Electronic Cloning and Validating of the Suppression Subtractive Hbridization EST ES274070 of Human Clorectal Aenocarcinoma

Chunle Zhang, Yao Chen

Dpartment of Anatomy, Basic medical and Legal Medical Institute, West China Medical Center, Sichuan University, Chengdu 610041, China

Abstract **Objective** To study the structure and function of the differently expressed EST sequences from suppression subtractive hybridization combining with cDNA microarray. **Methods** From the cDNA suppression subtractive library combining with cDNA microarray, collecting a EST sequence (accession number ES274070) and utilizing the bioinformatics to extend EST sequence in order to acquire full-length cDNA sequence. And then one pair of primer was designed to amplify it's ORF according to full-length cDNA sequence. Besides, the other primers were made out from ES274070 combined with 5′RACE-PCR technique to amplify really full-length cDNA sequence. **Results** ES274070 was validated expressing in HT-29 cell line and it's ORF was cloned successfully by PCR. Two fragments of EF611099 and EF606874 were cloned successfully by 5′ RACE PCR technique, which were new sequences located at human chrosome. **Conclusion** The method offers new thoughts and ways in experimental aspect of cloned novel genes.

Key words Bioinformatics; Expression sequence;Tag contig; Colorectal adenocarcinoma (CRA); HT-29 cell line

The initiation and development of colorectal adenocarcinoma is multisteps process which involves multi-gene change. During the process, cell division and differentiation are alnormally regulated by the many genes. It will be benefit to the early diagnosis, effective treatment and prevention that molecular pathological mechanism of human colorectal adenocarcinoma continues to be researched. 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. This study aims to screen and to clone the genes related to genesis of human colorectal carcinoma. On the basis of the EST sequence (accession number ES274070) of human colorectal adenocarcinoma cDNA, bioinformatics was utilized to extend EST sequence in order to acquire full-length cDNA and validating express by PCR [1]. It is being predicted that the EST′s chromosome locate, tissue distribute and protein function, furthermore, we also got open reading frame of ES274070 by PCR-sequence.

Correspondence to: Yao Chen
Email: xmxfh@263.com

MATERIALS AND METHODS

Reagents

HT-29 cell line was purchased from the ATCC Company, 1640 Culture medium was purchased form the Bao Xing Company, TRIzol Reagent and SupperScript III enzyme was purchased from Invitrogen Company, 5′RACE-PCR kit and LA-Taq DNA polymerase was purchased from TaKaRa Company; The primers were synthesized by Bao Sheng Wu Company; DNA sequencing was performed by Bao Sheng Wu Company.

Obtaining full-length cDNA

Draw one differential expressed EST fragment from the cDNA suppression subtractive library (accession number ES274070). The first step is to choose this EST

as seeded sequence and to find its matching sequence. According to American Genetics Research Institute, if two ESTs′ alkali bases overlap beyond 40bp and share 95% similarity in their overlapping zone, they are matching sequences. The second step is to extend this EST as long as possible through the extension method of matching sequences blast cycling. That is, assembling these matching sequences together to form a longer EST and then using the new EST to conduct blast retrieval to find more matching sequences. Repeat such process until no more matching sequences can be found. Thus, a contig is obtained.

Predict protein function

We analyze the obtained full −length cDNA sequence by softwares of bioinformatics and databases on network.[3-6](1)Search with the ORF Finder program: (http://www.ncbi.nlm.nih.gov/gorf/gorf.html)offered by NCBI to carry on the open read −frame ORF identification of contig sequence. (2) Directly use genome sequence to do genetic electron location, Net to (http://www.ncbi.nlm.nih.gov/genome/seq/page.cgi?F =HsBlast.html&&ORG =Hs)in order to acquire chromosome location. (3)Took the assembled full −length cDNA sequence as a probe and used the virtual Northern program (http://www.ncbi.nlm.nih.gov/SAGE/) provided by CGAP in NCBI as a tool. The expression quantum analysis was done by drawing genetic representative tag form the SAGE library. (4) The prediction and analysis of the protein structural domain and functional site were performed through Prosite software (http://www.expasy.org/prosite/).

HT−29 cell line and culture

HT −29 cell line was grown and routinely maintained in 1640 medium and was incubated in a

Table 1 PCR amplification and sequencing

Name	Oligo	Sequence
EST	Forward prime	5′−CAGAGATGGATGGGTAAATGGT−3′
	Reverse primer	5′ −TGGAAACAGTCACTTGGAGAGC−3′
ORF	Forward primer	5′ −TTGGAGCCCTGAGTATCTGTG−3′
	Reverse primer	5′ −TAATGGAACCTGGTGCTAAGTC−3′
5′RACE−PCR	RT primer	5′ −(P)TAAGTCACCCCGTA−3′
	Forward primer 1	5′ −CTGCTGCTTGCTTGTTTCTCC−3′
	Reverse primer 1	5′ −GTCTGATTCCCTCCTTCCCAA−3′
	Forward primer 2	5′ −TACCCATCCATCTCTGTCCCT−3′
	Reverse primer 2	5′ −GCATGGGTAGTCCTCACAGTTCT−3′

Name	PCR CYCLE
EST	94℃ 3 min; 94℃ 30 sec, 55℃ 30 sec, 72℃ 1 min, 30 cycles; 72℃ 5 min
ORF	94℃ 3 min; 94℃ 30 sec, 55℃ 30 sec, 72℃ 1min, 30 cycles; 72℃ 5 min
5′RACE−PCR	
1st pcr	94℃ 4 min; 94℃ 30 sec, 57℃ 30 sec, 72℃ 2 min, 35 cycles; 72℃ 7 min
2nd pcr	94℃ 4 min; 94℃ 30 sec, 57℃ 30 sec, 72℃ 2 min, 25 cycles; 72℃ 7 min

37℃ humidified atmosphere (95% air and 5% CO_2). When cells grew to 80% confluence of Petri dish, it was digested with 0.2% tropsin to make single cell suspension; cells were passaged into new dishes.

RNA isolation and Synthesis of first strand cDNA

Total RNA was isolated through the Invitrogen system of RNA isolation according to the manufacturer´s instructions. A reaction mixture of the synthesis cDNA containing 10μl ribonuclease −free water, 1μl 10mM dNTP, 1μl total RNA (0.1g/L) and 1uμl oligo (dT) primer, the reaction eppendorf was incubated at 65℃ for 3min and chilled on ice. Then added 4μl 5 ×reaction buffer, 1μl DTT, 1ul ribonuclease inhibitor and 1μl SupperScript III reverse transcriptase (200U/μl). The mixture was incubated at 50℃ for 60min. The reaction was terminated by heat at 70℃ for 15 min and then the mixture was kept at 4℃.

The PCR products were separated on 1.0% agarose gel and analyzed, these PCR fragments were cloned into PMD18T vectors and sequenced.

RESULTS

EST sequence assembling

Software DNAStar assemble the matching sequences together to form a 1721bp cDNA sequence.

ORF identification

A 1721bp sequence is obtained by assembling together the matching EST sequences. Search with ORF Finder program to discover that initiation codon ATG exist in the location of 948 −950bp of the sequence; a termination codon TGA exist in the location of 1281 −1283bp of the low reaches. Which constitutes the longest ORF at 948 −1283bp, with a length of 336bp. Termination codon TAA exist in the location of 885 −887bp of the upper reaches; all the three codons are in the same read−frame. There is poly A site signal AAAATAAATA at the 3´ end.

1 ATGACTTGGCCTT 13
1 M T W P
14
CAGATGCTTCCTACTCTGGCCTCCCTTTTTTC
TCCTAGGAATTGTTTCCAGGTAACCT 73
5 S D A S Y S G L L L F S P R N C F
Q V T
74
ACCATGTGCACTTCTCAGCGCTTCGCCACCTT
TGCCCACCTTCCCGTGACCCTGAGAG 133
25 Y H V H F S A A L L T F A H L P
V T L R
134
TACAGATCCGAATAATGTGGCCTGTGCAGAGC
TCAGAGAACTGTGAGGACTACCCATGCC 193
45 V Q I R I M W P V Q S S E N C E
D Y P C
194
TGTCAGACTCTGCTCAGGGACAGAGATGGATG
GGTAAATGGTGCTGTTGGAGACATTT 253

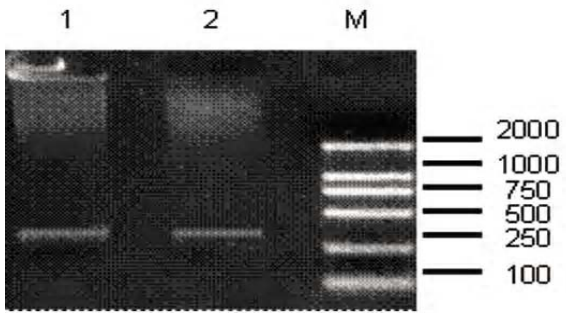

Fig.2A

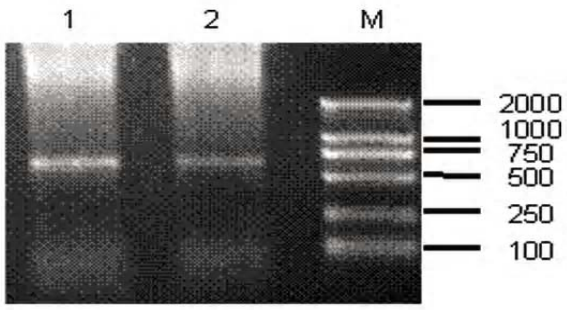

Fig.2B

Fig.2A : RT−PCR of EST, 315bp; Fig.2B: RT−PCR of ORF, 616bp; M:Marker(DNA Ladder 2000)

65 L S D S A Q G Q R W M G K W C C W R H F
254
ATCTTCATACCAGGCTCTGTTTCATCCCTGCA CCCCGCACCCCCACTCACTCCCCTGAGG 313
85 Y L H T R L C F I P A P R T P T H S P E
314 GGGTAAAATGCAGAGGTGGTTGA
105 G V K C R G G ⋆

Fig.1 ORF of clone full-length sequence

RT-PCR of EST and ORF

The PCR products were separated on 1.0% agarose gel.

EST-sequencing

Seeded sequence ES274070 was obtained form colorectal adenocarcinoma tissue. We amplified EST sequence of ES274070 from colorectal adenocarcinoma cell line HT-29 and validated the result by sequencing, It is being found that it is as identical as the original sequence. The result is as Fig.3

1
TCTTCATACCAGGCTCTGTTTCATCCCTGCAC CCCGCACCCCCACTCACTCCCCTGAGGG
61
GGTAAAATGCAGAGGTGGTTGACCTGAAGGGT CTGATTCCCTCCTTCCAAGCCTTAGGG
121
CCTACCCTGGAGTGCTGCAGTGTGTGAGAGCT GCTGCTTGCTTGTTTCTCCACTAGGCCT
181
GCTCCAAATGCTTAGCCAATCTCTGGAGCCGA CACAGTTGCCTACGGGGTGACTTAGCAC
241
CAGGTTCCATTAGCTCTCCAAGTGACTGTTTC CAA

Fig.3 Results of EST-sequencing

Results of ORF-sequencing

We also acquired the longest encoded sequence cloned by RT-PCR, then sequencing, it is being found that it is as coincident as predicted sequence. The result is as Fig.4

1
GTTGACTGTAAGAGGAGAGCGGCCTTCTTAAT TTTGCATTTCTTAATTTTTGTTTTTCTCC
61
ATTCCTGTTTTATGACTTGGCCTTCAGATGCT TCCTACTCTGGCCTCCTCCTTTTTTCTC
121
CTAGGAATTGTTTCCAGGTAACCTACCATGTG CACTTCTCAGCTGCTCTGCTCACCTTTG
181
CCCACCTTCCCGTGACCCTGAGAGTACAGATC CGAATAATGTGGCCTGTGCAGAGCTCAG
241
AGAACTGTGAGGACTACCCATGCCTGTCACAG ACTCTGCTCAGGGACAGAGATGGATGGGTA
301
AATGGTGCTGTTGGAGACATTTTTATCTTCAT ACCAGGCTCTGTTTCATCCCTGCACCCC
361
GCACCCCCACTCACTCCCCTGAGGGGGTAAAA TGCAGAGGTGGTTGACCTGAAGGGTCTG
421
ATTCCCTCCTTCCCAAGCCTTAGGGCCTACCC TGGAGTGCTGCAGTGTGTGAGAGCTGCT
481
GCTTGCTTGTTTCTCCACTAGGCCTGCTCCAA ATGCTTAGCCAATCTCTGGAGCCGACAC
541
AGTTGCCTACGGGGTGACTTAGCACCAGGTTC CATTAA

Fig. 4 Results of ORF-sequencing

5′RACE-PCR sequencing

We getted two sequences cloned by RACE-PCR, then sequencing, it is being found that they are not as expected. The result is as Fig.5.

1
TTTCTCCTTCCTTGACAGCAACATTCGGTATA GCTATGTGCACAAAAAGCTCTCCACTT
61
CTGCAGTCTACACTTCCTCCTTGACCTACCAC

CACCACCTAGCTTTCCTCTCACTGAGAC
121
CGTGACACCACAGTCATCAACAATTTTCTGTA
CAATTTTCTGTAGCCAGATCCATGGGAAACTT
CTTTATTC
181
TCAATCTCTTGGGCCACTCAGCTACATTTCAC
CCCATTGGCACCTCCTCTGTTCCAAAGC
241
ATTGCTTTTCTTGGTTTCTATGACCTCATTCC
TCTTGATTCTTATTCTACCTGTTTGCT
301
CTTCAATCTCTTTGCATAGACTCCTTTTCTAC
AGACCCTTTCAACTCTGATTTATATTAT
361
ATCATTTTGTCATTGTGCATTTATGTTTTTAT
GTGTCATGAGTGGACTGCATTCATCTCA
421
ACCTCCGTGGCAATTAGCACAGGGATTCAGAA
TATCCTAATGAGTGAATGAATAGAGGGA
481
TGAATGAATACACAAATGGATACACTTAGCAC
TGCTCATGAAGTTTTGGCCGATGAAAGA
541
GGGTGGAAGGCAAGGCTAGTATTGAAGGTGAC
AACTCACTCACATGTGAAGTACTCGGAC
601
ACTTTCTTGACTAATGTGAAAAAATTTCCAGG
AGGAGGTGGTAGAACCTGGGAGCGGGAA
661
GGAGGAAACCTGAATAGAACCAGACTAGGATC
TTGGCTTAGTCTTCAGGGCTCCTGATTC
721
TTTTCAGAATCTCAGTTTCCTATCATGTCCAA
TTTACCACAGTTAAGAAAGGAAAGGGAA
781
AGATGGACACAAAGAGGATCATTAGAGGAAAA
AGATGATAAGATGAAGAAGAGGGGAGAA
841 AC

Fig. 5A Results of RACE−PCR sequencing

1
TTAGCAATAAGGAGAGTAAAGCCTTAACAAAA
CTGGATACAGTTTGTTAAGCCCTCCAAA
61
CTGACATGTTAAGATGGAGATGTTTACCAGAC
ATCCAAGTCAAAATATCAAATAAGCAAT
121
TTAATGACTACAGCTCAAGAGAAATGGGTTTG
AGATAGAAATTTGGGTCTTCAGTATAGG
181
TAATAAAGCCTGGAAATGAATGTGGTCTCACT
AATAGAATGTTAGAGAAGTCCAAGAACA
241
AAACCTTGGTGATATGGTTTGGCTGTGTCCCC
ACCCAAGCCTCAACTTCAATTGCATCTC
301
CCAGAATTCCCGCGTGTTGTGGGAAGGACCAA
GGGGGAGGTAATTGAATCATGGGGGTCA
361
GTCTTTCCTGTGCTGTTCTCATGATAGTGAAT
AAGCCTCACGAGATCTGATAGATTTATC
421
AGGCTGCTTTTGCTTCTTCCTCATTTTCACTT
GCCTCCACCATGTAAGAAGTGCCTTTCA
481
CCTCCTGCCATGATTCTGAGACCTCCTCAGCC
ATGTGGAACTGTTAAGTCCAATTAAACC
541
TCTTTTTCTTCCCAGTATCAGGTATGTCTTTA
TCAGCCGCATGAAAACGGACTGATACAG
601
TATATTGGTACCAGGAGTGGGGTGCTGCTAAA
AAGATAACCAAAAATGTGGAAGCAACCT
661
TGGAACTGGGTAACAGGTAGAGGTTGGAACAG
TTTGGAGGGCTCAGAGGAAGACAGGAAA
721
ATGTGGGGAAGTTTGGAATTTCCCAGAGACTT
GTTGAATGGCTTTGCCCAAAATGCTGAT
781
ACCAATATGAACAATAAGGTCCAGGCTGAGGT
GGTCTCAGATGGAGATGAGGAAC

Fig. 5B Results of RACE−PCR sequencing

DISCUSSION

Colorectal adenocarcinoma is one of the most common malignant cancer, which is harmful to people´s health and its incidence is being grown year by year, but its pathogenesis is still not completely clear. Most studies indicated there was multi−gene and multi−stage change in the process of occurrence and development of human colorectal adenocarcinoma. Although it has been found that there are some proto−oncogene´s activation and anti−oncogene´s inactivation. For example some had known genes such as p53, DCC, APC inactivation and c−myc, k−ras activation, the many new genes and new pathologically steps were needed to be clarified. It is useful to study the molecular mechanism of occurrence and development of colorectal adenocarcinoma. So we first study differential expression gene fragments in the colorectal adenocarcinoma. Furthermore, we will clone and isolate some genes that is possibly related to the human colorectal adenocarcinoma.[1,2]

Bioinformatics has become a kind of frontier research method with the development of life science. During tumor research, bioinformatics is applied to establish tumor related database, and to develop new tumorous drugs and research. With bioinformatics, we could scan and screen related genes for colorectal adenocarcinoma in genomic level, we also could find out susceptible and invasive genes related to the varieties of diseases.

In this study, bioinformatics´ method were utilized, EST fragment from suppression subtractive hybrization library of colorectal adenocarcinoma was taken as seeded sequence and electronic clone to get full−length cDNA 1721bp. genomic sequence was inquested and it is being found that it locates in human chromosome 17q21.2, and it has no homology with other chromosomes. The ORF was predicted and its product is human hypothetical protein LOC90110 containing 111 amino acids, its molecular weight is 12817.7 Da,PI is 8.79. The SAGE (Serial analysis of gene expression)indicates that there is expression in oligodendroglioma, astrocytoma, prostatic carcinoma, ovarian cancer, and breast cancer, carcinoma of testis, sarcoenchondroma, and high expression in brain tumor tissue.

The two sequences were cloned by RACE−PCR, then sequencing, it is being found that it is completely different sequence as expected, but is really locates at the human chrosome and it is being shown that they are new human mRNA sequence. It is not found that same sequences as the two sequences mentioned above in GeneBank Nucleic Acid Database, so the two new sequences were submitted to the NCBI database and were validatedby NCBI, the accession number is EF611099 and EF606874), it is shown that it is a new approach to to find new genes. It is failed to acquire the full−long cDNA sequence for ES274070. Probably, there are non−specificity of primer in 5´race and the shortcomings of the RACE−PCR kit itself.

REFERENCE

1. Chen Yao, Zhang Yizeng, Zhou Zongguang, *et al.* Identification of differentially expressed genes in human colorectal adenocarcinoma. World Journal of gastroenterology, 2006, 12: 1025−1032.

2. Chen Yao, Zhang Yizeng. The construction of cDNA suppression subtractive library of human rectum adenocarinomas. U.S Chinese Journal of lymphology and oncology, 2002, 1: 9−14.

3. T K Attwood, D J party−smith. Introduction to Bioinformatics. Beijing: Beijing University Publisher, 2002, 97−126.

4. Zhang Chenggang, He Fuchu. Bioinformatics Methods and Practices. Beijing: Science Press, 2002.

5. Zhong Yang, Zhang Liang, Zhao Qiong. An Introduction to Bioinformatics. Beijing: Higher Education Press, 2001.

6. Zhao Guoping (chief editor). Bioinformatics, Beijing: Science Press. 2002.

7. Kong Ying, Chen Yao. Analysis of the Subtractive Fragment of Human Rectal Adenocarcinoma cDNA by Means of Bioinformatics. U.S Chinese Journal of lymphology and oncology, 2004, 3: 133−138.

8. Yue Wen *et al.* cloning and location of a human gene spindlin1 full cDNA. Progress in Natural Science. 2004, 14): 173−178.

9. Zhang Deli, Ding Peiguo, Ling Lunjiang, Chen Runsheng, Ma Dalong. A novel human gene C17orf32, electricity cloning and testifying the encoding sequence with RT−PCR. Progress in Biochemistry and Biophysics, 2002, 29: 543−548.

10. Zhang DeLi, Li YanDa, Ji Liang. Correction of Five Different Types of Errors of Model Refseqs Appeared in NCBI Human Gene Database Only by Using Two Novel Human Genes C17orf32 and ZNF362. Acta Genetica Sinica, 2004, 31: 325−334.