Identification of novel gene highly expressed in pancreatic cancer

Ai Hua Jin

Department of Medical Science

The Graduate School, Yonsei University

Identification of novel gene highly expressed in pancreatic cancer

Directed by Professor Si Young Song

The Master's Thesis
submitted to the Department of Medical Science,
the Graduate School of Yonsei University
in partial fulfillment of the requirements for the
degree of Master of Medical Science

Ai Hua Jin

June 2007

This certifies that the Master´s Thesis
of Ai Hua Jin is approved.

Thesis Supervisor : Si Young Song

Thesis Committee Member : Jin Sung Lee

Thesis Committee Member : Woo Jung Lee

The Graduate School
Yonsei University

June 2007

ACKNOWLEDGEMENTS

I thank Prof. Si Young Song (Department of Internal Medicine, Yonsei University College of Medicine), Prof. Jin Sung Lee (Department of Clinical Genetics, Yonsei University College of Medicine) and Prof. Woo Jung Lee (Department of surgery, Yonsei University College of Medicine) for helpful discussion and encouragements. I also thank Prof. Seung Woo Park, Jingu Gang Ph.D. and laboratory members Esther Dawoon Jung, Jieun Lee, Jinhui Bae, Jing Wen, Kyunghwa Park, Sun A Kim, Taeyun Oh, Kyoungsun Park, Jaewan Kim, Yeeshin Choo for technical and emotional supports.

.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Identification of novel gene highly

expressed in pancreatic cancer

Ai Hua Jin

Department of Medical Science

The Graduate School, Yonsei University

(Directed by Professor Si Young Song)

To identify novel genes highly expressed in pancreatic cancer, we used bioinformatic analysis of expressed sequence tags (ESTs). ESTs are an excellent source of data for such studies using bioinformatic approaches because of the rich libraries and tremendous amount of data now available in the public domain. Sixty four ESTs were selected depending on fold change and p-value that were highly expressed in pancreatic cancer tissues compared to normal tissues from analysis of Affymetrix Human Genome U133 GeneChip set. Eight ESTs were tried to construct their full-length cDNAs using full-length

cDNA library. As a result, one EST was successful for cloning, and then was sequenced to construct the full-length cDNA. In this way, a putative novel gene AT-32 highly expressed in pancreatic cancer was identified. The cDNA sequence was searched in several public databases to predict the open reading frame (ORF) and possible protein related information.

In this study, in silico screening and experimental expression analysis were combined to select ESTs that are highly expressed in pancreatic cancer and identified one putative novel gene through cloning EST.

Key words: pancreatic cancer, microarray, EST, novel gene, identification

Identification of novel gene highly expressed in pancreatic cancer

Ai Hua Jin

Department of Medical Science
The Graduate School, Yonsei University

(Directed by Professor Si Young Song)

I. INTRODUCTION

Pancreatic cancer is the fourth leading cause of cancer death in the United States and has one of the worst mortality rates among the common malignancies, with an overall 5-year survival rate of less than 1% and a median survival of approximately 5 to 6 months.[1] This is largely due to the lack of symptoms and diagnostic tools for the detection of this disease in the early stages as well as a deficiency of effective therapeutics for later-stage disease.[2]

Some approaches that combine surgery and chemotherapy based on gemcitabine or 5-fluorouracil, with or without radiation therapy, can improve the quality of life of patients. However, such treatments have a very limited effect on long-term survival.[3] Surgical resection offers the only possibility for cure at present, but 80-90% of patients who undergo curative surgery suffer from relapse and die due to the metastatic or disseminated disease.[4]

To overcome this situation, further understanding of the molecular carcinogenesis of pancreatic cancer and development of new markers for pancreatic cancer at an early stage or therapeutic targets through identification of novel genes highly expressed in pancreatic cancer will be the initial step. Global analysis of the gene expression patterns of pancreatic cancer have been used to identify a number of genes highly overexpressed in pancreatic cancer.[5-8]

The aim of this study was to identify the novel gene highly expressed in pancreatic cancer compared with normal tissue through cloning ESTs from microarray chip data for the better understanding of molecular characteristics of pancreatic cancer and the possible future target of the novel diagnostics and therapeutics.

Novel genes were identified using bioinformatic analysis of expressed sequence tags (ESTs). ESTs are an excellent source of data for such studies

using bioinformatic approaches because of the rich libraries and tremendous amount of data now available in the public domain.[9] EST analysis has been extensively used in large-scale cDNA sequencing projects to discover novel genes for exploring gene expression patterns and for identifying differentially regulated genes.[10]

The EST list that highly expressed in pancreatic cancer tissues compared with normal tissues was gained from Affymetrix U133 GeneChip data. Then we narrowed down research interests on 64 novel ESTs depending on fold change and p-value. These 64 ESTs´ cDNA sequences and exact function are not known yet. So firstly, cDNA sequence was amplified using target EST gene-specific primers through rapid amplification of cDNA ends polymerase chain reaction, and among 8 amplified ESTs, one EST was succeeded with cloning using human placenta full-length cDNA library. In this way, one putative novel gene AT-32 highly expressed in pancreatic cancer was identified. Followed, in silico analysis was carried out using public database search tools NCBI, Affymetrix, PredictProtein Server and CBS Prediction Site, for identifying potential open reading frame (ORF) and protein information.

Ⅱ. MATERIALS AND METHODS

1. Tissues

Samples of normal pancreas were collected from 14 surgical specimens. Samples of pancreatic ductal adenocarcinoma were collected from 8 surgical resection specimens from patients. In each case, the specimens were harvested within 10 minutes of resection from the patient and snap-frozen in liquid nitrogen before storage at -80℃. The resected cancer tissues were not microdissected because we were interested not only in identifying the genes expressed by neoplastic epithelial cells, but also the genes expressed as a result of the neoplastic cell-stroma interaction.[12]

2. Microarray sample preparation and Affymetrix GeneChip hybridization

Microarray procedure was performed according to the methods described in the Affymetrix GeneChip Expression Analysis Technical Manual (Affymetrix, Santa Clara, CA, USA). Total RNA was extracted from the powdered normal and neoplastic tissues using TRIzol (Life Technologies, Inc., Rockville, MD, USA). Total RNA was reversely transcribed to cDNA using the SuperScript Choice system (Invitrogen Life Technologies, Inc., Rockville, MD,

USA). The cDNA was purified and used as a template to synthesize biotin-labeled cRNA using Enzo BioArray HighYield RNA Transcript Labeling Kit (Affymetrix). The cRNA was fragmented before hybridization. The fragmented labeled cRNA was applied to Human Genome U133 Array (Affymetrix) and hybridized to the probes in the array. After washing and staining, the arrarys were scanned.

3. Statistical Data Analysis

For each gene fragment, the ratio of the geometric means of the expression intensities in the normal control tissues and the pancreatic cancer samples was calculated, and the fold change was then calculated on a per fragment basis.[8] For this analysis, the *t* test was used and significance was defined as $p < 0.01$.

4. Expression analysis

Expression values of tumor and normal tissues were obtained from the GeneExpress Oncology Datasuite™ of Gene Logic Inc. (Gaithersburg, MD, USA), based on the Affymetrix Human Genome U133 array set. We analyzed the expression profiles of normal and cancer tissue sets from the pancreas (14/8). Genes were included in the analysis if they had a p-value ≤ 0.01. Unigene was used to identify EST clusters matched with selected Affymetrix

fragments. Vector NTI®suite (Informax Inc., Bethesda, MD, USA) was used for the nucleotide analysis and multiple sequence alignments.

5. Cell lines

All used human pancreatic cancer cell lines AsPc-1, BxPc-3, CAPAN-1, CAPAN-2, CFPAC-1, Hpac, Mia PaCa-2, Panc-1 were purchased from the American Type Culture Collection (ATCC, Rockville, MD, USA). The ascites derived human pancreatic cancer cell line, AsPc-1 and BxPc-3 were cultured in RPMI-1640 medium (Invitrogen Life Technologies, Carlsbad, CA, USA) supplemented with 10% FBS (FBS; HyClone, Logan, Utah, USA) and 1% antibiotic-antimycotic solution (1,000 unit/mL penicillin; Gibco BRL, Grand Island, NY, USA). Capan-1 and CFPAC-1 were cultured in Iscove's modified Dulbecco's medium (Invitrogen Life Technologies, Carlsbad, CA, USA) supplemented with 20% and 10% FBS respectively. Capan-2 was grown in McCoy's 5α medium with 10% FBS, Panc-1 was grown in Dulbecco's modified Eagle's medium with 10% FBS, and Hpac was cultured in a 1:1 mixture of Dulbecco's modified Eagle's medium and Ham's F12 medium. Human pancreatic duct epithelial cell line (HPDE) immortalized by transduction with the E6/E7 genes of human papilloma virus 16 (HPV16) was kindly provided by Dr. Ming-Sound Tsao (University of Toronto, Ontario, Canada). HPDE was

cultured in keratinocyte serum-free medium supplemented with bovine pituitary extracts and human epidermal growth factor (Life Technologies, Grand Island, NY). hYGIC-6 was grown in waymouth medium (Sigma, St. Louis, MO, USA) with 10% FBS. All cells were maintained at 37℃ in an atmosphere of humidified air with 5% CO_2 .

6. Construction of full-length cDNA library

Two full-length cDNA libraries were kept in our lab. One is made by Marathon cDNA Amplification Kit (Clontech, Mountain View, CA, USA). mRNA was isolated from Capan-1 cells (5×10^5) directly using Oligotex Direct mRNA Mini Kit (Qiagen, Valencia, CA, USA). Then the extracted mRNA was reverse-transcribed to cDNA. Double-strand cDNA synthesis was performed according to Marathon cDNA Amplification Kit (Clontech, Mountain View, CA, USA). First-strand cDNA was synthesized with cDNA synthesis primer and second-strand was synthesized with Second-Strand Enzyme Cocktail. Followed, Marathon cDNA Adaptor was ligated to complete full-length cDNA library.

The second library is human placenta full-length cDNA library. It was purchased from company (Origene, Rockville, MD, USA) and contained mostly 500,000 human full-lengh cDNA clones in 96-well Master Plate, each

well has 5,000 clones.

7. RACE (rapid amplification of cDNA ends) - PCR

EST is a partial sequence of a gene. In order to obtain the full-length cDNA from the EST, 5′ and 3′ rapid amplification of cDNA ends (RACE) were used according to Marathon cDNA Amplification Kit (Clontech, Mountain View, CA, USA). Gene-specific primers (GSP) and adaptor primer 1 (AP1) were used to amplify the 5′ / 3′-end of EST using full-length cDNA library as a template (Figure 1). The gene-specific primers for RACE-PCR were:

AT-32 5′ RACE: 5′-GCCAACCTTTCAGATGAGTACCAC-3′,

AT-32 3′ RACE: 5′-GGGCACTCTCATTTCACTGCATAC-3′,

Adaptor Primer 1: 5′-CCATCCTAATACGACTCACTATAGGGC-3′.

Gene specific primers for RACE PCR were designed according to the each sequence of the target ESTs.

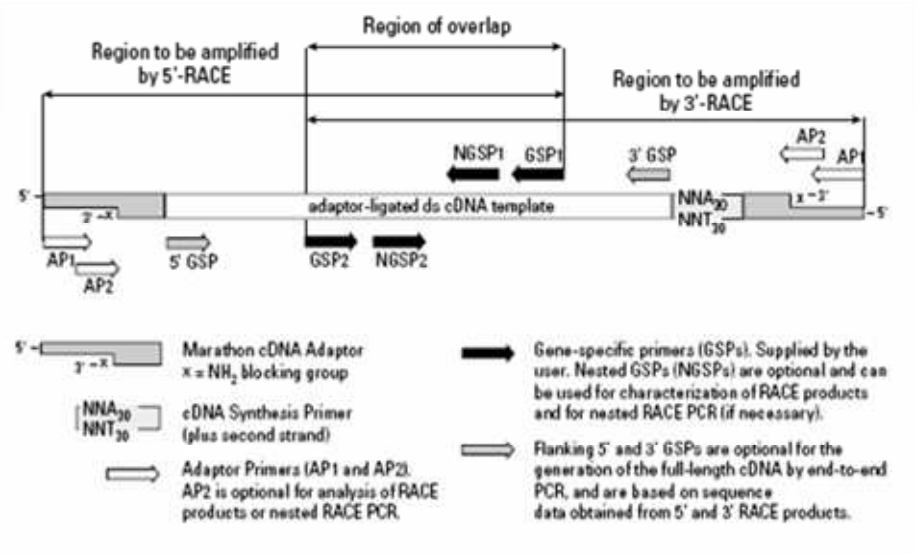


Figure 1. RACE-PCR principle and conditions. Touchdown PCR reactions were performed under hot-lid thermal conditions of 30 sec at 94℃ and then 5 cycles of 5 sec at 94℃ and 2 min 30 sec at 72℃, 5 cycles of 5 sec at 94℃ and 2 min 30 sec at 70℃, 25 cycles of 5 sec at 94℃ and 2 min 30 sec at 68℃; followed by a final 5 min extension at 72℃. (This figure was offered by Clontech web site.)

8. Cloning and Sequencing of RACE-PCR products

PCR products from RACE reactions were gel-extracted, cloned into T&A Cloning Vector (RBC, Taipei, Taiwan). The transformation procedure was followed according to the manufacturer′s direction with DH5α competent cells (RBC, Taipei, Taiwan). Cells were spread on Luria Broth plate containing 50mg/ml ampicillin. The LB plate was placed in 37℃ incubator overnight. Several white colonies were picked up to grown in Luria Broth media with ampicillin 16h in 37℃ shaking incubator. Plasmids were isolated from bacterial cultures with HiYield Plasmid Mini Kit (RBC, Taipei, Taiwan) and were then analyzed by PCR with M13 forward and reverse primers for insert cDNA fragments. Then sequenced at the DNA sequencing facility of Corebio company.

Sequencing primers were:

M13 forward: 5′- GTTTTCCCAGTCACGAC -3′,

M13 reverse: 5′- TCACACAGGAAACAGCTATGAC -3′.

9. Identification of target clone from human Placenta cDNA Library

The Master Plate (OriGene Technologies, Inc., Rockville, MD, USA) for each library panel contains DNA from 500,000 cDNA clones and is initially screened in one simple 96-well PCR. This first set of PCRs identifies the Sub-Plate, which contains the clone of interest. The Sub-Plates are contained in standard 96-well microtiter dishes. Each well cantains 50 cDNA clones with a single 96-well PCR to identify the positive well in the Sub-Plate. Cells from the positive well were then plated on LB/ampicillin agar and the resulting bacterial colonies were screened by PCR in order to obtain the desired clone. Then sequenced at the DNA sequencing facility of Corebio company. The sequencing primers were arranged in Table 1.

Table 1. Oligonucleotide primers for human placenta cDNA library clone

Primer name	Sequence 5′-3′
Primer for full-length cDNA clone	
AT-32 (F)	CACAAGGCATCATAGTAGCTGGAGG
AT-32 (R)	GGTCTTCCTTCCCTTGGTACTCAC
Primer for sequencing	
Vector primer 3	GCAGAGCTCGTTTAGTGAACC
Primer XL39	ATTAGGACAAGGCTGGTGGG
32 clone primer 1	GGTCTTCCTTCCCTTGGT
32 clone primer 2	GGTGATTTAGCATAGTGCC
32 clone primer 3	AATTGGGAAGTGGTATGT
32 clone primer 4	GGTCTGCTGGTGTATCTT
32 clone primer 5	TGATATGCTGAATGAGGC
32 clone primer (rev) 1	CTGGCTGAAAGGTTTGGAG
32 clone primer (rev) 2	GCCTAGCTGATTGAGATTG

10. Public database search

The integrated nucleotide and deduced amino acid sequences were analyzed with the software Vector NTI, ExPASy, SWISS-MODEL, CBS Prediction Site, NCBI and PredictionProtein Server to predict the open reading frame (ORF) and possible protein information**,** such as predicted protein sequence, structure, homology, function domain and so on.

11. Reverse transcriptase – polymerase chain reaction

Standard RT-PCR was performed using total RNA prepared from 8 pancreatic cancer cell lines, human pancreatic duct epithelial cell line, human YGIC-6, peripheral blood monocytes and bone marrow cells. Reverse transcription was carried out under hot-start conditions of 5 min at 94℃ and then 25 cycles of 30 sec at 94℃, 30 sec at 56℃ and 30 sec at 72℃; followed by a final 5 min extension at 72℃. Primers designed for AT-32 were:

AT-32RT forward: 5′- TGCACTCACAGGTAACATTG -3′,

AT-32RT reverse: 5′- GCCAACCTTTCAGATGAGTA -3′.

Primers designed for β–actin that was used as a loading control for RT-PCR reations were:

Forward: 5′- ATGATATCGCCGCGCTCGTCGTC -3′,

Reverse: 5′- CGCTCGGCCGTGGTGGTGAA -3′.

Ⅲ. RESULTS

1. Data Filtering

cRNA samples were hybridized to the complete Affymetrix Human Genome U133 GeneChip set for simultaneous analysis of 33,000 fragments corresponding to 6,000 ESTs. Firstly, 532 ESTs with a 1.2-fold greater increase in expression in the pancreatic cancer tissues compared with normal tissues were selected. The level of significance for each gene fragment ranged from less than p = 0.05 (modified Welch t test). Followed, among of these 532, 101 ESTs with a 2.0-fold greater and 6.0-fold less increase in expression were selected, of which 37 corresponded to known genes and 64 were novel ESTs. The level of significance for each gene fragment ranged from less than p = 0.01 (modified Welch *t* test). Finally, 64 novel ESTs were selected to continue followed experiments.

2. Amplification of ESTs using RACE-PCR

EST is a partial sequence of gene, it is a cDNA fragment. 5′ and 3′ RACE were performed using a commercially supplied Marathon cDNA Amplification Kit (Clontech, Palo Alto, CA, USA) following the manufacturer′s protocol. A gene-specific primer (GSP) and a RACE adaptor primer 1 (AP1) were used to

amplify the 5′- and 3′- end of the EST using human full-length cDNA library as a template (Fig. 1). The ESTs were randomly selected from 64 EST list to be carried out RACE-PCR and finally gained 8 ESTs RACE-PCR products (Table 2).

EST No.	Fold change	P-value	5′-RACE	3′-RACE
AT-12	2.2	<0.01	yes	yes
AT-26	2.0	<0.01	no	yes
AT-27	2.5	<0.01	yes	no
AT-32	4.4	<0.01	yes	yes
AT-40	2.6	<0.01	no	yes
AT-42	2.3	<0.01	yes	no
AT-44	2.6	<0.01	yes	no
AT-46	2.9	<0.01	no	yes

Table 2. RACE-PCR products of ESTs. RACE-PCR products of ESTs were randomly selected from the 64 novel EST list. These PCR products were amplified using a GSP (Gene-Specific Primer) and AP1 (Adaptor Primer1). AT-12 and AT-32 were carried out both 5′ and 3′ RACE. The others were amplified only one part of RACE.

3. Expression levels of AT-32 in normal and cancer tissues.

AT-32 expression levels in 14 normal tissues and 8 pancreatic cancer tissues from Affymetrix U133 GeneChip data were shown at Fig. 2A. AT-32 expression level was evaluated in five different organs′ cancer tissues and corresponding normal tissues also using Affymetrix U133 GeneChip set (Fig. 2B).

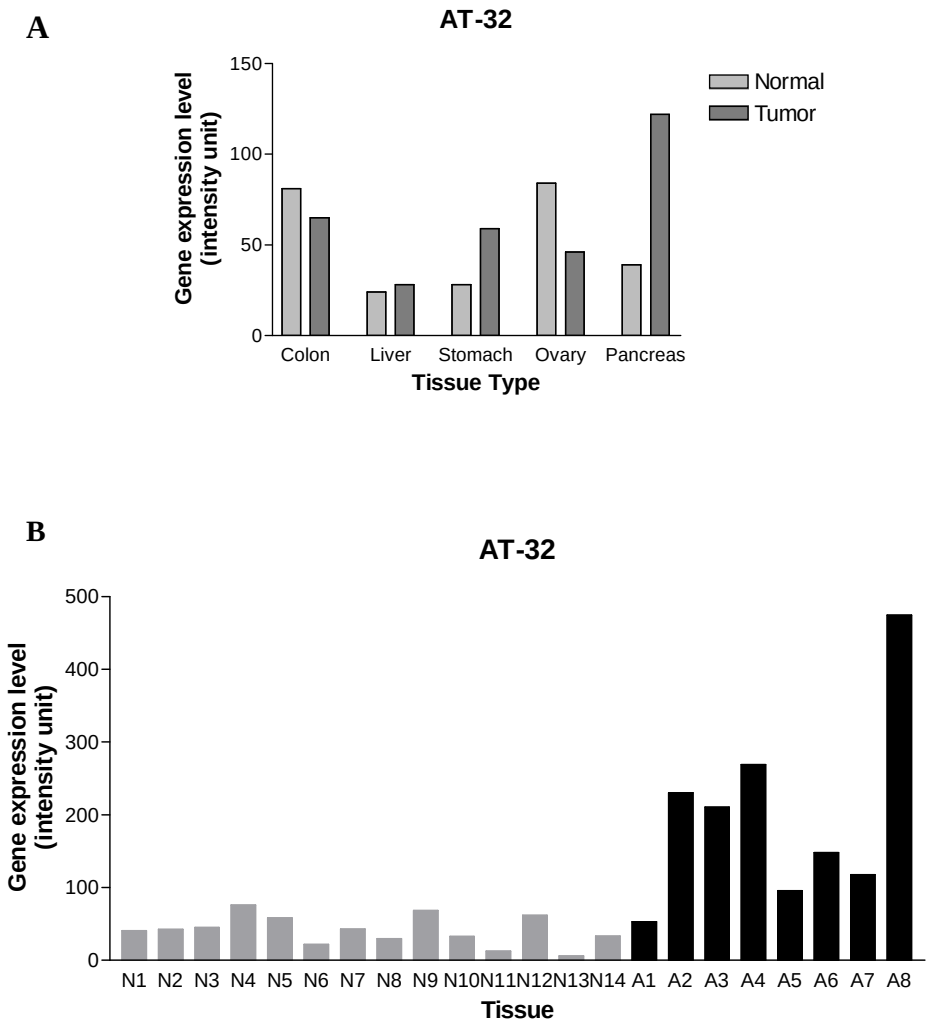
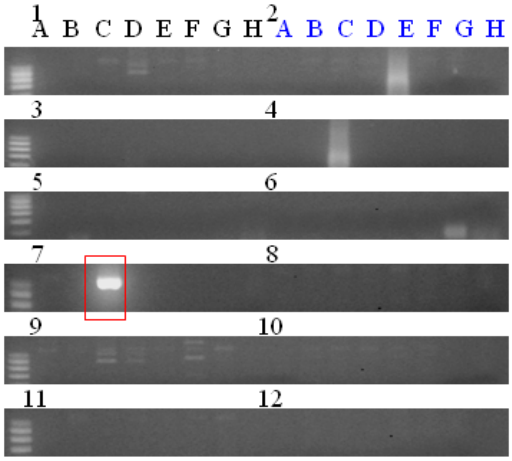


Figure 2. AT-32 expression levels in normal tissues and cancer tissues. A, AT-32 expression levels in 14 normal tissues and 8 pancreatic cancer tissues from Affymetrix U133 GeneChip data. B, AT-32 expression in human differential organs′ cancer and corresponding normal tissues. Black bar is adenocarcinoma and gray bar is normal tissue.

4. Identification of target clone from human placenta cDNA library

The Master Plate of human placenta cDNA library was screened by 96-well PCR using gene-specific primers and identified positive well by electrophoresis. AT-32 was identified in C9 well of Master Plate of human placenta cDNA library (OriGene Technologies, Inc., Rockville, MD, USA). Then, the Subplate of C9 well was purchased from Origene company. The C9 Subplate was also screened by 96-well PCR using the same primers and identified positive wells by electrophoresis. AT-32 was identified in C7 well of Subplate (Fig. 3). The bacteria from positive well of Subplate were plated onto LB agar + ampicillin plate. Eventually, most colonies were screened by colony PCR to identify the positive one and prepared plasmid DNA from the positive colony.



AT-32 Subplate 96-well PCR

Figure 3. 96-well PCR of AT-32 using Subplate library. 96-well PCR reactions were performed under hot-start conditions of 5 min at 95℃ and then 30 cycles of 30 sec at 95℃, 30 sec at 60℃ and 3 min at 72℃; followed by a final 5 min extension at 72℃. AT-32 was identified in C7 well of Subplate.

5. Full-length cDNA and putative amino acid sequence

AT-32 positive clone was identified from human placenta cDNA library (OriGene Technologies, Inc., Rockville, MD, USA). AT-32 clone containing insert of 6875bp was obtained. Vector NTI program was used to predict the putative ORF (from bp 72 to bp 1232) encoding a 387-amino acid polypeptide (Fig. 4). AT-32 full-length cDNA sequence was searched in NCBI (http://www.ncbi.nlm.nih.gov/), there were two matched linear genomic DNA sequences and no any matched mRNA sequence.

```
1    ggcacgagga tgaggaagag tgggacaggg tgtgggccaa tgtggggaag agcctgaact
61   gcattattgc tatggtggac aaactgattg aaagagatgg tggcagtgaa ggcagtggcg
                M  V  D  K  L  I  E  R  D  G  G  S  E  G  S  G
```

121 gcaacaatga tggagaaaag gaaccttcat taacagatgc cattccctct cacccaagag

G N N D G E K E P S L T D A I P S H P R

181 aggactggta tgaacagttg tatcccctca tccttaccct gaaggactgc atgggagaag

E D W Y E Q L Y P L I L T L K D C M G E

241 tggtgaaccg agccaagcag tccctgacat ttgtgctcct tcaggaactt gcgtacagct

V V N R A K Q S L T F V L L Q E L A Y S

301 tgccccagtg tctgatgctg acgctaagaa gagacatcgt cttcagccaa gcacttgctg

L P Q C L M L T L R R D I V F S Q A L A

361 gattggtttg tggttttatc atcaaattac agacaagtct gtatgaccca ggcttcctac

G L V C G F I I K L Q T S L Y D P G F L

421 agcagcttca cacagtgggg ttgatagtac aatatgaagg actgctaagt acatacagcg

Q Q L H T V G L I V Q Y E G L L S T Y S

481 atgaaattgg aatgctagag gacatggccg ttggcatttc cgatttaaag aaagtcgcat

D E I G M L E D M A V G I S D L K K V A

541 ttaaaataat tgaagccaaa tccaatgatg tgttgccagt tataacagga agacgagaac

F K I I E A K S N D V L P V I T G R R E

601 attacgtggt agaggtcaag cttccagcca gaatgtttga gtcactacct ctacagatta

H Y V V E V K L P A R M F E S L P L Q I

661 aagaaggaca gttgcttcat gtgtatccag tacttttaa tgttggaatc aatgaacagc

K E G Q L L H V Y P V L F N V G I N E Q

721 aaactctggc tgaaaggttt ggagatgtct ctttgcaaga aagtattaat caggaaaact

Q T L A E R F G D V S L Q E S I N Q E N

781 tcgaacttct acaagaatat tacaagatat ttatggaaaa gatgcctcct gattatattt

F E L L Q E Y Y K I F M E K M P P D Y I

841 cacattttca ggaacaaaat gatttaaaag cattgctaga aaatctcctt caaaatatcc

S H F Q E Q N D L K A L L E N L L Q N I

901 aatccaaaaa aagaaagaat gtagaaatta tgtggctggc tgcaacgatt tgccgcaaac

Q S K K R K N V E I M W L A A T I C R K

961 tgaatggtat tcgtttcacc tgttgtaaaa gtgccaaaga caggacatcg atgtcagtga

L N G I R F T C C K S A K D R T S M S V

1021 cacttgaaca atgctcaatc ttgagagatg agcaccagtt acacaaggac ttctttatcc

```
      T  L  E  Q  C  S  I  L  R  D  E  H  Q  L  H  K  D  F  F  I
1081  gagcgctgga ttgcatgaga agagaaggat gccgcataga gaatgtactg aagaatatca
      R  A  L  D  C  M  R  R  E  G  C  R  I  E  N  V  L  K  N  I
1141  aatgcagaaa gtatgctttc aacatgctac agctgatggc tttccccaag tactacagac
      K  C  R  K  Y  A  F  N  M  L  Q  L  M  A  F  P  K  Y  Y  R
1201  ctccagaggg gacttatgga aaagctgaca cctaagttta ccaacatgtt aataaacagg
      P  P  E  G  T  Y  G  K  A  D  T
1261  aacacaaata catttcagtt ggataatctt caccttggtc ttttttgttt gtttttattg
1321  tcatgaattt gtggtggggg agatgatcac agatgtttcc caaaatctag gaactacttc
1381  aattcatcat cagcatattg ctctgaaaga aatccatttt aaaaaatctt acatgattca
1441  gtactgtcat ttctagttct aagcttctat gtgttgagcc tagctgattg agattggttc
1501  acaaagtgta aatttcattc atgagtattt aaatgtttct gaggaaacat ggatcatttc
 …     …     …     …     …     …
6541  catttcactg catacataca atactttgta tacacatact tctcgtgtgg atattttata
6601  attttattta atctaacatt taacttttag gtttttgtgg tactcatctg aaaggttggc
6661  atttgttttc tttttatgtt atatatgtat tttagatgaa tcaataatgt gtattcctag
6721  aaatttttttc taggcataaa tagatcattt tatgttggca tttaagagat atcatttttg
6781  ttttattgtt ctgtgtaaac catattcatt gtacactgtg tgtatatttt tgtttaaata
6841  aatgattttt attttttaaa aaaaaaaaaa aaaaa
```

Figure 4. Nucleotide full-length sequence and putative open reading frame of AT-32 cDNA. The sequence was searched in Vector NTI, the putative ORF was underlined. The ATG start codon, TAA stop codon, and poly(A) tail were in bold. The deduced amino sequences were shown under the corresponding ORF.

6. In Silico Analysis

AT-32 full-length cDNA sequence was blasted against the Human BLAT Search (http://genome.ucsc.edu/) provided by the genome bioinformatics site of University of California, Santa Cruz. It was mapped to chromosome 4q31.21 (Fig. 5). The sequence was searched in CBS Prediction Site (http://cbs.dtu.dk/services/) and NCBI (http://www.ncbi.nlm.nih.gov/), neither significant homology with known proteins nor notable functional domains were identified from the peptide sequence. The putative ORF of AT-32 was analyzed in depth through several predict servers. The putative protein localization is any other one except chloroplast, mitochondrion, secretory pathway (Fig. 6). The predicted protein sequence was searched PredictProtein Server (http://cubic.bioc.columbia.edu/predictprotein/) to find if has high similar protein function domains with previously identified genes, there was nothing found.

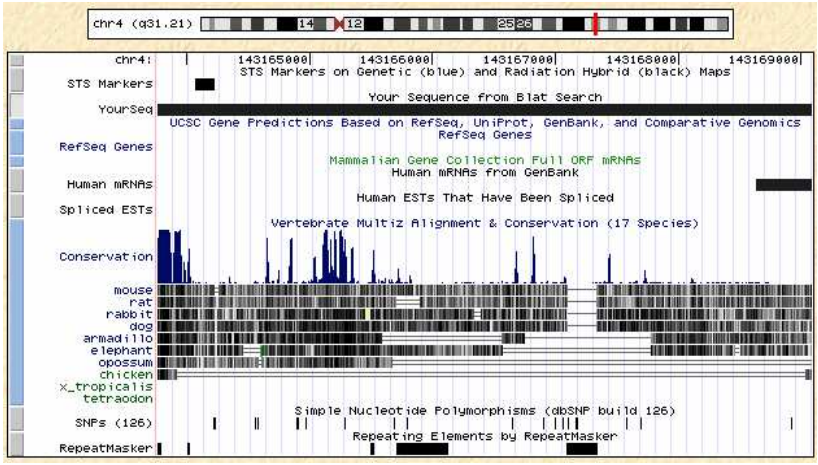
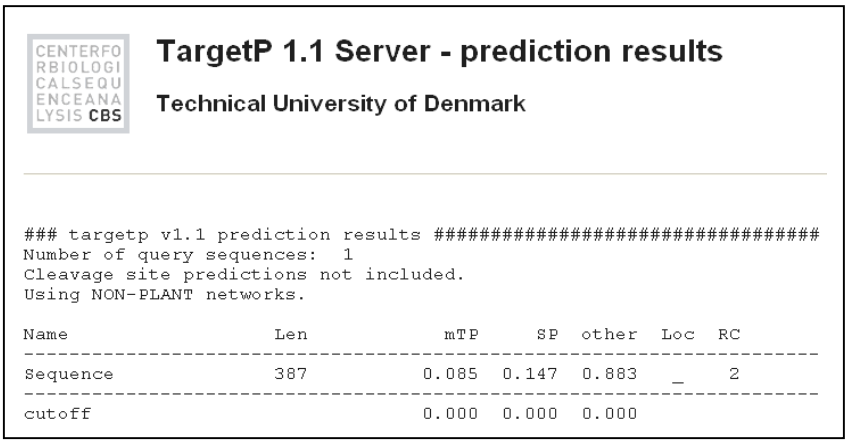


Figure 5. AT-32 chromosome mapping. Nucleotide sequence of the AT-32 cDNA was blasted against the Human BLAT Search (http://genome.ucsc.edu/) provided by the genome bioinformatics site of University of California, Santa Cruz. It was mapped to chromosome 4q31.21.



TargetP 1.1 Server - prediction results

Technical University of Denmark

```
### targetp v1.1 prediction results ###################################
Number of query sequences:  1
Cleavage site predictions not included.
Using NON-PLANT networks.

Name                   Len            mTP     SP  other  Loc  RC
----------------------------------------------------------------------
Sequence               387          0.085  0.147  0.883   _    2
----------------------------------------------------------------------
cutoff                              0.000  0.000  0.000
```

Figure 6. Analysis of AT-32 putative localization. We analyzed the putative ORF of AT-32 in CBS Prediction Site (http://cbs.dtu.dk/services/) TargetP. Loc – means any other one except chloroplast, mitochondrion, secretory pathway. RC means reliability class, from 1 to 5, 1 indicates the strongest prediction.

7. AT-32 is expressed in a various pancreatic cancer cell lines.

RT-PCR was carried out using 12 RNA samples including RNAs from 8 pancreatic cancer cell lines, human pancreatic duct epithelial cell lines, human YGIC-6, peripheral blood monocytes and bone marrow stromal cells to confirm the microarray data (Fig. 7).

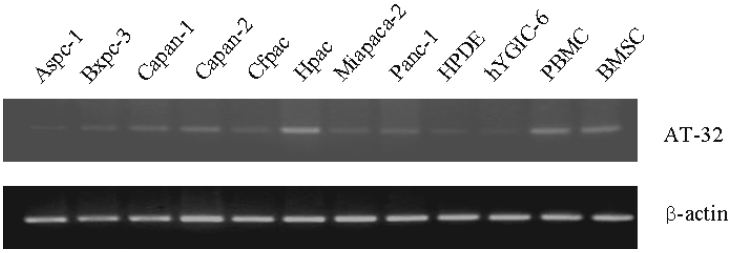


Figure 7. RT-PCR analysis of AT-32 mRNA expression in a various pancreatic cancer cell lines. Standard RT-PCR was conducted using RNA prepared from 8 pancreatic cancer cell lines, human pancreatic duct epithelial cell line, human YGIC6 cell, peripheral blood monocytes and bone marrow stromal cells. RT-PCR reactions were performed under hot-start conditions of 5 min at 94℃ and then 25 cycles of 30 sec at 94℃, 30 sec at 56℃ and 30 sec at 72℃; followed by a final 5 min extension at 72℃. The human β-actin gene was used as a control to confirm the integrity of the mRNA samples.

Ⅳ. DISCUSSION

Pancreatic cancer is a highly aggressive type of malignancy and the prognosis for disease presenting typically at a late stage is extremely poor.[13] Gene expression technologies are increasingly being used to understand the complexities of transcriptional regulation and dysregulation in normal and cancerous tissues. One important application of these methods is for the rapid identification of differentially expressed genes in neoplastic tissues as compared with their normal counterparts.[8]

The identification of genes differentially expressed in pancreatic cancer relative to normal tissues provides a basis for the development of novel strategies to detect and treat this highly lethal cancer.[7,11]

Affymetrix, Inc., has produced the UG-133 GeneChip based on Build 133 of the human genome. One particular advantage of the UG-133 chip is that many transcripts previously identified only as expressed sequence tags (ESTs) have now been characterized, allowing for a more comprehensive view of gene expression in samples studied.[11] As a result of progress on the human genome project, approximately 19,000 genes have been identified and tens of thousands more tentatively identified as partial fragments of genes termed

ESTs. Most of these genes are only partially characterized and the functions of the vast majority are as yet unknown.

In this study, bioinformatic analyses were used to characterize those ESTs that were found to be overexpressed in pancreatic adinocarcinomas. Total RNA was prepared from 14 normal pancreas and from 8 pancreatic cancers and hybridized to the complete Affymetrix Human Genome U133 GeneChip set for simultaneous analysis of 45,000 fragments corresponding to 33,000 known genes and 6,000 ESTs. 532 ESTs were overexpressed in cancer tissues and at least 50 ESTs out of them were identified that were expressed at levels at least 2-fold greater in the pancreatic cancer cell lines.[13] Some of these ESTs may represent diagnostically and therapeutically useful targets that might be missed using data solely from currently annotated databases. Subsequently, in order to pick the most relevant candidate genes for a more detailed analysis, full-length cDNA were isolated and sequenced. To search for unidentified cancer-causing genes, the mRNA expression profiles of pancreatic adenocarcinoma from patients' tissues were compared with the profiles form normal tissues. Comparisons of the gene expression profiles between pancreatic adenocarcinoma tissues with normal pancreas tissues showed significant and consistent expression changes.

Up-regulated and unknown genes were focused in order to uncover new

genes responsible for the process of tumor formation. 64 ESTs were selected from this intense screening using bioinformatics tools and several candidates were finally chosen for construct full-length cDNA.

AT-32 was identified from human placenta cDNA library (OriGene Technologies, Inc., Rockville, USA). The target clone was sequenced and constructed full-length cDNA. It contained insert of 6875bp, had one putative open reading frame. Putative ORF (from bp 72 to bp 1232) encoding a 387-amino acid polypeptide (Fig. 4). The predicted protein sequence was searched in several public databases such as NCBI, Human BLAT Search, SWISS-MODEL, ExPASy Proteomics Server, CBS Prediction Site and Predictprotein Server. AT-32 was mapped to chromosome 4q31.21 (Fig. 5) and the putative protein localization is any other one except chloroplast, mitochondrion, secretory pathway (Fig. 6). Neither significant homology with known proteins nor notable functional domains were identified from the peptide sequence. AT-32 was 4-fold up-regulated specifically in pancreatic cancer tissue, while no obvious difference or down-regulated in the other human tissues from genechip data (Fig.2A). This condition is beneficial to serve as a diagnostic marker of pancreatic cancer.

In this study, we attempted to identify some novel genes that highly expressed in pancreatic cancer tissues compared with normal tissues. As a

result, one putative novel gene AT-32 was identified. The identification of genes differentially expressed in pancreatic cancer is critical to the development of novel therapeutics and new markers to detect this disease at an earlier, potentially curable stage.[6]

Ⅴ. CONCLUSION

Among 8 ESTs were attempted to construct their full-length cDNAs which are highly expressed in pancreatic ductal adenocarcinoma, cloning of full-length cDNA from an EST was successful. In this way, one putative novel gene, AT-32, highly expressed in pancreatic cancer was identified. Even though further investigations such as northern blot analysis, Quantitative reverse transcriptase-polymerase chain reaction, protein expression, and functional analysis would be performed to know the role of pancreatic carcinogenesis, this novel gene may form the basis for the development of diagnostic marker or screening method to detect this disease at an earlier, potentially curable stage or may serve as a novel target for drug development or imaging.

REFERENCES

1. Li D, Xie K, Wolff R, Abbruzzese JL. Pancreatic cancer. Lancet 2004 Mar 27;363(9414):1049-57.

2. Wang H, Han H, Von Hoff DD. Identification of an Agent Selectively Targeting DPC4 (Deleted in Pancreatic Cancer Locus 4)-Deficient Pancreatic Cancer Cells. Cancer Res 2006 Oct 1;66(19):9722-30.

3. DiMagno EP, Reber HA, Tempero MA. AGA technical review on the epidemiology, diagnosis, and treatment of pancreatic ductal adenocarcinoma. American Gastroenterological Association. Gastroenterology 1999 Dec;117(6):1464-84.

4. Zervos EE, Rosemurgy AS, Al-Saif O, Durkin AJ. Surgical management of early-stage pancreatic cancer. Cancer Control 2004 Jan-Feb;11(1):23-31.

5. Iiizumi M, Hosokawa M, Takehara A, Chung S, Nakamura T, Katagiri T, et al. EphA4 receptor, overexpressed in pancreatic ductal adenocarcinoma, promotes cancer cell growth. Cancer Sci 2006 Nov;97(11):1211-6.

6. Crnogorac-Jurcevic T, Efthimiou E, Nielsen T, Loader J, Terris B, Stamp G, et al. Expression profiling of microdissected pancreatic adenocarcinomas. Oncogene 2002 Jul 4;21(29):4587-94.

7. Han H, Bearss DJ, Browne LW, Calaluce R, Nagle RB, Von Hoff DD.

Identification of differentially expressed genes in pancreatic cancer cells using cDNA microarray. Cancer Res 2002 May 15;62(10):2890-6.

8. Iacobuzio-Donahue CA, Maitra A, Shen-Ong GL, van Heek T, Ashfaq R, Meyer R, et al. Discovery of novel tumor markers of pancreatic cancer using global gene expression technology. Am J Pathol 2002 Apr;160(4):1239-49.
9. Logsdon CD, Simeone DM, Binkley C, Arumugam T, Greenson JK, Giordano TJ, et al. Molecular profiling of pancreatic adenocarcinoma and chronic pancreatitis identifies multiple genes differentially regulated in pancreatic cancer. Cancer Res 2003 May 15;63(10):2649-57.
10. Ryu B, Jones J, Blades NJ, Parmigiani G, Hollingsworth MA, Hruban RH, et al. Relationships and differentially expressed genes among pancreatic cancers examined by large-scale serial analysis of gene expression. Cancer Res 2002 Feb 1;62(3):819-26.
11. Tan ZJ, Hu XG, Cao GS, Tang Y. Analysis of gene expression profile of pancreatic carcinoma using cDNA microarray. World J Gastroenterol 2003 Apr;9(4):818-23.
12. Iacobuzio-Donahue CA, Ashfaq R, Maitra A, Adsay NV, Shen-Ong GL, Berg K, et al. Highly expressed genes in pancreatic ductal adenocarcinomas: a comprehensive characterization and comparison of the

transcription profiles obtained from three major technologies. Cancer Res 2003 Dec 15;63(24):8614-22.

13. Pao SY, Lin WL, Hwang MJ. In silico identification and comparative analysis of differentially expressed genes in human and mouse tissues. BMC Genomics 2006;7:86.

14. Liu PG, Yang Q. Identification of genes with a biocontrol function in Trichoderma harzianum mycelium using the expressed sequence tag approach. Res Microbiol 2005 Apr;156(3):416-23.

15. Cao D, Hustinx SR, Sui G, Bala P, Sato N, Martin S, et al. Identification of novel highly expressed genes in pancreatic ductal adenocarcinomas through a bioinformatics analysis of expressed sequence tags. Cancer Biol Ther 2004 Nov;3(11):1081-9; discussion 90-1.

16. Weindruch R, Kayo T, Lee CK, Prolla TA. Gene expression profiling of aging using DNA microarrays. Mech Ageing Dev 2002 Jan;123(2-3):177-93.

17. Missiaglia E, Blaveri E, Terris B, Wang YH, Costello E, Neoptolemos JP, et al. Analysis of gene expression in cancer cell lines identifies candidate markers for pancreatic tumorigenesis and metastasis. Int J Cancer 2004 Oct 20;112(1):100-12.

ABSTRACT (IN KOREAN)

췌장암 관련 신규 유전자의 발굴

<지도교수 송 시 영>

연세대학교 대학원 의과학과

김 애 화

췌장암은 1년 생존율이 10%에 불과하고 5년 생존율이 1%에 불과하여 인체에 발생하는 다양한 암들 중에서 예후가 가장 불량한 암중 하나이다. 서양에서는 암환자 사망률의 4위를 차지하고 한국에서도 최근 10년간 2배의 발생율 급증을 보이고 있다. 암세포의 경우 세포의 유전자 중 일부에 이상이 발생하여 이들 유전자의 산물인 단백질의 특성이 바뀌고, 그 결과로 세포 성장 조절에 이상이 발생한다. 따라서 암의 원인, 성장 및 전이에 관여하는 주된 요인의 분자 생물학적 특성을 규명하는 것이 중요한데, EST를 통한 신규 유전자 발굴을 통해 새로운 표적 분자를 발견할 가능성이 높을 것으로 생각된다. 본 연구는 췌장암 환자들의 유전자 발현패턴을 microarray 분석자료를 바탕으로, 췌장암에서 높게 발현되는 미지의 유전자 조각(EST)을 선

별, cloning하여 췌장암 관련 신규 유전자를 규명하는 것이다. 본 연구에서 수술로 절제된 정상조직과 암조직의 RNA와 Affymetrix GeneChip을 이용하여 oligonucleotide microarray를 시행하였다. Chip data를 MatLab 통계프로그램으로 p-value와 fold change 등을 계산하여, 유전자 발현 양상을 분석하였다. 532개의 EST 조각(p-value < 0.05, fold change > 1.2) 이 정상 췌장조직에 비해 췌장암에서 과발현 되였다. 이들 중 p-value가 0.01 이하이고, fold change가 2.0 이상, 6.0 이하인 EST 조각 101개를 선별하였다. 101개 EST 조각을 Affymetrix, Human Genome DataBase 와 NCBI database 의 검증을 거쳐, 현재까지 기능이 알려지지 않은 64개 EST를 최종 선택하였고, 이들의 full length cDNA cloning을 실시하였다. Human Placenta cDNA Library로부터 각각의 gene-specific primer을 이용하여 타겟 EST 클론을 찾았고, 시퀀싱하였다. 그 결과, 전체 cDNA length가 6875bp 되는, AT-32라 명명한 신규 유전자를 확보하였다. 여러 가지 public database와 생물정보학을 이용하여 후보 유전자 AT-32의 염기서열 분석 및 단백질 서열을 분석하여 그들의 예상 기능을 일차적으로 도출하였다. AT-32는 1161bp (from bp 72 to bp 1232) 크기의 open reading frame 을 갖고 있으며, 이는 387-amino acid 의 예상 단백질을 coding한다.

특히 췌장암의 경우, 현재까지 이들 질환을 조기에 진단하거나 효과적으로 치료하기 위한 뚜렷한 방법이 없기 때문에 보다 효과적인 조기진단법 및 치료법의 개발이 요망되고 있다. 본 연구에서 발굴한 신규 유전자 AT-32는 향후 다양한 기능 관련 연구가 진행될 것이고, 종양발생과 발전의 mechanism의 규명에 중요한 정보를 제공할 것이라 생각된다. 또한 췌장암 조기진단 marker나 췌장암 치료 target 개발에도 정보를 제공할 것이다.

핵심되는 말 : 췌장암, microarray, est, 신규 유전자, 발굴