a 5′ sequencing reaction on an additional 9,696 clones. Analysis of 25,277 sequences obtained from the library shows that the most abundant gene is *kallikrein 3*, which encodes the secreted protein prostate-specific antigen and is represented by 112 cDNA clones. In this work we report the properties of another cDNA that is very abundant in the library. It is derived from a previously uncharacterized gene located on chromosome 9 and is expressed in many cancers, including breast, prostate, colon, and pancreas. An analysis of the protein sequence suggests that the protein product of this gene, cytokeratin-associated protein in cancer (CAPC), is a membrane protein, but experimental results show that it is associated with cytokeratins and intermediate filaments and that its expression is increased in mitotic MCF7 breast cancer cells.

Results

Identification of the *CAPC* Gene. A breast and prostate cancer cDNA library enriched with genes that encode membrane and secreted proteins was generated from membrane-associated polyribosomal RNA isolated from four breast cancer cell lines (MCF7, ZR-75-1, SK-BR-3, and MDA-MB-231), one telomerase-immortalized normal breast cell line (hTERT-HME1), and a prostate cancer cell line (LNCaP). The MAPcL was subtracted with RNA derived from five normal libraries: brain, lung, liver, kidney, and skeletal muscle. To determine which genes are represented in the subtracted MAPcL, a single sequencing reaction from the 5′ end was performed on a total of 25,277 clones. Of the 25,277 sequences, the most abundant gene is *kallikrein 3*, which encodes prostate-specific antigen and is represented by 112 cDNA clones.

By aligning the 25,277 sequences with the human genome, unique transcripts for 14,397 genes were assembled and classified as either known or unknown by using a BLAST analysis; 7,813 transcripts represent known genes, and 6,584 transcripts represent unknown genes. From the 6,584 unknown genes, we selected 1,439 that had no expression in essential normal tissues as determined by overlap with expressed sequences tags (dbEST) (4, 5). We then used several criteria to select genes for further study: (*i*) those that contained an ORF with a predicted transmembrane domain; (*ii*) those expressed as ESTs derived from breast cancers and/or prostate cancers; and (*iii*) those with limited expression in essential normal tissues.

Using the criteria stated above, we identified a gene located on chromosome 9, designated as *CAPC*, which was very abundant and represented by 34 clones in the MAPcL. This number of clones is approximately one-third the level of *kallikrein 3*, the most abundant cDNA in the library. One clone, pKAE68h05,

Conflict of interest statement: No conflicts declared.

Abbreviations: CAPC, cytokeratin-associated protein in cancers; dbEST, expressed sequence tags database; MAPcL, membrane-associated polyribosomal cDNA library.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. DQ355157).

¶To whom correspondence should be addressed at: Laboratory of Molecular Biology, National Cancer Institute, 37 Convent Drive, Room 5106, Bethesda, MD 20892-4264. E-mail: pastani@mail.nih.gov.

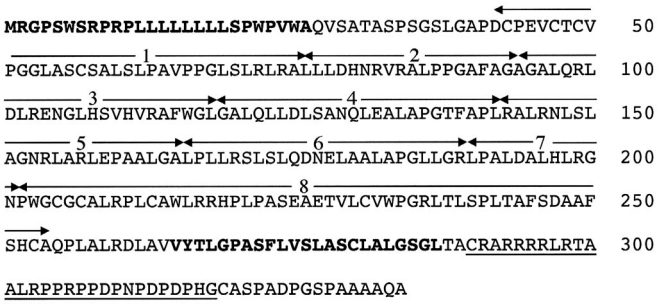

Fig. 1. CAPC amino acid sequence. Proposed signal peptide and transmembrane regions are shown in bold. Leucine-rich repeats are numbered 1–8 with N-terminal and C-terminal leucine-rich repeats labeled 1 and 8, respectively. The peptide used to raise polyclonal antibodies is underlined.

was selected for further analysis. The full-length insert of pKAE68h05 was sequenced and contains a complete ORF. This clone ends with the stop codon (TGA) and has no UTR at the 3′ end because when the MAPcL was generated, a NotI site (5′-GCGGCCGC-3′) was used as a cloning site on the 3′ end of the cDNA fragments, and unfortunately the *CAPC* gene contains a native NotI site adjacent to the stop codon. Therefore, pKAE68h05 did not contain the 3′ UTR of *CAPC*. A full-length cDNA was obtained by analyzing various ESTs from the dbEST and designing a primer complementary to the distal end of the putative 3′ UTR of *CAPC*. This primer and a primer complementary to the region upstream of *CAPC* were used to amplify the full-length transcript from breast cDNA. The resulting product was sequenced.

As shown in Fig. 1, *CAPC* has an ORF encoding a 334-aa protein. Analysis of the protein sequence by using SMART (6) predicted that the protein contains a signal sequence, a series of leucine-rich repeats, a single transmembrane domain, and a short intracellular portion. Its high level of expression in the library derived from cancer cell lines indicated that the new gene might be expressed frequently in human cancers and was worthy of further investigation. The recently renamed *LRTM1* (NCBI accession no. NM_020678) and *LRTM2* (NCBI accession no. NM_001039029) genes appear to be paralogs of *CAPC* because they are highly homologous and have genomic organizations similar to that of *CAPC*. Nothing is known about these genes except that they encode leucine-rich repeats and a predicted transmembrane sequence organized in a way similar to that of *CAPC*.

Expression of *CAPC* in Cell Lines, Breast Cancers, and Normal Tissues. Because the MAPcL was generated from pooled membrane-associated polyribosomal RNA derived from six cell lines, we determined which of the cell lines express *CAPC* using RT-PCR analysis (Fig. 2*A*). Total RNA from each of the six cell lines was used as a template to generate cDNA. The specific primers used for PCR are located in separate exons of *CAPC* and amplify a 430-bp fragment. *CAPC* was expressed in the five cancer cell lines examined (lanes 1, 4, 5, 7, and 8) but not the normal breast cell line hTERT-hME1 (data not shown). As a control for the quality of the generated cDNA, separate PCR analyses were performed by using primers to *actin* (Fig. 2*A*). To examine the frequency of *CAPC* expression in other cancer cell lines, RT-PCR was carried out by using RNA from many other cell lines. The data in Fig. 2*A* show that *CAPC* mRNA is expressed abundantly in many other cell lines, including two of three prostate cancer lines (lanes 1–3), one ovarian line (lane 10), a melanoma line (FEMX-I), some leukemia and lymphoma cell lines (HUT102, U937, Daudi, JD 38, and Raji),

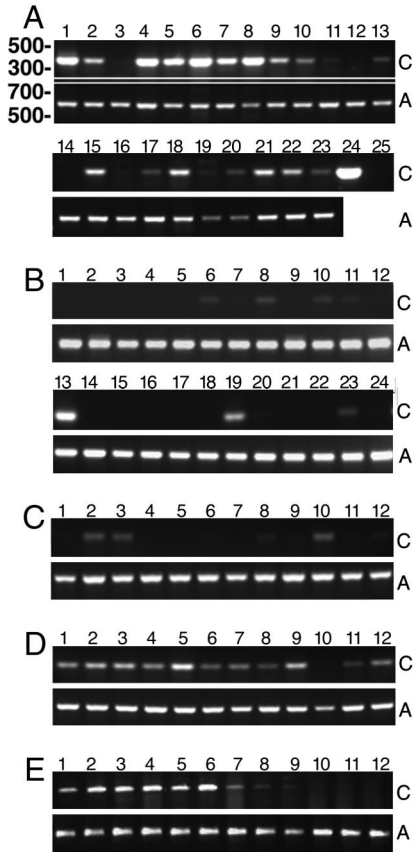

Fig. 2. Expression of *CAPC* in cell lines, cancers, and normal tissues. The positions of size markers are indicated on the left of the first lane of *A*. On right of each lane is a label: C for *CAPC* or A for *actin*. (*A*) Expression of *CAPC* in cancer cell lines. Lanes: 1, LNCaP; 2, PC-3; 3, DU 145; 4, CRL1500; 5, MCF7; 6, BT-474; 7, MDA-MB-231; 8, SK-BK-3; 9, MDA-MB-468; 10, OVCAR, 11, FEMX-I; 12, HUT102; 13, U937; 14, Daudi; 15, JD 38; 16, Raji; 17, A 172; 18, IMR-32; 19, COLO 205; 20, LOVO; 21, SW403; 22, SW480; 23, SW620; 24, plasmid for positive control; and 25, water for negative control. Expression levels were determined by RT-PCR using total RNA as a template for cDNA synthesis or purchased cDNAs. PCR was performed by using primers to *CAPC* that amplify a 430-bp fragment. The PCR products were analyzed on a 1.5% agarose gel with ethidium bromide staining. Separate PCRs were done by using *actin* primers, which amplify a 640-bp fragment to verify the quality of the generated cDNA. (*B*) Expression of *CAPC* in normal human tissues. PCRs were performed by using a rapid-scan gene expression panel containing cDNA samples from 24 different normal tissues: 1, brain; 2, heart; 3, kidney; 4, spleen; 5, liver; 6, colon; 7, lung; 8, small intestine; 9, muscle; 10, stomach; 11, testis; 12, placenta; 13, salivary gland; 14, thyroid; 15, adrenal gland; 16, pancreas; 17, ovary; 18, uterus; 19, prostate; 20, skin; 21, peripheral blood lymphocyte; 22, bone marrow; 23, fetal brain; and 24, fetal liver. (*C*) Expression of *CAPC* in normal breast. (*D*) Expression of *CAPC* in breast cancers. (*E*) Expression of *CAPC* in various cancers: colon (lanes 1–3), pancreas (lanes 4–6), and lung (lanes 7–12).

a glioblastoma line (A 172), a neuroblastoma line (IMR-132), and five of five colon cancer cell lines.

Fig. 2*B* shows expression in normal tissues. *CAPC* is expressed in a very limited set of normal tissues with high expression detected only in salivary gland and prostate and very weak expression in colon, pancreas, intestine, stomach, and fetal brain. This limited expression in normal tissues was supported by our analysis of the EST database (data not shown).

Because *CAPC* was identified in mRNA made from a pool containing mainly breast cancer cell line mRNAs, we next examined its expression in human breast cancers and normal breast. Fig. 2*D* demonstrates relatively strong expression in 12 of

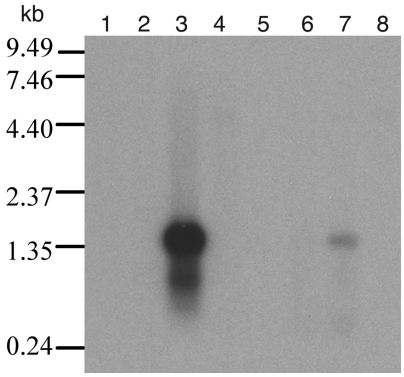

Fig. 3. Northern blot analysis of *CAPC* transcripts. The lanes contain 2 μg of mRNA from the following tissues: 1, spleen; 2, thymus; 3, prostate; 4, testis; 5, uterus; 6, small intestine; 7, colon; and 8, peripheral blood lymphocytes.

13 breast cancer samples [Human Breast Cancer Rapid-Scan panel (OriGene Technologies, Rockville, MD)]. High expression was also detected in 15 of 21 breast cancer samples obtained from the Cooperative Human Tissue Network (data not shown). In total, *CAPC* was expressed in 26 of 33 or 79% of the breast tumors analyzed. Fig. 2*C* shows weak expression in 5 of 12 normal breast samples. Separate PCR analyses were performed by using *actin* primers to verify the quality of the generated cDNA.

Fig. 2*E* shows that *CAPC* was also expressed in other cancers, including three of three colon, three of three pancreatic, and a few lung cancer specimens. In addition, *CAPC* was expressed in three of three prostate cancer samples, determined by performing RT-PCR analysis (data not shown).

To determine the transcript size of *CAPC* and to verify further the expression pattern of *CAPC* in normal tissues, a Northern blot analysis was performed. Membranes containing 2 μg of mRNA derived from 17 different normal tissues were probed with a 430-bp *CAPC* fragment. Results from eight tissues are shown in Fig. 3. Two bands were observed in prostate, with the most abundant transcript at $\approx$1.5 kb and a less abundant transcript at $\approx$1.2 kb (Fig. 3, lane 3). The predicted size of the *CAPC* cDNA is 1,212 bp (Fig. 1), which corresponds to the 1.5-kb transcript with the addition of poly(A). A weak 1.5-kb transcript was also observed in colon (Fig. 3, lane 7), and we have also observed a single band expressed in salivary gland in a separate blot (data not shown). *CAPC* transcripts were not detected in the other 14 mRNA samples (Fig. 3 and data not shown), including small intestine, testis, and pancreas. The failure to detect mRNA in intestine and pancreas probably reflects the very-low-level expression detected by RT-PCR.

Characterization of the CAPC Protein. As mentioned earlier, the nucleotide sequence of *CAPC* predicts that it encodes a 334-aa protein, and an analysis of the protein sequence by using the program SMART indicated that the protein consisted of a signal sequence (amino acids 1–26), a series of leucine-rich repeats (amino acids 42–254), a single transmembrane domain (amino acids 265–287), and a C-terminal peptide (Fig. 1). To characterize the protein encoded by *CAPC* and to determine its cellular location, a polyclonal antibody was made in rabbits against a peptide as described in *Materials and Methods*. The location of the peptide is shown in Fig. 1. Fig. 4 shows an immunoblot in which the antibody specifically detects an intense 37-kDa band in MCF7 cells that stably express a *CAPC* plasmid (lane 1) and a lighter band in untransfected MCF7 cells that are known to contain *CAPC* mRNA (lane 3). No 37-kDa band was detected in two control cell lines: 293T, which does not express *CAPC* mRNA (K.A.E., unpublished data) and the normal breast line MCF10A. Furthermore, the 37-kDa band was not detected by serum obtained before immunization (data not shown). Besides the specific 37-kDa band, a weak 25-kDa nonspecific band was detected in all cell lines examined whether or not they contained *CAPC* mRNA. In addition, a portion of the antiserum was purified on a column containing the peptide used to make the antibody, and the same 37-kDa band was observed in MCF7 cells (data not shown). These results establish the specificity of the antibody for the 37-kDa protein.

Cellular Localization of CAPC. We first examined the cellular location of CAPC by confocal immunofluorescence microscopy by using HeLa cells transfected with a plasmid (pcDNA3-*CAPC*-myc); 2 days after transfection the cells were fixed, permeabilized, and stained with an anti-myc antibody. Fig. 4*B* shows that CAPC-myc protein (green) is present in an intracellular location that is adjacent to filamentous actin fibers just beneath the cell membrane. CAPC-myc expression was often concentrated in the spreading edge of a cell (as shown by the arrow). In mitotic cells (Fig. 4*C*), CAPC-myc showed a clear multilayer fiber structure underneath filamentous actin (red). We have observed a similar staining pattern with CAPC-EGFP expression in HeLa cells (data not shown). To examine the location of the endogenous protein, MCF7 cells were stained with anti-CAPC antiserum with or without purification on a CAPC peptide column. The staining pattern (green) in both interphase cells (Fig. 4 *D*, *E*, *G*, and *H*) and in mitotic cells (Fig. 4 *D*, *E*, *J*, and *K*) is very similar to that observed in HeLa cells expressing CAPC-myc. In interphase cells, CAPC is expressed mostly in the spreading edge of the cell underneath actin fibers (Fig. 4 *D* and *E*, arrows). The specificity of staining of MCF7 cells was confirmed by a peptide competition experiment, in which addition of the CAPC peptide completely removed the signal of CAPC protein (data not shown). This result indicates that anti-CAPC specifically recognizes the CAPC protein as observed by confocal microscopy.

We then carried out a series of colocalization experiments to identify the precise location of CAPC. Because the sequence of CAPC indicated that it was a membrane protein, we initially examined markers for membrane proteins and found that CAPC did not colocalize with any of the proteins we initially examined: *trans*-Golgi marker, TGN46 (7); endoplasmic reticulum marker, disulfide isomerase (8); and plasma membrane and endocytic compartment marker, transferrin receptor (9). The close association with actin and the filamentous appearance, particularly in mitotic cells, suggested that it might colocalize with intermediate filaments (10). As shown in Fig. 4 *F–K*, CAPC colocalizes with most of the cytokeratin signal in both interphase cells and mitotic cells (Fig. 4 *F–K*). CAPC colocalization with cytokeratin is most striking in the cell spreading edge as shown by an arrow in Fig. 4*H*. Most interestingly, in mitotic cells, from prophase to telophase and cytokinesis (data not shown), CAPC showed a 5- to 10-fold elevated signal intensity as compared with interphase cells, where the actin signal (Fig. 4 *D* and *E*) and the cytokeratin signal (data not shown) did not increase in fluorescence intensity.

CAPC in Human Breast Cancers. Fig. 5 shows the localization of CAPC in a few typical breast cancers by using the affinity-purified anti-CAPC antibody. Note that the cells in normal ducts (arrows) show no distinct CAPC immunostaining, whereas cells in lobular neoplasia (*a*), invasive ductal carcinoma (*b* and *c*), and metastatic lesions (*d*) are positive for CAPC. Consistent with our observation in cell culture, we also observed a high level of CAPC expression in mitotic cells in metastatic lesions (data not shown).

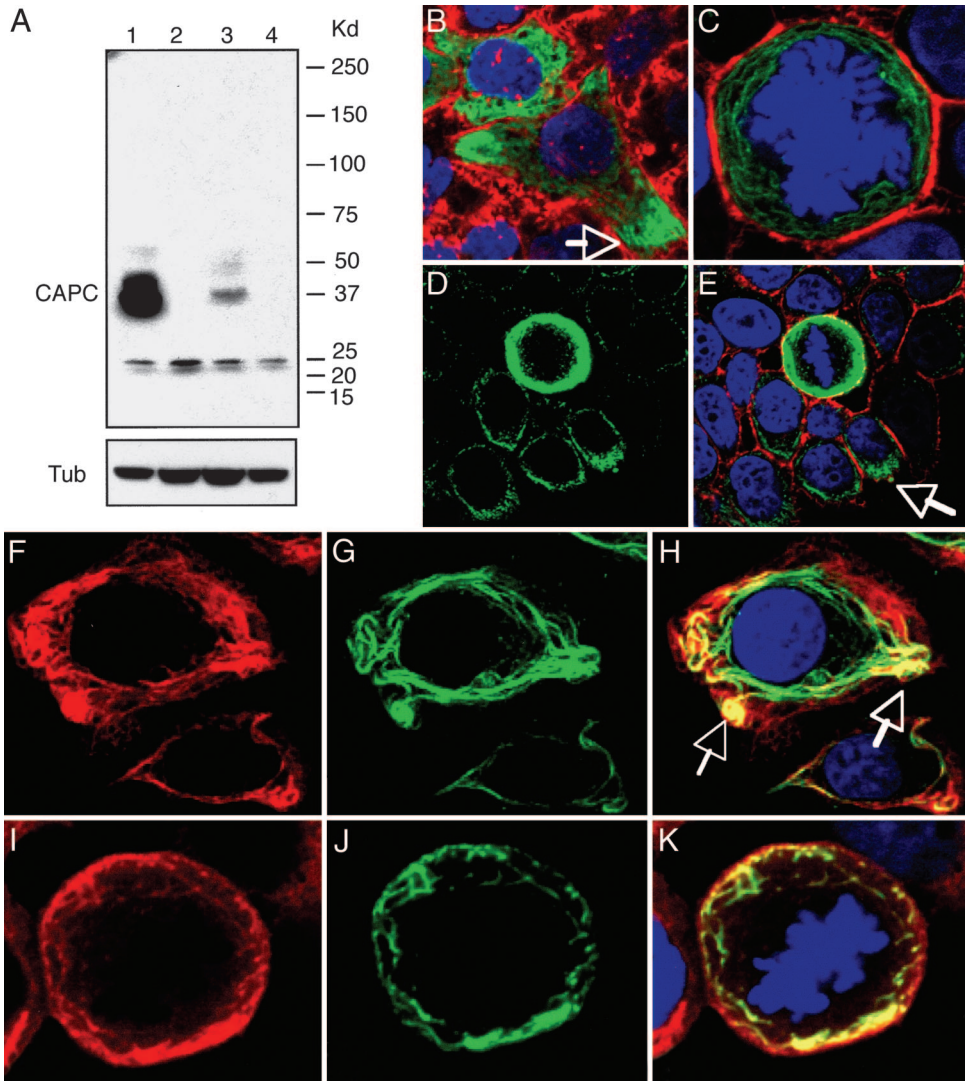

Fig. 4. CAPC protein expression and localization by confocal microscopy. (*A*) Western blot analysis of CAPC expression: 10 μg of lysate from CAPC-pcDNA3-transfected MCF7 cells, stable line (lane 1), or 20 μg of cell lysate from 293T (lane 2), MCF7 (lane 3), and MCF10A (lane 4) was analyzed by Western blotting using rabbit anti-CAPC antiserum (1:5,000; Sigma) as a loading control. (*B* and *C*) pcDNA3-CAPC-myc was transfected into HeLa cells. After 48 h of expression, cells were stained with anti-myc (green) and tetramethylrhodamine B isothiocyanate-labeled phalloidin (red), and merged images were stained with DAPI. (*B*) Interphase. (*C*) Metaphase. (*D–K*) MCF7 cells were stained with anti-CAPC (green in *D*, *G*, and *J*) and costained for actin (red in *E*) or with pan-cytokeratin antibody (red in *F* and *I*) and merged with DAPI (blue in *E*, *H*, and *K*). *D* and *E* are from the same field showing the level of expression of mitotic cells compared with interphase cells. *F–H* are the same fields showing the colocalization of CAPC with cytokeratin in interphase cells. *I–K* show the same cells for colocalization of the two proteins in mitosis. Arrows point to the cell spreading edge.

Discussion

We have found and characterized a previously undescribed gene, *CAPC*, which is expressed highly in breast cancer, pancreatic cancer, and colon cancer and many cancer cell lines. Expression is absent or low in normal tissues except for salivary gland and prostate, where it is also expressed highly. The existence of this gene was predicted from analysis of cDNAs prepared from polysome-associated mRNAs obtained from several breast cancers and one prostate cancer cell line.

The protein sequence indicates that most of the protein is made of leucine-rich repeats, but it also contains what are predicted to be a signal peptide at the N terminus and a transmembrane helix near the C terminus. Thus, the protein is predicted to be a transmembrane protein, as might be expected for the product of a gene that is represented abundantly in the MAPcL. Surprisingly, cellular localization studies using a specific antibody to CAPC clearly showed that CAPC is not associated with the plasma membrane or other intracellular membranes but instead is located in the cytoplasm and partially colocalizes with proteins stained by a pan-cytokeratin antibody. Because this location did not fit with the sequence analysis, we carried out several confirmatory experiments in which CAPC-myc or CAPC-EGFP was transfected into HeLa cells. The results of all of these experiments were in complete agreement with the localization results in MCF7 cells when using antibodies to CAPC.

Intermediate filaments are made up of cytokeratins, of which there are many different types (10). We used a pan-cytokeratin antibody that reacts with cytokeratins in MCF7 cells to show that CAPC colocalizes with cytokeratins. The specific cytokeratin with which CAPC interacts has not yet been identified. CAPC has an entirely different sequence than cytokeratins and is not a member of the cytokeratin family. Cytokeratins have an important role in the control of cell shape and cell–cell adhesion.

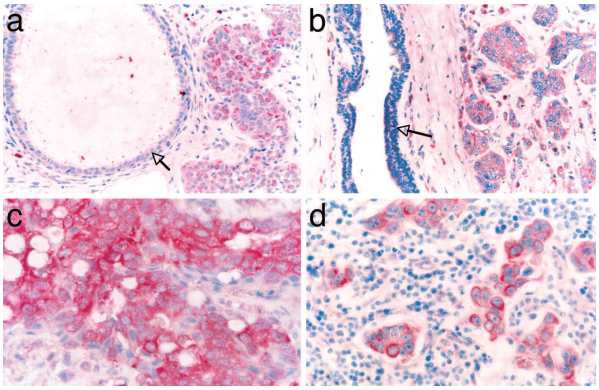

Fig. 5. Expression of CAPC in human breast specimens. Paraffin-embedded tissue sections from patients with coexisting preinvasive and invasive breast tumors were immunostained with the anti-CAPC antibody at a 1:50 dilution, by using an avidin–biotin complex detection kit and an alkaline phosphatase red substrate kit. Note that the cells in normal ducts (arrows) show no distinct CAPC immunostaining, whereas cells in lobular neoplasia (*a*), invasive ductal carcinoma (*b* and *c*), and metastatic lesions (*d*) are positive for CAPC. (Magnification: ×300.)

Recently, cytokeratin 18 expression has been shown to be correlated inversely with invasiveness and metastatic potential (11, 12). CAPC could affect cancer cell behavior by modifying some function of intermediate filaments. Cytokeratin intermediate filament networks are reorganized dynamically during the cell cycle (13). Given that the CAPC protein signal is intensified during the mitotic period, CAPC could also be involved in the disassembly and reassembly process of the intermediate filament network.

Materials and Methods

Primers. The primers used were KAE68h05C-For (5′-CTCTAGACGCGCTGCACCTGC-3′) and KAE68h05C-Rev (5′-CAGGCTTGGGCGGCAGCGG-3′). The primers were synthesized by Lofstrand Laboratories (Gaithersburg, MD). The primers actin-For (5′-GCATGGGTCAGAAGGAT-3′) and actin-Rev (5′-CCAATGGTGATGACCTG-3′) were purchased from OriGene Technologies.

Cell Culture. MCF7, SK-BR-3, ZR-75-1, MDA-MB-231, and LNCaP cell lines were maintained as recommended by American Type Culture Collection. The hTERT-HME1 cell line (Clontech) was maintained according to the mamufacturer's instructions.

Northern Blot Hybridization. The human MTN blot (multiple-tissue Northern blot) was purchased from Clontech. The blot contains 2 μg of mRNA per lane from eight different normal human tissues. The RNA was run on a formaldehyde/1.0% agarose gel and transferred to a nylon membrane. The 430-bp *CAPC* probe used for hybridization was generated by PCR amplification of the MAPcL clone pKAE68h05 by using primers KAE68h05C-For and KAE68h05C-Rev. The 430-bp PCR fragment was isolated by agarose gel electrophoresis, and the DNA fragment was purified by using the Qiagen Gel Extraction kit (Qiagen, Alameda, CA). The DNA fragment was labeled with ^{32}P by random primer extension (Lofstrand Laboratories), and the hybridization conditions were as described previously (14). The human MTN blot was exposed to film for 6 h.

RT-PCR and Rapid-Scan Gene Expression Panel Analysis. Total RNA was isolated from frozen breast tumor samples acquired from the Cooperative Human Tissue Network and tissue culture cell lines by using the StrataPrep Total RNA Miniprep kit (Stratagene) according to the manufacturer's instructions. To generate single-stranded cDNAs, total RNA (5 μg) was used with the First-Strand cDNA synthesis kit by using random hexamer priming according to the manufacturer's instructions (Amersham Pharmacia Biosciences). PCRs used to amplify *CAPC* were performed by using the following protocol: initial denaturation at 94°C for 3 min, 35 cycles of denaturation at 94°C for 1 min, annealing at 65°C for 1 min, and elongation at 72°C for 1 min with a final 5-min extension at 72°C. The PCR buffer conditions used to amplify *CAPC* were optimized by using the PCR Optimizer kit (Invitrogen) and buffer A with 10% DMSO. The same PCR conditions were used with the Human Rapid-Scan and Human Breast Cancer Rapid-Scan gene expression panels and the Human Multiple-Tissue cDNA panels I and II by BD Biosciences.

Generation of Polyclonal Antibodies Against CAPC. Polyclonal antibodies were generated in rabbits against a CAPC peptide Ac-CRARRRRLRTAALRPPRPPDPNPDPDPHG-amide coupled to keyhole limpet hemocyanin and after two immunizations boosted with the peptide coupled to ovalbumin. Bleed 4 was used for studies reported here. A portion of bleed 4 was affinity-purified on a column to which the peptide was conjugated.

Western Blots. MCF7 or 293T cells were washed once with PBS and then disrupted in lysis buffer (150 mM NaCl/50 mM Tris, pH 7.5/0.5% Triton X-100/0.5% sodium deoxycholate/1% Nonidet P-40/2 mM EDTA/10 μg/ml leupeptin/2 μg/ml aprotinin/20 μg/ml phenylmethylsulfonyl fluoride). Cell extracts were separated by SDS/PAGE. For Western blot analysis, gels were blotted to a poly(vinylidene difluoride) membrane. Blots were incubated sequentially with 5% milk for 1 h in Tris-buffered saline plus 0.1% Triton X-100, in primary antibody (anti-CAPC antiserum, 1:5,000) for 1.5 h and then in secondary antibody for 1 h. After blotting, signals were detected by enhanced chemiluminescence (Amersham Pharmacia Biosciences).

Immunofluorescence and Confocal Microscopy. Cells grown on a cover glass were washed with PBS and fixed with 4% paraformaldehyde for 30 min. After permeabilization with 0.5% Triton X-100 for 10 min, the samples were then blocked with 3% BSA for 30 min. Samples were exposed to primary antibodies (1:1,000 anti-myc, 1:1,000 anti-CAPC antiserum, or 1.5 μg/ml affinity-purified anti-CAPC antibody) for 1–2 h in PBS/BSA followed by Alexa Fluor-conjugated secondary antibody staining for 1 h. When filamentous actin was also stained, tetramethylrhodamine B isothiocyanate-labeled phalloidin (Sigma) was added with the secondary antibody incubation. Cells were washed four times with PBS after secondary antibody staining and then mounted on glass slides by using mounting medium with DAPI (VECTASHIELD; Vector Laboratories). All images were obtained by using the Zeiss LSM 510 confocal microscope.

Immunohistochemical Staining. Tissue blocks of formalin-fixed, paraffin-embedded human breast tumors (n = 100) with coexisting normal, hyperplastic, *in situ*, and invasive components were retrieved from the files of the Armed Forces Institute of Pathology. Consecutive sections at 4- to 5-μm thickness were cut, placed on positively charged microscopic slides, and subjected to morphological evaluation, as described previously (15). Immunohistochemical staining was carried out by using our published protocol (16). Briefly, sections were deparaffinized with xylene, washed with ethanol and water, and incubated in 10% normal serum to block nonspecific binding. Then, sections were incubated with the primary antibody at a 1:50 dilution or control

solution for 3–4 h at room temperature. After the incubation, sections were washed with PBS, and the antigen–antibody complex was detected by using an avidin–biotin complex detection kit (Vector Laboratories) and an alkaline phosphatase red substrate kit (Zymed Laboratories), according to the manufacturer's instructions.

The immunostained sections were reviewed independently by at least two investigators. A given cell was considered to be CAPC-positive if distinct substrate coloration was seen consistently in its cytoplasm or nucleus in at least two duplicates of the same immunostaining procedure with the same antibody concentration, whereas the negative controls were totally devoid of substrate coloration. A given case was considered to be CAPC-positive if >5% of its entire cell population showed distinct CAPC immunoreactivity.

Sequence Identification of *CAPC*. MAPcL sequences were clustered by using CAP3. A single consensus sequence was derived for each cluster, along with a count of the number of sequence members of that cluster. Individual MAPcL sequences and consensus sequences were aligned with the human genome by using BLAT. Overlap of MAPcL sequences with known genes and ESTs from dbEST were collected based on BLAT alignments with the human genome. Alignment of any MAPcL sequence with a known gene or EST derived from essential organs, such as brain, heart, kidney, or lung, excluded that sequence and its parent cluster from further consideration. The remaining MAPcL clusters were ranked by the number of member sequences as a measure of relative abundance. The cluster containing *CAPC* was the largest cluster that did not overlap a known gene or essential ESTs.

URLs. UniGene, RefSeq, dbEST, and GenBank sequence databases, the BLAST program, LocusLink, OMIM, and the CGAP project can be accessed from www.ncbi.nlm.nih.gov. The Gene Ontology Consortium database can be obtained from ftp://ftp.geneontology.org/pub/go. The genome build and annotation databases can be accessed from http://genome.ucsc.edu. All database versions were taken as a snapshot from public releases available March 14, 2002, except genome sequences and annotations, which were taken from the December 2001 UCSC human genome database.

We thank Mindy Jo Knudson for excellent technical assistance and Sookhee Bang and Yoonsoo Hahn for scientific help. We thank Susan Garfield and Steven Wincovis (Center for Cancer Research Confocal Microscopy Core Facility, Bethesda) for technical support. Tissue samples were provided by the Cooperative Human Tissue Network, which is funded by the National Cancer Institute. Other investigators may have received samples from these same tissues. This research was supported in part by the Intramural Research Program of the National Cancer Institute, National Institutes of Health, Center for Cancer Research and was also funded in part by the South Dakota 2010 Initiative through the South Dakota Signal Transduction Center Proposal.

1. Egland, K. A., Vincent, J. J., Strausberg, R., Lee, B. & Pastan, I. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 1099–1104.
2. Van Beijnum, J. R. & Griffioen, R. W. (2005) *Biochim. Biophys. Acta* **1752,** 121–134.
3. Liu, R., Hsieh, C. Y. & Lam, K. S. (2004) *Semin. Cancer Biol.* **14,** 13–21.
4. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
5. Boguski, M. S., Lowe, T. M. & Tolstoshev, C. M. (1993) *Nat. Genet.* **4,** 332–333.
6. Ponting, C. P., Schultz, J., Milpetz, F. & Bork, P. (1999) *Nucleic Acids Res.* **27,** 229–232.
7. Prescott, A. R., Lucocq, J. M., James, J., Lister, J. M. & Ponnambalam, S. (1997) *Eur. J. Cell Biol.* **72,** 238–246.
8. Frand, A. R., Cuozzo, J. W. & Kaiser, C. A. (2000) *Trends Cell Biol.* **10,** 203–210.
9. Dautry-Vast, A. (1986) *Biochimie* **68,** 375–381.
10. Moll, R., Franke, W. W. & Schiller, D. L. (1982) *Cell* **31,** 11–24.
11. Buhler, H. & Schaller, G. (2005) *Mol. Cancer Res.* **3,** 365–371.
12. Becker, M., Nitsche, A., Neumann, J., Junghahn, I. & Fichtner, I. (2002) *Br. J. Cancer* **87,** 1328–1335.
13. Windoffer, R. & Leube, R. E. (2001) *Cell Motil. Cytoskeleton* **50,** 33–44.
14. Essand, M., Vasmatzis, G., Brinkmann, U., Duray, P., Lee, B. & Pastan, I. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 9287–9292.
15. Tavassoli, F. A. & Man, Y. G. (1995) *Breast J.* **1,** 155–162.
16. Man, Y. G. & Burgar, A. (2003) *Pathol. Res. Pract.* **199,** 815–825.