

# Molecular Cloning of a Novel Human Gene on Chromosome 4p11 by Immunoscreening of an Ovarian Carcinoma cDNA Library

Liu-Ying Luo,\* † Antoninus Soosaipillai,\* and Eleftherios P. Diamandis\* †.¹

\*Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada; and † Department of Laboratory Medicine and Pathobiology, University of Toronto, 110 College Street, Toronto, Canada

Received December 1, 2000

In our efforts to identify immunoreactive antigens in ovarian cancer, we used the method of immunoscreening of an ovarian carcinoma cDNA expression library with ascites fluid from ovarian cancer patients. Among many positive clones, one was found to contain partial sequence of a novel gene. By searching expressed sequence tags (ESTs) and human genome project databases as well as by screening other cDNA libraries and by RT-PCR strategies, we were able to obtain the full-length cDNA sequence (1.4 kb) and establish the genomic organization of this new gene. We also identified two alternatively spliced forms, encoding for slightly different proteins. The longer form (1.4 kb) is predicted to encode for a 27.6 kDa protein of 245 amino acids. The shorter form (1.3 kb) encodes for a truncated protein of 20.7 kDa and 208 amino acids. These proteins are not significantly homologous to any known protein in the GenBank database. This gene is composed of nine exons and eight introns. By fluorescence in situ hybridization (FISH), it was mapped to chromosome 4p11. This gene is highly expressed in many tissues, including testis, brain, placenta, ovary, prostate, and mammary gland. The high level expression of the shorter form is restricted to the central nervous system, including brain, cerebellum, and spinal cord, suggesting that this form may have a unique function in the central nervous system. © 2001 Academic Press

Abbreviations used: PBS-T, phosphate-buffered saline with Tween-20; RT-PCR, reverse transcriptase polymerase chain reaction; SSC, saline-sodium citrate buffer; SDS, sodium dodecyl sulphate; BAC, bacterial artificial chromosome; DAPI, 4', 6-diamidin-2-phenylindol-dihydrochloride; FITC, fluorescein isothiocyanate; FISH, fluorescence in situ hybridization; bp, base pair; kDa, kilodal-

1 To whom correspondence should be addressed at Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada. Fax: 416-5868628. E-mail: ediamandis@mtsinai.on.ca.

Key Words: immunoscreening of expression libraries; ovarian cancer; splice variants; gene cloning; fluorescence in situ hybridization; chromosome 4p11.

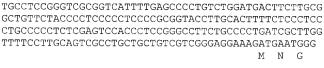
Cancer is one of the leading causes of death in humans. In North America, it is estimated that one in four people will die of cancer. The etiology of cancer is still not clear. Although some cancers have specific etiology, in general, it is thought that both genetic factors and environmental carcinogens contribute to disease pathogenesis.

In the human body, normal cellular components are recognized as "self" and they do not trigger immune responses. In cancer, it is now known that certain cellular components may act as immunogens, thus triggering autoimmune responses. The underlying mechanisms may include (a) Failure of self-tolerance of the immune system; (b) Aberrant expression of certain cellular components; (c) Expression of mutated protein forms. The circulating autoantibodies to such antigens have potential for cancer diagnosis, monitoring or prognosis, and for developing therapeutic cancer vaccines (1-4).

Autoimmune reactions have been frequently described in ovarian cancer. Two proteins have been investigated intensively, CA-125 and p53 (5-12). High titers of CA-125 and p53 autoantibodies have been detected in a significant proportion of ovarian cancer patients.

In this study, we have attempted to identify novel cellular components that may trigger immune responses in ovarian cancer patients, by screening an ovarian carcinoma cDNA expression library with ascites fluid from ovarian cancer patients. Among a few other positive clones, encoding for known genes, we identified and characterized a novel human gene, which resides on chromosome 4p11.





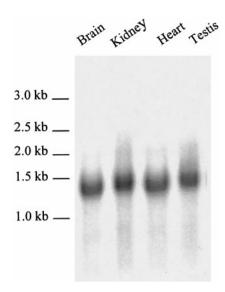
AGGGCTGATTTCGAGAGCCGAATGCAGAGGTTCCAAGACCAATTCCCCAC RADFREPNAEVPRPIPH ATAGGGCCTGATTACATTCCAACAGAGGAAGAAAGGAGAGTCTTCGCAGAA TGPDYTPTEEERRVFAE TGCAATGATGAAAGCTTCTGGTTCAGATCTGTGCCTTTGGCTGCAACAAGT NDESFWFRSVPLAATS ATGTTGATTACTCAAGGATTAATTAGTAAAGGAATACTTTCAAGTCATCCC M L I T O G L I S K G I L S S H P AAATATGGTTCCATCCCTAAACTTATACTTGCTTGTATCATGGGATACTTT YGSIPKLILACIMGYF GCTGGAAAACTTTCTTATGTGAAAACTTGCCAAGAGAAATTCAAGAAACTT AGKLSYVKTCOEKFKKL GAAAATTCCCCCCTTGGAGAAGCTTTACGATCAGGACAAGCACGACGATCT N S P L G E A L R S G O A R R S TCACCACCTGGGCACTATTATCAAAAGTCAAAATATGACTCAAGTGTGAGT P P G H Y Y O K S K Y D S S GGTCAATCATCTTTTGTGACATCCCCAGCAGCAGCAGCAACATAGAAATGCTT Q S S F V T S P A A D N I E M L HYEPIPFSSSMNESAP ACTGGTATTACTGATCATATTGTCCAAGGACCTGATCCCAACCTTGAAGAA G I T D H I V Q G P D P N L E E

> ggtagaaacttctcttgaaatgaa G R N F S \*

#### tttcaacattttatggtccaac

E S Y E V S L T Q K T D P S V  ATGCATGAAAGAGTGCCAAAAAAAAGAAGTCAAAGTAAACAAGTAT	<u>ATTACATATGAGGAATTAAGGAATAAGAACAGA</u>	TAC	TAT	AAAA	AAA	AAG	TAA	TCC	AG	
E S Y E V S L T Q K T D P S V  ATGCATGAAAGAGTGCCAAAAAAAGAAGTCAAAGTAAACAAGTAT	I T Y E E L R N K N R	Τ	I	N	K	R	K	Р	S	
ATGCATGAAAGAGTGCCAAAAAAAGAAGTCAAAGTAAACAAGTAT	GAGTCATATGAAGTATCTTTAACACAAAAGACTGACCCCTCAGTCAG									
	L T Q K T D P S V R P	Т	L	S	V	Ε	Y	S	Ε	
	AAAAAAGAAGTCAAAGTAAACAAGTATGGAGAT	AAA	AAA	rgcc	AGT	.AAG	TGA	GCA	TΑ	
M H E R V P K K E V K V N K Y (	K K E V K V N K Y G D	K	K	P	V	R	E	Н	M	

**FIG. 1.** Full-length cDNA sequence of the novel gene, alternatively spliced forms, and their open reading frames. The bold sequences in the  $3^\prime$  end indicate the polyadenylation signal (AATAAA) and the polyA tail. The underlined DNA and protein sequences denote the sequences which are present in the alternatively spliced

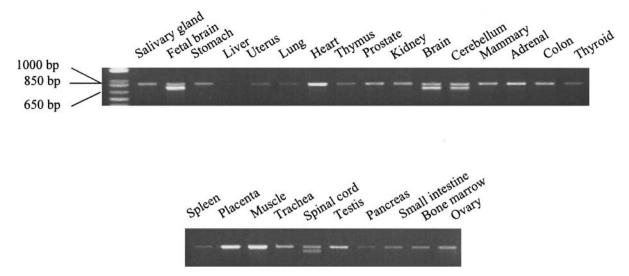


**FIG. 2.** Northern blot analysis of the novel gene to determine the transcript size in various human tissues. The position of the RNA marker is indicated on the left side of the picture.

## MATERIALS AND METHODS

Immunoscreening of an ovarian carcinoma cDNA expression library with ascites fluid from ovarian cancer patients. The Uni-ZAP XR premade ovarian carcinoma cDNA expression library was purchased from Stratagene (La Jolla, CA). Ovarian cancer ascites fluids were pooled from five different primary ovarian cancer patients. The cDNA library was plated on NZY agar plates at a density of 500 clones/15 cm plate. The plates were incubated at 42°C for 4 h to allow plaques to develop. Nitrocellulose filters soaked with IPTG were then laid on top of the plagues and incubated at 37°C for 4 h to transfer the plaques onto the membranes. The filters were blocked with 5% non-fat dried milk/PBS-T overnight, at 4°C. The PBS-T buffer contained 80 mM sodium orthophosphate, 20 mM sodium dihydrogen orthophosphate, 100 mM sodium chloride, and 0.1% Tween-20. To screen the library, the ascites were diluted 1:100 in 5% non-fat dried milk/PBS-T. The diluted ascites were first incubated with E. coli phage lysate (from Stratagene) for 2 h at room temperature to minimize the cross-reaction between the autoantibodies and the bacterial/phage proteins. Nitrocellulose filters were incubated with this preabsorbed ascites for 2 h at room temperature, to identify cellular proteins that react with the autoantibodies in the ascites. Following probing with the ascites, the filters were washed and further treated with goat anti-human IgG conjugated with alkaline phosphatase (Jackson Immunoresearch Laboratories, West Grove, PA) at 1:2000 dilution, in the blocking buffer for 1 h at room temperature, then proceeded to chemiluminescence detection with dioxetane-based substrate (Diagnostic Products Corporation, Los Angeles, CA). The plaques exhibiting immunoreactivity were excised from the plates and the phages were converted into the pBluescript phagemid form by in vivo excision with Exassist helper phage following the manufacturer's instructions (Stratagene). The excised phagemids were purified and subjected to automated DNA sequencing with M13 forward and reverse primers. The insert sequences

form A, but not in form B. The DNA and the protein sequences unique to form B are shown in italic and bold letters. Asterisks indicate stop codons. For more details, see text.



**FIG. 3.** Tissue expression of the novel gene, determined by RT-PCR. The various human tissues are shown on the top of each lane and the DNA marker is indicated on the left side of the picture. Note the alternatively spliced form (lower band) in brain, cerebellum, and spinal cord.

were compared to the known sequences in the GenBank database with the BLASTN alignment algorithm (13).

Northern blot analysis. The human multiple tissue blot was obtained from OriGene Technologies, Inc. (Rockville, MD). Anti-sense RNA probe labelled with digoxigenin was prepared by in vitro transcription using the T7 RNA polymerase kit (Roche Molecular Systems, Laval, Quebec, Canada) following the manufacturer's instructions. The blot was hybridized with the RNA probe in Ultrahyb (OriGene) hybridization buffer overnight at  $68^{\circ}\mathrm{C}$ , then subsequently washed with  $2\times\mathrm{SSC}$ , 0.1% SDS and  $0.2\times\mathrm{SSC}$ , 0.1% SDS at  $68^{\circ}\mathrm{C}$  for 15 min. To detect the signal, the blot was incubated with alkaline phosphatase-conjugated anti-digoxigenin antibody (Roche) at 1:2000 dilution for 1 h, then proceeded to chemiluminescence detection with dioxetane-based substrate.

Human tissue total RNA. Human tissue total RNA was purchased from Clontech Laboratories, Inc. (Palo Alto, CA).

Reverse transcriptase polymerase chain reaction (RT-PCR). Two micrograms of total RNA was converted to cDNA with Superscript Preamplification Kit (Gibco BRL, Gaithersburg, MD), according to the manufacturer's recommendations. The final volume was 20  $\mu$ l. The PCR primers are as follows, forward, 5'-CTTGCTTGTATCAT-GGGATAC-3'; reverse, 5'-CCAATTCAACATGCAGATATC-3'. PCR was carried out in a 20  $\mu$ l reaction mixture, containing 1  $\mu$ l cDNA, 10 mM Tris–HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 200  $\mu$ M dNTPs (deoxynucleoside triphosphates), 100 ng primers, and 2.5 units of Hotstar Taq DNA polymerase (QIAGEN Inc., Valencia, CA) on a Perkin-Elmer 9600 thermal cycler. The PCR conditions were 94°C for 15 min to activate the polymerase, following by 30 cycles of 94°C for 30 s, 58°C for 1 min, 72°C for 1 min and a final extension at 72°C for 10 min. The PCR products were then separated on a 2.5% agarose gel, stained with ethidium bromide, and visualized under UV light.

Chromosomal localization of the novel gene by fluorescence in situ hybridization (FISH). A BAC clone (clone identification: 19F13) containing the novel gene of interest was labeled with biotin by nick translation. This BAC clone was identified by screening the human male genomic BAC library RPCI-11 with the excised phagemid, as described elsewhere (14). The regional assignment of the BAC clone was determined by FISH to normal human lymphocyte chromosomes counterstained with propidium iodide and 4′, 6-diamidin-2-phenylindol-dihydrochloride (DAPI). Biotinylated probe was detected with avidin-fluorescein isothiocyanate (FITC), followed by

biotinylated anti-avidin antibody and avidin-FITC. Images of metaphase preparations were captured by a thermoelectrically cooled charge coupled camera (Photometrics, Tucson, AZ). Separate images of DAPI banded chromosomes and of FITC targeted chromosomes were obtained, pseudo colored blue (DAPI) and yellow (FITC) and merged electronically using Adobe PhotoshopA 3.0 software. The band assignment was determined by measuring the fractional chromosome length and by analyzing the banding pattern generated by the DAPI counterstained image.

## **RESULTS**

Identification of a novel ovarian carcinoma immunoreactive antigen by immunoscreening of an ovarian carcinoma cDNA expression library with ovarian cancer ascites. In order to identify cellular proteins that trigger immune responses in ovarian cancer, we screened a λ ZAP phage ovarian carcinoma cDNA expression library with ascites from ovarian cancer patients. Positive clones were excised from the plates and the  $\lambda$  ZAP phages were converted into the phagemid form. The phagemids were then sequenced with vector-specific primers. The identity of the insert sequences was examined using the BLASTN program and the GenBank databases. Using this method, we identified, among eight other clones encoding for known genes, one immunoreactive clone with insert sequence not matching with any known gene, suggesting that it encoded for a novel gene. The length of this insert sequence was about 800 bp. The polyA tail and the polyadenylation signal were found in the 3' end of this sequence (Fig. 1).

Identification of the full-length cDNA sequence of the novel gene. In order to obtain the full-length cDNA sequence of the novel gene, we searched the GenBank EST database with the initial 800 bp sequence and found a number of matching ESTs. With ESTs AA099273, W44679, and AA332589, we were able to



**FIG. 4.** Genomic organization and partial nucleotide sequence of the novel gene. The promoter and intron sequences are shown with lower case letters. Exon sequences are in capital and bold letters. Dotted lines represent sequence that is not shown. See our GenBank submission (pending) for detailed sequence information. For some introns, we do not have the full sequence. The intron/exon numbers and lengths are as indicated in base pairs (bp). The SP1, CREB binding sites, and splice junction donor/acceptor sites are underlined.

construct a 1.4 kb contiguous sequence (Fig. 1). To obtain more sequence, a testis cDNA library (OriGene Technologies, Inc.) was screened twice and six clones were found to contain this new gene, but when these clones were sequenced, no more new sequence was identified (data not shown).

Determining the transcript size of the novel gene in human tissues. To assess the transcript size of the novel gene and determine whether the 1.4 kb cDNA sequence obtained full length, we performed Northern blot analysis. An anti-sense RNA probe was prepared using in vitro transcription with the initial excised phagemid as a template. This RNA probe was then hybridized to a blot containing 2  $\mu g$  of polyA RNA. This method detected a distinct band of about 1.4 kb in length. No other band was identified (Fig. 2). This result suggested that the 1.4 kb cDNA sequence, very likely represented the full-length cDNA of the novel gene.

*Tissue specific expression of the novel gene.* We investigated the tissue expression pattern of the novel gene by RT-PCR. As shown in Fig. 3, this gene is highly

expressed in all tissues tested, except liver, with highest expression seen in cardiac and skeletal muscle and in placenta. Two PCR products were detected on the agarose gel with brain tissues, with approximate lengths of 850 bp and 700 bp. By cloning and sequencing of these two bands, we verified that the longer band had the same sequence as the cDNA sequence we had already obtained from the phagemid. The sequence of the smaller size band revealed that it represented an alternatively spliced transcript. We named the longer transcript as "alternatively spliced form A" and the shorter one as "alternatively spliced form B." We observed that only in the brain tissue (fetal and adult), spinal cord, and cerebellum, the expression of alternatively spliced form B is more predominant than form A. whereas, in the other tissues, form A is the only form expressed (Fig. 3).

The identified transcripts encode for novel proteins. We identified one open reading frame within the full-length cDNA sequence of the novel gene. Alternatively spliced form A encodes a protein of 245 amino acids with molecular mass of 27.6 kDa. Form B encodes for a

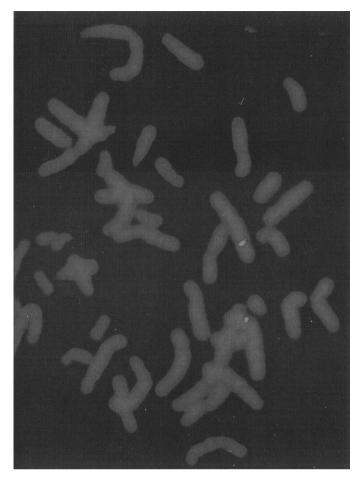
truncated protein of 208 amino acids with molecular mass of 20.7 kDa (Fig. 1). In order to determine the identity of these proteins, we used the BLASTP, BLASTP + BEAUTY (http://www.dot.imgen.bcm.tmc. edu.9331) and BLOCKS protein motif (http://www.blocks.fhctc.org) programs to search the databases. One putative protein submitted by the Drosophila genome project (GenBank Accession No. AAF46923) was found to have 30% identity and 46% similarity at the amino acid sequence level with these novel proteins. The function of this protein is still not known. We found no other significant homologies with any known proteins.

Genomic organization of the novel gene. To establish the genomic organization of the new gene, we searched the GenBank database with the cDNA sequence. Two BAC clones, 19F13 (Accession No. AC011956) and 15P24 (Accession No. AC011977.1) were found to contain the majority of the gene sequence. By aligning the cDNA sequence with the BACs, the intron/exon organization of this gene was established. As shown in Fig. 4, this gene consists of eight introns and nine exons (Exon VI and exon IX sequences were missing from the BACs and we considered them as continuous exons, but the splice sites could not be established). Exon I is untranslated. Among the eight introns, for only three of them, we have full-length sequences, including introns I, IV, and VII. For the rest of the introns, we have partial sequence. The intron-exon splice junctions are completely conserved, following the rule that the "GT" is in the splice donor site and the "AG" is in the splice acceptor site (15). Within the 200 bp region upstream from the cDNA start site, although a typical "TATA box" was not identified, the sequence in this region is GC rich, indicating that it may harbor a CpG island. In addition, a consensus SP1 binding site "GGGGCGGGC" (16) and a CREB binding site "TGAC" (17) were found within this region. These findings suggest that this region represents the potential promoter of the novel gene.

Chromosomal localization of the novel gene. To determine the chromosomal localization of this new gene, fluorescence *in situ* hybridization was performed using the BAC clone 19F13 as a probe. The new gene maps to chromosome 4p11 (Fig. 5).

## DISCUSSION

In this study, we used the method of immunoscreening of an ovarian carcinoma cDNA expression library with ovarian cancer ascites fluid to identify cellular components that may trigger immune responses in ovarian cancer. We identified one immunoreactive clone containing the partial sequence of a novel gene. We cloned the full-length sequence and characterized



**FIG. 5.** Chromosomal localization of the novel gene by fluorescence *in situ* hybridization. This gene is localized to chromosome 4p11.

the cDNA and genomic structure of this new gene. The encoded protein has no significant homology to any other known proteins with the exception of a Drasophila protein of unknown function.

Whether this gene triggers immune responses in a significant proportion of ovarian cancer patients is currently unknown. In the expression library used, the ovarian cDNA was fused with the LacZ gene, thus, the proteins expressed are fusion proteins. This type of expression libraries has been shown to be suitable for identifying immunoreactive antigens and this screening strategy is widely used (18–22). However, one limitation of this screening method is that the probing autoantibodies (in our case from cancer ascites) may recognize the fusion protein, but not necessarily the native protein, due to conformational differences.

There are two alternatively spliced forms of this gene (Figs. 1 and 3). Interestingly, the tissue distribution of these two forms is quite different. The longer form A is expressed in almost all human tissues tested (Fig. 3). The shorter form B, is only expressed abundantly in the central nervous system, including brain, cerebel-

lum, and spinal cord. We hypothesize that form B may have a different function than form A, and that this function may be unique to the central nervous system. Thus, it will be interesting to examine the possible utility of the short isoform as a specific biomarker of the central nervous system.

#### **ACKNOWLEDGMENT**

This work was supported by a grant to Dr. E. P. Diamandis from the Aventis-Pasteur Cancer Vaccine Program.

#### REFERENCES

- Abrams, M. B., Bednarek, K. T., Bogoch, S., Bogoch, E. S., Dardik, H. J., Dowden, R., Fox, S. C., Goins, E. E., Goodfried, G., Herrman, R. A., and et al. (1994) Cancer Detect. Prev. 18, 65–78.
- Amagai, M., Klaus-Kovtun, V., and Stanley, J. R. (1991) Cell 67, 869–877.
- 3. Tan, E. M. (1991) Cell 67, 841-842.
- 4. Zouali, M. (1994) Clin. Exp. Rheumatol. 12 (Suppl 11), S33-S36.
- Angelopoulou, K., and Diamandis, E. P. (1993) Cancer J. 6, 315–321.
- Angelopoulou, K., and Diamandis, E. P. (1997) Eur. J. Cancer 33, 115–121.
- Angelopoulou, K., Diamandis, E. P., Sutherland, D. J., Kellen, J. A., and Bunting, P. S. (1994) Int. J. Cancer 58, 480–487.
- 8. Hassapoglidou, S., and Diamandis, E. P. (1992) *Clin. Biochem.* **25,** 445–449.
- 9. Laurent-Puig, P., Lubin, R., Semhoun-Ducloux, S., Pelletier, G.,

- Fourre, C., Ducreux, M., Briantais, M. J., Buffet, C., and Soussi, T. (1995) *Gut.* **36**, 455–458.
- Mazurek, A., Niklinski, J., Laudanski, T., and Pluygers, E. (1998) Eur. J. Cancer Prev. 7, 23–35.
- O'Brien, T. J., Tanimoto, H., Konishi, I., and Gee, M. (1998) *Int. J. Biol. Markers.* 13, 188–195.
- 12. Rosenthal, A. N., and Jacobs, I. J. (1998) *Int. J. Biol. Markers.* **13,** 216–220.
- Altschul, Stephen, F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* 25, 3389–3402.
- Luo, L., Herbrick, J. A., Scherer, S. W., Beatty, B., Squire, J., and Diamandis, E. P. (1998) *Biochem. Biophys. Res. Commun.* 247, 580–586.
- 15. Mount, S. M. A. (1982) Nucleic Acids Res. 10, 459-472.
- Briggs, M. R., Kadonaga, J. T., Bell, S. P., and Tjian, R. (1986) Science 234, 47–52.
- Benbrook, D. M., and Jones, N. C. (1994) Nucleic Acids Res. 22, 1463–1469.
- Jager, D., Stockert, E., Scanlan, M. J., Gure, A. O., Jager, E., Knuth, A., Old, L. J., and Chen, Y. T. (1999) *Cancer Res.* 59, 6197–6204.
- Lin, B., White, J. T., Ferguson, C., Bumgarner, R., Friedman, C., Trask, B., Ellis, W., Lange, P., Hood, L., and Nelson, P. S. (2000) Cancer Res. 60, 858–863.
- 20. Old, L. J., and Chen, Y. T. (1998) J. Exp. Med. 187, 1163-1167.
- Sahin, U., Tureci, O., Schmitt, H., Cochlovius, B., Johannes, T., Schmits, R., Stenner, F., Luo, G., Schobert, I., and Pfreundschuh, M. (1995) Proc. Natl. Acad. Sci. USA 92, 11810–11813.
- Sahin, U., Tureci, O., and Pfreundschuh, M. (1997) Curr. Opin. Immunol. 9, 709–716.