Isolation of Differentially Expressed Sequence Tags from Human Breast Cancer

Mark A. Watson and Timothy P. Fleming

Division of Laboratory Medicine [M. A. W.], Departments of Ophthalmology and Visual Sciences [T. P. F.] and Genetics [T. P. F.], Washington University School of Medicine, St. Louis, Missouri 63110.

Abstract

Identification of quantitative changes in gene expression that occur in the malignant mammary gland, if sufficiently characterized, may yield novel molecular markers which may be useful in the diagnosis and treatment of human breast cancer. Using modifications of a previously documented technique, the differential display polymerase chain reaction, we describe the isolation of differentially expressed sequence tags, short complementary DNA fragments corresponding to mRNAs that are differentially expressed in breast cancer biopsies, as compared to normal breast tissue controls. Direct sequencing and expression analysis of two sequence tags demonstrate that they represent sequences which are overexpressed in a number of breast carcinoma cell lines. A paradigm for generating a catalogue of these sequence tags is discussed.

Introduction

The evolution of breast neoplasia is accompanied by a number of quantitative and qualitative changes in gene expression (1-3). Although most of these alterations have been characterized using breast carcinoma cell lines, it is reasonable to assume that many similar changes occur in vivo. However, application of genetic alterations seen in carcinoma cell lines as clinical markers in breast cancer may be of limited value if they do not truly represent genetic changes that occur in vivo. Similarly, while recent investigations have focused on genes that may be altered in a high percentage of multiple cancer types (4, 5), those changes that are specific to the pathobiology of breast carcinoma may have more relevance to disease-specific diagnosis and treatment. Therefore, identification of tumor-specific changes in gene expression that occur in vivo, if sufficiently characterized, may provide more useful information for predicting clinical course and patient outcome. Identification of such alterations may also aid in better understanding the biology of the mammary epithelium.

The DDPCR (1) and arbitrarily primed polymerase chain reaction are techniques previously used to identify changes in gene expression among several different cell lines (6-10). Unfortunately, the DDPCR technique generates a large number of spurious sequence fragments that do not represent differentially expressed genes (6-9), and the technique has been used largely to examine differences between only two differing cell line populations (6, 7). Few candidate oncogenes and tumor suppressor genes have been isolated using this method, but these investigations have involved the relatively time- and labor-intensive steps of subcloning, library screening, and cDNA sequencing of individual genes (11-13). On the other hand, creation of expressed sequence tag libraries is a rapid method used to identify or “tag” sequences that are expressed in specific tissues or cell lines (14). The advantage of this methodology, compared to isolation and sequencing of individual cDNAs, is that a large number of sequences can be “cataloged” with small amounts of sequencing data.

We have combined the concepts of differential display and expressed sequence tag libraries to describe the creation of a breast cancer-specific DEST catalogue. With modifications of techniques described previously (6-9), we have used patient biopsy specimens to isolate and directly sequence cDNA fragments that are uniquely present in breast cancer tissue as compared to normal breast tissue controls. In this initial report, we demonstrate the feasibility of this technique and show that the sequence tags generated are specifically expressed in neoplastic mammary epithelial cells. By analyzing a large number of biopsy specimens with this high throughput protocol and accompanying clinical data, we hope to identify sequence tags which have clinical correlations to disease parameters such as treatment response, relapse interval, and overall survival. These tags could then be used as a diagnostic “genetic fingerprint” to supplement currently used pathological markers.

Materials and Methods

Cell Culture, Tissue Procurement, and RNA Isolation. Cell lines were obtained from the American Type Culture Collection and grown in Dulbecco’s minimal essential medium supplemented with 10% fetal calf serum. Biopsy specimens 2410TU, C256, and C034 are infiltrating intraductal carcinomas and were obtained from the Cooperative Human Tissue Network (15). Samples C120K and C115R are normal breast tissue biopsies from reduction mammoplasty and were also obtained from the Cooperative Human Tissue Network. Cell pellets (~1 x 10^6 cells) and tissues (~50 mg) were homogenized in guanine isothyocyanate and RNA was prepared using standard methodology. RNA was treated with RNase-free DNase (Promega, Madison, WI), and a small aliquot of each RNA sample was run on a 1% agarose gel and stained with ethidium bromide to assess concentration and integrity.

Differential Display. One µg of total RNA from the indicated source was RT with AMV reverse transcriptase (Promega) and an oligo-T21 primer in a volume of 20-µl according to the manufacturer’s protocol. Each reaction was diluted 10-fold, and 1 µl or 10 µl (1/200 or 1/20, respectively, of the original reaction) were used for subsequent PCR reactions. PCR T1pMN primers, T1pMC, T1pMA, T1pMG, and T1pMT (where M = A, G, and C), were end-labeled in a reaction volume of 10 µl containing 1µM of primer, 1 µM [γ-^32P]ATP (3000 Ci/mmol) and T4 polynucleotide kinase according to the manufacturer’s protocol (Promega). PCR reactions were carried out with indicated primers F1 (5’-CTGATCCATG-3’) or F2 (5’-CTTGATTGCC-3’) at 1.5 µM, indicated T1pMN primer at 1.5 µM, corresponding labeled T1pMN primer at 25 nM, dNTPs at 200 µM, MgCl2 at 2 mM, Taq DNA polymerase reaction buffer, and 0.3 units Taq DNA polymerase (Perkin Elmer Cetus). Reactions were performed in a Perkin Elmer 9600 thermal cycler at 94°C for 1 min, and then at 94°C for 20 s, 40°C for 1 min and 72°C for 30 s for a total of 40 cycles. Samples were precipitated with ethanol, resuspended in formamide sequencing dye, and run on a 6% acrylamide 7.5 M urea sequencing gel. The gel was dried without fixation and subject to overnight autoradiography on Kodak XAR film.

Received 6/23/94; accepted 7/13/94.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

1 Supported in part by Dept. Ophthalmology and Visual Sciences, Washington University School of Medicine.

2 To whom requests for reprints should be addressed, at the Department of Ophthalmology, Box 8096, Washington University School of Medicine, 660 South Euclid Ave., St. Louis, MO 63110.

3 The abbreviations used are: DDPCR, differential display polymerase chain reaction; cDNA, complementary DNA; DEST, differentially expressed sequence tag; RT, reverse transcribed; PCR, polymerase chain reaction; GADPH, glyceraldehyde 3-phosphate dehydrogenase; SSC, standard saline citrate (1 x SSC = 150 mM NaCl, 15 mM sodium citrate, pH 7.0).
Sequencing and Expression Analysis. Indicated bands were excised from the sequencing gel, eluted, and precipitated as described previously (6, 7). Eluted DNA was reamplified using the primers PF1 (5'-GGAATTCGATCCGCCGCGCCTGATCCATG-3') or PF2 (5'-GGAATTCGATCCGCCCGCGCCTGATCCATG-3') and TT7 (5'-TAATACGACTCACTATAGGAATTCGGATCCGCGGCCGCGATCCATG-3'). Conditions were identical to those described above except for the omission of 32P-labeled primer and an annealing temperature of 45°C rather than 40°C. PCR products were run on a 2% agarose gel, visualized by ethidium bromide staining, excised, and purified using the GeneClean procedure (Bio101, La Jolla, CA). PCR products were sequenced directly using primer PF1, PF2, or the appropriate T10MN oligonucleotide with the fmol sequencing kit (Promega) according to the manufacturer’s protocol. cDNA probes were synthesized using the Megaprime kit (Amersham, Arlington Heights, IL) and [α-32P]dCTP according to the manufacturer’s protocol. RNA probes were synthesized using T7 RNA polymerase (Promega) and [α-32P]UTP according to the manufacturer’s protocol. Ten μg of indicated total RNA were electrophoresed on a 1.2% formaldehyde agarose gel and transferred to Hybond N+ membrane (Amersham). Equivalent sample loading was assessed by ethidium bromide staining and visualization of ribosomal RNA bands. Northern analysis was carried out using Rapid-Hyb buffer (Amersham), followed by hybridization at 55°C (65°C for RNA probes), and washings with 0.2x SSC at 65°C (72°C for RNA probes). RT-PCR analysis of DEST1 expression was carried out essentially as described above using two sequence specific primers 1F (5'-CATCTGGCACATGTACCTCTC-3') and IR (5'-GCTGTATATATTCAT-3'). Conditions were identical to those described above except for the use of 3-phosphate dehydrogenase control amplifications were performed with sequence specific primers IF (72°C for RNA probes). RT-PCR analysis of DEST0 expression was carried out essentially as described above using two sequence specific primers 1F (5'-CATCTGGCACATGTACCTCTC-3') and IR (5'-GCTGTATATATTCAT-3') and a PCR-annealing temperature of 55°C. Glyceraldehyde 3-phosphate dehydrogenase control amplifications were performed with sequence specific primers according to the manufacturer’s protocol (Stratagene).

Results

Differential Display of Breast Carcinoma mRNA. Using the modifications discussed above, differential displays were created from three independent breast carcinoma biopsies and two nonmalignant breast biopsies from reduction mammoplasty. Combinations of the forward primers F1 or F2 and the reverse T10MN primers T10MA, T10MC, T10MG, or T10MT produced strikingly different patterns when used in the DDPCR reactions (data not shown). This observation has been documented by others (6-9). However, when multiple RNA samples from the same tissue specimen were subjected to reverse transcription and PCR, very similar patterns were observed (not shown). Fig. 1 shows examples obtained from one of the eight primer pairs used, F1 and T10MC. The indicated bands were retrospectively labeled DEST001, C001, and DEST002. DEST001 and DEST002 were specifically displayed in tumor specimens C256 and 2410TU, respectively, while fragment C001 was present in both carcinoma and normal breast displays and was further analyzed with the expectation that it would represent a well-characterized, constitutively expressed gene that could be readily identified in a gene database search (see below). Other tissue-specific bands were identified and excised, but were not amenable to further sequence analysis (see below).

Compared to previous reports, the differential display obtained using this protocol was considerably less complex (see Fig. 1 versus Ref. 6-8, and 12). This difference may be attributable to the use of the end-labeled T10MN primer that: (a) labeled only a single strand of the double-stranded PCR product; and (b) labeled only PCR products formed by priming of the T10MN oligonucleotide. Products formed by priming of the 10-mer oligonucleotide in both the forward and reverse orientation would not be visualized using this protocol. Together with the use of a longer T10MN primer, our intent was to bias amplification of authentic 3' ends of mRNAs rather than random amplification of contaminating genomic DNA, ribosomal RNA, and multiple regions of a single mRNA. Although this protocol may have eliminated amplification of potentially interesting sequences, it may have also eliminated a large number of the artifacts associated with DDPCR (see below).

Fig. 1. Differential display of breast carcinoma mRNAs. Total RNAs from indicated biopsy specimens were subjected to differential display PCR as described in “Materials and Methods.” (A) Carcinomas C256 and C034 were compared to normal breast tissue specimens C120R in the presence and absence (-RT lane) of reverse transcriptase using PCR primers F1 and T10MC. (B) Carcinoma 2410TU was compared to normal breast specimens C120R and C115R in the presence and absence (-RT lane) of reverse transcriptase using PCR primers F1 and T10MC. Fragments which were excised for further analysis are indicated.

Direct Sequencing of DESTs. A total of 23 fragments from all 5 tissue samples were excised and subjected to reamplification using the PF1 or PF2 and TT7 primers. Only 14 of 23 fragments were successfully reamplified. Those fragments that did not amplify on secondary PCR could not be amplified after a tertiary PCR or when secondary PCR was performed with the appropriate T10MN primer rather than the TT7 primer. Using both PF1 and T10MC oligonucleotides as sequencing primers, clear sequence could be obtained from both ends of three PCR fragments (C001, DEST001, and DEST002), as shown in Fig. 2a. In the remaining 11 cases where sequence obtained was unreadable, we were able to demonstrate that each fragment contained homologous ends composed of either PF1 sequence or a polyadenylic acid tract (not shown). Thus, simultaneous priming of both ends generated two coincident sequencing reads which were unintelligible.

Approximately 100-200 base pairs of sequence were obtained from each of the three fragments and compared to the National Center for Biotechnology Information Genbank data base using the BLASTN
algorithm (16, 17). DEST001 and DEST002 sequences yielded no matches in the Genbank data base. Matches were defined as greater than 95% homology to a known human sequence or greater than 85% homology to a known sequence from any other mammalian species. No homology of greater than 70% was identified for either fragment. Because these sequence tags presumably represent 3’ untranslated regions of their corresponding genes, it is possible that they are derived from human homologues of genes previously characterized from other species but that there is insufficient interspecies homology between 3’ untranslated regions to properly identify them as such. Conversely, because these sequence tags contain AT-rich regions characteristic of 3’ untranslated sequence, numerous matches between small (20–30 base pair) sequence tracts interspersed among much longer runs of disparate sequences were frequently identified. None of these comparisons, however, showed sufficient overall homology to be considered a match.

Unlike the DEST001 and DEST002 tags, C001, the tag isolated from both carcinoma and normal breast samples, demonstrated a perfect match (100% homology) with a previously identified human gene. As predicted from its pattern of nondifferential display, tag C001 represented a 3’ fragment of the mRNA for human L26 ribosomal protein (18). Fig. 2b shows the partial sequence of the L26 cDNA and the corresponding sequence of tag C001 obtained using the PF1 primer in a single sequencing run. Further sequencing with the T19MC primer (not shown) confirmed that the entire C001 fragment was identical to the reported L26 sequence. This result demonstrated that at least one of the isolated fragments represented an authentic mRNA and that direct sequencing provided sufficient information to establish its identity.

Confirmation of DEST Differential Expression. One disadvantage of the differential display technique is that a large number of bands are generated which do not represent differentially expressed mRNAs (7–9). To determine that the initial DESTs isolated in this study were not artifacts, we examined their expression in the three carcinomas and two nonmalignant tissues used in the original differential display. Each DEST fragment was used as a template to generate a randomly labeled cDNA probe for Northern hybridization experiments. In addition, secondary amplification was performed using the TT7 primer (see “Materials and Methods”) so that all DEST fragments contained a TT7 promoter that could also be used to generate more sensitive antisense RNA probes for use in Northern hybridization or RNase protection experiments. This latter property proved helpful for expression analysis (see below).

As shown in Fig. 3, the C001/L26 fragment detected an appropriately sized transcript (~1.0 kilobase) in all five biopsy specimens, consistent with the expression pattern seen in the PCR experiment from which it was isolated (Fig. 1) and in keeping with its general function as a “housekeeping gene.” Furthermore, despite the existence of a highly homologous rat L26 gene (18), no transcript was detected in rat brain RNA. This suggested that, despite its high AT sequence content and common sequence motifs found in 3’ ends of many mRNAs, C001 (and presumably other DESTs isolated by this method) could serve as a very specific probe for expression analysis.

When DEST001 was used as a probe in an identical experiment, no signal was detected in any of the five breast tissues, even after prolonged exposure times (not shown). Other investigators have also reported the isolation of apparently differentially expressed fragments that did not appear to be expressed on subsequent Northern analysis (7–9). However, when DEST001 was used as a transcriptional template to generate a more sensitive antisense RNA probe, a specific, albeit faint, transcript of ~2.2 kilobase could be seen in tumor C256, the biopsy specimen from which DEST001 was derived (Fig. 3B). A DEST002 cDNA probe could also detect a relatively more abundant ~1.2 kilobase transcript in its parental tissue specimen, 2410TU, but not in any of the other four specimens (Fig. 3C).

DEST Expression in Breast Carcinoma Cell Lines. Although identification of in vivo changes in gene expression may have the greatest relevance to mammary biology, there are many disadvantages in using patient tissues as opposed to homogeneous cell lines. The strongest criticism of such an approach is that DESTs isolated from tissue may represent differences in gene expression due to cell heterogeneity between samples rather than differences in malignant versus nonmalignant breast epithelial cells. As a first approximation to determine whether the DESTs isolated by differential display had any relevance to mammary neoplasia, we examined expression of DEST001 and DEST002 in a panel of breast carcinoma cell lines.

Because of the low expression level of DEST001 expression in
Fig. 3. Expression of differentially displayed products. Total RNA from indicated breast biopsy specimens or rat brain was subjected to Northern analysis with cDNA probes C001 and DEST002, and RNA probe DEST001. Single lines, 28S and 18S ribosomal bands. Arrowhead, probe specific signal in each panel. C001 and DEST001 blots were exposed for a 24-h period; DEST002 blot was exposed for 72 h. Each lane contains an equivalent amount of total RNA as judged by ethidium bromide staining of ribosomal RNA bands.

Fig. 4. DEST001 and DEST002 are expressed in malignant mammary cell lines. (A) Total RNAs from indicated tissue biopsies or carcinoma cell lines were subjected to RT-PCR analysis with DEST001-specific primers (top) or GADPH-specific primers (bottom). Results show ethidium bromide staining of resulting PCR products. Expected size is indicated for each. (B) Total RNAs from indicated cell lines were subjected to Northern analysis with cDNA probe DEST002. Single lines, 28S and 18S ribosomal bands. Arrowhead, probe specific signal. Each of the two blots were exposed for an equivalent 24-h time period. Each lane contains an equivalent amount of total RNA as judged by ethidium bromide staining of ribosomal RNA bands.

Discussion

This report describes modifications of a previously documented technique, the DDPCR, and demonstrates how this methodology may be used to isolate sequence tags representing differentially expressed genes from breast carcinoma biopsy specimens. Three of several initial tags isolated have been characterized. Direct PCR sequencing, database analysis, and identification of a control tag, C001, as the cDNA of ribosomal protein L26 demonstrates the feasibility of this technique. Two other tags represent apparently novel genes that are truly differentially expressed in carcinoma tissue versus normal breast tissue and are each specifically overexpressed in different human breast carcinoma cell lines. As such, both tags represent excellent candidate oncogenes involved in breast cancer pathogenesis. Currently, many more DESTs are being isolated from a large number of patient biopsy specimens using this high throughput protocol.

It is important to note a number of limitations of this methodology. Because biopsy specimens are used rather than homogeneous cell lines, a large number of isolated DESTs may represent differences in gene expression due to tissue heterogeneity between specimens rather than malignant versus nonmalignant changes. Furthermore, contamination of tumor specimens with normal breast tissue may obfuscate...
any cancer-specific loss of expression. In fact, no DEST which is lost or underexpressed in carcinoma has yet been isolated. The data in this report demonstrate that, despite potential cell heterogeneity, sequence tags of genes differentially expressed in malignant breast epithelial cells can be successfully isolated. Future attempts using microdissected biopsy specimens should further enhance isolation of genes relevant to the mammary epithelium.

A second concern also raised by others (6–8) is that, in using a small number of primer pairs (eight in this study), only a small fraction of expressed sequences is displayed. In fact, it has been calculated that a total of ~100 primer pairs would be needed to create a complete display of expressed messages (6–8). Therefore, any catalogue of sequences created by the methodology described in this report would be a partial list at best. The development of methods to incorporate a subtraction step (19) followed by differential display should allow for the simultaneous use of multiple primer pairs and a more enriched display of differentially expressed genes.

Finally, this methodology will only evaluate qualitative changes in gene expression. Because the PCR reaction is not quantitative, small changes in gene expression between specimens may not be apparent (7, 8). Furthermore, mutations that cause functional changes at the protein level would not be detected by this system. While such functional changes have been documented in breast carcinoma (20), our expectation is that even these changes will lead to qualitative changes in gene expression in downstream genes that would be detected by this system.

Despite such potential limitations, the ultimate goal of this project is to implement a rapid method for characterizing changes in gene expression in vivo and correlating these alterations with available clinical data. This concept is illustrated in Fig. 5. By creating a large catalogue of DESTs and accompanying patient data, it may be possible to identify sequence tags which correlate with treatment efficacy, relapse interval, and overall survival. This approach, as opposed to isolation of single genes or cDNAs, will have the advantage of allowing for a rapid search through many candidate sequences with minimal effort and simultaneous identification of multiple sequence tags that may collectively characterize a tumor phenotype. Such sequence tags could then be used as a panel of probes to predict the clinical course of patients in future prospective clinical studies and to isolate their corresponding biologically relevant, full-length genes whose characterization may open new and unforeseen windows into the pathobiology, treatment, and prevention of human breast cancer.

References

Isolation of Differentially Expressed Sequence Tags from Human Breast Cancer

Mark A. Watson and Timothy P. Fleming


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/54/17/4598

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.