

Bioinformatics Analysis and Validation of the Expressed Sequences Tag in Human Colorectal Adenocarcinoma

Yao Chen^{a, b}, Chun Le Zhang^a, Yong Qiang Shen^a, Li Cheng Wang^a

Abstract

Background: This study was to investigate some new pathological genes in colorectal adenocarcinoma of human.

Methods: Human colorectal adenocarcinoma tissues and normal colorectal tissues were taken and suppression subtractive hybridization (SSH) and cDNA microarray techniques were employed. From differentially expressed 86 expressed sequence tags (EST), 10 EST of the SSH were selected as seed sequence for bioinformatics analyses, semi-quantitative RT-PCR and PCR-sequencing. Each lane of semi-quantitative RT-PCR was analyzed by Q1 software.

Results: Among these 10 EST, it has been found that ES274070, ES274071, ES274077, ES274076 and ES274081 may play role in the onset of colorectal adenocarcinoma in human.

Conclusions: The ES274070, ES274071, ES274076 and ES274081 are related to the onset of human colorectal adenocarcinoma.

Keywords: Bioinformatic analyses; Expressed sequence tag; Colorectal adenocarcinoma

Introduction

The initiation and development of colorectal adenocarcinoma are multistep process that is involved multi-gene

changes. During this process, cell division and differentiation are abnormally regulated by the many genes [1]. It will be benefit to the early diagnosis, effective treatment and prevention if molecular pathological mechanism of human colorectal adenocarcinoma are elucidated. With the accomplishment of human genome program, genome research has entered into a new phase of gene abstraction and data analysis. Cloning novel genes by means of bioinformatics has become a new strategy [2].

This study aims to screen and clone the genes related to genesis of human colorectal carcinoma and further study their function and relationship with the colorectal adenocarcinoma. We selected 10 expressed sequence tags (EST) from suppression subtractive hybridization (SSH) combining cDNA microarray as seed sequences to extend them in order to acquire full-length cDNA and validate expression by RT-PCR. It will be helpful for us to further study their function; moreover, this will make us find early diagnostic and curative targets.

Materials and Methods

The normal human colorectal tissues and the human colorectal adenocarcinoma (HCRAC) tissue samples from HCRAC patients in West China Hospital of Sichuan University were snap-frozen in liquid nitrogen immediately after surgery and stored at -80°C. Nine pairs of HCRAC and normal colorectal tissues had been used, isolation of RNA was performed by the TRIzol method (Invitrogen) according to manufacturer's instructions. The SSH combined with cDNA microarray was performed [3]. The 10 EST that include ES274070, ES274071, ES274073, ES274075, ES274076, ES274077, ES274081, ES274083, ES274084 and ES274085 were selected sequences for further bioinformatics analysis. The first step of bioinformatics analysis is to choose the EST as seed sequence and to find its matching sequence. The second step is to extend those EST as long as possible through the extension method of matching sequences blast cycling. That is to assemble these matching sequences together to form a longer EST and then use the new EST to conduct blast retrieval to find more matching sequences. This process

Manuscript accepted for publication April 3, 2009

^aDepartment of Anatomy, Basic Medical and Legal Medical Institute of Sichuan university, Chengdu, Sichuan, China. 610041

^bCorresponding author: mxmfh@263.net

doi:10.4021/gr2009.04.1287

Table 1. PCR conditions

Accession number	PCR condition
ES274070	94°C 3 min; 94°C 30 sec, 54°C 30 sec, 72°C 1 min, 30 cycles; 72°C 5 min
ES274071	94°C 4 min; 94°C 30 sec, 52.5°C 30 sec, 72°C 2 min, 30 cycles; 72°C 7 min
ES274076	94°C 4 min; 94°C 30 sec, 52°C 30 sec, 72°C 2 min, 33 cycles; 72°C 7 min
ES274073	94°C 4 min; 94°C 30 sec, 57°C 30 sec, 72°C 2 min, 33 cycles; 72°C 7 min

is repeated until no more matching sequences can be found. Thus, a contig is obtained. Then we analyzed the obtained full-length cDNA sequence by bioinformatics softwares and databases on network. Semi-quantitative RT-PCR and PCR-sequencing were done (Table 1, 2). Each lane of semi-quantitative RT-PCR was analyzed by Q1 software. Then t test was done using SPSS13.0 statistic software.

Results

We obtained 1260 differentially expressed clones by SSH. The cDNA microarray showed total 86 EST being identified were more than half past one fold or two-fold differentially expressed genes between two tissues. After they had been sequenced and analyzed by bioinformatics analysis, the sequence data were accepted by Genebank and were identified as EST and got genebank accession number and dbEST number. The 1721bp length cDNA sequence for ES274070 was formed by DNASTar Software and it located in human chromosome 17q21.2. The open reading frame (ORF) was predicted and its product contains 111 amino acids, its molecular weight is 12.8KD, PI is 8.79. On NCBI website, it is found that it has the homology 1649bp sequence (BC042012). The Serial analysis of gene expression

(SAGE) indicates that there is expression in oligodendroglioma, astrocytoma, prostatic carcinoma, ovarian cancer, breast cancer, and carcinoma of testis, sarcoenchondroma, and high expression in brain tumor tissue. Semi-quantitative RT-PCR results for its' ORF were showed that its mRNA in colorectal cancer tissue was significantly higher than that of normal colorectal tissues ($P = 0.011$), (Fig. 1).

The 2790bp length cDNA sequence for ES274071 was also formed. The sequence extends 1550bp from 5'end and 820bp from 3'end. It is located in human chromosome 9q34, ORF's length is 834bp. Its coded protein is SET translocation, we found the registered sequence NM003011 which is homologous with the assembled sequence, the length is 2577bp. We assembled the full-length cDNA by extending 309bp compared with NM003011 by RT-PCR. The result of ES274071 ORF was shown by sequencing and bioinformatics analysis, it was extended 16bp compared with NM003011 and we got accession number EF534308 (Fig. 2).

We also found that ES274073 corresponding full-length sequence were human mitochondrial genome (NM183047). Semi-quantitative RT-PCR results of NM183047 showed there was no significant difference in normal tissues and colorectal cancer tissues ($P > 0.05$). The ES274076 corresponding full-length sequence was human zinc finger protein (NM002644), semi-quantitative RT-PCR results of its ORF

Table 2. PCR primers and Products

Accession number	Primer	Product
ES274070	Forward 5' -TTGGAGCCCTGAGTATCTGTG -3' Reverse 5' -TAATGGAACCTGGTGCTAAGTC -3'	616bp
ES274071	Forward 5'-CCTTCGCCTTCCCTTCTC-3' Reverse 5'-CGACTGAGCACAAAGAGGGA-3'	700bp
ES274076	Forward 5' -GCACTGCCAAGATAGACAA -3' Reverse 5' -CTGGAACCTGCTACGAAT -3'	512bp
ES274073	Forward 5' -TACCACCTACCTCCCTCA -3' Reverse 5' -ATGGGCTGGGTTTACTA -3'	791bp

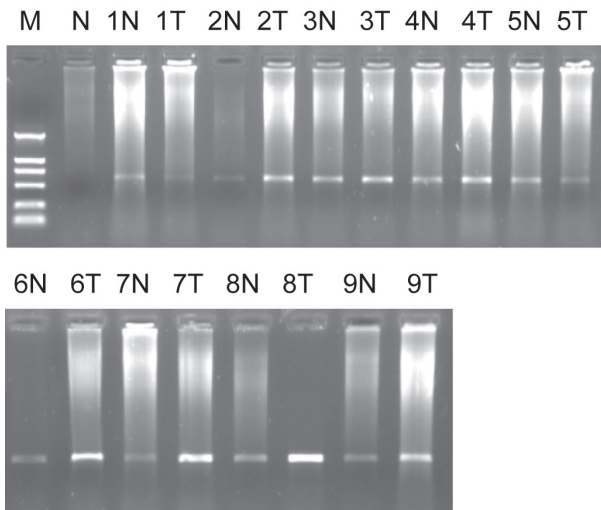


Figure 1. Marker: DL2000, 1% agarose ORF of ES274070 semi-quantitative RT-PCR. 1, 2.....9: sample number; N: normal colorectal tissues; T: colorectal adenocarcinoma tissues.

showed that its ORF in colorectal cancer tissue was lower than that of normal organizations ($P = 0.044$), (Fig. 3).

In fact we found that two EST (ES274075 and ES274084) are the same sequence. Its coded protein may be human collagen protein type. ES274077 located in human chromosome 3q24.3. Using the software DNASTar to assemble the matching sequences together to form a 4503bp cDNA sequence. The ORF was predicted and its product contains 268 amino acids. Its product is ELL associated factor-1 (EAF-1). The SAGE indicates that there is expression in oligodendroglioma, astrocytoma, prostatic carcinoma, ovarian cancer, breast cancer, carcinoma of testis, sarcoenchondroma, and high expression in brain tumor tissue. ES274081 located in hu-

man chromosome 20q13.1. The ORF was predicted and its product contains 432 amino acids. Its product is hypothetical protein LOC388969. The SAGE indicates that there is expression in cervical tumor, gastrointestinal tumor, glioma, head and neck tumor, leukemia, lung tumor, ovarian tumor, pancreatic tumor, and high expression in head and neck tumor tissue. ES274083 is located in human chromosome 2q11.2. The ORF was predicted and its product contains 268 amino acids. Its product is S-adenosylhomocysteine hydrolase (SAHH). The SAGE indicates that there is expression in breast (mammary gland) tumor, cervical tumor, colorectal tumor, leukemia, liver tumor, lung tumor, lymphoma, retinoblastoma, and high expression in retinoblastoma tissue. ES274085 is located in 13 chromosome (103-313bp) and 14 chromosome (1-103bp).

Discussion

The ORF of ES274070 located in BC042012 code the kind of hypothetical protein LOC90110, until now, there is its function report [4]. In our study, semi-quantitative RT-PCR results showed that ES274070 in colorectal cancer tissues was up-regulated; so, further study is needed to identify the its relationship with human colorectal adenocarcinoma. The coded protein of ES274071 was found to be SET translocation. SET is a suppressor of PP2A and NM23. PP2A is a kind of phosphorylase and is a tumor suppressor protein, its over-expression could lead to disorder of cell cycle, and it contributes to cell proliferation, growth and division. It plays some role in the cell apoptosis mediated by caspases. SET has a non-dependent nucleosome assembly protein; its NAP activity can enhance the affinity of chromatin [5, 6]. ES274073 located in the 8590-9000 of human mitochondria, its' ORF of gene is mitochondrially encoded ATP synthase

```

1  CCTTCGCGCTTCCTTCTC]CCCCCTCCCGCTCCCCCCCCGACCGCGGAGCAGCACCATGTCGG
64  CGCCGGCGGCCAAAGTCAGTAAAAAGGAGCTCAACTCCAACCACGACGGGGCCGACGAGACCT
127 CAGAAAAAGAACAGCAAGAAGCGATTGAACACATTGATGAAGTACAAAATGAAATAGACAGA
189 CTTAATGAACAAGCCAGTGAGGAGATTTTGAAGTAGAACAGAAATATAACAAACTCCGCCA
251 ACCATTTTTTTCAGAAGAGGTCAGAATTGATCGCCAAAATCCCAAATTTTTGGGTAACAACAT
313 TTGTCAACCATCCACAAGTGTCTGCACTGCTTGGGGAGGAAGATGAAGAGGCACTGCATTAT
375 TTGACCAGAGTTGAAGTGACAGAATTTGAAGATATTAATCAGGTTACAGAATAGATTTTTTA
437 TTTTGATGAAAATCCTTACTTTGAAAATAAAGTTCTCTCAAAGAATTCATCTGAATGAGA
499 GTGGTGATCCATCTTCGAAGTCCACCGAAATCAAATGGAAAATCTGGAAAAGGATTTGACGAAA
561 CGTTTCGAGTCAAACGCAGAATAAAGCCAGCAGGAAGAGGCAGCATGAGGAACCAGAGAGCTT
623 CTTTACCTGGTTTACTGACCATTCTGATGCAGGTGCTGATGAGTTAGGAGAGGTCATCAAAG
685 ATGATATTTGGCCAAACCCATTACAGTACTACTTGGTTCCCGATATGGATGATGAAGAAGGA
747 GAAGGAGAAGAAGATGATGATGATGAAGAGGAGGAAGGATTAGAAGATTTGACGAAGA
809 AGGGGATGAGGATGAAGGTGAAGAAGATGAAGATGATGATGAAGGGGAGGAAGGAGAGGAGG
871 ATGAAGGAGAAGATGACTAAATAGAACACTGATGGATTCCAACCTTCCTTTTTTTAAATTTT
933 CTCCAGTCCCTGGGAGCAAGTTGCAGTCTTTTTTTTTTTTTTTTTTTTTTTTTCCTTGTGCTCAGTCG
    
```

Figure 2. The result of ES274071 ORF sequencing.

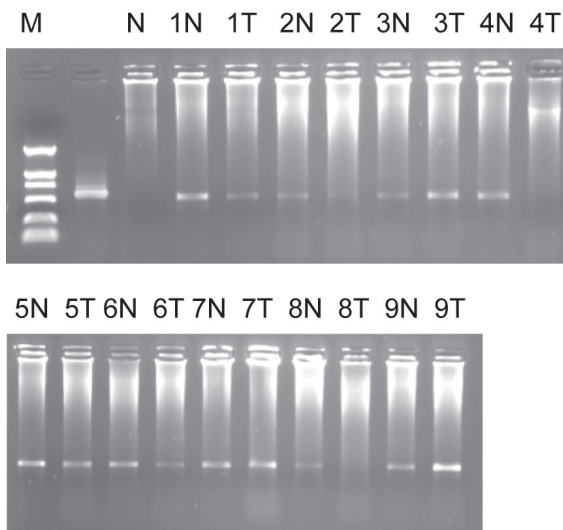


Figure 3. Marker: DL2000, 1% agarose ORF of ES274076 semi-quantitative RT-PCR. 1, 2.....9: sample number; N: normal colorectal tissues; T: colorectal adenocarcinoma tissues.

6. Some authors reported that the mutation and down-regulated expressed mitochondrially encoded ATP synthase 6 is related to heart disease and disorder of neural system [7-9]. In our study, ES274073 was not significantly different between normal colorectal tissues and colorectal cancer tissues ($P > 0.05$). The protein encoded by ES274076's gene is a receptor for activated C-kinase (RACK) protein. The encoded protein has been shown to bind in vitro to activated protein kinase C beta I. In addition, this protein is a cutaneous T-cell lymphoma-associated antigen. Finally, the protein contains a bromodomain and two zinc fingers, and is thought to be a transcriptional regulator. Multiple transcript variants encoding several different isoforms have been found for this gene [10-12]. Semi-quantitative RT-PCR results showed that ES274076 in colorectal cancer tissues was down-regulated; probably, activated protein kinase C beta I (PKC β I) will be reduced, so the sensitivity of cancer cells for anti-tumor factor (TNF α and others) will decrease and promote the growth of cancer cells and inhibit apoptosis of cancer cells.

ES274077 encoded protein is ELL associated factor-1 (EAF-1), which is a positive regulation protein for ELL and RNA polymerase II. EAF-1 has been found to have a dominant effect in leukemogenesis [13-15]. The function of the hypothetical protein LOC388969 for ES274081 is unclear now. But it is shown that it expresses in various types of human tumors, this knowledge is helpful to guide our further study. The protein production for ES274083 is S-adenosyl-homocysteine hydrolase (SAHH). Although SAGE shows that there is expression in many human tumor tissues, it has not been found that there is relationship between colorectal adenocarcinoma (CRA) and SAHH [16, 17]. The other two EST, ES274084 and ES274075, after analysis, we found that

they are the same sequence, its coded protein may be human collagen protein type I [18].

In conclusion, with the development of molecular biology, more and more data need to be analyzed, this is possible because of the development of computer and the other sciences and techniques, so the bioinformatics have become very useful tool and new approaches. The colorectal adenocarcinoma is a common malignant and a multi-genes and multi-steps disease. Through the bioinformatics analyses, we are able to screen the data from cDNA microarray, thus we avoid wasting of time and manpower. It is very helpful for us to search the pathologic genes associated with the development of human colorectal adenocarcinoma. Furthermore, ES274070, ES274071 and ES274076 and ES274081 were shown related to onset of human colorectal adenocarcinoma.

Acknowledgements

This study was supported by a Fund from the Chinese Education Ministry, No. 2008101-3-3. The authors declare no commercial associations or conflict of interests related to this article.

References

1. Frattini M, Balestra D, Suardi S, Oggionni M, Alberici P, Radice P, Costa A, et al. Different genetic features associated with colon and rectal carcinogenesis. *Clin Cancer Res* 2004;10:4015-4021.
2. Bayat A. Science, medicine, and the future: Bioinformatics. *BMJ* 2002;324:1018-1022.
3. Chen Y, Zhang YZ, Zhou ZG, Wang G, Yi ZN. Identification of differently expressed genes in human colorectal adenocarcinoma. *World J Gastroenterol* 2006;12:1025-1032.
4. Strausberg RL, Feingold EA, Grouse LH, Derge JG, Klausner RD, Collins FS, Wagner L, et al. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc Natl Acad Sci U S A* 2002;99:16899-16903.
5. Trotta R, Ciarlariello D, Dal Col J, Allard J, 2nd, Nevisani P, Santhanam R, Mao H, et al. The PP2A inhibitor SET regulates natural killer cell IFN-gamma production. *J Exp Med* 2007;204:2397-2405.
6. Fan Z, Beresford PJ, Oh DY, Zhang D, Lieberman J. Tumor suppressor NM23-H1 is a granzyme A-activated DNase during CTL-mediated apoptosis, and the nucleosome assembly protein SET is its inhibitor. *Cell* 2003;112:659-672.
7. Craig K, Elliott HR, Keers SM, Lambert C, Pyle A, Graves TD, Woodward C, et al. Episodic ataxia and hemiplegia caused by the 8993T->C mitochondrial DNA

- mutation. *J Med Genet* 2007;44:797-799.
8. Baracca A, Sgarbi G, Mattiazzi M, Casalena G, Pagnotta E, Valentino ML, Moggio M, et al. Biochemical phenotypes associated with the mitochondrial ATP6 gene mutations at nt8993. *Biochim Biophys Acta* 2007;1767:913-919.
 9. Moslemi AR, Darin N, Tulinius M, Oldfors A, Holme E. Two new mutations in the MTATP6 gene associated with Leigh syndrome. *Neuropediatrics* 2005;36:314-318.
 10. Cesaro P, Raiteri E, Demoz M, Castino R, Baccino FM, Bonelli G, Isidoro C. Expression of protein kinase C beta1 confers resistance to TNFalpha- and paclitaxel-induced apoptosis in HT-29 colon carcinoma cells. *Int J Cancer* 2001;93:179-184.
 11. Zhu GH, Wong BC, Slosberg ED, Eggo MC, Ching CK, Yuen ST, Lai KC, et al. Overexpression of protein kinase C-beta1 isoenzyme suppresses indomethacin-induced apoptosis in gastric epithelial cells. *Gastroenterology* 2000;118:507-514.
 12. Gilhooly EM, Morse-Gaudio M, Bianchi L, Reinhart L, Rose DP, Connolly JM, Reed JA, et al. Loss of expression of protein kinase C beta is a common phenomenon in human malignant melanoma: a result of transformation or differentiation? *Melanoma Res* 2001;11:355-369.
 13. Charlet-Berguerand N, Feuerhahn S, Kong SE, Ziserman H, Conaway JW, Conaway R, Egly JM. RNA polymerase II bypass of oxidative DNA damage is regulated by transcription elongation factors. *EMBO J* 2006;25:5481-5491.
 14. Banks CA, Kong SE, Spahr H, Florens L, Martin-Brown S, Washburn MP, Conaway JW, et al. Identification and Characterization of a *Schizosaccharomyces pombe* RNA Polymerase II Elongation Factor with Similarity to the Metazoan Transcription Factor ELL. *J Biol Chem* 2007;282:5761-5769.
 15. Polak PE, Simone F, Kaberlein JJ, Luo RT, Thirman MJ. ELL and EAF1 are Cajal body components that are disrupted in MLL-ELL leukemia. *Mol Biol Cell* 2003;14:1517-1528.
 16. Steere JA, Honek JF. Synthesis and biological activity of novel S-adenosyl-L-homocysteine hydrolase inhibitors. *Bioorg Med Chem* 2003;11:3229-3236.
 17. Turner MA, Yuan CS, Borchardt RT, Hershfield MS, Smith GD, Howell PL. Structure determination of selenomethionyl S-adenosylhomocysteine hydrolase using data at a single wavelength. *Nat Struct Biol* 1998;5:369-376.
 18. Guo P, Imanishi Y, Cackowski FC, Jarzynka MJ, Tao HQ, Nishikawa R, Hirose T, et al. Up-regulation of angiopoietin-2, matrix metalloprotease-2, membrane type 1 metalloprotease, and laminin 5 gamma 2 correlates with the invasiveness of human glioma. *Am J Pathol* 2005;166:877-890.