

Prediction of Gene Function by Genome-Scale Expression Analysis: Prostate Cancer-Associated Genes

Michael G. Walker,^{1,2,3} Wayne Volkmuth,¹ Einat Sprinzak,¹ David Hodgson,¹ and Tod Klingler¹

¹*Incyte Pharmaceuticals, Palo Alto, California 94304 USA;* ²*Department of Medicine, Stanford University, Stanford, California 94305 USA*

We wish to identify genes associated with disease. To do so, we look for novel genes whose expression patterns mimic those of known disease-associated genes, using a method we call Guilt-by-Association (GBA), on the basis of a combinatoric measure of association. Using GBA, we have examined the expression of 40,000 human genes in 522 cDNA libraries, and have discovered several hundred previously unidentified genes associated with cancer, inflammation, steroid-synthesis, insulin-synthesis, neurotransmitter processing, matrix remodeling, and other disease processes. The majority of the genes thus discovered show no sequence similarity to known genes, and thus could not have been identified by homology searches. We present here an example of the discovery of eight genes associated with prostate cancer. Of the 40,000 most-abundant human genes, these 8 are the most closely linked to the known diagnostic genes, and thus are prime targets for pharmaceutical research.

[The sequence data described in this paper have been submitted to the GenBank data library under accession nos. AF109298–AF109303.]

INTRODUCTION

Genes that are differentially expressed in disease states are candidates for pharmaceutical intervention. Previous researchers have collected expression data for up to 10,000 genes simultaneously (Lockhart et al. 1996; Lashkari et al. 1997), have identified genes differentially expressed in cancer (DeRisi et al. 1996; Fannon 1996; Zhang et al. 1997; Vasmatazis et al. 1998), and have identified clusters of coexpressed genes (Eisen et al. 1998; Michaels et al. 1998; Wen et al. 1998; Tamayo et al. 1999). Previous work has focused on differential expression, for example, in healthy versus diseased tissue (Greller and Tobin 1999); the joint expression of novel genes with known disease genes has rarely been examined. In addition, previous work has examined a small fraction of the total genome (typically 10,000 genes or less) and has used linear or monotonic measures of correlation, which fail to detect many known gene associations.

To identify genes that are candidate therapeutic or diagnostic targets, we look for novel genes whose expression patterns mimic those of known disease-associated genes. For the analyses presented here, we examined the expression of 40,000 human genes in 522 cDNA libraries in Incyte's LifeSeq database. The libraries were prepared from a diverse set of human anatomic and pathologic samples, representing most

major tissue categories and many of the major pathologies.

RESULTS

We present an example of the application of Guilt by Association (GBA) to identify genes associated with prostate cancer. Each year in the United States, prostate cancer kills >40,000 men, and >200,000 new cases are diagnosed, making it the second most common cancer and the second most common cause of cancer deaths among males (Parker et al. 1996; Presti and Carroll 1996; Foster 1998). Unfortunately, the best available diagnostic tests are substantially <100% sensitive and specific, and many men have incurable prostate cancer at the time of diagnosis (Whittemore et al. 1995; Richie and Kaplan 1996; Stamey 1996).

The standard molecular diagnostic marker for prostate cancer is prostate-specific antigen (PSA), a protease produced in the prostate (Morris et al. 1998); however, ~20% of men who undergo prostatectomy for prostate cancer have normal levels of PSA (Presti et al. 1996). Prostatic acid phosphatase (PAP) was used widely in diagnostic tests for prostate cancer prior to the development of the more reproducible PSA test (Bostwick 1998). Kallikrein is a protease expressed in the prostate that has 80% sequence similarity with PSA (Corey et al. 1997) and is differentially expressed in prostate cancer. Several groups are evaluating kallikrein for use as a diagnostic test for prostate cancer and as a measure of response to therapy (Charlesworth

³Corresponding author.
E-MAIL mwalker@incyte.com; FAX (650)855-0572.

et al. 1997; Eerola et al. 1997; Mikolajczyk et al. 1998). Two other proteins that have been linked to the disease are seminal-plasma protein and prostate-specific transglutaminase. Seminal-plasma protein is a prostate-secreted protein with inhibin-like activity (Mbikay et al. 1987). Inhibins are members of the transforming growth factor β (TGF β) superfamily of growth factors (Thomas et al. 1998) that modulate prostate tumor growth (Perry et al. 1997; Guo and Kyprianou 1998). Prostate-specific transglutaminase catalyzes post-translational protein cross-linking, and exhibits differential expression in prostate cancer cell lines (Dubink et al. 1996). Other proteins have been identified as potential markers of prostate cancer (Cramer et al. 1998; Rentzepis et al. 1998), and are candidates for use in the analysis method described here. Despite the current standard treatments for prostate cancer (surgery, radiation, and chemotherapy), men with this disease still have significant mortality and morbidity rates.

By use of GBA as described in the Methods section, we identified eight novel genes (*IPCA-1* through *IPCA-8*) that show a strong association with the known prostate cancer genes (PSA, PAP, kallikrein, seminal-plasma protein, and prostate-specific transglutaminase). Table 1 shows that, for PSA, the most closely coexpressed genes are glandular kallikrein, three novel genes, prostate seminal protein, PAP, a fourth novel gene, prostate transglutaminase, a fifth novel gene, and neuropeptide Y. (*IPCA-9*, *IPCA-10* and *IPCA-11* are coexpressed with PSA but appear to be 3' untranslated sequences.) Table 2 summarizes the coexpression of the eight novel genes with the five known prostate cancer genes, and, for comparison, two unrelated genes (myosin light chain and elongation factor 1 α). The four values for each gene in Table 1 are the values used in the GBA probability calculations (see Methods; Table 6, below). Each of the eight novel genes is coexpressed with at least one known prostate cancer gene with a *P* value <10 E-06.

By use of 522 libraries from diverse tissues, we in-

crease the sample size available for the statistical tests, but we run the risk of detecting associations that simply indicate that two genes are expressed in the same tissue, rather than being more closely linked in their function. We wish to test that the observed associations are not simply due to coexpression in the same tissue. Hence, we performed the same GBA analysis on a set of 51 male-reproductive tissue libraries, with the results shown in Table 3. The associations detected in the male-reproductive tissue libraries support the conclusions reached with all 522 libraries, specifically, that the same set of genes show close association even within the tissue type. The *P* values are not as small as with 522 libraries, because of the smaller sample size, but are still much less than the *P* values for the unrelated genes (*ef1- α* and *myosin*).

Four other known genes (beside the five used in the GBA analysis) are coexpressed with at least one of the known prostate cancer genes and are among the ten genes most closely coexpressed with that prostate cancer gene. *MAT8* is coexpressed with prostatic seminal protein. It has been reported to be differentially expressed in breast cancer (Morrison et al. 1995; Schiemann et al. 1998). Neuropeptide Y is coexpressed with PSA. It has been reported to be associated with prostate cancer (Minth et al. 1984; Mack et al. 1997). Sorbitol dehydrogenase is coexpressed with prostate transglutaminase and with kallikrein. Significant alterations in its activity in toxin-damaged male reproductive tissues have been reported (Pant et al. 1995); we found no previous report of an association with cancer. *ZN- α -2-glycoprotein* is coexpressed with prostate transglutaminase. It has been reported to be associated with prostate cancer (Gagnon et al. 1990) and breast cancer (Hurlimann and van Melle 1991; Freije et al. 1993; Lopez-Boado et al. 1994). Our analysis did not detect genes that are significantly underexpressed when the prostate cancer-associated genes are overexpressed.

Seven of the eight *IPCA* genes showed distant or no sequence similarity to genes known at the time of

Table 1. Ten Genes Most Closely Coexpressed with PSA

<i>P</i> value	Coexpressed gene	Number of libraries in which			
		both genes occur	only PSA occurs	only coexpressed gene occurs	neither gene occurs
1.53E-31	glandular kallikrein	26	12	3	481
1.65E-26	<i>IPCA-3</i>	22	16	2	482
7.48E-25	<i>IPCA-4</i>	26	12	14	470
8.12E-25	<i>IPCA-9</i>	23	15	6	478
3.38E-24	Prostate seminal protein	23	15	7	477
1.89E-23	PAP	24	14	11	473
6.87E-18	<i>IPCA-10</i>	19	19	9	475
9.01E-18	Prostate transglutaminase	14	24	0	484
4.61E-14	<i>IPCA-11</i>	27	11	66	418
1.58E-13	neuropeptide Y	16	22	11	473

Table 2. Coexpression ($-\log p$) of Eight Novel Genes with Five Known Prostate Cancer Genes and Two Unrelated Genes (ef1- α and Myosin Light Chain) in 522 Libraries

No.	Gene	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	IPCA-1														
2	IPCA-2	2													
3	IPCA-3	8	7												
4	IPCA-4	8	6	23											
5	IPCA-5	3	1	10	8										
6	IPCA-6	4	5	9	8	2									
7	IPCA-7	4	1	6	9	0	4								
8	IPCA-8	3	3	7	6	1	3	3							
9	PSA	8	6	26	24	8	9	11	6						
10	PAP	8	6	27	20	10	8	6	6	23					
11	kallikrein	10	6	30	23	9	9	8	7	31	30				
12	glutaminase	6	2	16	14	4	7	6	5	17	18	19			
13	seminal protein	9	4	22	21	5	8	8	7	23	19	21	14		
14	ef1- α	1	0	1	0	0	0	0	1	0	1	0	0	1	
15	myosin	1	1	4	3	1	2	1	1	4	3	3	2	4	1

the analysis. Gene IPCA-3 exhibits significant sequence similarity to several serine proteases. Subsequent to the identification of these genes by GBA and submission of their sequences to GenBank in November 1998, the sequence of a gene with >99% identity to IPCA-3 was reported in the GenBank database. This gene is prostase, an androgen-regulated serine protease with prostate-restricted expression (Nelson et al. 1999).

DISCUSSION

We have analyzed the pairwise coexpression patterns of 40,000 genes in >500 libraries (the largest such expression analysis reported to date), and have identified several hundred disease-associated genes using a novel coexpression algorithm, GBA. We identified eight novel genes associated with prostate cancer. Of the 40,000 most-abundant human genes, these 8 are the most closely linked to the known prostate cancer diagnostic genes, and thus are prime targets for pharmaceutical research.

The method of analysis presented here is complementary to traditional correlation measures in that GBA is better able to detect nonlinear relationships in data with high variability; for genes whose relationships are linear or monotonic, we would expect correlation analysis to provide similar results, possibly even with smaller sample sizes than are required for GBA.

GBA is able to detect potential disease-associated genes that have no sequence similarity to known genes, and does not require that expression be measured in both diseased and healthy tissue (as is required by differential expression analysis); thus, it offers opportunities to discover the functions of genes that cannot be readily identified by other means.

METHODS

For the method of analysis described here, we reduce each expression datum to a binary variable (present or absent), rather than analyzing expression as a continuous variable using linear or rank correlation. Before we chose this binary-

Table 3. Coexpression ($-\log p$) of Eight Novel Genes with Five Known Prostate Cancer Genes and Two Unrelated Genes (ef1- α and Myosin Light Chain) in 51 Male Reproductive Tissue Libraries

No.	Gene	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	IPCA-1														
2	IPCA-2	1													
3	IPCA-3	5	2												
4	IPCA-4	6	2	8											
5	IPCA-5	1	0	3	3										
6	IPCA-6	3	2	3	3	1									
7	IPCA-7	4	0	2	5	1	1								
8	IPCA-8	1	1	2	2	0	1	2							
9	PSA	6	2	10	9	3	3	5	2						
10	PAP	4	1	9	5	2	3	2	2	8					
11	kallikrein	6	2	10	8	3	3	3	2	12	10				
12	glutaminase	2	0	4	4	1	2	3	1	5	5	5			
13	seminal protein	5	1	8	9	1	3	5	3	9	7	7	4		
14	ef1- α	0	0	1	1	1	0	0	1	1	1	1	0	1	
15	myosin	0	0	2	1	1	1	0	1	1	1	1	1	1	0

encoding method to identify coexpressed genes, we evaluated Pearson linear correlation and Spearman rank correlation using continuous values. Whereas these correlation methods sometime identified known relationships well, they often performed unsatisfactorily, possibly for several reasons. Many genes that are known to be associated do not exhibit the simple linear or monotonic relationships assumed by these methods. Libraries may in some cases be normalized or subtracted to increase complexity. The quantitative measurement of expression has sufficiently high variability (inability to accurately distinguish two- to threefold changes, particularly at low expression levels) that correlation measures may not reliably distinguish true associations from spurious correlation.

For the purpose of this analysis, we consider a gene to be present (expressed) if cDNA corresponding to that gene is detected in the sample from that library. We consider a gene to be absent (not expressed) when no cDNA for that gene is detected in the library. To determine whether two genes, A and B, have similar expression patterns, we examine their occurrences in the 522 cDNA libraries, as shown in Table 4. A 0 indicates that the gene was not detected in the library; a 1 indicates that it was detected.

For a given pair of genes, the expression data in Table 4 can be summarized in a 2 by 2 contingency table. Table 5 presents such a coexpression contingency table for the hypothetical genes A and B in a total of 30 libraries; Table 6 presents the same data as variables that we will use shortly. We determine the probability that the coexpression shown in Table 5 occurs by chance with a counting method, as follows. We take as our null hypothesis that there is no association between gene A and gene B. Under the null hypothesis, the marginal counts in Tables 5 and 6 are fixed, the expected count in each cell is a function of the marginals, and deviations from the expected count are random. The number of ways that k occurrences of a gene can be distributed in r libraries is $\binom{r}{k}$, that is, the combinatoric choose function. From Table 6, we can calculate the probability of observing n_{11} counts using the hypergeometric distribution, as in a Fisher Exact test (Agresti 1990). From the hypergeometric distribution, the probability of observing exactly n_{11} counts is $p(n_{11}) = \binom{n_1}{n_{11}} \times \binom{n_2}{n_{21}} / \binom{n}{n_{11}}$.

To determine whether there is association (lack of independence) between the genes, we calculate the sum of all the (hypergeometric) probabilities for outcomes at least as extreme as the observed outcome. As a concrete example, consider the n_{11} count in the cell (Gene A present and Gene B present) in Table 5. We can calculate the probability of observing a count of exactly 8 using the hypergeometric distribution, that is, $p(n_{11} \text{ is } 8) = \binom{10}{8} \times \binom{20}{2} / \binom{30}{10}$. To test the null hypothesis, we are interested not only in the case in which we observe a count of exactly 8 in the cell, but also the cases in which we observe more extreme values of n_{11} ,

Table 4. Expression Data for Two Hypothetical Genes A and B

Gene	Library				522
	1	2	3	...	
A	1	1	0	...	0
B	1	0	1	...	0

Table 5. Summary of Coexpression for Genes A and B from Table 4

No. of libraries	Gene A present	Gene A absent	Total
Gene B present	8	2	10
Gene B absent	2	18	20
Total	10	20	30

subject to the constraints of the marginals. Hence, we sum the probability of the observed count and of the more extreme possible counts ($n_{11} = 8, 9,$ and 10) to determine the total probability of counts at least as extreme as those observed. In the case of Table 5, the probability that the observed coexpression is due to chance is $P = 0.0003$.

This method of estimating the probability for coexpression of two genes makes several assumptions that do not hold strictly. Because more than one library may be obtained from a single patient (for example, both tumor and nontumor tissue), libraries are not completely independent. In addition, because we perform multiple statistical tests on each gene, we must consider the question of statistical significance and interpretation of the P values. One method to correct for multiple comparisons in determining a suitable P value is to apply a Bonferroni correction (dividing the desired α , say, $P = 0.01$, by the number of comparisons performed). For n genes, we perform $n(n-1)/2$ pairwise comparisons; thus 40,000 genes yield 8×10^8 pairwise comparisons, requiring a Bonferroni-corrected P value of $0.01/(8 \times 10^8)$ or $\sim 10^{-11}$. With such a correction, some, but not all, of the eight identified genes still show significant association with the known genes. However, the Bonferroni correction is extremely conservative, yielding almost no false positives at the price of failing to detect many real associations. For this reason, we implemented an alternative to the Bonferroni correction, which we describe next.

To analyze 40,000 genes, we perform 8×10^8 pairwise comparisons. Suppose that the genes in this set are not, in fact, associated. If none of the genes in the 40,000 were associated, we would still expect, by chance, to see $8 \times 10^8 \times 10^{-6} = 800$ pairs of genes with a P value $< 10^{-6}$. Empirically, when we perform the 8×10^8 pairwise comparisons, we observe $>250,000$ pairs with a P value $< 10^{-6}$, which is consistent with the notion that many pairs of genes do, in fact, have related function and therefore have some similarity in their expression patterns. We expect 800 by chance, but observe $>250,000$; this result suggests that, at a P value of 10^{-6} , a very large proportion of all associations are due to true biological relationships.

For practical interpretation, we can only claim that, of the 40,000 genes examined, the genes identified here are the most closely associated with the known prostate cancer genes.

Table 6. Variables Representing Counts of Gene Occurrences

No. of libraries	Gene A present	Gene A absent	Total
Gene B present	n_{11}	n_{12}	n_1
Gene B absent	n_{21}	n_{22}	n_2
Total	n_1	n_2	n

The observed expression patterns demonstrate association with, but do not prove direct involvement in, prostate cancer. The known genes in this analysis are all far downstream in the process of prostate cancer progression; we speculate that some of the novel genes, whose expression patterns do not exactly match those of the known genes, may be closer to the origin of the disease, and perhaps involved in regulation of the progression of the cancer.

ACKNOWLEDGMENTS

We thank our colleagues at Incyte and our editor, Lyn Dupre, for their support and assistance in this research. We thank the reviewers for several suggestions that improved the quality and clarity of the paper.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Agresti, A. 1990. *Categorical data analysis*. Wiley, New York, NY.
- Bostwick, D. 1998. Pathology of prostate cancer. In *Prostate cancer*, (ed. M. Ernstoff, J. Heaney, and R. Peschel), pp. 15. Blackwell Science, Malden, MA.
- Charlesworth, M.C., C.Y. Young, G.G. Klee, M.S. Saedi, S.D. Mikolajczyk, J.A. Finlay, and D.J. Tindall. 1997. Detection of a prostate-specific protein, human glandular kallikrein (hK2), in sera of patients with elevated prostate-specific antigen levels. *Urology* **49**: 487–493.
- Corey, E., K.R. Buhler, and R.L. Vessella. 1997. Cross-reactivity of ten anti-prostate-specific antigen monoclonal antibodies with human glandular kallikrein. *Urology* **50**: 567–571.
- Cramer, S., E. Blind, and D. Peehl. 1998. Detection of prostate specific antigen (PSA)-cleaved parathyroid hormone-related protein (PTHrP) in human serum. *Proc. Am. Assoc. Cancer Res.* **39**: 105.
- DeRisi, J., L. Penland, P.O. Brown, M.L. Bittner, P.S. Meltzer, M. Ray, Y. Chen, Y.A. Su, and J.M. Trent. 1996. Use of a cDNA microarray to analyze gene expression patterns in human cancer. *Nat. Genet.* **14**: 457–460.
- Dubbink, H.J., N.S. Verkaik, P.W. Faber, J. Trapman, F.H. Schroder, and J.C. Romijn. 1996. Tissue specific and androgen-regulated expression of human prostate-specific transglutaminase. *Biochem. J.* **315**: 901–908.
- Eerola, R., T. Piironen, K. Pettersson, J. Lovgren, M. Vehniainen, H. Lilja, B. Dowell, T. Lovgren, and M. Karp. 1997. Immunoreactivity of recombinant human glandular kallikrein using monoclonal antibodies raised against prostate-specific antigen. *Prostate* **31**: 84–90.
- Eisen, M.B., P.T. Spellman, P.O. Brown, and D. Botstein. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci.* **95**: 14863–14868.
- Fannon, M.R. 1996. Gene expression in normal and disease states—identification of therapeutic targets. *Trends Biotechnol.* **14**: 294–298.
- Foster, H. 1998. Diagnostic evaluation of prostate cancer. In *Prostate cancer*, (ed. M. Ernstoff, J. Heaney, and R. Peschel), pp. 48. Blackwell Science, Malden, MA.
- Freije, J.P., A. Fueyo, J.A. Uria, G. Velasco, L.M. Sanchez, Y.S. Lopez-Boado, and C. Lopez-Otin. 1993. Human Zn-alpha 2-glycoprotein: Complete genomic sequence, identification of a related pseudogene and relationship to class I major histocompatibility complex genes. *Genomics* **18**: 575–587.
- Gagnon, S., B. Tetu, J.Y. Dube, and R.R. Tremblay. 1990. Expression of Zn-alpha 2-glycoprotein and PSP-94 in prostatic adenocarcinoma. An immunohistochemical study of 88 cases. *Am. J. Pathol.* **136**: 1147–1152.
- Greller, L.D. and F.L. Tobin. 1999. Detecting selective expression of genes and proteins. *Genome Res.* **9**: 282–296.
- Guo, Y. and N. Kyprianou. 1998. Overexpression of transforming growth factor (TGF) beta1 type II receptor restores TGF-beta1 sensitivity and signaling in human prostate cancer cells. *Cell Growth Differ.* **9**: 185–193.
- Hurlimann, J. and G. van Melle. 1991. Prognostic value of serum proteins synthesized by breast carcinoma cells. *Am. J. Clin. Pathol.* **95**: 835–843.
- Lashkari, D.A., J.L. DeRisi, J.H. McCusker, A.F. Namath, C. Gentile, S.Y. Hwang, P.O. Brown, and R.W. Davis. 1997. Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc. Natl. Acad. Sci.* **94**: 13057–13062.
- Lockhart, D.J., H. Dong, M.C. Byrne, M.T. Follettie, M.V. Gallo, M.S. Chee, M. Mittmann, C. Wang, M. Kobayashi, H. Horton et al. 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays [In Process Citation]. *Nat. Biotechnol.* **14**: 1675–1680.
- Lopez-Boado, Y.S., I. Diez-Itza, J. Tolivia, and C. Lopez-Otin. 1994. Glucocorticoids and androgens up-regulate the Zn-alpha 2-glycoprotein messenger RNA in human breast cancer cells. *Breast Cancer Res. Treat.* **29**: 247–258.
- Mack, D., G.W. Hacker, C. Hauser-Kronberger, J. Frick, and O. Dietze. 1997. Vasoactive intestinal polypeptide (VIP) and neuropeptide tyrosine (NPY) in prostate carcinoma. *Eur. J. Cancer.* **33**: 317–318.
- Mbakay, M., S. Nolet, S. Fournier, S. Benjannet, P. Chapdelaine, G. Paradis, J.Y. Dube, R. Tremblay, C. Lazure, N.G. Seidah et al. 1987. Molecular cloning and sequence of the cDNA for a 94-amino-acid seminal plasma protein secreted by the human prostate. *DNA* **6**: 23–29.
- Michaels, G.S., D.B. Carr, S. Fuhrman, X. Wen, and R. Somogyi. 1998. Cluster analysis and data visualization of large-scale gene expression data. In *Pacific symposium on biocomputing*, (ed. R. Altman, A. Dunker, L. Hunter, and T. Klein), pp. 42. World Scientific, Singapore.
- Mikolajczyk, S.D., L.S. Millar, A. Kumar, and M.S. Saedi. 1998. Human glandular kallikrein, hK2, shows arginine-restricted specificity and forms complexes with plasma protease inhibitors. *Prostate* **34**: 44–50.
- Mintz, C.D., S.R. Bloom, J.M. Polak, and J.E. Dixon. 1984. Cloning, characterization, and DNA sequence of a human cDNA encoding neuropeptide tyrosine. *Proc. Natl. Acad. Sci.* **81**: 4577–4581.
- Morris, D.L., P.W. Dillon, D.L. Very, P. Ng, L. Kish, J.L. Goldblatt, D.J. Bruzek, D.W. Chan, M.S. Ahmed, D. Witek et al. 1998. Bayer Immuno 1 PSA Assay: An automated, ultrasensitive method to quantitate total PSA in serum. *J. Clin. Lab. Anal.* **12**: 65–74.
- Morrison, B.W., J.R. Moorman, G.C. Kowdley, Y.M. Kobayashi, L.R. Jones, and P. Leder. 1995. Mat-8, a novel phospholemmann-like protein expressed in human breast tumors, induces a chloride conductance in *Xenopus* oocytes. *J. Biol. Chem.* **270**: 2176–2182.
- Nelson, P.S., L. Gan, C. Ferguson, P. Moss, R. Gelinis, L. Hood, and K. Wang. 1999. Molecular cloning and characterization of prostase, an androgen-regulated serine protease with prostate-restricted expression. *Proc. Natl. Acad. Sci.* **96**: 3114–3119.
- Pant, N., A.K. Prasad, S.C. Srivastava, R. Shankar, and S.P. Srivastava. 1995. Effect of oral administration of carbofuran on male reproductive system of rat. *Hum. Exp. Toxicol.* **14**: 889–894.
- Parker, S.L., T. Tong, S. Bolden, and P.A. Wingo. 1996. Cancer statistics, 1996. *CA Cancer J. Clin.* **46**: 5–27.
- Perry, K.T., C.T. Anthony, T. Case, and M.S. Steiner. 1997. Transforming growth factor beta as a clinical biomarker for prostate cancer. *Urology* **49**: 151–155.
- Presti, J.C., Jr. and P.R. Carroll. 1996. Use of prostate-specific antigen (PSA) and PSA density in the detection of stage T1 carcinoma of the prostate. *Semin. Urol. Oncol.* **14**: 134–138.
- Presti, J., M. Stoller, and P. Carroll. 1996. Urology. In *Current medical diagnosis and treatment*, (ed. L. Tierney, S. McPhee, and M. Papadakis), pp. 822–857. Lange, Stamford, CT.
- Rentzepis, M., P. Gumerlock, and R. deVere White. 1998. Future

- directions in prostate cancer research. In *Prostate cancer*, (ed. M. Ernstoff, J. Heaney, and R. Peschel), pp. 203–209. Blackwell Science, Malden, MA.
- Richie, J. and I. Kaplan. 1996. Screening for prostate cancer: The horns of a dilemma. In *Prostate cancer: A multidisciplinary guide*, (ed. P. Kantoff, K. Wishnow, and K. Loughlin), pp. 1–10. Blackwell Science, Malden, MA.
- Schiemann, S., M. Schwirzke, N. Brunner, and U.H. Weidle. 1998. Molecular analysis of two mammary carcinoma cell lines at the transcriptional level as a model system for progression of breast cancer. *Clin. Exp. Metastasis* **16**: 129–139.
- Stamey, T.A. 1996. Lower limits of detection, biological detection limits, functional sensitivity, or residual cancer detection limit? Sensitivity reports on prostate-specific antigen assays mislead clinicians. *Clin. Chem.* **42**: 849–852.
- Tamayo, P., D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E.S. Lander, and T.R. Golub et al. 1999. Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci.* **96**: 2907–2912.
- Thomas, T.Z., S.M. Chapman, W. Hong, C. Gurusinghe, S.L. Mellor, R. Fletcher, J. Pedersen, and G.P. Risbridger. 1998. Inhibins, activins, and follistatins: Expression of mRNAs and cellular localization in tissues from men with benign prostatic hyperplasia. *Prostate* **34**: 34–43.
- Vasmataz, G., M. Essand, U. Brinkmann, B. Lee, and I. Pastan. 1998. Discovery of three genes specifically expressed in human prostate by expressed sequence tag database analysis. *Proc. Natl. Acad. Sci.* **95**: 300–304.
- Wen, X., S. Fuhrman, G.S. Michaels, D.B. Carr, S. Smith, J.L. Barker, and R. Somogyi. 1998. Large-scale temporal gene expression mapping of central nervous system development. *Proc. Natl. Acad. Sci.* **95**: 334–339.
- Whittemore, A.S., C. Lele, G.D. Friedman, T. Stamey, J.H. Vogelman, and N. Orentreich. 1995. Prostate-specific antigen as predictor of prostate cancer in black men and white men. *J. Natl. Cancer Inst.* **87**: 354–360.
- Zhang, L., W. Zhou, V.E. Velculescu, S.E. Kern, R.H. Hruban, S.R. Hamilton, B. Vogelstein, and K.W. Kinzler. 1997. Gene expression profiles in normal and cancer cells. *Science* **276**: 1268–1272.

Received April 26, 1999; accepted in revised form September 23, 1999.