Identification of Differentially Expressed Genes in Human Prostate Cancer Using Subtraction and Microarray

Jiangchun Xu, John A. Stolk, Xinquin Zhang, Sandra J. Silva, Raymond L. Houghton, Masazumi Matsumura, Thomas S. Vedvick, Kevin B. Leslie, Roberto Badaro, and Steven G. Reed


ABSTRACT

We have identified human prostate cancer- and tissue-specific genes using cDNA library subtraction in conjunction with high throughput microarray screening. Subtracted cDNA libraries of prostate tumors and normal prostate tissue were generated. Characterization of subtracted libraries showed enrichment of both cancer- and tissue-specific genes. Highly redundant clones were eliminated by colony hybridization. The remaining clones were selected for microarray to determine gene expression levels in a variety of tumor and normal tissues. Clones showing overexpression in prostate tumors and/or normal prostate tissues were selected and sequenced. Here we report the identification of two genes, P503S and P504S, from subtracted libraries and a third gene, P510S, by subtraction followed by microarray screening. Their expression profiles were further confirmed by Northern blot, real-time PCR (TaqMan), and immunohistochemistry to be overexpressed in prostate tissues and/or prostate tumors. Full-length cDNA sequences were cloned, and their subcellular locations were predicted by a bioinformatic algorithm, PSORT, to be plasma membrane proteins. The genes identified through these approaches are potential candidates for cancer diagnosis and therapy.

INTRODUCTION

Prostate cancer is the second leading cause of death among men in the United States. The American Cancer Society estimated that there would be approximately 179,000 new cases diagnosed with prostate cancer and 37,000 deaths in 1999 (1). Little is known about the genetic events in malignant transformation of prostatic cells. This is due in part to the cellular heterogeneity of the prostate tissues and a lack of genetic information from systematic analysis. Tissue- and/or cancer-specific genes can be used as markers for screening, diagnosis, prognosis, therapeutic monitoring, and/or follow-up for early indication of relapse. These genes or gene products can also be targeted by various agents, e.g., antibodies or drugs, for cancer therapy. Furthermore, therapeutic vaccines may be developed from these. The most commonly used marker for diagnosis and prognosis of prostate cancer is serum PSA. This has now replaced the previously used but less reliable comparison of gene expression between tissue types. Expression cloning using cancer patient serum (19) or T-cell lines or clones (20) from patients has identified a panel of genes that may be immunologically relevant, although none have yet been validated, and the process is not very efficient. We have analyzed the genetic composition of prostate and prostate cancer by combining a cDNA library subtraction method (21) and a microarray high throughput screening procedure (22). Here we report the discovery of three prostate tissue- and/or cancer-specific genes using these approaches.

MATERIALS AND METHODS

Tumor Samples and RNA Preparation. All normal and tumor tissue samples obtained from various clinical sources were accompanied by clinical information and pathological reports and were histologically confirmed by pathologists. The tissues were frozen in liquid nitrogen and homogenized with a polytron (Kinematica). Total RNA was prepared using Trizol reagent (Life Technologies, Inc.) poly(A)+ RNA was then purified using a Qiagen oligotex spin column mRNA purification kit.

cDNA Library Subtraction. cDNA library subtraction was performed using the protocol described by Hara et al. (21), with modifications. Briefly, cDNA libraries were first constructed with the Superscript Plasmid System for cDNA Synthesis and Plasmid Cloning Kit (Life Technologies, Inc.) using poly(A)+ RNA from prostate tumors, normal prostate tissue, and normal pancreatic tissue. The control cDNA library (70–150 μg; driver) was digested with EcoRI, NotI, and SfiI, followed by filling in with the DNA polymerase Klenow fragment. Driver DNA was labeled with Photobiotin (Vector Laboratories) and dissolved in 23 μl of H2O. To prepare tracer (tracer) DNA, cDNA libraries of normal or tumor prostate were digested with BamHI and XhoI, phenol chloroform-extracted, passed through Chroma spin-400 columns (Clontech), ethanol-purified, and dissolved in 5 μl of H2O. Tracer DNA was mixed with 15 μl of driver DNA and 20 μl of 2× hybridization buffer [1.5 M NaCl, 10 mM EDTA, 50 mM HEPES (pH 7.5), and 0.2% SDS], overlaid with mineral oil, heat-denatured, and incubated at 68°C for 20 h. The reaction mixture was then incubated with streptavidin and extracted with phenol/chloroform four times. This hybridization process was repeated with an additional 8 μl of driver DNA at 68°C for 2 h. Subtracted cDNA was ligated into the chloramphenicol-resistant PBC SK+ (Strategene) and transformed into ElectroMax Escherichia coli DH10B cells by electroporation to generate a tracer-specific subtracted cDNA library.

Library PCR. Library DNA (50 ng) was used as a template for PCR amplification of human β-actin at 18, 23, 28, and 33 cycles. PCR products were run on a 1% agarose gel and stained with ethidium bromide. Primers used for β-actin were as follows: (a) 5′ primer, 5′-AACCCTGCCTGCTGACC; and (b) 3′ primer, 5′-AGGAAGGAGGCGCTGAAAGAT.

Colonies Hybridization and Northern Blot Analysis. For colony hybridization, lifts were prepared with randomly picked colonies from subtracted libraries using 132 mm Hybond-N filters (Amersham Pharmacia Biotech). For

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.

Received 9/1/99; accepted 1/19/00.

1 Supported in part by National Cancer Institute Grant CA80518.

2 To whom requests for reprints should be addressed, at Corixa Corporation, 1124 Columbia Street, Suite 200, Seattle, WA 98104. Phone: (206) 754-5798; Fax: (206) 754-5917.

3 The abbreviations used are: PSA, prostate-specific antigen; BPH, benign prostatic hyperplasia; EST, expressed sequence tag; HGK-1, human glandular kallikrein 1; mAb, monoclonal antibody; poly(A)+ RNA, polyadenylated RNA; TCR, T-cell receptor; CCOII, cytochrome c oxidase subunit II.

1677
Northern blot analysis. 10 μg of total RNA were run out on formaldehyde denaturing gel, transferred to Hybond-N membrane, cross-linked, stained with methylene blue, and photographed. Radioactive 32P-labeled cDNA probes were prepared using a Ready-To-Go DNA labeling kit (Amersham). Filters were prehybridized in hybridization solution [1% BSA, 1 mM EDTA, 0.5 M NaHPO4 (pH 7.2), and 7% SDS] for 30 min at 65°C and replaced with fresh hybridization solution containing radioactive cDNA probe and hybridized overnight at 65°C. Washes were also carried out at 65°C with wash solution 1× SCP (0.1 M NaCl, 30 mM Na 2 HPO 4 , 1 mM EDTA, 1% N-lauroylsarcosine).

**Microarray.** mRNA expression of cDNA clones from subtracted libraries was determined using a high throughput microarray approach (22). Colonies that were negative from prescreening were randomly picked from subtracted libraries and PCR-amplified for 30 cycles with vector-specific primers according to a protocol suggested by Incyte. PCR products were then arrayed onto glass slides using Incyte patented chemistry. The arrayed cDNA clones were hybridized with a 1:1 mixture of Cy3- or Cy5-labeled first-strand cDNAs generated from poly(A)+ RNA from various tissues, including both normal and tumor tissues, using the protocol provided by Incyte. The fluorescence intensity was scanned, and data were analyzed using Incyte GEMTOOLS software. Results were also analyzed by normalizing fluorescence intensities between experiments using a subset of cDNA clones. This enabled us to compare data between different experiments.

**Quantitative Real-Time PCR (TaQMan).** Total RNA was treated with DNase I (Ambion) in the presence of RNasin (Promega) to remove DNA contamination before cDNA synthesis. cDNA was synthesized with oligodeoxythymidylic acid primer (Boehringer Mannheim) and Superscript II reverse transcriptase (Invitrogen). Real-time PCR (TaQMan) analysis was performed on a Perkin-Elmer/Applied Biosystems 7700 Prism. Matching primers and fluorescence probes (see below) were designed for each of the genes (P503S, P504S, and P510S) according to the Primer Express program provided by Perkin-Elmer/Applied Biosystems. Primer and probe concentrations were optimized with a pool of cDNAs from prostate tumors. For P503S and P504S, both forward and reverse primers were 900 nM. For P510S, both forward and reverse primers were 300 nM. In all cases, the final probe concentration was 160 nM. The PCR reaction was performed in 25 μl with dATP, dCTP, and dGTP at 0.2 mM and dUTP at 0.4 mM; 0.625 unit of Amplitaq Gold; 0.25 unit of Amperase uracil-N-glycosylase (Perkin-Elmer/Applied Biosystems); 5 mM MgCl2; trace amounts of glycerol, gelatin, and Tween 20 (Sigma); and 2 μl of cDNA template. β-Actin primers and probes were obtained from Perkin-Elmer/Applied Biosystems.

The following primers and probes were used: (a) P503S, TGCCCTCTG-GACGTTCCTT (forward primer), TTCTTCTTAGGGCAACCTACACT (reverse primer), and CACCACATGGCTGACATCTTGTGA (probe); (b) P504S, AAAATGGTTAATTACCCCCAGTTTGA (forward primer), TTCTTTCTAGACTAGAACCCTTACA (reverse primer), and TATCAAGCAAACTGAAAGGCGAATAACTCATAATT (probe); and (c) P510S, TTGAGACCTACCTAGAGTCTTATT (forward primer), GCAGAAGCAAGGATGTGT (reverse primer), and TTGATGGACACGTGTTTAAGGCCACACCACA (probe).

To quantitate the amount of specific mRNA in the samples, a standard curve was generated for each run using the plasmid containing the gene of interest (dilutions ranging from 20 to 2 copies). In addition, a standard curve was generated for β-actin ranging from 200 fg to 2 ng. This enabled standardization of the initial RNA content of a tissue relative to the amount of β-actin.

**Bioinformatic Analysis.** Protein localization was predicted by the PSORT algorithm using the amino acid sequences of P503S, P504S, and P510S.

**RESULTS**

cDNA Library Subtraction Is an Effective Way to Enrich Tissue- and/or Cancer-specific Genes. We generated subtracted cDNA libraries to enrich prostate tissue- and/or cancer-specific genes (Table 1). Subtraction one was done by subtracting prostate tumor (Gleason score = 7, 25 ng/ml PSA) with normal pancreas. Subtraction two was done by subtracting prostate tumor with normal prostate. Subtraction three was done by subtracting normal prostate with normal pancreas. Subtractions four and five were designed to enrich less abundant genes by spiking several abundant sequences identified from previous subtractions (e.g., PSA) into the driver. Subtraction four was also performed with a pool of three prostate tumors (tumor 1, Gleason score = 7, 100 ng/ml PSA; tumor 2, Gleason score = 8, 35 ng/ml PSA; tumor 3, Gleason score = 7, 59 ng/ml PSA) that differed from those used for previous subtractions. Normal pancreas was used in several subtractions as the driver control because both prostate and pancreas are secretory organs with glandular structures. Use of pancreas as a driver control should be effective in eliminating genes shared by both tissues.

Fig. 1A shows subtraction one, where several discrete bands are

![Fig. 1. cDNA library subtraction. A, subtraction one. The subtracted cDNA library was digested with BamHI and XhoI and run on a 1% agarose gel (Lane 2). Lane 1, 1-kb DNA ladder. B, subtraction five. The subtracted cDNA library was digested with BamHI and XhoI and run on a 1% agarose gel (Lane 4). Lane 1 is a 1-kb DNA ladder; Lanes 2 and 3 are prostate tumor and normal pancreas cDNA libraries, respectively.](image-url)

![Fig. 2. Subtraction efficiency showed by depletion of human β-actin in subtracted libraries. DNAs from nonsubtracted and subtracted cDNA libraries were PCR-amplified for 18, 23, 28, and 33 cycles with human β-actin-specific primers. PCR products were run out on a 1% agarose gel and stained with ethidium bromide. S-1, subtraction one; S-2, subtraction two; S-3, subtraction three; S-4, subtraction four; S-5, subtraction five.](image-url)
Sequence analysis of this library showed abundant clones to be PSA, HGK-1, CCOII, and TCR germ-line g chain that comigrate with the enriched bands. Less abundant tissue- and/or cancer-specific genes were enriched in the subtraction five library (Fig. 1B), where the three most abundant species were depleted by spiking them into the driver. Several discrete bands are seen in the subtraction five library (Fig. 1B, Lane 4). Abundant clones that comigrate with the enriched bands are tumor expression enhanced gene (TEEG), autonomously replicating sequence (ARS), a novel gene described in another study,4 and the two genes P503S and P504S.

The efficiency of subtraction was also demonstrated by depletion of the abundant housekeeping gene β-actin (Fig. 2). In nonsubtracted libraries, β-actin-specific PCR product is visible by eighteenth cycle of amplification and becomes saturated at 28–33 cycles. Subtracted libraries require a higher number of amplification cycles for β-actin to be detected, indicating that β-actin is preferentially depleted in subtracted cDNA libraries. Depletion of β-actin is more significant in the subtraction one, two, and three libraries, suggesting that these three subtractions are more complete.

Microarray Screening of cDNA Clones from Subtracted Libraries for Assessment of Tissue and Cancer Specificity. Subtracted cDNA libraries were prescreened by colony hybridization with cDNAs (PSA, HGK-1, CCOII, ARS, TEEG, aldehyde dehydrogenase 6, TCR germ-line γ chain, and sequence 3 from patent number 5565323) to eliminate redundant cDNA clones. A total of 500 prescreen negative clones, including P503S and P504S, were selected, and their mRNA expression levels were determined by microarray analysis. Clones showing overexpression in normal and/or tumor prostate tissues were sequenced to determine their identity.

Fig. 3 shows mRNA expression levels determined by microarray

![Fig. 3. Prostate gene expression determined by microarray. cDNA clones were PCR-amplified, arrayed onto glass slides, and probed with a 1:1 mixture of Cy3-labeled probe 1 and Cy5-labeled probe 2. Prostate tumors used had Gleason scores of 3–8 with a PSA value of 1.7–35 ng/ml. Fluorescence scans represented in pseudocolor correspond to hybridization intensities. Fluorescence intensity values were normalized using a subset of cDNA clones. Background fluorescence value is 215 ± 238, which is defined by the fluorescence values of empty spots (spots with no DNA target on glass slides, n = 8) for all probe pairs (n = 22).](image)

![Fig. 4. Northern blot analysis of prostate gene expression in prostate and other tissues. Ten μg of total RNA were run on a formaldehyde denaturing gel, blotted, and probed with random-primed cDNA probes of P503S, P504S, and P510S. Lanes 1–4, prostate tumors; Lanes 5 and 6, normal prostate; Lanes 7 and 8, BPH; Lane 9, normal colon; Lane 10, normal kidney; Lane 11, normal liver; Lane 12, normal lung; Lane 13, normal pancreas; Lane 14, normal skeletal muscle; Lane 15, normal brain (P510S, Lane 15, normal skeletal muscle); Lane 16, normal stomach; Lane 17, normal testis; Lane 18, normal small intestine; Lane 19, normal bone marrow. Loading control is the RNA gel stained with methylene blue showing 18S and 28S bands.](image)
P503S, P504S, and another gene, P510S, which was identified through microarray screening. P503S is overexpressed in prostate tumors, BPH, and normal prostate. It is also expressed in normal colon and is slightly elevated in normal kidney and bladder. Other normal tissues tested had low or undetectable levels of P503S expression. P504S is overexpressed in about 30% of prostate tumors and is low to undetectable in normal tissues tested, including normal prostate, suggesting that P504S is a prostate cancer-specific gene. P510S is primarily overexpressed in a subset of prostate tumors and BPH, whereas the expression levels in nonprostate normal tissues are low to undetectable. Although microarray is a rapid and effective way to screen hundreds to thousands of clones in parallel, expression values may be compressed due to the limitation of detection sensitivity for rare message and the saturation of signal for highly abundant genes. Therefore, other independent methods are necessary to give a more accurate assessment of mRNA expression levels.

Tissue and/or Cancer Specificity Was Determined by Northern Blot Analysis. Consistent with the microarray analysis, P503S was expressed in prostate tumors, normal prostate tissues, and normal colon and to a lesser extent in normal kidney and lung (Fig. 4). It was undetectable in other normal tissues tested including liver, pancreas, skeletal muscle, brain, stomach, testis, small intestine, and bone marrow. P504S was overexpressed in 50% (two of four) of prostate tumors but was low or undetectable in other tissues tested, including normal prostate. P510S was overexpressed in 50% of normal and tumor prostate tissues but was absent from other tissues tested. Therefore, Northern blot analysis confirmed the expression patterns observed with the microarray approach, supporting the use of microarray as a valid high throughput screening approach.

Gene Expression Measured by Quantitative PCR (TaqMan) Analysis. An additional more sensitive and quantitative approach, real-time PCR (TaqMan), was used to assess the overexpression of prostate candidates (Fig. 5). Using real-time PCR (TaqMan), P503S is highly expressed in prostate tumors and normal prostate, as well as in a subset of other tumors (breast tumors, ovarian tumors, and colon tumors but not lung tumors). It was also expressed by normal kidney and stomach but detected at a very low level in other normal tissues. P504S is overexpressed in about 60% (four of seven) of prostate tumors. With the exception of normal liver that shows a low level of message, other normal tissues including normal prostate show little expression of P504S. P510S was overexpressed in three of eight prostate tumors tested. It was expressed at lower levels in two other prostate tumors and two of seven normal prostate tissues. Other normal tissues did not have detectable P510S message. Thus, quan-

Fig. 5. Real-time PCR (TaqMan) analysis of P503S, P504S, and P510S. PCR was performed on a Perkin-Elmer/Applied Biosystems 7700 Prism instrument using cDNA reverse-transcribed from RNA of various tissues. β-Actin was used as an internal reference control. Tumors have Gleason scores of 5–8 and PSA levels of 8.1–66 ng/ml. Results are representative of three experiments for each gene.

Fig. 6. Immunohistochemistry staining for P503S and P504S. Prostate tumor (A) and BPH (B) were stained with rabbit mAb against P503S. Prostate tumor (C) and BPH (D) were stained with rabbit mAb against P504S. A total of five prostate carcinomas, grade 3–5, and one BPH have been examined with these antibodies, and results are similar on all prostate carcinomas.
Protein Expression Determined by Immunohistochemistry. mRNA expression profiles of P503S and P504S in prostate tumors and normal prostate tissues were confirmed by immunohistochemistry using rabbit mAbs raised against P503S and P504S proteins. mAb against P503S reacted with both prostate tumors and BPH (Fig. 6), whereas P504S mAb stained prostate tumors but not BPH. Immunohistochemistry also indicated that P503S staining was cell surface and cytoplasmic, whereas P504S showed strong granular cytoplasmic staining.

Full-length cDNAs and Prediction of Protein Structures. Using various approaches, including bioinformatics and colony hybridization, we cloned full-length sequences for all three genes. Localization of protein in different subcellular compartments was predicted by PSORT. The full-length cDNA of P503S, with 1228 bp and an open reading frame of 241 amino acids, was identified as human tetraspan NET-1 (GenBank accession number 3152700), a member of the tetraspan/TM4SF family shown to be involved in cancer metastasis (24). Consistent with what has been shown for other family members, P503S is predicted to be a type IIIa plasma membrane protein with two transmembrane spans and a cleavable signal sequence. P504S is a 1621-bp cDNA with an open reading frame of 382 amino acids. It has been identified as human α-methylacyl-CoA racemase (GenBank accession number 4204097). Although P504S is predicted to be a type Ib plasma membrane protein with one transmembrane span and no signal sequence, the rat homologue has been shown to be expressed in the cytosol and mitochondria (25). P510S has been identified as human ABC transporter MOAT-B (GenBank accession number 3335173). It is predicted to be a plasma membrane protein with nine potential transmembrane spans with no NH2-terminal signal sequence.

Discussion
Coupling subtraction and microarray is a rapid and efficient way to identify differentially expressed genes. cDNA library subtraction is a powerful approach to enrich tissue- or cancer-specific genes. The efficiency of subtractions was shown by depletion of the highly and ubiquitously expressed gene β-actin. Furthermore, the presence of other known prostate-specific genes (e.g., PSA and HGK-1) at high frequency in subtracted libraries also suggests that the subtraction process worked effectively. Enriched tissue- and/or cancer-specific libraries can serve as ideal target sources for microarray screening. Target selection is especially important for microarray. It can be very inefficient to use nonnormalized or nonsubtracted libraries. On the other hand, arrays with a limited number of preselected genes, which only constitute a small percentage of the human genome, will be neither comprehensive nor systematic. Other problems associated with microarray include the use of cell line RNAs as probes that may not reflect natural biological processes, and the use of a limited number of probes that may not cover enough cell types and tissues from different organs in body. In our study, we used primary tissues, both normal and tumor. We also used a broad spectrum of normal tissues to determine the tissue distribution of genes. We are currently performing studies with a larger number of tumor samples with well-documented histology information to further correlate gene expression and tumor histology.

Vasmatzis et al. (15) have reported the discovery of three genes specifically expressed in human prostate by using an electronic subtraction approach with EST databases. Three of the five most abundant genes discovered by electronic subtraction, PSA, HGK-1, and the TCR γ chain C region gene, are also abundantly present in our subtracted libraries. This indicates that the two approaches could generate similar but not identical results. It is postulated that the reasons for not identifying prostate-secreted seminal plasma protein and prostatic acid phosphatase gene in our experiment are as follows: (a) tumor samples used in our studies did not contain these two genes at high levels; and (b) normal pancreas may also express these genes or their homologues at some level; therefore, they were eliminated by subtraction process. Electronic subtraction is a rapid and powerful approach to identify potential differentially expressed genes; however, as the authors pointed out, it has some disadvantages. The biggest limitation is that it depends on the availability of EST sequences. The incompleteness of the EST database and the random nature of EST clones may cause the specificity determined by electronic subtraction to be false positive or false negative. Information regarding specificity is especially unreliable when gene expression is low. Furthermore, the algorithm the authors used can generate false EST clusters because one cluster can contain several different genes, and one gene can have several clusters. This may also cause specificity information to be inaccurate. Therefore, information obtained by electronic subtraction needs to be confirmed by experimental approaches.

The cDNA library subtraction approach described here systematically compares cDNAs from two different tissue types and preferentially depletes sequences common in both tissue types, thereby enriching sequences specific for one tissue type. This procedure eliminates the problems associated with electronic subtraction such as incompleteness of the EST database and false clustering. We have also performed additional subtractions that incorporated the following modifications to the procedure: (a) pooling and using more tumor samples to identify cancer-associated genes present only in a small subpopulation of cancer patients; (b) adjusting the driver/tracer ratio to identify less abundant genes and reduce the redundancy of subtracted libraries; and (c) using a pool of normal tissues as drivers, especially the ones that share similar cellular components with prostate tumor to eliminate nonspecific clones more efficiently; subtractions performed in this manner generated results similar to those reported herein. Alternative subtraction approaches, such as PCR-based subtraction (Clontech), have also been used to identify less abundant genes.

Tissue and cancer specificity is one of the key requirements for a marker to be used for diagnosis and therapy. We have shown that genes identified by subtraction and microarray approaches can prime human immunological responses, including both humoral and cell-mediated responses. Therefore, these genes can be used as potential vaccine candidates for cancer immunotherapy. Our approaches provide alternative methods for cancer antigen identification. This is extremely important for cancers other than melanoma, where cancer immunogenicity is weak, and it is difficult to use immunological approaches for cancer antigen identification.

Acknowledgments
We gratefully acknowledge Incyte for help with microarray technology. We thank Dr. Michael Kalos for valuable suggestions regarding the manuscript. Some tissue samples were obtained from the Cooperative Human Tissue Network, which is funded by the National Cancer Institute, and from the National Disease Research Interchange.

References


1681


Identification of Differentially Expressed Genes in Human Prostate Cancer Using Subtraction and Microarray

Jiangchun Xu, John A. Stolk, Xinqun Zhang, et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/60/6/1677

Cited articles
This article cites 24 articles, 16 of which you can access for free at:
http://cancerres.aacrjournals.org/content/60/6/1677.full.html#ref-list-1

Citing articles
This article has been cited by 38 HighWire-hosted articles. Access the articles at:
/content/60/6/1677.full.html#related-urls

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.