Peter S. Nelson[1,2,3]
David Han[2]
Yvan Rochon[2]
Garry L. Corthals[4]
Biaoyang Lin[2]
Adam Monson[2]
Vilaska Nguyen[2]
B. Robert Franza[2,5]
Stephen R. Plymate[3]
Ruedi Aebersold[2]
Leroy Hood[2]

[1]Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
[2]Department of Molecular Biotechnology, University of Washington, Seattle, WA, USA
[3]Department of Medicine, University of Washington, Seattle, WA, USA
[4]Garvan Institute for Medical Research, Sydney, NSW, Australia
[5]Department of Cell Systems Initiative, University of Washington, Seattle, WA, USA

Comprehensive analyses of prostate gene expression: Convergence of expressed sequence tag databases, transcript profiling and proteomics

Several methods have been developed for the comprehensive analysis of gene expression in complex biological systems. Generally these procedures assess either a portion of the cellular transcriptome or a portion of the cellular proteome. Each approach has distinct conceptual and methodological advantages and disadvantages. We have investigated the application of both methods to characterize the gene expression pathway mediated by androgens and the androgen receptor in prostate cancer cells. This pathway is of critical importance for the development and progression of prostate cancer. Of clinical importance, modulation of androgens remains the mainstay of treatment for patients with advanced disease. To facilitate global gene expression studies we have first sought to define the prostate transcriptome by assembling and annotating prostate-derived expressed sequence tags (ESTs). A total of 55 000 prostate ESTs were assembled into a set of 15 953 clusters putatively representing 15 953 distinct transcripts. These clusters were used to construct cDNA microarrays suitable for examining the androgen-response pathway at the level of transcription. The expression of 20 genes was found to be induced by androgens. This cohort included known androgen-regulated genes such as prostate-specific antigen (PSA) and several novel complementary DNAs (cDNAs). Protein expression profiles of androgen-stimulated prostate cancer cells were generated by two-dimensional electrophoresis (2-DE). Mass spectrometric analysis of androgen-regulated proteins in these cells identified the metastasis-suppressor gene NDKA/nm23, a finding that may explain a marked reduction in metastatic potential when these cells express a functional androgen receptor pathway.

Keywords: Prostate / Transcriptome / Proteome / Androgen / Microarray EL 3957

1 Introduction

The development and subsequent progression of human prostate carcinoma is propelled by the accumulation of genetic alterations and influenced by environmental factors. One pivotal mediator of prostate carcinogenesis is the androgen receptor (AR) pathway. The majority of prostate cancers initially require androgens for growth, and the elimination of AR-ligands by surgical or chemical castration leads to marked tumor regression through a mechanism of programmed cell death [1]. The manipulation of the AR pathway has been used in clinical medicine since the 1940s as the primary treatment of advanced prostate cancer. However, this therapy is palliative and eliminates the potential beneficial effects of androgen-induced cellular differentiation. Surviving cancer cells lose their dependence on androgens over time and are capable of prolifertion in the absence of serum androgens. The molecular events leading to androgen independence (AI) have not been defined, but potential mechanisms include overexpression of the AR, mutations in the AR gene leading to promiscuous ligand binding, and the activation of the AR or downstream regulatory molecules by other endocrine or paracrine growth factors [2, 3].

Until recently, biological investigations have almost entirely focused on the study of individual genes and proteins. This has partly been due to the submicroscopic nature and transient existence of relevant molecules, combined with a lack of quantitative technology capable of providing accurate comprehensive views of biological complexity. Significant advances have been achieved studying individual genes, proteins and small numbers of molecular interactions. However, conclusions made on the basis of the study of an individual gene may have limited relevance as to how the gene and gene product function in the context of the whole cell, tissue, or organism. Progress in understanding complex molecular processes

Correspondence: Peter Nelson, M.D., Division of Human Biology, Mailstop D4-100, 1100 Fairview Avenue, Seattle, WA 98109-1024, USA
E-mail: pnelson@fhcrc.org
Fax: +206-667-2917

Abbreviations: AR, androgen receptor; **cDNA**, complementary DNA; μ**LC**, microcapillary LC; **PEDB**, prostate expression database; **PSA**, prostate specific antigen

 0173-0835/00/0909-1823 $17.50+.50/0

has been hampered by the lack of a complete inventory or "tool-set" of all genes and their cognate proteins that are necessary for defining phenotypes of normal and pathological cellular states.

The completion of the Human Genome Project will provide a foundation for a thorough description of this molecular inventory. More specifically, the gene inventory or tool set required for studies of prostate carcinogenesis is that portion of the human genome used or expressed in the human prostate gland. The subset of genes transcribed or expressed in a given cell or tissue type such as the prostate may be defined as the "transcriptome", the dynamic link between the genome, the proteome, and the cellular phenotype associated with physical characteristics [4]. Once a transcriptome has been described, the next objective is to understand the relationships of the genes and their protein products in terms of a complex system, *e.g.*, biological pathways and networks, that may define health and disease. With this aim, novel technologies for comprehensively assessing genomes and patterns of gene expression have recently been developed.

Our initial efforts have focused on defining the prostate transcriptome through the production and assembly of expressed sequence tags (ESTs) derived from prostate complementary DNA (cDNA) libraries representing a wide sprectrum of normal and neoplastic states. These EST assemblies have been used to construct cDNA microarrays suitable for interrogating the transcriptome in experiments designed to examine specific biological pathways that may be involved in prostate carcinogenesis. The molecular pathway mediating androgenic hormone action on prostate cells is a specific focus of our work. The functional architecture of prostate gene networks is furth elucidated by our next level of analysis that incorporates studies of the prostate proteome. Analysis of the transcriptome facilitates proteome studies by providing a comprehensive prostate sequence database for identifying and annotating known and unknown proteins displayed by two-dimensional gel electrophoresis (2-DE) and analyzed by mass spectrometry (MS). Our objectives for delineating the molecular network(s) influenced by AR activation are to identify specific targets that promote the differentiation and apoptotic potential of prostate cancer cells while reducing their ability to proliferate.

2 Materials and methods

2.1 Assembly of a prostate transcriptome: Prostate Expression Database (PEDB)

A prostate transcriptome was assembled from ESTs derived from cDNA libraries representing a wide sprectrum of normal, benign, and malignant prostate tissues. A detailed description of the assembly and annotation procedure is described elsewhere [5]. Briefly, individual ESTs, detailed cDNA library information, and sequence annotations were loaded into a relational database (Oracle Corp.) termed the Prostate Expression Database (PEDB). Prostate ESTs used for the assembly were generated in our laboratory as previously described [6]. Additional public domain ESTs of prostate origin were obtained from Genbank (http://www.ncbi.nlm.nih.gov/Entrez/batch.html), the NCI Cancer Genome Anatomy Project (CGAP) [7], and The Institute for Genome Research (TIGR) (http://www.tigr.org). Each EST was examined for sequence homology to cloning vectors, *Escherichia coli*, and repetitive DNA sequences using a core program called AnalDemon (http://www.mbt.washington.edu/PEDB/software). AnalDemon first employs Cross_Match (http://bozeman.mbt.washington.edu/phrap.docs/general.html); a program based on the Smith-Waterman-Gotoh algorithm, to screen for vector and *E. coli* genome contamination. ESTs are then examined for interspersed repeats and regions of low sequence complexity using Repeatmasker (http://ftp.genome.washington.edu/RM/RepeatMasker.html). Specific portions of EST sequences exhibiting homology to any of these unwanted elements are masked in order to eliminate the sequence from contributing to an assembly process. CAP2 [8], a multiple alignment program based on a variant of the Smith and Waterman algorithm, was used for assembling ESTs into homologous groups or clusters. Clustering is based on maximal scoring of overlapping alignments and allows for general substitutions resulting from sequencing errors, insertions, and deletions. CAP2 produces a consensus sequence and allows varying sensitivity and overlap parameters. Each group or cluster of ESTs exhibiting significant homology with one another is termed a species. Thus, a species is a sequence or group of sequences that is unique relative to the nucleotide sequence of other groups of sequences, and each is given a unique PEDB Species Identification number (SID). The SID provides a means to analyze gene expression across the entire set of assemblies, and can be used to provide a library-by-library species-specific differential expression profile. Each distinct species from the clustering process was annotated by searching the Unigene (ncbi.nlm.nih.gov in /pub/schuler/unigene), Genbank (ncbi.nlm.nih.gov/blast/db/nt.Z), and EST databases (ncbi.nlm.nih.gov/blast/db/est.Z) using BLASTN (http://blast.wustl.edu). Annotations were assigned automatically using the program SmartBlast (http://www.mbt.washington.edu/PEDB/software) by selecting the database match with the lowest p value and the highest blast score where the maximum p value is e^{-20} and the minimum blast score is 500.

2.2 Prostate transcriptome analyses by cDNA microarray

2.2.1 Microarray fabrication

A nonredundant set of 1500 prostate-derived cDNA clones was identified from the prostate transcriptome archived in PEDB. Individual clone inserts were amplified by the PCR using 2 μL of bacterial transformant culture as template with primers BL_m13F (5′-GTAAAACGACGGCCAGTGAATTG-3′) and BL_m13R (5′-ACACAGGAAACAGCTATGACCATG-3′ as previously described [6]. PCR products were purified through Sephacryl S500 (Amersham Pharmacia Biotech, Uppsala, Sweden), mixed 1:1 with dimethylsulfoxide, and spotted in duplicate onto coated Type VII glass microscope slides (Amersham Pharmacia Biotech) using a Molecular Dynamics (Sunnyvale, CA, USA) GenII robotic spotting tool. After spotting, the glass slides were air-dried and UV-cross-linked with 500 mJ of energy and then baked at 95°C for 30 min.

2.2.2 Probe construction and microarray hybridization

Total RNA was isolated from the androgen-responsive LNCaP prostate cancer cells [9] at time points of 0, 4, 8, 24, and 72 h after androgen depletion or supplementation using TRIzol (Life Technologies, Paisley, UK) according to the manufacturer's directions. Fluorescence-labeled probes were made from 30 μg of total RNA in a reaction volume of 20 μL containing 1 μL anchored oligo-dT primer (Amersham Pharmacia Biotech), 0.05 mM Cy3-dCTP (Amersham Pharmacia Biotech), 0.05 mM dCTP, 0.1 mM each dGTP, dATP, dTTP, and 200 U Superscript II reverse transcriptase (Life Technologies). Reactants were incubated at 42°C for 120 min followed by heating to 94°C for 3 min. Unlabeled RNA was hydrolyzed by the addition of 1 μL of 5 N NaOH and heating to 37°C for 10 min. One μL of 5 M HCl and 5 μL of 1 M Tris-HCl, pH 7.5, were added to neutralize the base. Unincorporated nucleotides and salts were removed by chromatography (Qiagen, Chatsworth, CA, USA), and the cDNA was eluted in 30 μL dH2O. One μg of dA/dT 12–18 (Amersham Pharmacia Biotech) and 1 μg of human Cot1 DNA (Life Technologies) were added to the probe, heat-denatured at 94°C for 5 min, combined with an equal volume of 2 × microarray hybridization solution (Amersham Pharmacia Biotech) and prehybridized at 50°C for 1 h. The mixture was then placed onto a microarray slide with a coverslip and hybridized in a humid chamber at 52°C for 16 h. The slides were washed once with 1 × sodium chloride and sodium citrate (SSC), 0.2% SDS at room temperature for 5 min and then twice with 0.1 × SSC, 0.2% SDS at room temperature for 10 min. After washing, the slide was rinsed in distilled water to remove trace salts and dried.

2.2.3 Image acquisition and data analyses

Fluorescence intensities of the immobilized targets were measured using a laser confocal microscope (Molecular Dynamics). Intensity data were integrated at a pixel resolution of 10 μm using approximately 20 pixels per spot, and recorded at 16 bits. Quantitative data were obtained with the SpotFinder Version 2.4 program written at the University of Washington. Local background hybridization signals were subtracted prior to comparing spot intensities and dtermining expression ratios. For each experiment, each cDNA was represented twice on each slide, and the experiments were performed in duplicate producing four data points per cDNA clone and hybridization probe. Intensity ratios for each cDNA clone hybridized with probes derived from androgen-stimulated LNCaP and androgen-starved LNCaP cells were calculated (stimulated intensity/starved intensity). Gene expression levels were considered significantly different between the two conditions if all four replicate spots for a given cDNA demonstrated a ratio > 2 or < 0.5, and the signal intensity was greater than two standard deviations above the image background. We have previously determined that expression ratios less than 1.5 are not reproducible in our system (datas not shown).

2.3 Prostate proteome analyses by 2-DE and MS

2.3.1 2-DE

LNCaP prostate cancer cells were grown under conditions of androgen stimulation or androgen starvation as described above. M12AR cells, a highly metastatic prostate cancer cell line derived from the serial passaging of SV40 immortalized prostate epithelial cells [10] and transfected with the AR were grown in serum-free DMEM high-glucose media (Life Technologies) supplemented with insulin, transferrin, selenium, and dexamethasone as previously described [11]. Cells were allowed to reach 80% confluency and then treated for 24 h with the same media supplemented with 10 nM R1881. Cells were washed once with PBS, scraped from plates with a rubber policeman and pelleted by centrifugation. Protein was harvested as described by Garrels and Franza [12]. Briefly, cell pellets were lysed in a buffer containing 0.3% SDS, 1% β-mercaptoethanol, and 50 mM Tris-HCl, pH 8.0, 100 μg/mL DNAase I, 50 μg/mL RNAase A, 5 mM $MgCl_2$, and heated for 1 min at 100°C. Harvested protein was lyophilized, resuspended in isoelectric focusing (IEF) gel rehydration solution, and stored at –80°C. Soluble proteins were run in the first dimension by using a commercial flatbed electrophoresis system (Multiphor II; Amersham Pharmacia Biotech). Nonlinear immobilized pH gradient (IPG) dry strips ranging from 3.0 to 10.0 (Amer-

sham Pharmacia Biotech) were used for the first-dimensional separation. Forty micrograms of protein from whole-cell lysates were mixed with IPG strip rehydration buffer (8 M urea, 2% Nonidet P-40, 10 mM dithiothreitol), and 250–380 μL of solution (13 and 18 cm IPGs, respectively) was added to individual lanes of an IPG strip rehydration tray (Amersham Pharmacia Biotech). The strips were rehydrated at room temperature for 1 h. The samples were run at 300 V, 10 mA, 5 W for 2 h, ramped to 3500 V, 10 mA, 5 W over a period of 3 h, and then kept at 3500 V, 10 mA, 5 W for 15–19 h. Following IEF (60–70 kVh), the IPG strips were first reequilibrated for 8 min in a solution of 2% w/v dithiothreitol, 2% w/v SDS, 6 M urea, 30% w/v glycerol, 0.05 M Tris-HCl (pH 6.8) and subsequently for 4 min in a solution of 2.5% w/v iodoacetamide, 2% w/v SDS, 6 M urea, 30% w/v glycerol, 0.05 M Tris-HCl (pH 6.8) with a trace of bromophenol blue added for color. Following reequilibration, the strips were transferred and apposed to 10% polyacrylamide second-dimensional gels. Polyacrylamide gels were poured in casting stand with 10% acrylamide-2.67% piperazine diacrylamide-0.375 M Tris, pH 8.8, 0.1% w/v SDS, 0.05% w/v ammonium persulfate, 0.05% TEMED (*N,N,N′,N′*-tetramethylethylenediamine) in Milli-Q water (Millipore, Bedford, MA, USA). Second-dimensional gels (0.1 × 20 × 20 cm) were run in an apparatus supplied by Oxford Glycosciences (Abington, UK). Once the IPG strips were apposed to the second-dimensional gels, they were immediately run at a constant current of 50 mA at 500 V and 85 W for 20 min, followed by a constant current of 200 mA at 500 V and 85 W until the buffer front was 10–15 mm from the bottom of the gel. Gels were removed and silver stained according to the procedure of Blum *et al.* [13].

2.3.2 Protein identification by tandem mass spectrometry

Protein spots from gels were identified by tandem mass spectrometry (MS/MS) as previously described [14]. Spots from silver-stained gels were excised and in-gel tryptic peptides were separated by microcapillary LC (μLC) coupled to a tandem mass spectrometer (TSQ 7000; Finnigan, San Jose, CA). Peptide fragmentation spectra were generated in a data-dependent fashion. Spectra were searched against the composite OWL protein sequence database by using the computer program SEQUEST [15] and against the PEDB. A protein match was determined by comparing the number of peptides identified and their respective cross-correlation scores. Protein identifications were verified by comparison with theoretical molecular weights and isoelectric points.

3 Results and discussion

3.1 Prostate gene expression analyses: EST assemblies and annotation

ESTs produced from cDNA libraries derived from normal and neoplastic human prostate tissue samples were entered into the PEDB, an Oracle relational database running on a Sun SPARC workstation. The most recent PEDB build was assembled starting with 55 000 prostate ESTs produced from 42 cDNA libraries. Portions of EST sequences with homology to cloning vector, *E. coli* genomic DNA, and human repetitive DNA sequences

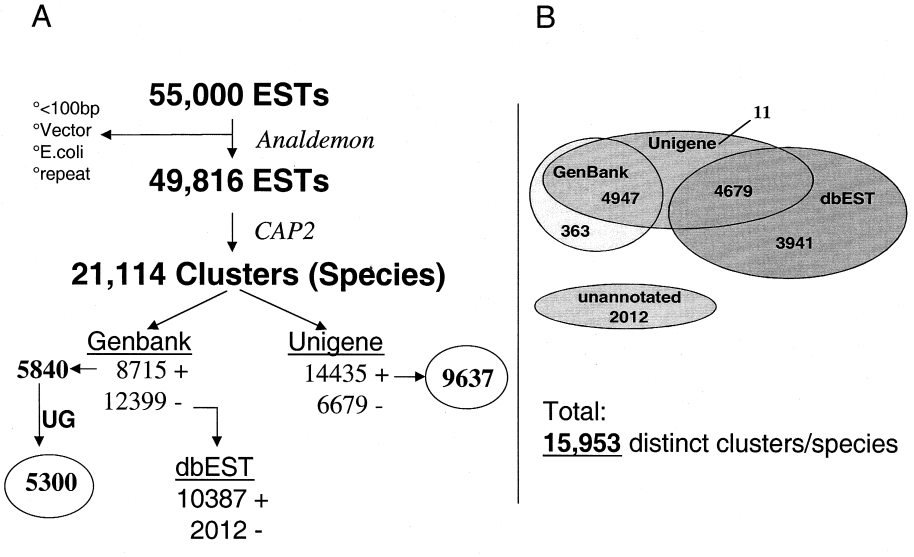


Figure 1. Assembly of a prostate transcriptome. (A) 55 000 prostate ESTs were examined for "junk" sequences leaving 49 816 high quality ESTs suitable for assembly. Clustering the ESTs into groups of high homology produced a set of 21 114 clusters that were annotated against nucleotide and protein sequences in the public sequence databases. Clusters exhibiting homology to Genbank sequences were also examined for homology to Unigene sequences (UG) to further collapse clusters into homologous groups. (B) Following clustering, database annotations and reclustering, a total of 15 953 distinct prostate EST species were identified. More than 2000 prostate species did not have homology to nonprostate-derived sequences in the public databases (unannotated).

were masked and ESTs with > 100 bp of high quality sequence were admitted to the assembly process (Fig. 1A). A total of 49 816 high quality ESTs were assembled using the sequence assembly program CAP2 to produce 21 114 clusters. Each cluster was annotated by searching the Unigene, Genbank, and dbEST databases with the CAP2-generated cluster consensus sequences using BLASTN. Clusters annotating to the same database sequence were joined to further reduce the number of distinct clusters to 15 953 (Fig. 1B).

Studies in the 1970s using reassociation kinetics to estimate the number of different transcripts indicate that between 10 000 and 30 000 distinct mRNAs are present in mammalian cells or organs [16, 17]. Recent data produced using the method of Serial Analysis of Gene Expression (SAGE) suport these estimates of transcript diversity in mammalian epithelial cells with estimates of 14 000–20 000 different mRNAs per cell [18]. Although the identification of alternatively spliced transcripts and highly homolgous gene family members may increase or decrease these estimates slightly, they nevertheless provide a rough estimate of the complexity of cellular gene activity. Based upon these data, the 15 953 prostate EST clusters that we have assembled should characterize roughly 50–75% of the prostate transcriptome. It is likely that this assembled dataset comprises all of the abundant and most of the moderately abundant prostate transcripts [6]. Ongoing work involves the acquisition of the remaining low abundance transcripts. Approaches to achieving this goal involve the construction of cDNA libraries from highly selected purified cell populations such as luminal epithelial and neuroendocrine cells, and from prostate tissues at different stages of development (*e.g.*, fetal prostate) or under different hormonal influences (*e.g.*, androgen stimulation). Another useful strategy involves the iterative removal of abundant and previously identified cDNAs in order to select for rare species. A high-throughput method using cDNA array-based technology has been developed to facilitate this process [19].

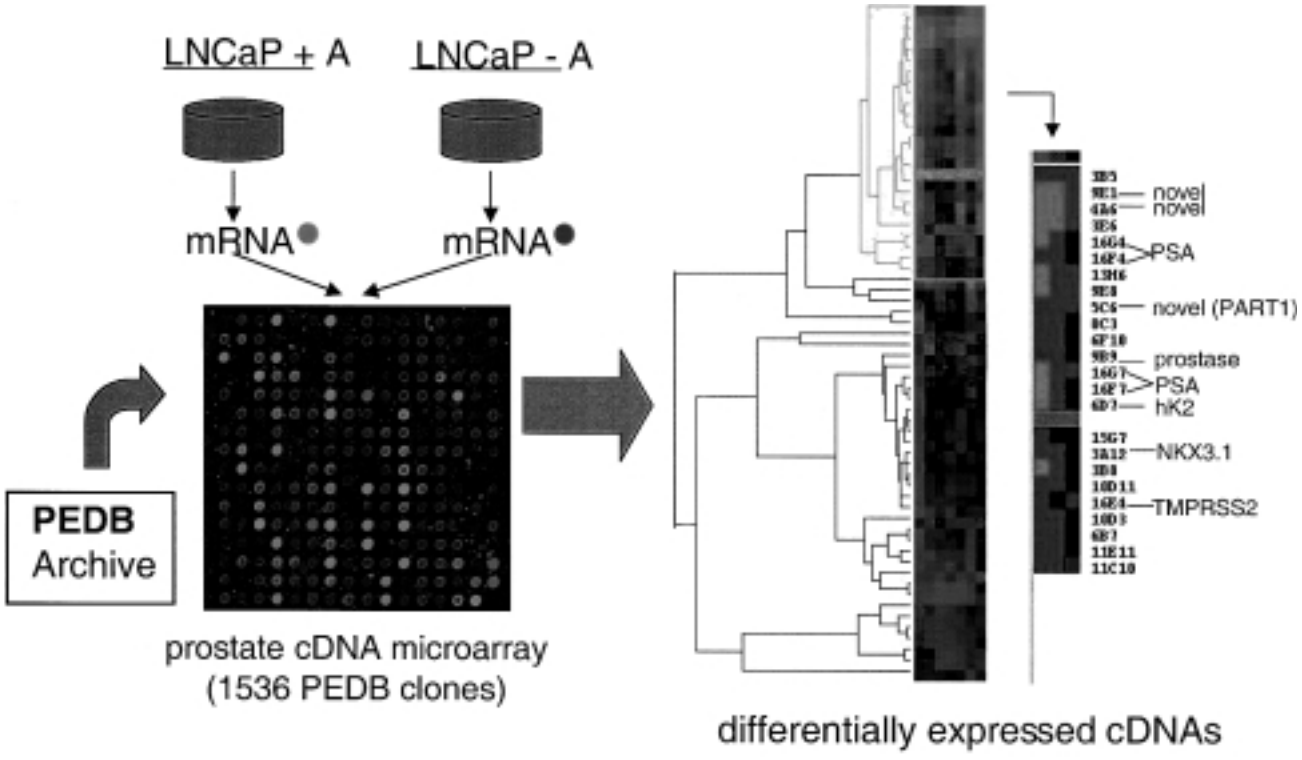


Figure 2. cDNA microarray analysis of prostate androgen-regulated gene expression. A nonredundant clone set comprised of 1536 cDNAs was hybridized with Cy3-labeled (red) cDNA from androgen-stimulated LNCaP cells and Cy5 labeled (green) cDNA from androgen-starved LNCaP cells. The expression ratio for each cDNA was determined and the ratios for all cDNAs with signal intensities 2.33-fold above the standard deviation of the background signal were clustered according to transcript levels over time. The Cluster and TreeView software programs availabe at the Stanford genome web site was used for the analysis (http://rana.Stanford.EDU/software/). Twenty genes were identified with increased expression after androgen stimulation.

3.2 Prostate gene expression analyses: cDNA microarray

Microarrays comprised of 1500 distinct prostate-derived cDNAs were hybridized with fluorescently labeled total cDNA probes produced from androgen-stimulated and androgen-starved LNCaP prostate cancer cells. No cDNAs were identified whose expression level decreased with androgen stimulation. In contrast, the hybridization ratios of 20 different cDNAs were consistently increased by > 2-fold in androgen-stimulated relative to androgen-starved cells (Fig. 2). This group included cDNAs encoding the human glandular kallikrein 2 (hK2) and human glandular kallikrein 3 (hK3), also known as prostate-specific antigen (PSA). The regulation of hK2 and PSA has previously been shown to be mediated by androgens through a mechanism involving androgen-response element (ARE) binding sites in the promoter regions of these genes [20, 21].

In addition to hK2 and PSA, we identified several other genes previously shown to be androgen-regulated, including the prostate homeobox gene NKX3.1 [22], the serine protease prostase/PRSS17 [23], and two genes involved in lipid metabolism. The microarray analysis also indicated that the expression of the membrane-bound serine protease TMPRSS2 [24] was regulated by androgen. We subsequently confirmed the androgen regulation of TMPRSS2 by Northern analysis, identified a putative ARE in the TMPRSS2 promoter region, and demonstrated that TMPRSS2 is highly expressed in the prostate gland relative to other human tissues [25]. Several cDNAs corresponding to uncharacterized genes also exhibited transcriptional regulation by androgen (Fig. 2). We have cloned the full-length cDNA and confirmed the androgen regulation of one of these novel sequences and designated it as PART-1, for Prostate Androgen-Regulated Transcript-1, as it lacks significant homology to nucleotide or protein sequences in the nonredundant subdivision of the GenBank and SWISS-Prot databases [26]. Interestingly, the tissue pattern of PART-1 expression is also essentially restricted to the prostate. The cloning and characterization of the other identified androgen-regulated cDNAs is in progress.

We anticipate that expanding these studies to include a greater portion of the prostate transcriptome coupled with experiments designed to determine direct *versus* indirect transcriptional regulation, and ultimately translational and post-translational regulation of these genes, will establish a framework for understanding the cellular functions mediated by androgens. Despite the important influence of androgenic hormones on prostate cancer growth, relatively few downstream targets of the AR pathway have been described. Studies designed to identify genes regu-

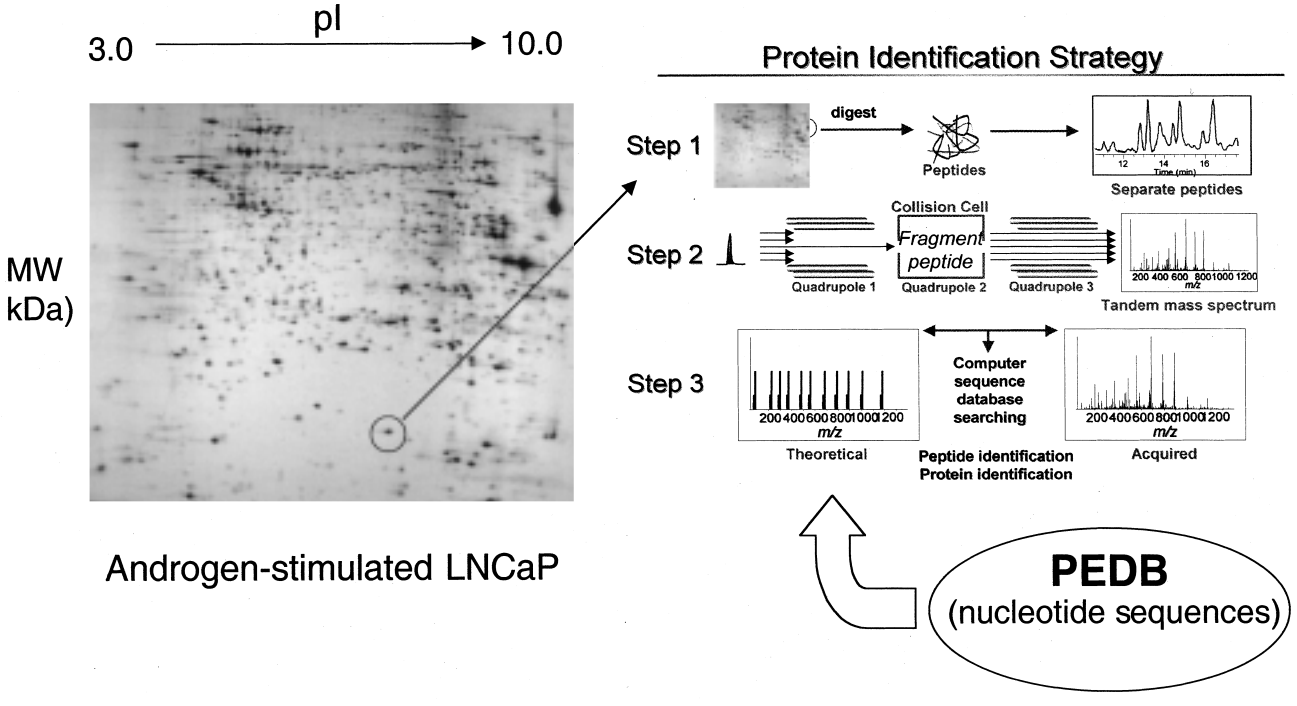
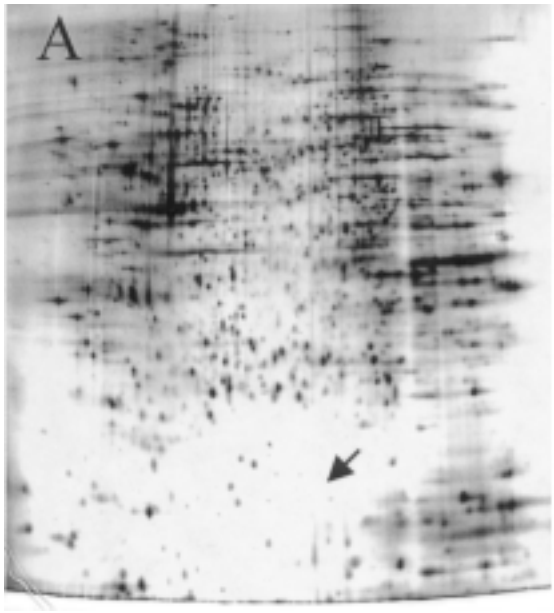
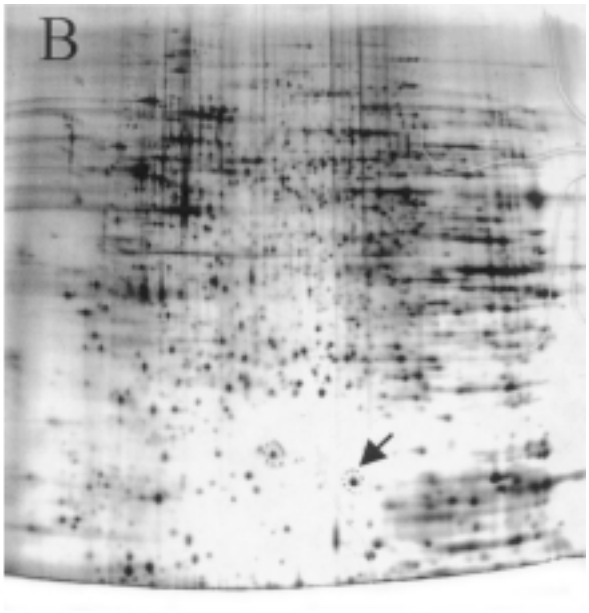


Figure 3. (Left) LNCaP 2-DE protein expression profile with androgen stimulation. (Right) Three-step schema for protein identification using MS and computer sequence database searching.

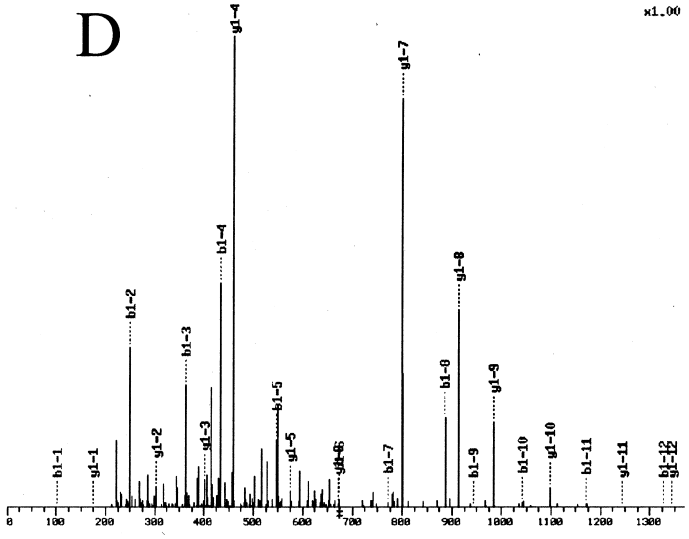




C

Seq	#	b	y	(+1)
T	1	102.0	1344.7	12
F	2	249.1	1243.7	11
I	3	362.2	1096.6	10
A	4	433.2	983.5	9
I	5	546.3	912.5	8
K	6	674.4	799.4	7
P	7	771.5	671.3	6
D	8	886.5	574.3	5
G	9	943.5	459.2	4
V	10	1042.6	402.2	3
Q	11	1170.6	303.2	2
R	12	1326.7	175.1	1

D



E

#	Rank/Sp	(M+H)+	Cn	deltCn	C*10^4	Ions	Reference	Peptide
1.	1 / 1	1344.8	1.0000	0.0000	2.9296.	14/22	NDKA_HUMAN	(R)TFIAIKPDGVQR
2.	2 / 30	1344.6	0.4673	0.5327	1.3689	9/20	RELB_HUMAN	(R)GMPFRYECEGR
3.	3 / 6	1344.8	0.4663	0.5337	1.3661	9/18	VE2_HPV25	(K)RRSRSRSRRK
4.	4 / 12	1344.8	0.4590	0.5410	1.3448	9/24	HIV1U16769	(R)GTKALTEVIALTK
5.	5 / 69	1345.7	0.4519	0.5481	1.3240	7/24	RU2A_HUMAN	(K)TFNPGAGLPTDKK
6.	6 / 8	1345.7	0.4189	0.5811	1.2273	9/22	HUMPRA	(K)GESGQSWPRLTK

Figure 4. Identification of an androgen-regulated protein from metastatic prostate cancer cells by 2-DE and MS. M12AR cells were (A) starved or (B) stimulated for 24 h with the synthetic androgen R1881 and total cell lysates (40 μg each) were subjected to 2-DE. Protein expression profiles were compared and proteins demonstrating a qualitative expression level differences were subjected to in-gel trypsin digestion, and identified by μLC-MS/MS analysis. (C), (D), MS/MS spectrum of identified peptide, peptide sequence, and identified ion series. (E) Results from correlation of acquired peptide fragmentation spectra with database entries (using SEQUEST software). The MS/MS spectrum in (D) was identified as NDKA_HUMAN (nm23) taken from the selected 2-D gel spot. Two additional peptides were identified from this protein in a single run.

lated by androgens in the rat prostate determined that androgens increase the transcription of about 56 genes and decrease the transcription of less than 10 genes [27]. From a therapeutic standpoint, it would be extremely useful to distinguish and subsequently modulate the relevant molecules in the AR program that mediate the divergent processes of cellular proliferation, cellular differentiation, and apoptosis.

3.3 Prostate gene expression analyses: 2-DE and MS

To complement our prostate transcriptional data and provide a more complete picture of prostate gene expression, we have undertaken a comprehensive analysis of that portion of the prostate proteome regulated by androgenic hormones. Reference protein expression profiles were produced for the LNCaP and M12AR prostate cancer cell lines using 2-DE protein separation techniques under steady-state conditions (Fig. 3). Protein expression profiles from cell lysates under conditions of androgen stimulation and androgen starvation have also been generated. A comparison of 2-DE protein profiles under these various conditions yielded a proteomic signature characterized by a subset of proteins with qualitative and quantitative changes. Individual proteins were identified using a sequential process of in-gel trypsin digestion and extraction, peptide separation by μLC, generation of MS/MS spectra, and database correlation with the acquired peptide fragmentation pattern (Fig. 3).

A comprehensive analysis of androgen-induced proteomic signatures is ongoing and our initial experiments demonstrate the utility of this approach in identifying molecules of potential importance in understanding androgen-mediated regulation of prostate cancer progression and metastasis. Figure 4 depicts a portion of the 2-DE protein profile from androgen-starved and androgen-stimulated M12AR prostate cancer cells with a differentially expressed protein spot that is upregulated in M12AR cells after exposure to androgens. This protein was identified as human nucleoside diphosphate kinase A (NDKA/nm23), a well-characterized gene with tumor metastasis suppressor activity in several different human tumors including melanoma, breast, ovary and prostate [28, 29]. Transfection of the DU-145 prostate cancer cell line with NDKA/nm23 inhibited the adhesion to cell matrix and impaired colony growth in soft agar [29].

The M12 prostate cancer cell line is highly tumorigenic when implanted into nude mice and metastasizes to different anatomical sites. Transfection of these cells with a functional androgen receptor (M12AR) markedly decreases the proliferation rate, tumor growth, invasiveness, and *in vivo* metastatic potential when these cells are injected into the prostate glands of nude mice (S. Plymate, unpublished observation). NDKA/nm23 transcripts have been shown to increase rapidly in prostate cancer cell lines after the administration of androgens, though no functional ramifications of this increased expression were described [30].

A possible mechanism for the decreased tumorigenic and metastatic capability of M12AR cells compared with M12 cells lacking the AR involves the upregulation of NDKA/nm23 by androgens through a functional androgen-response program restored by the AR transfection and expression. Such an observation has direct clinical relevance. Both human and *in vitro* studies suggest that there may be a survival benefit from maintaining an androgen responsive cohort of prostate tumor cells [31–33]. This concept has been studied in the LNCaP cell system by comparing the rate of tumor growth in castrated mice implanted with LNCaP cells with subsequent tumor growth (i) without further therapy, or (ii) followed by intermittent androgen replacement. The rate of tumor growth as measured by serum PSA was slower in animals treated with intermittent androgen supplementation compared to those maintained in the castrated state [31].

4 Concluding remarks

The results presented here demonstrate the utility of global expression studies to simultaneously identify multiple genes and gene products of biological relevance that participate in specific metabolic pathways. Both known and unknown genes are rapidly identified. Notable advantages of the microarray-based transcript profiling approach include the ability to perform detailed time-course or variable drug-dose experiments in a robust economical fashion. Controlled replicate experiments can determine system and procedural errors. However, this approach is absolutely dependent upon the identification of diverse clone sets for array construction that are biologically relevant to the system under study. In addition, a significant limitation of transcript profiling methods is the lack of a tight correlation between gene activity as measured by mRNA level, and protein abundance [34]. Global protein analyses focus on the actual biological effector molecules, but are restricted by difficulties in detecting low abundance proteins, accurately measuring the differences in protein levels between two samples, and a dependency on comprehensive annotated sequence databases for protein identification.

Integrating the assembly and annotation of sequence databases with transcript profiling and proteome analyses combines complementary robust approaches that capital-

ize on the strengths and avoid the limitations of relying on one method. The further expansion of this work to include the analysis of the entire prostate transcriptome coupled with quantitative proteome studies should enable the characterization of gene networks and cellular pathways that can be exploited for therapeutic intervention.

This work was supported in part by the CaPCURE Foundation, and a grant (CA75173) from the National Cancer Institute, grants (DAMD 17-98-1-8499 and PC991274) from the Department of Defense to PSN, a grant (R01-DK52683) and Veterans Affairs Merit Review Program to SRP, the Science Technology Center for Molecular Biotechnology, and a grant (NIH CRR11823) to RA. We wish to thank Joy Ware for the M12 prostate cancer cell line, Steve Lasky and the UW Molecular Biotechnology sequencing center for DNA sequencing support, and Roger Bumgarner, Nigel Clegg, Victoria Hawkins, and Burak Eroglu for bioinformatics support.

Received December 8, 1999

5 References

[1] Denmeade, S. R., Lin, X. S., Isaacs, J. T., *Prostate* 1996, *28*, 251–265.

[2] Isaacs, J. T., Wake, N., Coffey, D. S., Sandberg, A. A., *Cancer Res.* 1982, *42*, 2353–2371.

[3] Bruchovsky, N., Brown, E. M., Coppin, C. M., Goldenberg, S. L., Le Riche, J. C., Murray, N. C., Rennie, P. S., *Progr. Clin. Biol. Res.* 1987, *239*, 347–387.

[4] Velculescu, V. E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M. A., Bassett, D. E., Hieter, P., Vogelstein, B., Kinzler, K. W., *Cell* 1997, *88*, 243–251.

[5] Hawkins, V., Doll, D., Bumgarner, R., Smith, T., Abajian, C., Hood, L., Nelson, P. S., *Nucleic Acids Res.* 1999, *27*, 204–208.

[6] Nelson, P. S., Ng, W. L., Schummer, M., True, L. D., Liu, A. Y., Bumgarner, R. E., Ferguson, C., Dimak, A., Hood, L., *Genomics* 1998, *47*, 12–25.

[7] Strausberg, R. L., Dahl, C. A., Klausner, R. D., *Nature Genet.* 1997, *15*, 415–416.

[8] Huang, X., *Genomics* 1996, *33*, 21–31.

[9] Webber, M. M., Bello, D., Quader, S., *Prostate* 1997, *30*, 58–64.

[10] Bae, V. L., Jackson-Cook, C. K., Maygarden, S. J., Plymate, S. R., Chen, J., Ware, J. L., *Prostate* 1998, *34*, 275–282.

[11] Bae, V. L., Jackson-Cook, C. K., Brothman, A. R., Maygarden, S. J., Ware, J. L., *Int. J. Cancer* 1994, *58*, 721–729.

[12] Garrels, J. I., Franza Jr., B. R., *J. Biol. Chem.* 1989, *264*, 5283–5298.

[13] Blum, H., Beier, H., Gross, H., *Electrophoresis* 1987, *8*, 93–99.

[14] Corthals, LG., Gygi, S. P., Aebersold, R., Patterson, S. D., in: Rabilloud, T. (Ed.), *Proteome Research: 2D Gel Electrophoresis and Detection Methods*, Springer, New York 1999, pp. 197–232.

[15] Yates III, J. R., Eng, J. K., McCormack, A. L., Schieltz, D., *Anal. Chem.* 1995, *67*, 1425–1436.

[16] Hastie, N. D., Bishop, J. O., *Cell* 1976, *9*, 761–774.

[17] Bishop, J. O., Morton, J. G., Rosbash, M., Richardson, M., *Nature* 1974, *250*, 199–204.

[18] Zhang, L., Zhou, W., Velculescu, V. E., Kern, S. E., Hruban, R. H., Hamilton, S. R., Vogelstein, B., Kinzler, K. W., *Science* 1997, *276*, 1268–1272.

[19] Nelson, P. S., Hawkins, V., Schummer, M., Bumgarner, R., Ng, W. L., Ideker, T., Ferguson, C., Hood, L., *Genet. Anal.* 1999, *15*, 209–215.

[20] Riegman, P. H., Vlietstra, R. J., van der Korput, J. A., Brinkmann, A. O., Trapman, J., *Mol. Endocrinol.* 1991, *5*, 1921–1930.

[21] Murtha, P., Tindall, D. J., Young, C. Y., *Biochemistry* 1993, *32*, 6459–6464.

[22] He, W. W., Sciavolino, P. J., Wing, J., Augustus, M., Hudson, P., Meissner, P. S., Curtis, R. T., Shell, B. K., Bostwick, D. G., Tindall, D. J., Gelmann, E. P., Abate-Shen, C., Carter, K. C., *Genomics* 1997, *43*, 69–77.

[23] Nelson, P. S., Gan, L., Ferguson, C., Moss, P., Gelinas, R., Hood, L., Wang, K., *Proc. Natl. Acad. Sci. USA* 1999, *96*, 3114–3119.

[24] Paoloni-Giacobino, A., Chen, H., Peitsch, M. C., Rossier, C., Antonarakis, S. E., *Genomics* 1997, *44*, 309–320.

[25] Lin, B., Ferguson, C., White, J. T., Wang, S., Vessella, R., True, L. D., Hood, L., Nelson, P. S., *Cancer Res.* 1999, *59*, 4180–4184.

[26] Lin, B., White, J. T., Ferguson, C., Bumgarner, R., Friedman, C., Trask, B., Ellis, W., Lange, P., Hood, L., Nelson, P. S., *Cancer Res.*, in press.

[27] Wang, Z., Tufts, R., Haleem, R., Cai, X., *Proc. Natl. Acad. Sci. USA* 1997, *94*, 1299–3004.

[28] Freije, J. M., MacDonald, N. J., Steeg, P. S., *Biochem. Soc. Symp.* 1998, *63*, 261–271.

[29] Lim, S., Lee, H. Y., Lee, H., *Cancer Lett.* 1998, *133*, 143–149.

[30] Yoshimura, I., Wu, J. M., Chen, Y., Ng, C., Mallouh, C., Backer, J. M., Mendola, C. E., Tazaki, H., *Biochem. Biophys. Res. Commun. 1995, 208, 603–609.*

[31] Sato, N., Gleave, M. E., Bruchovsky, N., Rennie, P. S., Goldenberg, S. L., Lange, P. H., Sullivan, L. D., *J. Steroid Biochem. Mol. Biol.* 1996, *58*, 139–146.

[32] Grossfeld, G. D., Small, E. J., Carroll, P. R., *Urology* 1998, *51*, 137–144.

[33] Oliver, R. T., Williams, G., Paris, A. M., Blandy, J. P., *Urology* 1997, *49*, 79–82.