

## Human calgizzarin; one colorectal cancer-related gene selected by a large scale random cDNA sequencing and Northern blot analysis

Masaki Tanaka<sup>a</sup>, Kyo Adzuma<sup>b</sup>, Miki Iwami<sup>b</sup>, Katsuhiko Yoshimoto<sup>b</sup>,  
Yasumasa Monden<sup>a</sup>, Mitsuo Itakura<sup>\*b</sup>

<sup>a</sup>*The Second Department of Surgery, School of Medicine, The University of Tokushima, Kuramoto, Tokushima-city, Tokushima 770, Japan*

<sup>b</sup>*Otsuka Department of Clinical and Molecular Nutrition, School of Medicine, The University of Tokushima, Kuramoto, Tokushima-city, Tokushima 770, Japan*

Received 24 November 1994; revision received 21 December 1994; accepted 22 December 1994

---

### Abstract

A cDNA library was constructed from COLO 205 and 1056 clones randomly selected from this library were partially sequenced. Two hundred and two (38.4%) out of 526 independent genes had more than 80% similarity to the genes reported in GenBank. In Northern blot analysis, 96 out of 98 genes were shown to be expressed at the same level in colon and lung carcinoma cell lines and control fibroblasts. Only two clones, including human synovial phospholipase A-2 and a homologue to rabbit calgizzarin, were expressed at different levels among these cell lines. The full sequence of human calgizzarin was determined and its expression was remarkably elevated in colorectal cancers compared with that in normal colorectal mucosa.

**Keywords:** Human calgizzarin; Random cDNA sequencing

---

### 1. Introduction

A mammalian cell is assumed to express as many as 15 000 to 30 000 species of genes. Some genes expressed in cancer cells are expected to be different from those expressed in normal cells, because the morphology and function of cancer cells

are highly different from those of normal cells. Little is known about how many and what genes are differently expressed in cancer cells compared with normal cells. In cancers of colon, lung, breast, etc. point mutations or deletions of p53 and *ras* genes were detected [1]. Genetic changes of tumor cells are supposed to include not only such mutated genes, but also many other genes of which expression is different from that in normal cells.

Recently, large scale random cDNA sequencing

---

\* Corresponding author, Tel.: +81 886 319804; Fax: +81 886 319476.

has become a valuable approach to examine expressed genes in various cells and tissues by determining expressed sequence tags [2–5]. We performed a random cDNA sequencing approach to a human colon carcinoma cell line to obtain clones expressed in carcinoma cell lines. Northern blot analysis was performed to select those genes of which expression was different in different carcinoma cell lines and cultured fibroblasts were used as a control.

We report the result of random cDNA sequencing of a total of 1056 clones from a colon carcinoma cell line of COLO 205 [6] and the result of Northern blot analysis of 140 independent clones.

## 2. Materials and methods

### 2.1. Cell preparation

COLO 205 (a human colon carcinoma cell line), CW2 (a human colon carcinoma cell line), hNCI-H520 (a human lung squamous carcinoma cell line) and TIG-1 (a human fibroblast cell line) [7] were cultured up to 70–80% confluency in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum (FCS).

### 2.2. cDNA library construction

RNA was extracted from COLO 205 cells by acid guanidinium thiocyanate-phenol-chloroform (AGPC) method [8]. Five micrograms of poly(A)<sup>+</sup> RNA were incubated with 1 nmol of *Xho*I-(dT)<sub>15</sub> primer (ATTAACACTCG-AGTTT-TTTTTTTTTTTT), 1 mM 2-mercaptoethanol, 500 μM each of dNTPs, 200 units of RNasin (Promega), 100 units of AMV reverse transcriptase (Seikagaku Corporation), 50 mM Tris-HCl (pH 8.5), 8 mM MgCl<sub>2</sub>, 30 mM KCl at 37°C for 1 h. Then, it was incubated with 50 units of *Escherichia coli* DNA polymerase I, 3 units of RNase H, 500 μM each of dNTPs, 50 mM Tris-HCl (pH 7.5), 10 mM MgCl<sub>2</sub>, 1 mM 2-mercaptoethanol and 90 mM KCl at 14°C for 1 h, then 22°C for 1 h. It was incubated with 10 units of T4 DNA polymerase at 37°C for 10 min. The product was eluted from a Chroma spin column S-400 (CLONTECH) after centrifuging at 700 × g for 5 min.

An *Eco*RI adaptor was ligated to cDNA under the conditions recommended by Takara. After di-

gestion with *Xho*I, the product was eluted with a Chroma spin column. The 5' end of cDNA was phosphorylated with 10 units of T4 polynucleotide kinase and 100 μM of ATP at 37°C for 30 min. One microgram of λ-ZAPII arms (Stratagene) double-digested with *Xho*I and *Eco*RI was ligated to the cDNA using a Takara ligation kit. Phage packaging was done with the Gigapack II Gold (Stratagene), following the protocol recommended by the manufacturer.

### 2.3. Converting phage clones en masse to pBluescripts

The *E. coli* strain XL1-Blue infected with about 10 000 phage clones were incubated in LB top agar at 37°C overnight. Eluted λ-ZAPII phages were converted en masse to pBluescript SK (-), using R408 helper phage, following the protocol recommended by Stratagene.

### 2.4. Random cDNA sequencing

Randomly selected colonies were incubated in LB medium with 50 μg/ml of ampicillin at 37°C overnight. The pBluescript DNA was extracted with an alkaline lysis method. A total of 1056 randomly selected clones were partially sequenced both from 3' and 5' end of the cDNA with ABI 373A autosequencer (Perkin Elmer Cetus; Applied Biosystems Division) with a dye-labeled T7 and T3 primer, following the manufacturer's protocol, respectively.

### 2.5. Northern blot analysis

RNA was extracted by AGPC method from COLO 205, CW2, hNCI-H520, TIG-1. RNA was also extracted from human colorectal mucosa and colorectal carcinoma tissues, which were surgically resected at our department during the period from April 1991 to September 1994 with informed consent. Twelve micrograms each of total RNA were electrophoresed through a 1% denaturing agarose gel and were transferred to Hybond-N (Amersham). To make probes, each of diluted pBluescript inserts was amplified by PCR at 94°C for 1 min, 50°C for 1 min, 72°C for 2 min, after 40 cycles with 30 pmol of SK and T7 primers. Free nucleotides were removed using glass powder. The labeling by α-<sup>32</sup>P dCTP was performed by the

multiprime labeling system (Amersham). The hybridization was performed in 50% formamide, 5 × SSPE buffer, 5 × Denhardt's reagent and 0.5% SDS at 42°C overnight. Membrane was washed with 1 × SSPE and 0.1% SDS at 65°C for 30 min three times and exposed for autoradiography on Kodak XAR film for 6–72 h at –80°C.

### 3. Results

#### 3.1. Analysis of randomly sequenced cDNA

A cDNA library was constructed from a COLO 205 cell line. The sizes of inserts of randomly se-

lected clones ranged from zero to about 3500 base pairs (bp) with an average of 700 bp. The randomly selected clones were bidirectionally sequenced both from the 3'- and 5'-end with a dye-labeled T7 and T3 primer, respectively. The average length of compiled sequences of inserts were 280 bp. Finally, a total of 1056 clones were sequenced from the COLO 205 library. The sequence similarities of 1056 clones were firstly compared with each other and secondly with the DNA data base in GenBank encompassing mammalian genes, using a FASTA program [9,10]. We obtained 526 independent species. Genes which have more than

Table 1

Summary of the random cDNA sequencing. Representative known genes with more than 80% similarity and their frequencies (fq.) of appearance in the random cDNA sequencing are listed

Genes	fq.	Genes	fq.
<b>Secretory protein (2 species)</b>		<b>Protein synthesis (51 species)</b>	
synovial phospholipase A-2	5	18S ribosomal RNA	38
pancreatic trypsinogen III	2	initiation factor 4A1	10
<b>Cytoskeleton (10 species)</b>		ribosomal protein L35	9
cytokeratin 18	5	ribosomal protein L41	7
β-actin	3	ribosomal protein S11	7
cytokeratin 8	3	ribosomal protein L19	6
cytoskeletal type II keratin	2	ribosomal protein S6	5
α-tubulin	2	ribosomal protein L27	5
<b>Membrane protein (11 species)</b>		<b>Cytoplasm and others (66 species)</b>	
MHC protein homologous to chicken	10	ubiquitin	17
B complex protein		BBC1 mRNA	16
laminin-binding protein	3	α-enolase	8
cathepsin D	1	T-cell cyclophilin	5
connexin 26	1	polyA binding protein	4
<b>Nuclear protein (7 species)</b>		glyceraldehyde-3-phosphate dehydrogenase	4
snRNP protein B	6	valosin-containing protein	4
hnRNP core protein A1	3	glutaminyl-tRNA synthetase	3
DNA-binding protein B	2	menadione oxidoreductase	3
75-kDa autoantigen (PM-Scl)	2	calreticulin	3
cAMP response element regulatory protein	1	23 kDa highly basic protein	2
<b>Mitochondria (23 species)</b>		71 kDa heat shock cognate protein	2
16S rRNA	65	calpactin 1 light chain	2
cytochrome oxidase subunit 3	17	calmodulin	2
cytochrome c oxidase subunit 2	16	glutathione peroxidase-G1	2
12S rRNA	12	argininosuccinate synthetase	1
cytochrome b	11	calcium dependent protease	1
cytochrome oxidase subunit 1	7	aldolase A	1
ATPase 6	4	ferritin heavy subunit	1
		ornithine decarboxylase	1
		F1-ATPase β subunit	1

90%, 80% and 70% similarity with mammalian genes reported in GenBank (Release 82, June, 1994) were 171 (32.5%), 202 (38.4%) and 226 (43.0%) out of 526 species, respectively; 366 (69.6%) out of 526 species appeared once in the random sequencing. The number of genes which appeared more than 5 and 10 times in random sequencing were 31 (5.9%) and 14 (2.7%) out of 526 species, respectively.

Known genes with more than 80% similarity were classified into secretory protein, membrane protein, nuclear protein, mitochondria, protein synthesis and cytoplasm and others which included 2, 11, 7, 23, 51 and 66 species, respectively. Representative genes are listed in Table 1.

### 3.2. Northern blot analysis

Gene expression and sizes of transcripts in COLO 205, CW-2, hNCI-H520 and TIG-1 were examined by Northern blot analysis (Fig. 1A). A total of 140 randomly selected genes from 526 species were used as <sup>32</sup>P-labeled probes. A total of 96 probes (68.6%) showed positive signals within 72 h of autoradiographic exposure. Signal intensity of Northern blot was compared among COLO 205, CW-2, hNCI-H520 and TIG-1. Most genes hybridized to four cell lines at the same intensity. Only two clones showed signals at different intensity. Clone 0133 showed stronger signals to COLO 205 and TIG-1 than to CW-2, but no signals to hNCI-H520. Clone 0634 showed a much stronger

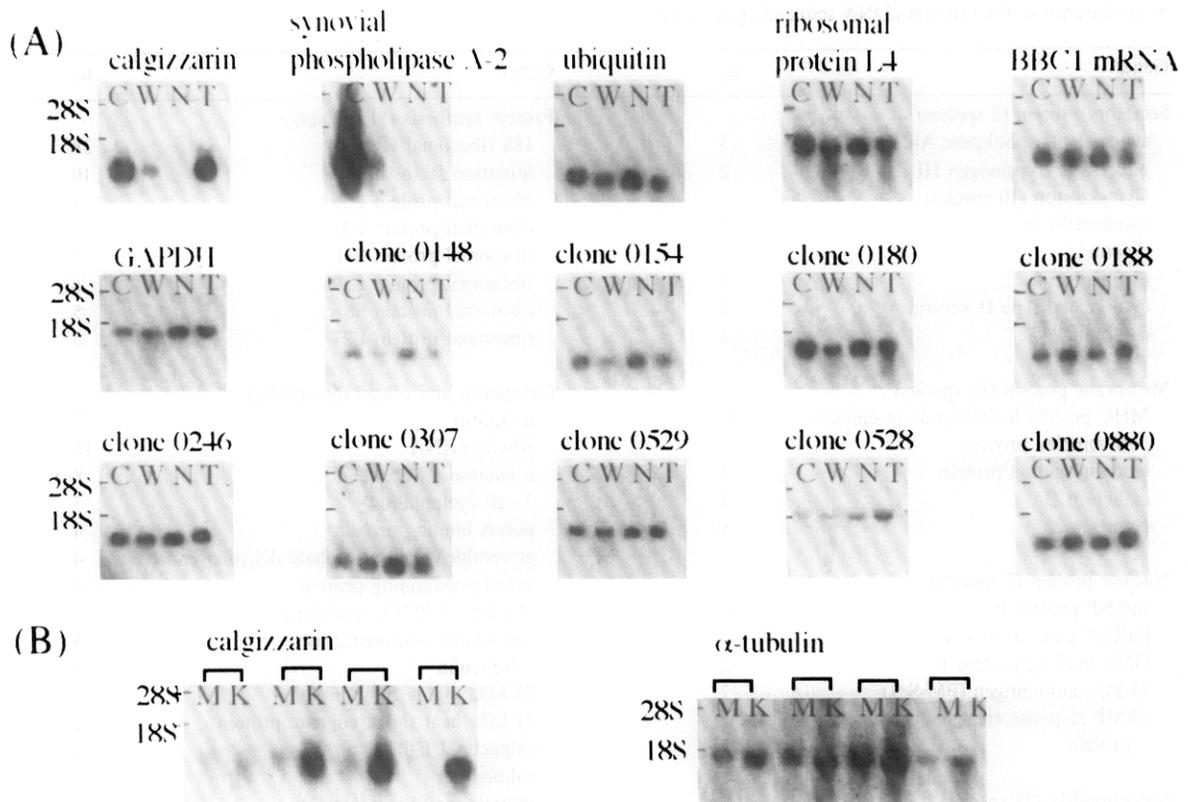


Fig. 1. Northern blot analysis to obtain cancer cell-related clones. (A) 140 independent clones were hybridized with 12 mg of total RNA of COLO 205, CW-2, hNCI-H520 and TIG-1, as abbreviated, C, W, N and T, respectively. Exposure was for 2–72 h. A representative 15 clones are shown. Calgizzarin and synovial phospholipase A-2 were differently expressed in different cells, although others showed signals at the same intensities. GAPDH stands for glyceraldehyde-3-phosphate dehydrogenase. (B) Calgizzarin was hybridized with total RNA of four cases of human normal mucosa and colorectal cancer tissues, abbreviated as M and K, respectively. As a control,  $\alpha$ -tubulin was rehybridized. The signals in cancer tissues were much stronger than those in normal colorectal mucosa in three out of four different colorectal cancer samples with corresponding mucosa.

(A)

```

GGGCAAGGCTGGGCCGGGAAGGGCGTGGGTTGAGGAGAGGCTCCAGACCCGACGCCC86 60
CGCACAGAGCTCTCAGCCGCTCCAGCCACAGACTCCCGC8CCTCGCTCAGTCCAAAC 120
ATGGCAAAAATCTCCAGCCCTACAGAGACTGAGCGGTGCATC8AGTCCCTGATTCGTCTC 180
M A K I S S P T E T E R C I E S L I A V
TTCCAGAAGTATGCTGGAAAGGATGGTTATAACTACACTCTCTCCAAGACAGAGTTCCTA 240
F Q K Y A G K D G Y N Y T L S K T E F L
AGCTTCATGAATACAGAACTAGCTGCTCCACAAAAGACAGAAGGACCCCTGGTGCCTT 300
S F M N T E L A A F T K N Q K D P G V L
GACCGCATGATGAAGAACTGGACACCAACAGTGTGGTGCAGTATTTCTCAGAAATT 360
D R M M K K L D T N S D G Q L D F S E F
CTTAATCTGATGGTGGCTAGCTATGGCTTGCATGACTCCTTCTCAAGGCTGTCCCT 420
L N L I G G L A M A C H D S F L K A V P
TCCAGAAAGCGGACTGAGGACCCCTTGGCCCTG8CCTTCAAACCCACCCCTTCTCCTC 480
S Q K R T *
CAGCCTTTCTGTCATCATCTCCAGAGCCACCCATCCCTCGAGCAGACTAACCACCTCAT 540
GCAGGCCCCACCTGCCAATAGTAAATAAGCAATGTCACTTTTTTAAACATGAAAAAAA 600

```

(B)

```

human MAKISSPTETERCIESLIAVFKYAGKDGYNLYLSKTEFLSFMNTELAFTKNQKDP6VL
      . * .*****.*****.*****
rabbit NSRPTETERCIESLIAVFKYAGKDGHSVTLKTEFLSFMNTELAFTKNQKDP6VL

human DRMMKKLDLNSDGLDFEFLNLIIGGLAMACHDSFLKAVPSQKRT
      *****
rabbit DRMMKKLDLNSDGLDFEFLNLIIGGLAVACHESFVKAAPPQKRF

```

Fig. 2. A nucleotide sequence and predicted amino-acid sequence of human calgizzarin. (A) Nucleotide sequence of human calgizzarin is shown and a poly adenylation signal is underlined. A predicted amino-acid sequence is described in single-letter amino-acid nomenclature with \* for a stop codon. (B) Amino-acid sequence of human calgizzarin was compared with that of a rabbit.

signal to COLO 205 than to CW-2, without any signals to hNCI-520 and TIG-1. Clone 0634 was human synovial phospholipase A-2 [11] and clone 0133 had 80% homology to rabbit mRNA for calgizzarin [12].

Clone 0133 of human calgizzarin contained 592 bp excluding (dA)<sub>32</sub> and poly A adenylation signal (Fig. 2) and was registered in the DNA data bank of Japan with the accession number D38583. It was more abundantly expressed in colorectal cancer tissues than in normal colorectal mucosa (Fig. 1B).

#### 4. Discussion

Five hundred and twenty-six species of various

genes were obtained in our random sequencing of 1056 clones, though some genes appeared multiple times. Except ribosomal RNAs, most genes appeared once and genes found more than a few times in random sequencing were extremely rare. It is suggested that a great variety of genes are expressed in a cancer cell in relatively low abundance in vivo.

Northern blot analysis was performed to examine the difference of gene expression in other cell lines. Only two (2.1%) out of 96 genes were differently expressed among colon carcinomas, a lung squamous carcinoma and a fibroblast cell line. Based on this frequency, if we assume that a cell expresses 20 000 species of genes, about 400 genes are estimated to be differently expressed in each cell line. Most other genes are supposed to be house-keeping genes. This is consistent with the concept that transcripts specific to a certain cell are relatively small in number and that most genes of cancer cells are expressed in a similar way to normal cells.

The expression of calgizzarin was much higher in colorectal cancer than in normal colorectal mucosa. Calgizzarin, which is a calcium binding protein with an EF-hand structure [12], may be related not only to rapid growth and the accelerated metabolism in cancer cells but also to an unknown specific aspect of colorectal carcinomas, though its detailed function is not clear. Synovial phospholipase A-2 was expressed much more in COLO 205 than in CW2, though they were both derived from colorectal cancers. It is suggested that the level of gene expression of synovial phospholipase A-2 is different among different cell lines and among different cases of colorectal cancers. Elevated expression of synovial phospholipase A-2 can be used as a tumor marker of colorectal cancers, because its peptide is secreted extracellularly in contrast to weak expression in normal colorectal mucosa.

Gene expression of all unknown genes found in the random cDNA sequencing is now being compared among different cancer cell lines, fibroblasts, normal colorectal mucosa and colorectal cancer tissues by Northern blot analysis. The genes, which are expressed only in cancer cells but not in normal mucosa, will be further studied in regard to their functions in addition to full-length

sequencing. It is important to obtain a group of genes of which expression in cancer cells is different from either normal fibroblasts or corresponding normal tissues to study oncogenesis and to find the therapeutic target of cancers. A random cDNA sequencing combined with Northern blot analysis should be useful to obtain and understand the roles of a group of novel cancer-related genes.

### Acknowledgments

This study was supported in part by a grant from Otsuka Pharmaceutical Factory Inc. We thank S. Hamaguchi, Y. Shuganaka and H. Kimura for their technical help.

### References

- [1] Fearon, E.R. and Vogelstein, B. (1990) A genetic model for colorectal tumorigenesis. *Cell*, 61, 759–767.
- [2] Fukushima, A., Hori, N., Kojima, Y., Matoba, R., Matsubara, K., Niiyama, T. and Okubo, K., (1992) Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. *Nature Genet.*, 2, 173–179.
- [3] Hori, N., Okubo, K., Matoba, R., Matsubara, K. and Niiyama, T. (1991) A novel system for large-scale sequencing of cDNA by PCR amplification. *DNA Sequence*, 2, 137–144.
- [4] Adams, M.D., Fields, C., Kerlavage, A.R. and Venter, C. (1993) 3,400 new expressed sequence tags identify diversity of transcripts in human brain. *Nature Genet.*, 4, 256–267.
- [5] Adams, M.D., Gocayne, M.D., Kelley, J.M., Kerlavage, A.R., McCombie, W.R., Merril, C.R., Moreno, R.F., Olde, B., Polymeropoulos, M.H., Venter, C., Wu, A. and Xiao, H. (1991) Complementary DNA sequencing: Expressed sequence tags and human genome project. *Science*, 252, 1651–1656.
- [6] Moore, G.E., Quinn, L.A., Semple, T.U. and Woods, L.K. (1978) Tumor and lymphoid cell lines from a patient with carcinoma of the colon for a cytotoxicity model. *Cancer Res.*, 38, 1345–1355.
- [7] Ayusawa, D., Kaneda, S., Koyama, H., Seno, T., Shimizu, K. and Takeishi, K. (1986) Cell-cycle-directed regulation of thymidylate synthase messenger RNA in human diploid fibroblasts stimulated to proliferate. *J. Mol. Biol.*, 190, 559–567.
- [8] Rothblum, L.I. and Xie, W. (1991) Rapid, small scale RNA isolation from tissue culture cells. *BioTechniques*, 11, 125–127.
- [9] Lipman, D.J. and Pearson, W.R. (1998) Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA*, 85, 2444–2448.
- [10] Lipman, D.J. and Pearson, W.R. (1985) Rapid and sensitive protein similarity searches. *Science*, 227, 1435–1441.
- [11] Johnson, L.K., Kloss, J., Miller, J.A., Plant, S., Pruzanski, W., Seilhamer, J.J. and Vadas, P. (1989) Cloning and recombinant expression of phospholipase A2 present in rheumatoid arthritic synovial fluid. *J. Biol. Chem.*, 264, 5335–5338.
- [12] Ando, Y., Hiroyoshi, H., Minami, H., Todoroki, H. and Watanabe, M. (1991) Molecular cloning and sequencing of a cDNA clone encoding a new calcium binding protein, named calgizzarin, from rabbit lung. *Biochem. Biophys. Res. Commun.*, 181, 644–649.