

# Comprehensive Gene Expression Analysis of Prostate Cancer Reveals Distinct Transcriptional Programs Associated with Metastatic Disease<sup>1</sup>

Eva LaTulippe, Jaya Satagopan, Alex Smith, Howard Scher, Peter Scardino, Victor Reuter, and William L. Gerald<sup>2</sup>

Departments of Pathology [E. L., V. R., W. L. G.], Epidemiology and Biostatistics [J. S., A. S.], Urology [P. S.], and Genitourinary Oncology Service, Department of Medicine [H. S.], Memorial Sloan-Kettering Cancer Center, New York, New York 10021, and Department of Medicine, Joan and Sanford I. Weill College of Medicine of Cornell University, New York, New York 10021 [H. S.]

## ABSTRACT

The identification of genes that contribute to the biological basis for clinical heterogeneity and progression of prostate cancer is critical to accurate classification and appropriate therapy. We performed a comprehensive gene expression analysis of prostate cancer using oligonucleotide arrays with 63,175 probe sets to identify genes and expressed sequences with strong and uniform differential expression between nonrecurrent primary prostate cancers and metastatic prostate cancers. The mean expression value for >3,000 tumor-intrinsic genes differed by at least 3-fold between the two groups. This includes many novel ESTs not previously implicated in prostate cancer progression. Many differentially expressed genes participate in biological processes that may contribute to the clinical phenotype. One example was a strong correlation between high proliferation rates in metastatic cancers and overexpression of genes that participate in cell cycle regulation, DNA replication, and DNA repair. Other functional categories of differentially expressed genes included transcriptional regulation, signaling, signal transduction, cell structure, and motility. These differentially expressed genes reflect critical cellular activities that contribute to clinical heterogeneity and provide diagnostic and therapeutic targets.

## INTRODUCTION

Carcinoma of the prostate, the most common cancer in the United States, was expected to affect ~198,000 individuals in 2001 (1). In recent years, there has been a dramatic increase in the proportion of patients diagnosed with tumors that are seemingly confined to the gland. This is the result of increased public awareness and the more widespread application of detection strategies based on measurements of the level of prostate-specific antigen in the blood (2). Early disease is clinically heterogeneous because many patients have an indolent course that does not significantly affect an individual patient's survival. In contrast, once metastatic disease is documented on an imaging study, the majority of patients die from their tumors as opposed to other causes. This has led us to consider the disease as a series of "states" that include clinically localized tumors, and those that have metastasized, as a framework to assess the clinical and biological factors that are associated with specific phenotypes and outcomes (3). Understanding the biological basis for this clinical heterogeneity is critical to assess prognosis, to select therapy, and to assess treatment effects.

Tumor metastasis is the most clinically significant event in prostate cancer patients. Development of metastases requires that a cancer cell must complete a series of steps involving complex interactions between tumor cells and the host (4). Cells from primary tumors must detach, invade stromal tissue, and penetrate vessels by which they disseminate. They must survive in the circulation to reach a secondary site in which they arrest because of physical size or binding to specific

tissues. To form clinically significant tumors, metastatic cells must proliferate in the new microenvironment and recruit a blood supply. Those tumor cells growing at metastatic sites are then continually selected for growth advantage. This is a complex and dynamic process that is expected to involve alterations in many genes and transcriptional programs. Identification of genes, gene expression profiles, and biological pathways that contribute to metastasis will be of significant benefit to improved tumor classification and therapy. To address this question, we performed a genome-wide expression analysis and identified genes differentially expressed in primary and metastatic prostate cancers. These genes reflect the distinct clinical phenotype of these two cohorts, and provide insight into the biology of prostate cancer progression.

## MATERIALS AND METHODS

**Samples.** Tissues from 3 noncancerous prostates, 23 primary prostate cancers, and 9 metastatic prostate cancers, obtained from therapeutic or diagnostic procedures performed as part of routine clinical management at MSKCC<sup>3</sup> from 1993 to 1999, were analyzed. Clinical and pathological features of prostate cancer cases from which samples were obtained are presented in Table 1. Samples were snap-frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ . Each sample was examined histologically using H&E-stained cryostat sections. Cells of interest were manually dissected from the frozen block, trimming away other tissues. Care was taken to remove nonneoplastic tissues from tumor samples. All of the studies were conducted under MSKCC Institutional Review Board-approved protocols.

**Gene Expression Analysis.** RNA was extracted from frozen tissues by homogenization in guanidinium isothiocyanate-based buffer (RNAeasy; Qiagen, Valencia, CA) and evaluated for integrity by denaturing agarose gel. Complementary DNA was synthesized from total RNA using a T7-promoter-tagged oligo(dT) primer. RNA target was synthesized by *in vitro* transcription and labeled with biotinylated nucleotides (Enzo Biochem, Farmingdale, NY). Labeled target was assessed by hybridization to Test2 arrays (Affymetrix, Santa Clara, CA) and detected with phycoerythrin-streptavidin (Molecular Probes, Eugene, OR) amplified with antistreptavidin antibody (Vector, Burlingame, CA). Gene expression analysis was carried out using Affymetrix U95 human gene arrays with 63,175 features for individual gene/EST clusters using instruments and protocols recommended by the manufacturer. The U95 set consists of five distinct microarrays (A through E), each containing probes for about 12,000 unique gene/EST transcripts. Two response measures, the Average Difference and Absolute Call, were extracted for each gene on every sample, as determined by default settings of Affymetrix Microarray Suite 4.0. Average Difference was used as the primary measure of expression level, and Absolute Call was retained as a secondary measure. Expression values on each array were multiplicatively scaled to have an average expression of 2500 across the central 96% percent of all genes on the array.

**Data Analysis.** For oligonucleotide arrays, scanned image files were visually inspected for artifacts and analyzed using Microarray Suite v4.0 (Affymetrix). Differential expression was evaluated using several measures. Final ranking to obtain genes uniformly and strongly differentially expressed was determined by the following approach (fully described in supplementary

Received 3/20/02; accepted 6/3/02.

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.

<sup>1</sup> Supported by NIH Grant U01 CA84999 (to W. L. G.).

<sup>2</sup> To whom requests for reprints should be addressed, at Department of Pathology, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021. Phone: (212) 639-5905; Fax: (212) 639-4559; E-mail: geraldw@mskcc.org.

<sup>3</sup> The abbreviations used are: MSKCC, Memorial Sloan-Kettering Cancer Center; EST, expressed sequence tag.

Table 1 Clinical and Pathological Features of Prostate Carcinomas

ID <sup>a</sup>	Race	Age	InitPSA	GLSum	pT	pN	pM	Rec <sup>b</sup>	Site	PreTx
Primary prostate carcinomas										
P1	W	66	6.0	7	2+	0	0	NO		
P2	W	58	5.6	6	2B	0	0	NO		
P3	W	58	4.7	7	2B	0	0	NO		
P4	W	62	6.6	7	2+	0	0	YES		
P5	W	64	2.6	7	2B	0	0	NO		
P6	W	62	53.0	7	2B	0	0	YES		
P7	B	42	11.9	7	2+	0	0	NO		
P8	W	56	9.0	6	3A	0	0	YES		
P9	W	53	11.9	8	4	0	0	YES		
P10	W	49	11.8	9	2+	1	0	YES		
P11	B	61	25.8	7	3B	0	0	NO		
P12	W	68	7.9	7	3B	1	0	YES		
P13	W	60	6.4	7	4	0	0	NO		
P14	W	67	8.2	7	3B	1	0	YES		
P15	W	66	20.0	9	3B	0	0	NO		
P16	W	63	29.0	8	3A	0	0	NO		
P17	W	60	5.5	7	3A	0	0	YES		
P18	W	75	NA	7	3A	0	0	NO		
P19	W	68	19.0	7	3A	0	0	NO		
P20	W	69	22.0	7	3A	0	0	NO		
P21	W	68	7.5	9	2+	0	0	NO		
P22	W	59	7.1	9	3B	1	0	YES		
P23	W	64	21.0	9	2+	0	0	NO		
Metastatic prostate carcinomas										
M1	W	52	NA	NA					LN	YES
M2	B	47	9.9	8					LN	NO
M3	W	64	14.0	7					ST	NO
M4	W	65	NA	8					ST	YES
M5	W	82	NA	NA					BONE	NO
M6	W	54	33.4	8					LN	YES
M7	W	52	NA	8					LN	YES
M8	W	82	100+	9					BONE	YES
M9	W	69	NA	7					LUNG	YES

<sup>a</sup> ID, sample unique identifier; InitPSA, serum PSA in ng/ml at diagnosis; GLSum, Gleason histological score of cancer in prostate biopsy; pT, pN, pM: tumor, node, and metastasis grading per criteria of American Joint Commission on Cancer stage; Rec, biochemical or clinical recurrence; PreTx, prior antineoplastic therapy; W, white; B, black; NA, not available; LN, lymph node; ST, soft tissue.

<sup>b</sup> Mean follow-up was 42 months.

information).<sup>4</sup> The expression data set was first filtered to include only those probe sets detecting genes with mean expression values that differed by at least 3-fold between the two groups (absolute base-ten logarithm of the ratio of the means  $\geq 0.4771$ ). Probes were then ranked based on the relative magnitude of the difference (*t* test) between the means of the two sample sets. Genes with expression differences likely attributable to contaminating nonneoplastic tissues were removed from the ranking. Datasets used for hierarchical clustering were normalized by standardizing each gene and array to mean = 0 and variance = 1. Hierarchical clustering and result display was performed using Cluster and TreeView software (5). Gene expression that was attributable to nonneoplastic tissues was identified by comparison with expression levels in nonneoplastic prostate samples (supplementary information). Specific genes corresponding to Unigene clusters were identified by GenBank accession number of the clone used to produce the oligonucleotide probe set and annotated through review of internet resources.<sup>5</sup>

**Immunohistochemistry.** Multitissue blocks of formalin-fixed, paraffin-embedded tissue corresponding to the samples used in this analysis were prepared using a tissue arrayer (Beecher Instruments, Silver Spring, MD). The blocks contained three representative 0.6 mm cores from diagnostic areas for each case. Immunohistochemical detection of Ki67 (mib1; Dako, Carpinteria, CA; 1:200) was carried out with standard streptavidin-biotin peroxidase methodology using formalin-fixed, paraffin-embedded tissue, and microwave antigen retrieval as described previously (6).

**Quantitative Reverse Transcriptase-PCR.** Quantitative reverse transcriptase-PCR was performed using the LightCycler thermal cycler system (Roche Diagnostics, Basel, Switzerland). Fifty to 100 ng of total RNA were used as template with the LightCycler RNA SYBR Green I system (2.7×

concentration; Roche Diagnostics) according to the manufacturer's instructions. A typical protocol included reverse transcription at 61°C for 20 min and a denaturation step at 95°C for 2 min followed by 45 cycles with 95°C denaturation for 5 s, 60°C annealing for 5 s, and 72°C extension for 8 s. Detection of the fluorescent product was performed at the end of the extension period. A melting curve analysis was performed at the end of the PCR at 95° for 5 s, 65° for 15 s, and a final denaturation at 95° for 0 s. Negative controls were run to confirm that the samples were not cross-contaminated. Data were analyzed with the LightCycler analysis software. A standard curve was created by serial dilution of the appropriate RNA template. Primers for *TAGLN* were: forward, 5'-GTCATTGGCCTTCAGATG-3', and reverse, 5'-ACACCTCAAAGCTTG-3'; for *MSMB*: forward, 5'-TGTTTCTACACCTGTGGG-3', and reverse, 5'-TGATAGGCATGGCTACAC-3'; and for 18S RNase: forward, 5'-GACATTGACCTCACCAAG-3', and reverse, 5'-ATCTTCTTCAGTCGCTCC-3'.

**RESULTS**

**Gene Expression Analysis Using Oligonucleotide Arrays.** We analyzed gene expression in prostate samples using hybridization of RNA target to oligonucleotide microarrays with 63,175 features for gene/EST clusters. Three nonneoplastic prostate samples, 23 primary prostate carcinomas and 9 metastatic prostate cancers were evaluated. Expression of genes corresponding to 5992 probe sets were reliably detected (absolute call of present) in all 23 primary prostate carcinomas; 34,518 probe sets detected gene expression in at least some cases; and 22,665 did not detect expression (absolute call of absent) in any of these samples. Gene expression corresponding to 7,713 probe sets were detected in all 9 metastatic prostate cancers, whereas 29,784 genes/ESTs were variably expressed, and 25,678 genes/ESTs were not detected in any. The highest proportion of expressed genes was represented by probe sets on the U95A microarray corresponding to

<sup>4</sup> Supplementary information for this article is available at *Cancer Research Online* (<http://cancerres.aacrjournals.org>).

<sup>5</sup> Internet resources: Genecards, <http://nciarray.nci.nih.gov/cards/>; Locus Link, <http://www.ncbi.nlm.nih.gov/LocusLink/index.html>; Online Mendelian Inheritance in Man, <http://neptune.nlm.nih.gov/entrez/query.fcgi?db=OMIM>; Gene Ontology Browser, <http://cgap.nci.nih.gov/Genes>; and publications identified in PubMed, <http://www.ncbi.nlm.nih.gov/PubMed/>.



Fig. 1. Representative gene expression clusters enriched for genes differentially expressed between primary (blue boxes) and metastatic (orange boxes) prostate carcinomas. Clusters were selected from a hierarchical clustering dendrogram of gene expression data from all of the 12,559 probe sets of U95A array (rows) and 32 samples of prostate carcinoma (columns). Pseudocolored red, expression levels with transcript levels below the median for that gene across all of the samples; pseudocolored green, expression levels with transcript levels above the median for that gene across all of the samples. Color saturation is proportional to the magnitude of expression. NON-NEOPLASTIC PROSTATE, a cluster of genes corresponding to high-level expression in nonneoplastic prostate.

Table 2 100 highest ranked differentially expressed genes and functional classification

Rank <sup>a</sup>	MET/PRI <sup>b</sup>		
Cell cycle regulation, DNA replication and repair, and mitosis			
5	5.70	<i>DEEPEST</i>	Deepest
7	4.93	<i>KNTC1</i>	Kinetochore associated 1
8	6.84	<i>FEN1</i>	RAD2 ( <i>Schizosaccharomyces pombe</i> ) homologue, flap structure-specific endonuclease 1
9	14.73	<i>TK1</i>	Thymidine kinase
11	7.85	<i>TOP2A</i>	Topoisomerase (DNA) II $\alpha$ ( $M_r$ 170,000)
12	9.58	<i>CDKN3</i>	Cyclin-dependent kinase inhibitor 3 (CDK2-associated dual specificity phosphatase)
16	5.43	<i>RFC5</i>	Human replication factor C, $M_r$ 36,000 subunit
23	3.13	<i>RFC4</i>	Replication factor C (activator 1) 4 ( $M_r$ 37,000)
28	4.96	<i>MAD2L1</i>	MAD2 (mitotic arrest deficient, yeast, homologue)-like 1
35	5.86	<i>KNSL2</i>	Kinesin-like 2
36	8.84	<i>CDC2</i>	Cell division cycle 2, cell division control protein 2 homologue (EC 2.7.1) (p34 protein kinase) [cyclin-dependent kinase 1 (cdk1)]
42	3.09	<i>MPHOSPH9</i>	M-phase phosphoprotein 9
44	5.83	<i>RNASEHI</i>	Ribonuclease HI, large subunit
50	4.58	<i>CCNE2</i>	cyclin E2
54	3.20	<i>MCM7</i>	Minichromosome maintenance deficient ( <i>Saccharomyces cerevisiae</i> ) 7
56	3.81	<i>BUB1B</i>	Budding uninhibited by benzimidazoles 1 (yeast homologue), $\beta$
57	5.61	<i>SMC4L1</i>	SMC4 (structural maintenance of chromosomes 4, yeast)-like 1
58	10.02	<i>STK15</i>	Serine/threonine kinase 15
59	3.12	<i>ZWINT</i>	ZW10 interactor
70	3.41	<i>MGC1780</i>	DDA3; p53-regulated DDA3
73	5.33	<i>CCNB2</i>	cyclin B2
83	3.77	<i>TTK</i>	TTK protein kinase
84	3.08	<i>RPA3</i>	Replication protein A3
95	13.61	<i>CCNB1</i>	cyclin B1
96	3.31	<i>CDC25B</i>	Cell division cycle 25B
Signaling and signal transduction			
15	7.29	<i>STK11</i>	Serine/threonine kinase 11 (Peutz-Jeghers syndrome)
17	32.82	<i>JAG1</i>	Jagged 1
21	0.14	<i>DUSP1</i>	Dual specificity phosphatase 1
26	0.31	<i>KIAA0135</i>	Similar to PIM-1 proto-oncogene
34	3.91	<i>EDN3</i>	Endothelin 3
49	3.97	<i>CSNK1G2</i>	Casein kinase 1, $\gamma$ 2
51	6.37	<i>INPP4A</i>	Inositol polyphosphate-4-phosphatase
64	3.83	<i>TMPO</i>	Thymopoietin
67	6.05	<i>ESRRB</i>	Estrogen-related receptor $\beta$
76	0.21	<i>GPR68</i>	Ovarian cancer G protein-coupled receptor 1; member of the G protein-coupled receptor family
77	0.27	<i>LTBP1</i>	Latent transforming growth factor $\beta$ -binding protein 1
78	0.20	<i>IL8RA</i>	Interleukin 8 receptor, $\alpha$
79	4.49	<i>CIT</i>	Citron (rho-interacting, serine/threonine kinase 21)
89	0.11	<i>MSMB</i>	Microseminoprotein $\beta$
92	0.17	<i>PPP3CB</i>	Protein phosphatase 3 (formerly 2B), catalytic subunit, $\beta$ isoform (calcineurin A $\beta$ )
Transcriptional regulation, chromatin modification, RNA processing and protein synthesis and modification			
2	8.55	<i>USP13</i>	Ubiquitin-specific protease 13 (isopeptidase T-3)
6	4.33	<i>SMARCD1</i>	SWI/SNF related, matrix-associated, actin-dependent regulator of chromatin, subfamily d, member 1
13	4.99	<i>MYBL2</i>	v-myb avian myeloblastosis viral oncogene homologue-like 2
18	28.60	<i>SF3A2</i>	Splicing factor 3a, subunit 2, $M_r$ 66,000
20	5.40	<i>HOXB5</i>	HOX2, homeo box B5
22	0.14	<i>ZFP36</i>	Tristetraprolin
24	11.88	<i>EZH2</i>	Enhancer of zeste ( <i>Drosophila</i> ) homologue 2
25	0.27	<i>SATB1</i>	Special AT-rich sequence binding protein 1
32	0.10	<i>FOS</i>	v-fos FBJ murine osteosarcoma viral oncogene homologue
33	0.12	<i>NR4A1</i>	Nuclear receptor subfamily 4, group A
43	36.85	<i>UBCH10</i>	Ubiquitin carrier protein E2-C
45	7.72	<i>PTTG1</i>	Pituitary tumor-transforming 1, securin
48	7.18	<i>CTRL</i>	Chymotrypsin-like
52	3.53	<i>EP300</i>	E1A binding protein p300
53	4.28	<i>E2EPF</i>	Ubiquitin carrier protein
62	0.22	<i>JUNB</i>	jun B proto-oncogene
68	11.05	<i>FOXM1</i>	Forkhead box M1, hepatocyte nuclear factor
72	6.06	<i>PRO1</i>	Prophet of Pit1, paired-like homeodomain transcription factor
75	4.54	<i>CNAP1</i>	Chromosome condensation-related SMC-associated protein 1
85	6.23	<i>HOXC5</i>	Homeo box C5
88	4.75	<i>U5-100K</i>	prp28, U5 snRNP $M_r$ 100,000 protein
91	3.37	<i>CGGBP1</i>	CGG triplet repeat binding protein 1
94	0.29	<i>CEBPD</i>	ccat/enhancer binding protein $\delta$ (c/ebp $\delta$ ) (nuclear factor nf- $\text{il6}$ - $\beta$ )
99	3.38	<i>PLOD2</i>	Procollagen-lysine, 2-oxoglutarate 5-dioxygenase
Cell adhesion, migration, cytoskeleton, and extracellular matrix			
1	0.15	<i>TAGLN</i>	Transgelin actin, $\alpha$ 2
3	0.23	<i>ACTA2</i>	$\alpha$ -actin
30	0.13	<i>FHL1</i>	Four and one-half LIM domains 1
46	0.23	<i>ITGA8</i>	integrin, $\alpha$ 8
47	7.70	<i>HMMR</i>	Hyaluronan-mediated motility receptor (RHAMM)
61	0.15	<i>TPM3</i>	Tropomyosin 3 (nonmuscle)
66	0.29	<i>ITGA7</i>	Integrin, $\alpha$ 7
80	4.43	<i>THBS2</i>	Thrombospondin 2
82	0.19	<i>RARRES1</i>	Retinoic acid receptor responder (tazarotene induced) 1
97	4.19	<i>CEACAM7</i>	Carcinoembryonic antigen-related cell adhesion molecule 7

Table 2 *Continued*

Rank <sup>a</sup>	MET/PRI <sup>b</sup>		
Metabolism, biosynthesis, and molecular transport			
4	0.30	ABCB1	P glycoprotein 1/multiple drug resistance 1
10	4.06	PGK1	Phosphoglycerate kinase 1
19	10.82	FTH1	FERRITIN HEAVY CHAIN 1; FTH1
27	4.50	NUP155	Nucleoporin <i>M</i> <sub>v</sub> 155,000
29	0.20	ABCA8	ATP-binding cassette, subfamily A (ABC1), member 8
38	0.16	FXYD3	FXYD domain-containing ion transport regulator 3
41	4.33	SLC29A1	Solute carrier family 29 (nucleoside transporters), member
60	4.02	FMO3	Flavin-containing monooxygenase 3
74	3.17	PDK1	Pyruvate dehydrogenase kinase, isoenzyme 1
86	5.70	SCD	stearoyl-CoA desaturase ( $\delta$ -9-desaturase)
87	3.71	KPNA2	Karyopherin $\alpha$ 2 (RAG cohort 1, importin $\alpha$ 1)
98	0.17	APOD	Apolipoprotein D
100	0.25	ABCC4	ATP-binding cassette, subfamily C (CFTR/MRP), member 4
Unclassified			
14	6.95	KIAA0101	None available
31	4.49	KIAA0906	None available
37	4.98	CGTHBA	Conserved gene telomeric to $\alpha$ globin cluster
39	6.91	KIAA0186	None available
40	12.69	KIAA0543	None available
55	0.23	DKFZP434N043	None available
63	6.08	R32184	None available
65	7.67	DKFZP434N093	None available
69	3.20	KIAA0008	None available
71	3.39	DKFZP564C152	None available
81	3.59	MEST	Mesoderm specific transcript (mouse) homologue
90	5.54	DKFZP586L151	None available
93	0.30	EYA1	Eyes absent ( <i>Drosophila</i> ) homologue 1

<sup>a</sup> Rank, rank order of U95A probe sets based on *t*-statistic.

<sup>b</sup> MET/PRI, mean expression value of metastatic prostate cancers divided by mean expression value of primaries.

genes with near full-length cDNAs, most of which are named. Analyses of the gene/EST clusters queried by the U95A array are described in detail below. Identical algorithms were applied to data from arrays U95B-E which are primarily comprised of probe sets for uncharacterized ESTs. Those results are available in the supplementary data.

**Identification of Differentially Expressed Genes.** An unsupervised analysis of gene expression in all 32 prostate cancers based on the hybridization results for the U95A array revealed a strong tendency for primary and metastatic tumors to have distinct expression profiles based on an average linkage hierarchical clustering algorithm (Fig. 1). This analysis revealed several groups of genes with distinct expression profiles for these two major tumor subdivisions. As expected, some of the gene expression differences that distinguished primary and metastatic tumors were contributed by the small amount of contaminating nonneoplastic prostate tissue present in the primary tumor samples (Fig. 1, *bottom panel*). These were further confirmed by a comparison with expression data from nonneoplastic prostate (see supplementary information). The remaining gene expression differences are expected to be intrinsic to tumor cells and reflect biological distinctions.

The high level of discrimination between primary and metastatic carcinomas based on an unsupervised analysis of expression profiles suggested that specific genes responsible for biological differences could be identified. Nine of the 23 primary sample patients experienced a recurrence. We compared the subset of 14 primary tumors from patients that did not recur with the 9 metastatic prostate cancers to identify differentially expressed genes in these clinically distinct groups. Primary tumors from patients who subsequently developed metastases were not included in the comparison. Genes that were uniformly and strongly differentially expressed were selected as described under "Data Analysis" in the "Materials and Methods" section. The expression data set was first filtered to include only those probe sets detecting genes with mean expression values that differed by at least 3-fold between the two groups. Probes were then ranked based on the relative magnitude of the difference (*t* test) between the means of the two sample sets. Genes with expression differences that

were likely attributable to contaminating nonneoplastic tissues were removed from the ranking. A total of 3,436 of the 63,175 probe sets detected tumor-related differential gene expression of at least 3-fold. Of the U95A probe sets for near-full-length genes, 132 were overexpressed in primary tumors and 360 in metastases. The 100 most highly ranked tumor-intrinsic, U95A genes based on the *t* test statistic are listed in Table 2. The entire list of 3,436 differentially expressed sequences is available in the supplementary material.

**Functional Attributes of Differentially Expressed Genes.** We reviewed available sources and assigned a general molecular or biological function to each gene in Table 2. The functional attributes of differentially expressed genes are expected to reflect biological differences between early-stage and advanced tumors. In keeping with this concept, 26 of the 100 most highly ranked genes are believed to play a role in some aspect of cell cycle regulation, DNA replication and repair, or mitosis including many genes, such as *RFC5*, *TOP2A*, *RFC4* and *MAD2L1*, that are known to be up-regulated in highly proliferative cells (7). This finding correlated with the increased proliferation index of metastatic tumors used for gene expression analysis based on the immunohistochemical assessment of *MKI67* (Fig. 2). Fifteen of the 100 highly ranked genes correspond to products potentially involved in signaling and signal transduction and 9 others may contribute to cell adhesion, cell migration, or extracellular matrix. These include *HMMR*, which encodes an extracellular matrix-binding protein believed to play a role in cell motility through the RAS-ERK signaling pathway (8), and *INPP4A*, the substrates of which are intermediates in pathways regulated through the *AKT* proto-oncogene (9).

An unexpectedly large proportion of highly differentially expressed genes are believed to be involved in the regulation of gene expression and gene product function. For example, of the 100 highest-ranked differentially expressed U95A genes, 13 encode products that are expected to function as transcription factors, components of the transcriptional complex, or other proteins contributing to the regulation of transcription; 3 others encode products believed to participate in RNA splicing or metabolism; 3 gene products are thought to participate in

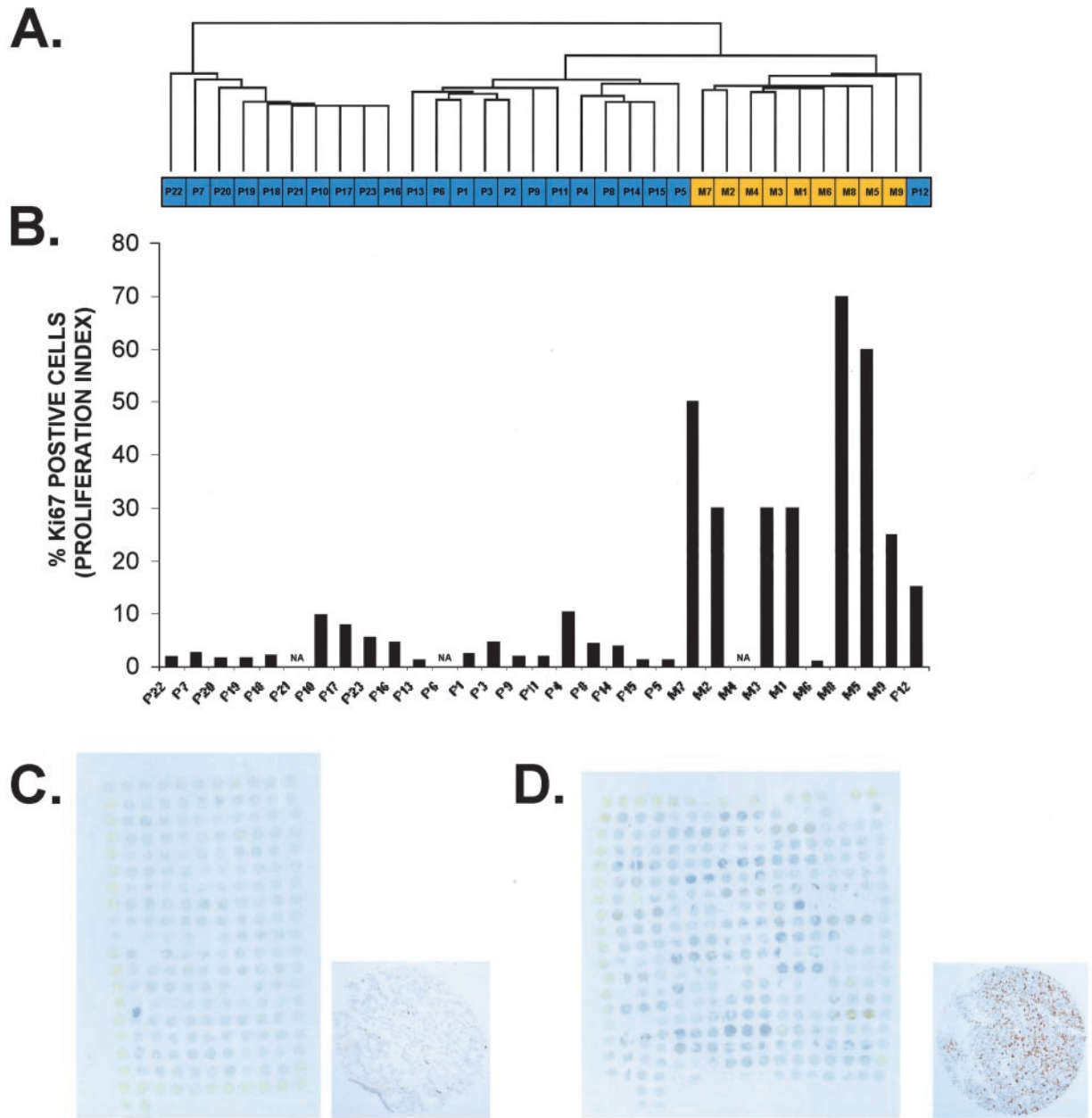


Fig. 2. Proliferation indices for prostate carcinoma samples. *A*, hierarchical dendrogram of samples showing overall similarity of gene expression profiles. *Blue boxes*, primary tumors; *orange boxes*, metastatic tumors. *B*, bar graph, proliferation indices based on Ki67 immunohistochemistry. *C*, section of multitissue block of primary prostate carcinomas immunostained for Ki67. *Inset*, high-power magnification of representative core tissue section. *D*, section of multitissue block of metastatic prostate carcinomas immunostained for Ki67. *Inset*, high-power magnification of representative core tissue section.

chromatin modifications that may also impact on transcriptional regulation; and 5 genes contribute to posttranscriptional regulation of protein function. These findings suggest that the development and progression of prostate cancer metastases are associated with many gene expression changes related to cell proliferation, interactions with the microenvironment, properties that might contribute to cell motility, activated signal transduction pathways, and the regulation of gene product synthesis and function.

**Validation.** Gene expression values that were determined by oligonucleotide array were validated in several ways. Some gene transcripts were represented by more than one probe set, and each detected similar relative levels of expression. An analysis of Cancer Genome Anatomy Project, serial analysis of gene expression, and other published data verified expression in prostate sample for many genes (10–14). Some differentially expressed genes were

selected for the measurement of transcript levels using a quantitative reverse transcriptase-PCR technique in the same samples used for microarray-based expression analysis. The results were in good agreement with the relative levels of expression as determined by oligonucleotide arrays (examples shown in Fig. 3). We also performed immunohistochemical analysis of tumor samples and established that protein expression of some gene products correlated with mRNA levels [e.g., *MKI67* (Fig. 2) and *PSA*, *AR* (data not shown)]. Several of the genes that we identified as highly differentially expressed have been previously implicated in progression of cancer and help to validate our experimental approach. In addition to those mentioned above, other examples include *STK11* (15), *PTTG1* (16), *STK15* (17), and *MYBL2* (18). *STK15* was overexpressed in metastatic prostate cancers and has previously been implicated in aggressive prostate cancer (17).

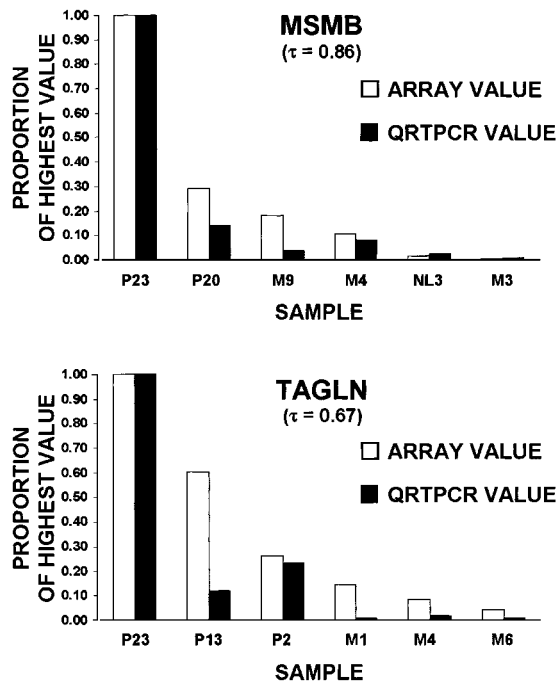


Fig. 3. Comparison of relative expression values for selected differentially expressed genes based on quantitative reverse transcriptase-PCR (QRT-PCR) and oligonucleotide array analysis (ARRAY). The Kendall rank correlation coefficient ( $\tau$ ) is shown.

## DISCUSSION

The heterogeneity of prostate cancers extends within clinical states and across states. At the extreme ends of the clinical spectrum are prognostically favorable localized tumors with a low biological potential for metastasis and tumors with a high propensity for early dissemination that are invariably lethal. These clinical phenotypes are related in part to the intrinsic biology of tumor cells and are reflected in the pattern of expression of specific genes. Using comprehensive gene expression analysis of tumor samples representing the nonmetastatic and metastatic phenotypes, we identified genes that were consistently and strongly differentially expressed and represent common and valid biological differences underlying clinical heterogeneity.

Few prior studies have used high-throughput gene expression analysis to study prostate cancer metastases. One reason is that well-preserved surgical tissue samples of metastatic prostate cancer are rare, which limits the availability of appropriate samples. Oligonucleotide arrays with 6800 probe sets were used to compare expression in three metastases to eight primary tumors (18). Five of the nine genes found to be commonly differentially expressed in that study also have average expression values for our sample sets that are in agreement with that data. In addition, two (*EGR2* and *EGR3*) were differentially expressed at least 3-fold in the present study. In the other published analysis of metastatic tumors (12), the hierarchical clustering of cDNA array data was used to identify coordinately expressed groups of genes that were specifically differentially expressed in metastatic *versus* primary prostate cancer. Two of the genes in clusters that were overexpressed in metastatic prostate carcinomas (*MTA1* and *MYBL2*) were also overexpressed in the present study. In fact, *MYBL2* met our stringent criteria for strongly and commonly overexpressed genes in metastases. In a related study to identify genes associated with aggressive primary tumors, Singh *et al.* (19) identified five genes that were commonly used in models to predict the recurrence of carcinoma after radical prostatectomy. Two of the genes that were overexpressed in recurrent cases (*HOXC6* and *PDGFRA*)

were also overexpressed in our samples of metastatic carcinoma, and one gene overexpressed in nonrecurrent cases (*SIATI*) was more strongly expressed in primaries, although none achieved the 3-fold difference filter used here. The partial agreement between these studies is encouraging in that, despite very different methodologies, genes that participate in the process of metastasis are being identified.

The predicted function of these differentially expressed genes provides a glimpse into the biology of prostate cancer progression. Although functional assignment is based on the limited published data, it is encouraging that many differentially expressed genes reflect biological distinctions and functional pathways previously implicated in aggressive disease. Included are cell cycle regulators and DNA replication and repair proteins that may drive cell proliferation; transcriptional regulators believed to play a role in development and differentiation; and proteins that play a role in signal transduction, cell structure, and cell interactions with environmental factors. The distinct expression patterns of these genes help to validate their role in the clinical phenotype of aggressive metastatic disease.

Some of the specific genes that are differentially expressed may identify critical functional pathways. One example is the *MYBL2* gene, which was overexpressed in many metastatic tumors. *MYBL2* activates *CDC2* gene expression in proliferating fibroblasts (20). *CDC2* is a catalytic subunit of a protein kinase complex that induces entry into mitosis. Cyclin E modulates the functional activity of these genes. All of them were overexpressed in our metastatic samples, which suggests that this pathway may be a critical component of cell cycle regulation in many metastatic prostate cancers. Identifying these genes in specific tumor samples may, therefore, provide diagnostic information and serve as a therapeutic target.

In addition to the limited number of genes discussed here, our analysis identified hundreds of poorly characterized EST clusters that likely represent novel genes of unknown function that were highly differentially expressed between primary and metastatic prostate cancers. The biological activity of these uncharacterized genes may be inferred from attributes of known genes with shared expression patterns. Therefore, many are likely to play important biological roles similar to those predicted for the known gene products discussed here. All of them warrant further study and may provide new insights into prostate cancer biology. Additional study of the complete molecular phenotype and patterns of variation will lead to a more in-depth understanding of the clinical heterogeneity of prostate cancers.

## ACKNOWLEDGMENTS

We thank Lishi Chen, David Kuo, Sandra Levcovici, Faye Taylor, and Michelle Pappas for technical and database assistance.

## REFERENCES

- Greenlee, R. T., Hill-Hamon, M. B., Murray, T., and Thun, M. Cancer statistics, 2001. *CA Cancer J. Clin.*, 51: 15–36, 2001.
- Potosky, A., Feuer, E., and Levin, D. Impact of screening on incidence and mortality of prostate cancer in the United States. *Epidemiol. Rev.*, 23: 181–186, 2001.
- Scher, H. I., and Heller, G. Clinical states in prostate cancer: toward a dynamic model of disease progression. *Urology*, 55: 323–327, 2000.
- Fidler, I. J. Molecular biology of cancer: invasion and metastasis. In V. T. Devita, S. Hellman, and S. A. Rosenberg (eds.), *Cancer Principles and Practice of Oncology*, pp. 135–152. Philadelphia: Lippincott-Raven, 1997.
- Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA*, 95: 14863–14868, 1998.
- Lee, S. B., Kolquist, K. A., Nichols, K., Englert, C., Maheswaran, S., Ladanyi, M., Gerald, W., and Haber, D. A. The EWS-WT1 translocation product induces PDGF-A in desmoplastic small round cell tumour. *Nat. Genet.*, 17: 309–314, 1997.
- Ross, D. T., Scherf, U., Eisen, M. B., Perou, C. M., Rees, C., Spellman, P., Iyer, V., Jeffrey, S. S., Van de Rijn, M., Waltham, M., *et al.* Systematic variation in gene expression patterns in human cancer cell lines. *Nat. Genet.*, 24: 227–235, 2000.

8. Zhang, S., Chang, M. C. Y., Zylka, D., Turley, S., Harrison, R., and Turley, E. The hyaluronan receptor RHAMM regulates extracellular-regulated kinase. *J. Biol. Chem.*, *273*: 11342–11348, 1998.
9. Vyas, P., Norris, F. A., Joseph, R., Mjerus, P. W., and Orkin, S. Inositol polyphosphate 4-phosphatase type 1 regulates cell growth downstream of transcription factor GATA-1. *Proc. Natl. Acad. Sci. USA*, *97*: 13696–13701, 2000.
10. Welsh, J. B., Zarrinkar, P. P., Sapinoso, L. M., Kern, S. G., Behling, C. A., Monk, B. J., Lockhart, D. J., Burger, R. A., and Hampton, G. M. Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. *Proc. Natl. Acad. Sci. USA*, *98*: 1176–1181, 2001.
11. Luo, J., Duggan, D. J., Chen, Y., Sauvageot, J., Ewing, C. M., Bittner, M. L., Trent, J. M., and Isaacs, W. B. Human prostate cancer and benign prostatic hyperplasia: molecular dissection by gene expression profiling. *Cancer Res.*, *61*: 4683–4688, 2001.
12. Dhanasekaran, S., Barrette, T., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Plenta, K., Rubin, M., and Chinnalyan, A. Delineation of prognostic biomarkers in prostate cancer. *Nature (Lond.)*, *412*: 822–826, 2001.
13. Waghray, A., Schober, M., Feroze, F., Yao, F., Virgin, J., and Chen, Y. Identification of differentially expressed genes by serial analysis of gene expression in human prostate cancer. *Cancer Res.*, *61*: 4283–4286, 2001.
14. Nelson, P. S., Clegg, N., Eroglu, B., Hawkins, V., Bumgarner, R., Smith, T., Hood, L. The prostate expression database (PEDB): Status and enhancements in 2000. *Nucleic Acids Res.*, *28*: 212–213, 2000.
15. Hamminki, A., Markie, D., Tomlinson, I., Avizienyte, E., Roth, S., Loukola, A., Bignell, G., Warren, W., Aminoff, M., Höglund, P., *et al.* A serine/threonine kinase gene defective in Peutz-Jeghers syndrome. *Nature (Lond.)*, *391*: 184–187, 1998.
16. Zhang, X., Horwitz, G. A., Prezant, T. R., Valentini, A., Nakashima, M., Bronstein, M. D., and Melmed, S. Structure, expression, and function of *human pituitary tumor-transforming gene (PTTG)*. *Mol. Endocrinol.*, *13*: 156–166, 1999.
17. Zhou, H., Kuang, J., Zhong, L., Kuo, W., Gray, J. W., Sahin, A., Brinkley, B. R., and Sen, S. Tumour amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation. *Nat. Genet.*, *20*: 189–193, 1998.
18. Sala, A., and Watson, R. B-Myb protein in cellular proliferation, transcription control, and cancer: latest developments. *J. Cell. Physiol.*, *179*: 245–250, 1999.
19. Singh, D., Febbo, P., Ross, K., Jackson, D., Manola, J., Ladd, C., Tamayo, P., Renshaw, A., D'Amico, A., Richie, J., Lander, E., Loda, M., Kantoff, P., Golub, T., and Sellers, W. Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell*, *1*: 203–209, 2002.
20. Oh, I-H., and Reddy, E. The *myb* gene family in cell growth, differentiation and apoptosis. *Oncogene*, *18*: 3017–3033, 1999.