Identification of differentially expressed genes in normal mucosa, adenoma and adenocarcinoma of colon by SSH

Min-Jie Luo and Mao-De Lai

Department of Pathology, School of Medicine, Zhejiang University, Hangzhou 310031, P.R.China.

This study is supported by Science Foundation of the Education Department of Zhejiang Province.
Correspondence to: Mao-De Lai, M.D., Department of Pathology, School of Medicine, Zhejiang University, Hangzhou 310031, P.R. China. lmd@sun.zju.edu.cn.

Received 2001-07-05 Accepted 2001-08-16

Abbreviation CRC colorectal cancer. SSH suppression subtractive hybridization. LD PCR long distance polymerase chain reaction. HLA human leukocyte antigen. IGFBP insulin-like growth factor binding protein. GN guanylin. EF1 elongation factor 1

Abstract

AIM To construct subtracted cDNA libraries and further identify differentially expressed genes that are related to the development of colorectal carcinoma (CRC).

METHODS Suppression subtractive hybridization (SSH) was done on cDNAs of normal mucosa, adenoma and adenocarcinoma tissues from the same patient. Three subtracted cDNA libraries were constructed and then hybridized with forward and backward subtracted probes for differential screening. Positive clones from each subtracted cDNA library were selected for sequencing and BLAST analysis. Finally, virtual Northern Blot confirmed such differential expression.

RESULTS By this way, there were about 3-4×102 clones identified in each subtracted cDNA library, in which about 85% positive clones were differentially screened. Sequencing and BLAST homology search revealed some clones containing sequences of known gene fragments and several possibly novel genes showing few or no sequence homologies with any known sequences in the database.

CONCLUSION All results confirmed the effectiveness and sensitivity of SSH. The differentially expressed genes during the development of CRC can be used to shed light on the pathogenesis of CRC and be useful genetic markers for early diagnosis and therapy.

Subject headings colonic neoplasms genetics; adenoma/adenocarcinoma genetics; intestinal mucosa metabolism; nucleic acid hybridization

Luo MJ, Lai MD. Identification of differentially expressed genes in normal mucosa, adenoma and adenocarcinoma of colon by SSH. World J Gastroenterol, 2001;7(5):726-731

INTRODUCTION

Colorectal carcinoma (CRC) is one of the most common forms of malignancy and the major cause of mortality in the population. The adenoma–carcinoma sequence has been accepted as the main pathway involving a series of molecular changes, which include activation of oncogenes, inactivation of tumor suppressor genes, etc. in colorectal carcinogenesis for several years[1]. However, due to the complexity of carcinogenesis, there are still many problems unsolved for the model of CRC. Not only are all the known genetic alterations not enough to explain fully the progression toward malignancy, but also there is no one genetic marker which is entirely tissue and/or cancer specific. Hence, an attempt was made to find new genes that are related to the development of CRC. Our goal was to clone and identify new cancer-predisposing genes.

In almost any area of biology or medicine, questions arise from the differential gene expression. Transformations in the normal mucosa-adenoma-carcinoma sequence result from the expression changes of specific gene(s) too. The identification and characterization of human genes expressed exclusively or preferentially in tumor tissue will hopefully shed light on the mechanisms of CRC development and provide useful genetic markers for screening, diagnosis, prognosis, therapeutic monitoring and even follow-up development of therapeutic vaccines.

There are many techniques that AIM at producing an inventory of differential transcripts between two populations of mRNAs. Although all these methods have proven successfully isolating differentially expressed genes, they all had some intrinsic drawbacks: need for plenty of initiating materials, difficulty in obtaining rare transcripts, significant incidence of false positives, labor intensive, etc. Suppression subtractive hybridization (SSH), a PCR-based cDNA subtracted method[2], combines normalization and subtraction in a single procedure and generates subtracted cDNA libraries. This method overcomes many drawbacks mentioned above.

In this research, we conducted SSH to identify genes differentially expressed among cDNAs of normal mucosa, adenoma and adenocarcinoma tissues from the same patient. Three subtracted cDNA libraries were constructed and numerous sequences were isolated.

MATERIALS AND METHODS

Tissue sample preparation

All normal (N), adenoma (A) and adenocarcinoma (T) tissue samples obtained from the same patient were histologically confirmed by a pathologist. The tissues were frozen in liquid nitrogen and homogenized with mortar. Total RNA was prepared using Trizol reagent (Gibco BRL, USA). RNA integrity was checked on 1% formaldehyde agarose gel. Poly (A)+ RNA was then purified using PolyATtract mRNA cDNA system kit (Promega, USA) as per manufacturer’s instructions.
SMART cDNA synthesis
0.5 µg mRNAs from each sample were reversely transcribed using MuLV SuperScript II reverse transcriptase (Gibco BRL, USA). SMART II oligonucleotides and CDS primers from SMART cDNA synthesis kit (Clontech, USA). The first strand cDNA mixtures were amplified by LD PCR. Before amplification, the number of PCR cycles was optimized (19 for N, 17 for both A and T) to ensure that cDNA products remain in the exponential phase of amplification to reduce nonspecific amplification.

SSH
After cDNAs were produced, they were purified by passing through Chroma spin-400 columns (Clontech, USA). Each purified tester cDNA was digested with Rsal to give an average insert length approximately longer than that of 600 base pairs. Then they were phenol chloroform-extracted, ethanol-precipitated, and dissolved in 6 µL ddH2O. The tester cDNAs were subdivided into two equal groups and then ligated to adaptor 1 and 2R in separate ligation reactions. Ligation efficiency was tested. Subtractive hybrization was performed by annealing an excess of driver cDNAs with each sample of adaptor-ligated tester cDNAs. The cDNAs were heat-denatured and incubated in 68°C for 8 h. After the first hybridization, the two samples were mixed together and hybridized again with freshly heat-denatured driver cDNAs for 20 h at 68°C. Two rounds of hybridization would generate a normalized population of tester-specific cDNAs with different adaptors on each end.

After filling in the ends, two rounds of PCR amplification were performed to enrich desired cDNAs containing both adaptors by exponential amplification of these products. The optimized cycles for the first and second PCR were 26 and 12 respectively to increase representation and reduce redundancy of subtracted cDNA libraries.

Secondary PCR products were used as templates for PCR amplification of human G3PDH at 18, 23, 28 and 33 cycles to assure subtraction efficiency. PCR products were run on 1.8% agarose gel. Both forward and backward SSH were performed. The difference between them was that the tester sample in forward SSH was used as driver sample in backward SSH.

Cloning and differentially screening of subtracted cDNA libraries
Products from the secondary PCR were T-A cloned into pGEM-T Easy Vector using pGEM-T Easy Vector System II kit (Promega, USA). The ligation products were transformed into DH5α competent cells. The transformed cells were plated on 2-YT agar plates containing ampicillin, X-Gal and IPTG allowed for color selection of colonies. Transformation efficiency was calculated and white clones were selected for further characterization.

Randomly selected individual colonies were grown for 8 h in 96-well culture plates and then used for clone PCR amplification. The primers used were nested Primer1 and 2R of adaptors. After 27 cycles amplification, the PCR products were subjected to a differential screening process. Briefly, 3 µL of amplified products of each candidate gene was mixed with DNA dilution buffer (50 µg/mL herring sperm DNA, 10 mM Tris. Cl, pH 8.0, 1 mM EDTA), spotted onto duplicate Hybrid N+ Membranes (Boehringer Mannheim, Germany) and UV cross-linked for 1 min. The forward and backward subtracted cDNAs were used as probes. These cDNAs were DIG-labeled by PCR amplification using PCR DIG probe synthesis kit (Boehringer Mannheim, Germany), digested with Rsal, Eag1 and Smal1 to remove the adaptors and purified with column. After prehybridization, the duplicate membranes were hybridized with forward and backward subtracted cDNA libraries probes respectively at 42°C for 16 h-20 h in hybridization solution with 50% deionized formamide, 0.1% (w/v)-Laurylsarcosinex, 2% blocking reagent, 0.02% (w/v) SDS, 2 µg/mL NP1, NP2R and CDS primer sequences. The membranes were then washed for 2×5 min with 2×SSC, 0.1% SDS at room temperature and for 2×15 min in 0.5×SSC, 0.1% SDS at 68°C. Then, they were blocked by gently agitating in blocking solution at 37°C for 1 h, incubated in buffer I with Anti-DIG (1:4 000) at 37°C for 30 min, washed in buffer I for 2×15 min and equilibrated in buffer III for 5 min. Finally, the membranes were incubated in mixture of NBT and BCIP for about 15 min. The clones giving at least 5-fold higher hybridization signals with the forward probes were selected for further analysis.

Sequencing and BLAST homology search
Candidate positive clones from each subtracted cDNA library were selected for sequencing. Cycle sequencing reactions were conducted with Thermo Sequenase cycle sequencing core kit (Amersham Life Science, USA). Each clone was sequenced on both 5’ and 3’ end. The BLASTN program was used to search for the DNA sequence homology of the isolated clones in the database of Genbank.

Virtual Northern blot
High yields of full-length cDNAs were generated from 1µg total RNA of normal mucosa, adenoma and adenocarcinoma using SMART PCR cDNA Synthesis kit (Clontech, USA). 0.2 µg cDNAs were electroposed on a 1.0% agarosegel, transferred onto Hybond N+Membrane (Boehringer Mannheim, Germany) and hybridized with DIG-labeled clone probes. These probes were generated by PCR amplification of the relevant inserts, adaptor-cutting and column-purification. Hybridization and washing conditions were the same as described above in screening of cDNA libraries. DIG-labeled G3PDH probes were used as control.

RESULTS
Subtracted cDNA library construction by SSH
In this research, SSH was done to identify differentially expressed genes among cDNAs of normal mucosa, adenoma and adenocarcinoma tissues from the same patient. Three subtracted cDNA libraries were constructed. The subtracted A-N cDNA library used normal mucosa as tester cDNA and adenoma tissue as driver cDNA; the subtracted T-N cDNA library used adenocarcinoma tissue as tester cDNA and normal mucosa as driver cDNA; the subtracted T-A cDNA library used adenocarcinoma tissue as tester cDNA and adenoma tissue as driver cDNA. Figure 1 shows subtracted cDNA libraries after secondary PCR amplification. They usually looked like smears with or without several discrete bands. However, the patterns among them were different.

Subtraction efficiency analysis showed the effectively reduced abundance of non-differentially expressed genes. In nonsubtracted cDNA libraries, housekeeping gene G3PDH PCR products were visible after 18 cycles of amplification and became saturated after 23-28 cycles. However, subtracted libraries required a higher number of amplification cycles for G3PDH to be detected. Depletion of G3PDH was more significant in the subtracted A-N and T-A cDNA libraries, suggesting that these two subtractions were more complete.

Differential screening of subtracted cDNA libraries
About 4×10²,3×10² and 3.5×10² clones were respectively identified in subtracted A-N, T-N and T-A cDNA libraries with average insert size of 0.2-1 kb. These clones were
screened by hybridization with both forward and backward subtracted cDNA library probes. Clones corresponding to truly differentially expressed genes would mainly hybridize with the forward subtracted cDNA probe, but not, or only faintly, with the backward subtracted cDNA probe, as shown by signal comparison between two blots. By this way, more than 85% positive clones in three subtracted cDNA libraries were differentially screened.

**Sequencing and homology search**

Automated sequencing proved that most clones were isolated no more than two times. BLASTN homology search revealed that some clones contained substantial sequence homologies with known gene fragments and several possibly novel genes, showed few sequence homologies with any known sequences in the GenBank database. The results of homology search are shown in Table 1.

**Virtual Northern Blot**

The cDNAs of selected clones were mainly expressed in tester tissue but not, or to a lesser extent, in driver tissue as Figure 4 shows. Virtual Northern Blot showed such differential expression in most clones. These results were also confirmed by RNA Dot blot (data not shown).

| Table 1 | Characterization of randomly selected differentially expressed clones from the three subtracted cDNA libraries screening of forward subtraction. Clone name, insert size, sequence identity and redundancy are shown |
|--------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Clone  | Insert (bp) | Sequence identity | Redundancy |
|--------|-------------|--------------------|------------|
| Subtracted A-N cDNA library | DNA sequence from clone RP5-991C6 on chromosome 6q14 | 1 |
| a1     | 850         |                     | 1 |
| a2     | 850         | Mitochondrion genome | 1 |
| a3     | 850         | DNA sequence from clone 75K24 on chromosome 6q13-1 | 1 |
| a4     | 550         | No match            | 1 |
| a5     | 500         | No match            | 1 |
| a6     | 550         | No match            | 1 |
| a7     | 500         | Guanylin mRNA, complete cds | 1 |
| a8     | 500         | DNA sequence from clone GSI-279B7 on chromosome 1q25.1-31.1 | 1 |
| a9     | 700         | beta 2-syntrophin mRNA | 1 |
| a10    | 500         | mRNA fragment coding for HLA-1 | 1 |
| a11    | 500         | mRNA for KIAA0212 gene | 1 |
| a12    | 650         | No match            | 1 |
| a13    | 800         | PAC clone RP5-1131G17 from 7p15.1-1p14 | 1 |
| a14    | 600         | BRX mRNA            | 1 |
| Subtracted T-N cDNA library | mRNA fragment coding for HLA DR | 2 |
| b1     | 480         | Insulin-like growth factor binding protein 7 | 1 |
| b2     | 750         | mRNA for KIAA0494 protein | 1 |
| b3     | 750         | mRNA for immunoglobulin kappa light chain | 1 |
| b4     | 760         | Complement component 3 gene, exons 1-30 | 1 |
| b5     | 500         | mRNA for anti-HbsAg immunoglobulin | 1 |
| b6     | 620         | mRNA for fibronectin | 2 |
| b7     | 600         | BAC clone RG158017 from 7q22-q31.1 | 1 |
| b8     | 770         | Ribosomal S25(RPS25) mRNA | 1 |
| b9     | 620         | No match            | 1 |
| b10    | 650         | Mitochondrion, cytochrome oxidase subunit II | 1 |
| b11    | 400         | Eukaryotic translation elongation factor1 gamma | 1 |
| b12    | 450         | No match            | 1 |
| b13    | 400         | Ribosomal S11(RPS11) | 1 |
| Subtracted T-A cDNA library | Ribosomal S12(RPS12) mRNA | 2 |
| r1     | 700         | No match            | 1 |
| r2     | 520         | Clone P13 anti-phospholipid immunoglobulin 1 | 1 |
| r3     | 550         | mRNA for heat shock protein 90 | 1 |
| r4     | 500         | Chromosome 14 DNA sequence | 2 |
| r5     | 700         | Chromosome 18, clone hRPK | 1 |
| r6     | 450         | HLA-DR alpha-chain mRNA | 1 |
| r7     | 350         | Novel human mRNA from chromosone 1, which has similar to BAT2 genes | 1 |
| r8     | 750         | mRNA for immunoglobulin kappa light chain | 1 |
| r9     | 520         |                     | 1 |

Species is human.

**DISCUSSION**

It was learned in the last two decades that genetic changes lie at the root of all cancers. In response, the Cancer Genome Anatomy Project (CGAP) will unite the newest technologies to identify all the genes responsible for the initiation and growth of cancer. In this research, we used SSH to identify candidate genes that were differentially expressed during CRC development.

The novel technique called SSH is an ideal subtractive system that combines high subtraction efficiency with normalized representation of differentially expressed genes, and is based on suppression PCR that permits exponential amplification of genes differing in abundance, at the same time that amplification of sequences of identical abundance genes in two populations is suppressed. Therefore, differentially expressed genes of low abundance can be cloned, while differentially expressed genes of high abundance are not
excessively isolated\textsuperscript{3,4}. It was reported that SSH achieved more than 1000-fold enrichment of low-abundance genes but only required 2 µg mRNA instead of the 10-20 µg required in other methods. This method has been successfully used to isolate significant genes in many researches. Here we report the efficiency of SSH technique in identifying differentially expressed genes in normal mucosa, adenoma and adenocarcinoma from the same patient. All of the results confirmed the effectiveness and sensitivity of this method.

Figure 1 The results of secondary PCR amplification: Lane 1: Subtracted A-N cDNA library; Lane 2: Unsubtracted A cDNA library; Lane 3: Subtracted T-A cDNA library; Lane 4: Unsubtracted T cDNA library; Lane 5: Subtracted T-N cDNA library; Lane 6: Control

Figure 2 Analysis of subtraction efficiency: Lanes 1-4 were performed on subtracted cDNA libraries, Lanes 5-8 were performed on unsubtracted cDNA libraries. Cycles for 1&5, 2&6, 3&7 and 4&8 were 18, 23, 28 and 33 respectively.

Figure 3 Clone PCR amplification with an average insert size of 0.2-1 kb.
PCR analysis of the subtraction efficiency after subtractive hybridization was showed that the abundance of the non-differentially expressed gene G3PDH was effectively reduced. The difference of subtraction efficiency among the three subtracted libraries might be due to the diversity of starting materials for comparison. As we know, no subtraction technique is 100% effective. Although SSH greatly improved the subtraction efficiency, some false positive clones were still present in the studied material. Therefore, non-radioactive hybridization with forward and backward subtracted cDNA library probes was used to differentially screen these subtracted libraries before Virtual Northern blot confirmation. About more than 85% positive clones in each subtracted cDNA library were differentially screened. von Stein OD et al. found about 94% positive rate in their research, so they considered confirmation of differential expression by Northern blot analysis for each clone obtained was probably unnecessary. In our research, Virtual Northern Blot and RNA Dot Blot were still performed to con firm such differential expression before further analysis of the clones.

By this way, we constructed three subtracted cDNA libraries, A-N, T-N and T-A, as mentioned above. About 3-4×10^3 clones were identified in each subtracted cDNA library through differential screening. After sequencing and BLASTN homology search, these clones could be divided into three groups: known genes previously described as differentially expressed genes in cancer and closely related with carcinogenesis of CRC; known genes never before associated with carcinogenesis; and unknown genes that remain to be fully characterized. The identification of the first group of known genes attested to the validity of this method. In addition, no more than only two times isolation of most clones also proved the high efficiency of SSH.

HLA-DR, identified in both subtracted T-N and T-A cDNA libraries, was reported to be abundantly expressed in CRC but less so in adenoma and normal mucosa[6,7]. However, the relationship of HLA-DR expression with prognosis is still disputed[6,8]. Other molecules such as invariant chain may take part in this mechanism. So it is very important to conduct comprehensive research on the role of these molecules.

Heat shock protein (Hsps) is thought to play an important role in the cell cycle s and various processes of carcinogenesis. It can be divided into two types: Hsp90α and Hsp90β[9]. The former may play a role in cell proliferation; the later is inversely related to differentiation. Both of them are present in normal tissue but are more highly expressed in neoplasm. More important, Hsp90 was recently reported to interact with HLA-DR. In this study, Hsp90 was screened in the subtracted T-A cDNA library two times, which showed its high expression in carcinoma.

IGFBP-rp1 belongs to the IGFBP family with low affinity with IGFs. It had been shown to decrease expression with disease progression in breast carcinomas[10] and in the worsening from the benign to malignant prostate epithelial cells[11,12]. Methylation with down-regulated expression was found to be associated with liver tumorigenesis[13]. So it is inferred that IGFBP-rp1 may have a suppressive effect on cancer development. Surprisingly, in this study, IGFBP-rp1 was overexpressed in CRC in contrast to their down-regulation in carcinoma, which was also shown in prostate cancer tissue[14]. It remains to be further investigated what the role of IGFBP-rp1 is and whether such different expression is individual- specific, tissue- specific or even cell-specific.

Two cDNA clones were shown to have more than 90% homology with Ig κ gene, one screened in the subtracted T-N
cDNA library and the other in T-A. Hu et al [15] reported a new-transforming gene Tx showed very high (99.5%) homology with the C region of Ig k. A literature survey revealed Ig or Ig-like material expression in carcinoma [16]. It is generally considered that Ig is mainly secreted by B-lympho cytes, so the relationship between Ig k and carcinogenesis of CRC should be further confirmed.

Guanylin (GN) is the endogenous peptide ligand for guanylyl cyclase C (GCC) and can regulate intestinal Cl-secretion. It is robustly expressed in normal intestinal epithelium of colon but is absent in CRC [17]. So, it is possible that loss of GN activity leads to or is a result of adenocarcinoma formation. In this study, GN was expressed at higher levels in adenoma than in normal mucos a but not in CRC. It is uncertain if such increase of GN activity in adenoma means a kind of reaction to carcinogen stimulation.

Several genes shown to be dysregulated in neoplasm cells such as EF1-γ [18], mitochondrion, RPS12, RPS 22/25, RPSII [19] were also identified. All of them were increased in cancer cells, as previously reported and some of them were related to the robust growth characteristics.

Besides the known genes, a very important finding in this study was the identification of several new genes that were previously undescribed in Genbank. Both Virtual Northern blot and RNA dot blot hybridization has showed such different expression. Among three subtracted cDNA libraries, A-N showed the highest rate of novel genes isolation. Although it remains to be explored the full-length sequence of these new genes by RACE or other methods and their function in carcinogenesis of CRC, these results really provide clues for us to understand newly proposed mechanisms of cancer formation.

In the past years, much of cancer research laid particular stress on analysis of single gene expressed differentially in tumor cells as compared with their normal counterparts, but little stress on comprehensive study of gene expression in cancer. It was interesting that in this study several genes including anti-HbsAg and IGFBP-rp1 were also differentially expressed in carcinoma. Based on available reports, anti-HbsAg gene was mainly related with the individual character, while IGFBP-rp1 was even supposed to be an anti- onco gene. So we propose that such diversity and specificity of gene expression may contribute to the heterogeneity in biological properties of cancer, which may be individual-specific, tumor-specific, process-specific or even cell-specific. Literature also indicated that there were differences of gene expression among cancers. Therefore, two novel concepts can be brought forward on carcinogenesis: one is the model formed by a total of individual gene expression in cancer, the other is the module formed by interaction among molecules [20]. Cellular functions, such as signal transduction, are carried out by modules. Connecting different levels of analysis from molecules, through modules, to models is essential for an understanding of carcinogenesis. Differentially expressed genes between tissues comprise the molecular basis for understanding model and module. So, despite the limitations of tissue from the same patient, this approach has provided interesting results for further research.

This is the first time that differentially expressed genes in colorectal adenoma and adenocarcinoma were isolated by the new technique SSH. Now, microarray detection of these differentially expressed genes as model and module to hybridize with other tumor samples is being investigated [21]. Comprehensive study of these genes will be helpful for us to create theoretical models of some systems and verify that such predictions match reality, which, in turn, will be useful for screening, detecting, classifying and therapy of CRC.