Analysis of the Subtractive Fragment of Human Adenocarcinoma cDNA by Bioinformatics Technique

Zhenni Ji, Yao Chen

The department of anatomy Biomedical and legal Institute of West China Medical Center of Sichuan University, chengdu, sichuan 610041, china

Abstract **Objective** Analyses the subtractive Fragment of human adenocarcinoma cDNA by bioinformatics technique. **Methods** electronic-clone the EST Fragment (Accession number BM360870) in cDNA Suppression Subtractive library of Human Rectum Adenocarcinomas in internet through bioinformatics technology. **Results** get full-length cDNA, predict the structure of protein which may be translated from it preliminarily, and primarily predict the function. **Conclusion** This method avoids some troublesome lab working, save time and expenditure, give direction and guidance to the next step at the same time.

Key Words rectum adenocarcinomas; bioinformatics; Full-length cDNA; Expression sequence tag

Combination and development of biology and computer technology result in a new subject--bioinformatics, it's materials are biologic data, it's techniques come from varies computer techniques[1,2]. The results and data of Human Genome Project, Human Protein Project, and gene chip techniques enrich the bioinformatics resource in internet greatly. Correct use of these materials will help to look for new gene interrelated with human rectum adenocarcinoma. This investigation clone the cDNA fragment (GENEBANK ID BM360870 the 18cDNA) which comes from the cDNA Suppression Subtractive library of Human Rectum Adenocarcinomas set up by Chen-Yao [3], and predict the structure and function of the protein which it may encode.

MATERIALS AND METHODS

cDNA fragment of Human Rectum Adenocarcinomas

This fragment comes from the cDNA Suppression Subtractive library of Human Rectum Adenocarcinomas. The sequence was detected by JiKang Company in Shanghai.

Materials in internet and analysis software's of bioinformatics

Materials in internet
http://www.ncbi.nlm.nih.gov/Blast
http://www.ncbi.nlm.nih.gov/uniGene/index.html
http://www.ncbi.nlm.nih.gov/SAGE
http://www.ncbi.nlm.nih.gov/ ORF Finder
http://www.sbc.su.se/~miklos/DAS/maindas.html
http://pasteur.fr/recherche/externe-en.html
http://www.expasy.org/
http://cubic.bioc.columbia.edu/Predictprotein/
http://srs6.ebi.ac.uk/srsbin/cgi-bin/wgetz?-page+applSelect+-newId
http://molbiol.soton.ac.uk/compute/GOR.html

Analysis software of bioinformatics

DNA Star

Fishing and assembly of EST sequence

Use BM360870 as a probe, select the EST sequences similar to this probe (base pair overlap > 40bps, similarity>95%) by searching human EST library of gene bank through Blastn programme.Then use seqmag programme in software DNAstar to patch up the similar sequences to a longer one. Repeat these steps with the longer sequence, until there is no similar EST be found.

Identify ORF

Have a six phasic identifying with the ultimate EST in ORF programme (http//:www.ncbi.nlm.nih.gov/gorf/gorf.html) of NCBI.

Gene orientation on chromosome

Compare this sequence with human genome map in NCBI (http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi), and analyzes their similarity.

Expression of this gene in human tissue

Analysis its expression in human tissue in library of SAGE (http://www.ncbi.nlm.nih.gov/SAGE) in NCBI.

Structure of encoded protein analysis

Predict the structure of protein whose sequence

may encode with internet materials Predictprotein of Columbia University Bioinformatics Center (http://cubic.bioc.columbia.edu/Predictprotein/)、GOR Secondary Structure Prediction (http://molbiol.soton.ac.uk/compute/GOR.html), TOOLS of SRS(http://srs6.ebi.ac.uk/srsbin/cgi-bin/wgetz) [5], Deep View Swiss-PdbViewer of Swiss-Model (http://www.expasy.org/spdbv)[6].

RESULTS

Results of searches and patching up

We found four ESTs of high similarity with BM360870 in Blastn, they are BF793243、BQ421725、BF6990020 and BG818274. We assembly them into a 1211bp cDNA sequence by software DNA star.

Identifying of ORF

There is a 147aa reading frame at +1 reading frame, initiation code is from 535 to 537 base pair, stop code from 976 to 978 base pair, the 1041-1046 base pairs are signal of poly-A (Fig.1).

Gene orientation on chromosome

Compare this sequence with human genome map, we orientate this sequence on the 19th chromosome 19p13.1-13.2(Fig.2), it has high similarity with fragment NT_011295 on the 19th[t] chromosome.

Expression in human tissue

SAGE show that this gene has high expression in ovary carcinoma, colon adenocarcinoma, mammary gland carcinoma, brain well differentiated astrocytoma. Just as table 1 shows us:

Analysis of protein structure

The protein which this sequence encode has 147 aa, molecular weight is 16.54KD.

86 AA in this protein make up α helix, they are 15-26, 32-64, 73-86, 121-147, 65.6% of all AA; 34AA residues make up β sheet, they are 2-5, 65-69, 87-94, 97-104, 108-118, 26.0% of all AA residues; the others are turns and coils. 13-16, 28-31, 130-133 are Casein kinase II phosphorylation site, 91-96 are N-myristoylation site, 19-27 are Elongation factor 1 beta/beta´/delta chain signature 1, 136-147 are Elongation factor 1 beta/beta/delta chain signature 2, 61-65、91-100 are potential menbrane traversing site.

Analysis of superfamily tells us that the 59-147 AA residues are conservative domain of this protein, homologous with eukaryotic translation elongation factor 1 beta, analysis of sequence similarity and domain also shows that this protein has high similarity with human, rat, mice, yeast´ translation elongation factor(>35%)[1], all above show that this protein is homologous protein of eukaryotic translation elongation factor.

We analyse its 3D structure by programme Deep View Swiss-PdbViewer (http://www.expasy.org/spdbv)of Swiss-Mode, Figure 3 shows its 3D structure.

```
535 atgcgccaagtggagcccccagccaagaagccagccacaccagca
    M  R  Q  V  E  P  P  A  K  K  P  A  T  P  A
580 gaggatgacgaggatgatgacattgacctgtttggcagtgacaat
    E  D  D  E  D  D  D  I  D  L  F  G  S  D  N
625 gaggaggaggacaaggaggcggcacagctgcgggaggagcggcta
    E  E  E  D  K  E  A  A  Q  L  R  E  E  R  L
670 cggcagtacgcggagaagaaggccaagaagcctgcactggtggcc
    R  Q  Y  A  E  K  K  A  K  K  P  A  L  V  A
715 aagtcctccatcctgctggatgtcaagccttgggatgatgagacg
    K  S  S  I  L  L  D  V  K  P  W  D  D  E  T
760 gacatggcccagctggaggcctgtgtgcgctctatccagctggac
    D  M  A  Q  L  E  A  C  V  R  S  I  Q  L  D
805 gggctggtctgggggcttccaagctggtgcccgtgggctacggt
    G  L  V  W  G  A  S  K  L  V  P  V  G  Y  G
850 atccggaagctacagattcagtgtgtggtggaggacgacaaggtg
    I  R  K  L  Q  I  Q  C  V  V  E  D  D  K  V
895 gggacagacttgctggaggaggagatcaccaagtttgaggagcac
    G  T  D  L  L  E  E  E  I  T  K  F  E  E  H
940 gtgcagagtgtcgatatcgcagctttcaacaagatctga 978
    V  Q  S  V  D  I  A  A  F  N  K  I  *
```

Fig.1 Result of translation of ORF

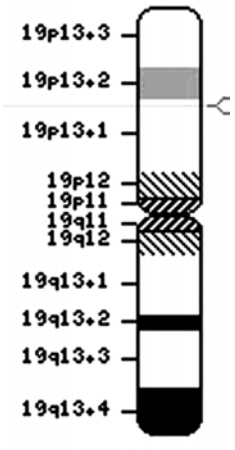

Fig.1 The position of 18cDNA

Table 1. The expression pattern of SAGE tag of 18cDNA

Library name	Tags per million	Tag counts	Total tags
SAGE OVCA432 2 ovary carcinoma SAGE CGAP non-normalized SAGE library method cell line	2429	7	2881
SAGE Duke H1126 brain well differentiated astrocytoma non-normalized bulk EST	1527	43	28159
AGE HX pancreas epithelium ductal normal cell line short term culture CGAP non-normalized SAGE library method cell line	1861	60	32226
SAGE SciencePark MCF7 estradiol 3h mammary gland carcinoma non-normalized cell line estrogen EST	1063	64	60162
SAGE RKO adenocarcinoma colon SAGE CGAP non-normalized SAGE library method cell line	2361	123	52094

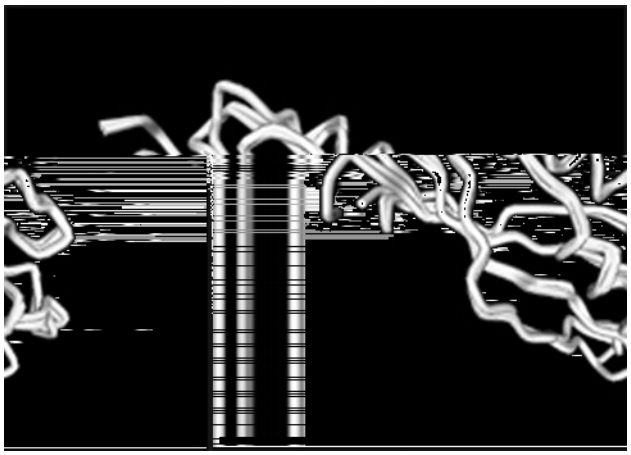

Fig. 3 3-D structure of protein

DISCUSSION

We study the subtractive Fragment BM360870 which come from human rectum adenocarcinoma cDNA library by bioinformatics technique, obtain its complete length cDNA, predict its expression and structure of protein that it may encode primarily. Primary predict of function, conclude that the subtractive Fragment BM360870 may have somgthing with the human rectum adenocarcinoma be taken bad, the study will be help to next step, to saving time and outlay. Bioinformatics technique is fit to our researchers and miniature labs of our nation[7]. From 90´s of last century, data of identified genes and parsed proteins has surprisingly increased, speed of identifying and parsing has greatly quickened [1], traditional method of single gene study in lab has become deficient to the accelerated increasing of bioinformatics study. Scientists either internal or overseas have analyzed various kinds of tumor genes by method of combination of lab studies and bioinformatics techniques, most of them have got satisfying results[6-11], we can say that bioinformatics technique is a new dependable experiment method. Rectum adenocarcinoma has correlation with several genes, its pathogenesis and every step of development is controlled and affected by different genes, while bioinformatics technique can not only find new gene but also study rectum adenocarcinoma at holistic genomic level in internet, bioinformatics technique will consequentially accelerate the development of studies of rectum adenocarcinoma and other diseases connected with several genes.

After all bioinformatics has its shortcoming as a new science, its base is database, its methods is

computer techniques, the veracity and integrality of the database and the use qualification of computer technique will restrict function of bioinformatics.

While bioinformatics has resulted in a new work mode to researchers, compare data quantity and its increment and produce and complement correlative software of these days with that of last several years, we can say that this new science has long-range foreground, it will become one of the most important research methods of biology.

REFERENCES

1. Zhao Guoping. Bioinfromatic; Beijing: Science publishing company, 2002, 4.
2. Chen Runsheng;Bioinformatic And Its Developmemt of Research. Medical Reserrch Communication, 2002, 31(12): 2-5.
3. chen Yao, Zhang Yi Zheng. The construction of cDNA suppression subtractive library of human rectum Adenocarcinomas. U.S Chinese Journal of lymphology and oncology, 2002, 1(1):9—14.
4. Joseph Sambrook, David W. Russell, Molecular Cloning: A laboratory manual, 3rd ed. Scinnce, Cold Spring Harbor Laboratory Press, 2001.
5. Sun Jiyong, Huang Qiang, Wang Aidong et al; Informatics analysis of malignant development associated gene of gliomas. Cancer, 2003, 22(3) 225-229.
6. Bi Meixia, Han Weidong, Lu Shixin; Cloning and Identifying of oesophagus carcinoma associated gene ECRG-4; acta biochimica et biophysica Sinica, 2001, 33(3) 257-261.
7. Li Hong, Wang Mengwei, Wang Gangshi; Fast Clonging and analysis of expression of fragment associated with stomach carcinoma; Prog. Biochem. Biophys, 2002, 29(4): 604-609.
8. Fu Junjiang, Lu Guangxiu, Li Luyun et al. Cloning of apoptosis associated gene TSARG1 and mtsarg1 of spermatogenic cell and analysis of expesion of mtsarg1; acvta genetica sinica, 2003, 30(1): 25-29.
9. Ben-Dor A, Bruhn L, Friedman N, et al., Tissue classification with gene expression profiles. J Comput Biol, 2002, 7(3-4):559-583.
10. Wang K,gan L, Jeffery, et al. Mornitoring gene expression profile changes in ovarian carcinomas using cDNA microarray, Gene, 1999, 229(1-2):101-108.
11. Alon U, Barkai N, Notterman DA, et al. Broad Patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc Natl Acad Sci U S A, 1999, 96 (12): 6745-6750.