

# Genomic Profiling Reveals Alternative Genetic Pathways of Prostate Tumorigenesis

Jacques Lapointe,<sup>1,2</sup> Chunde Li,<sup>6</sup> Craig P. Giacomini,<sup>1</sup> Keyan Salari,<sup>1,3</sup> Stephanie Huang,<sup>1</sup> Pei Wang,<sup>5</sup> Michelle Ferrari,<sup>2</sup> Tina Hernandez-Boussard,<sup>4</sup> James D. Brooks,<sup>2</sup> and Jonathan R. Pollack<sup>1</sup>

Departments of <sup>1</sup>Pathology, <sup>2</sup>Urology, <sup>3</sup>Genetics, and <sup>4</sup>Biochemistry, Stanford University, Stanford, California; <sup>5</sup>Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington; and <sup>6</sup>Department of Oncology-Pathology, Karolinska Institutet, and Department of Oncology, Karolinska University Hospital, Stockholm, Sweden

## Abstract

**Prostate cancer is clinically heterogeneous, ranging from indolent to lethal disease. Expression profiling previously defined three subtypes of prostate cancer, one (subtype-1) linked to clinically favorable behavior, and the others (subtypes-2 and -3) linked with a more aggressive form of the disease. To explore disease heterogeneity at the genomic level, we carried out array-based comparative genomic hybridization (array CGH) on 64 prostate tumor specimens, including 55 primary tumors and 9 pelvic lymph node metastases. Unsupervised cluster analysis of DNA copy number alterations (CNA) identified recurrent aberrations, including a 6q15-deletion group associated with subtype-1 gene expression patterns and decreased tumor recurrence. Supervised analysis further disclosed distinct patterns of CNA among gene-expression subtypes, where subtype-1 tumors exhibited characteristic deletions at 5q21 and 6q15, and subtype-2 cases harbored deletions at 8p21 (*NKX3-1*) and 21q22 (resulting in *TMPRSS2-ERG* fusion). Lymph node metastases, predominantly subtype-3, displayed overall higher frequencies of CNA, and in particular gains at 8q24 (*MYC*) and 16p13, and loss at 10q23 (*PTEN*) and 16q23. Our findings reveal that prostate cancers develop via a limited number of alternative preferred genetic pathways. The resultant molecular genetic subtypes provide a new framework for investigating prostate cancer biology and explain in part the clinical heterogeneity of the disease.** [Cancer Res 2007;67(18):8504–10]

## Introduction

Prostate cancer is the most frequently diagnosed cancer among men in the United States, with one in six men being diagnosed in their lifetime (1). Prostate cancers display a range of clinical behavior, from slow-growing tumors of no clinical significance to aggressively metastatic and lethal disease. Although therapies have been well documented to improve prostate cancer survival, there is a growing consensus that a large subset of patients do not require aggressive treatment (2). A critical question facing patients and their treating physicians is which localized prostate cancers

require aggressive therapy, with the attendant risks of urinary and sexual dysfunction, and which can be followed safely. The observed clinical heterogeneity of prostate cancer is likely attributable in large part to underlying molecular heterogeneity, and an improved understanding of molecular variation might, in the future, inform decisions of whether and how aggressively to treat.

Numerous molecular alterations have been implicated in prostate cancer initiation or progression, and some correlate with clinical outcome and have been proposed as surrogate markers of disease risk or prognosis. Germ-line mutations of several genes have been found in a relatively small subset of hereditary prostate cancers, although most do not seem to be commonly involved in sporadic tumors (3). Loss of expression of pi-class glutathione transferase, due to extensive promoter hypermethylation, is found in most prostate cancers and is also seen in many prostate cancer precursor lesions [prostatic intraepithelial neoplasia (PIN); ref. 4]. Likewise, allelic loss of chromosome 8p has been observed frequently in localized prostate cancers and has also been reported in PIN. Recently, fusion of the androgen-regulated gene *TMPRSS2* to ETS family oncogenic transcription factors (most frequently *ERG*) has been identified as a common rearrangement in prostate cancer (5) and may be associated with aggressive tumor features (6).

DNA copy number alterations (CNA) have been extensively characterized in prostate cancers and include recurrent regions of loss, indicating the locations of either known or candidate prostate cancer tumor suppressor genes (TSG), such as 13q (*RBI*), 10q (*PTEN*), 16q (*CDH1*; E-cadherin), 6q, 5q (*APC*), 2q, and 18q (*SMAD4*), as well as regions of recurrent gains at 8q (*MYC*) and 7, among others (3, 7). Overexpression of the androgen receptor (AR) and *AR* gene amplification have also been observed in late, hormone-refractory cancer (8). Most of these molecular and cytogenetic alterations have been found to occur with increased frequency in tumors of high histologic grade, advanced stage, and with clinical disease progression. Based on these observations, prostate carcinogenesis has been proposed to occur as a result of a stepwise accumulation of genetic alterations that correlate with disease progression (7).

In previous studies, we explored molecular variation in prostate tumors by gene-expression profiling (9). Unsupervised hierarchical cluster analysis identified three subtypes of prostate cancer not previously recognized or distinguishable histologically, based on different gene expression patterns. One of the subtypes, subtype-1, was notable for its clinically favorable behavior, where scoring two immunohistochemical (IHC)-based surrogate markers for this subtype (i.e., AZGP1+/MUC1–) predicted favorable tumor-free survival in a separate cohort of patients, independent of pathologic stage, Gleason grade, and preoperative serum prostate-specific antigen (PSA) levels. Whether these subtypes represent stages along a linear model of prostate carcinogenesis and progression or

**Note:** Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

J.D. Brooks and J.R. Pollack share senior authorship.

Current address for J. Lapointe: Urology Division, Department of Surgery, McGill University, Montreal, Quebec, Canada.

**Requests for reprints:** Jonathan R. Pollack, Department of Pathology, Stanford University School of Medicine, CCSR-3245A, 269 Campus Drive, Stanford, CA 94305-5176. Phone: 650-736-1987; Fax: 650-736-0073; E-mail: pollack1@stanford.edu.

©2007 American Association for Cancer Research.

doi:10.1158/0008-5472.CAN-07-0673

distinct molecular genetic entities that arise independently is currently unknown. To further investigate the biological underpinnings of disease heterogeneity, and in particular of the three gene-expression subtypes, we did array-based comparative genomic hybridization (array CGH) profiling of genomic DNA CNAs at submegabase resolution.

## Materials and Methods

**Specimens.** Freshly frozen prostate surgical specimens, scalpel-dissected to enrich for tumor cells, were described in Lapointe et al. (9), where detailed pathologic and clinical data are also available. Specimens were collected with Institutional Review Board approval. Genomic DNA was isolated from 64 samples, including 55 primary tumors and 9 unmatched pelvic lymph node metastases (from aborted surgeries), by back-extraction from the TRIzol (Invitrogen) organic phase according to the manufacturer's protocol.

**Array CGH.** cDNA microarrays for CGH were obtained from the Stanford Functional Genomics Facility and included 39,632 human cDNAs, representing 22,279 mapped human genes [18,049 UniGene clusters (10), together with 4,230 additional mapped expressed sequence tags not assigned UniGene IDs]. Array CGH was done according to our published protocols (11, 12). Briefly, 4  $\mu$ g of genomic DNA from each tumor specimen was random primer labeled with Cy5 and cohybridized to the microarray along with 4  $\mu$ g of Cy3-labeled normal male leukocyte reference DNA from a single donor. Following overnight hybridization and washing, arrays were imaged using a GenePix 4000B scanner (Molecular Devices). Fluorescence ratios were extracted using GenePix Pro software, and the data were uploaded into the Stanford Microarray Database (13) for storage, retrieval, and analysis. The complete microarray data set is accessible at the Stanford Microarray Database<sup>7</sup> and at GEO<sup>8</sup> (accession GSE6469).

**Microarray data analysis.** For array CGH analysis, background-subtracted fluorescence ratios were normalized by mean centering genes for each array. We included for subsequent analysis only well-measured genes with Cy3 reference-channel fluorescence signal intensity at least 1.4-fold above background in at least 50% of samples. Map positions for arrayed cDNA clones were assigned using the National Center for Biotechnology Information (NCBI) genome assembly, accessed through the University of California at Santa Cruz (UCSC) genome browser database (NCBI Build 35). For genes represented by multiple arrayed cDNAs, the average fluorescence ratio was used. DNA gains and losses were identified using the CLuster Along Chromosomes method (CLAC; ref. 14). Briefly, the CLAC algorithm builds a hierarchical cluster-style tree along each chromosome, such that neighboring genes with positive and negative ratios are separated into different clusters. DNA gains and losses are then called significant based on the height and width of clusters, and a false discovery rate (FDR) is estimated by comparison to normal-normal hybridization data. Hierarchical cluster analysis (15) of CNAs was done with Pearson correlation and average linkage clustering using Gene Cluster software.<sup>9</sup> Significant associations between CNAs and gene-expression subtypes were identified using the Significance Analysis of Microarrays (SAM) two-class method (16), which is based on a modified *t* statistic and uses random permutations of class labels to estimate a FDR. For the hierarchical cluster and SAM analyses, the ~22,000 mapped human genes were first collapsed into the 803 cytoband loci (boundaries defined by NCBI Build 35) represented by those genes. By integrating information across neighboring genes, binning provides a useful balance between minimizing noise and maximizing mapping resolution (17). For each specimen, cytobands displaying gain or loss were defined as those harboring at least two genes with gain or loss (respectively) called by CLAC. Frequency plots of CNAs were drawn using CGH Explorer (18). Kaplan-Meier survival analysis was done using WinSTAT (R. Finch software).

**Gene set enrichment analysis.** Gene set enrichment analysis (GSEA; ref. 19), done as described (20), was applied to a set of 4,741 variably expressed named genes from our expression-profiling data set (9). Analyzed gene sets (see Supplementary Table S1) included MYC up-regulated targets (Myc Target Gene Database; ref. 21) and cell-cycle-regulated genes (22). Genes with putative ETS binding sites within the first 1-kb upstream promoter sequence (ref. 23; obtained from the UCSC Genome Browser<sup>10</sup>) were defined using MATCH software (ref. 24; default setting to minimize false positives), applied to a TRANSFAC (25) binding site matrix (V\$ELK1\_02) for the ETS-family gene *ELK1*. Although an ERG binding site has not yet been empirically defined, all ETS factors share a common core binding site, GGAA/T (26). To assess enrichment of ETS binding sites, the absolute value of the GSEA metric (Pearson correlation) was used to consider both up-regulated and down-regulated targets.

**Fluorescence *in situ* hybridization.** Fluorescence *in situ* hybridization (FISH) was carried out on paraffin sections of paired tumor-normal tissue from selected cases, arrayed in TMA format as described previously (27). Probe labeling and FISH was done using Vysis reagents according to the manufacturer's protocols. Locus-specific BACs mapping to 6q16.1 (CTC-387L6; BACPAC Resources Centre) and 8p22 (RP11-105D13; Invitrogen), residing within subtype-characteristic deletions, were labeled with SpectrumOrange, and cohybridized with SpectrumGreen-labeled centromere control probes (Vysis) CEP6 and CEP8, respectively. Chromosomal locations of BACs were validated using normal metaphase slides (data not shown). Slides were counterstained with 4',6-diamidino-2-phenylindole (DAPI) and imaged using an Olympus BX51 fluorescence microscope with Applied Imaging Cytovision 3.0 software. For each tumor-normal pair, the average ratio of locus specific to centromere FISH signals was tabulated for 50 tumor and 50 normal interphase nuclei, and deletion was defined by a statistically significant difference (Student's *t* test).

## Results

To investigate prostate cancer heterogeneity at the genomic level, we profiled DNA CNAs by CGH on cDNA microarrays representing ~22,000 mapped genes. Genomic DNA was available for 64 prostate tumors, including 55 primary tumors and 9 unmatched therapy-naïve pelvic lymph node metastases, from a set of 71 tumor specimens previously profiled for gene expression (9). Overall, we observed numerous recurrent CNAs (summarized in Supplementary Fig. S1), the spectrum of which was consistent with published chromosome-based CGH studies (7), with deletions outnumbering gains (average fractional genome loss and gain of 6.4% and 2.6%, respectively;  $P < 0.001$ , Student's *t* test). The most frequent aberrations included gains at 8q (27% of cases) and losses at 13q (52%), 8p (47%), 6q (38%), 10q (28%), and 12p (27%).

To probe the organization of CNAs across the specimen set, we did unsupervised hierarchical clustering of prostate samples in the space of CNAs (Fig. 1A). Cluster analysis accentuated common recurrent aberrations, including gain at 8q22-q24 (*MYC*) and losses at 8p11-p23 (*NKX3-1*), 10q23 (*PTEN*), 12p13 (*CDKN1B*), and 13q14 (*RBI*). Notably, cluster analysis also highlighted a sample group with deletion on 6q (smallest shared deletion at 6q15), which was significantly associated with subtype-I gene expression patterns ( $P < 0.001$ ,  $\chi^2$  test), as well as a decreased tumor recurrence rate (Fig. 1B;  $P = 0.028$ , log-rank test).

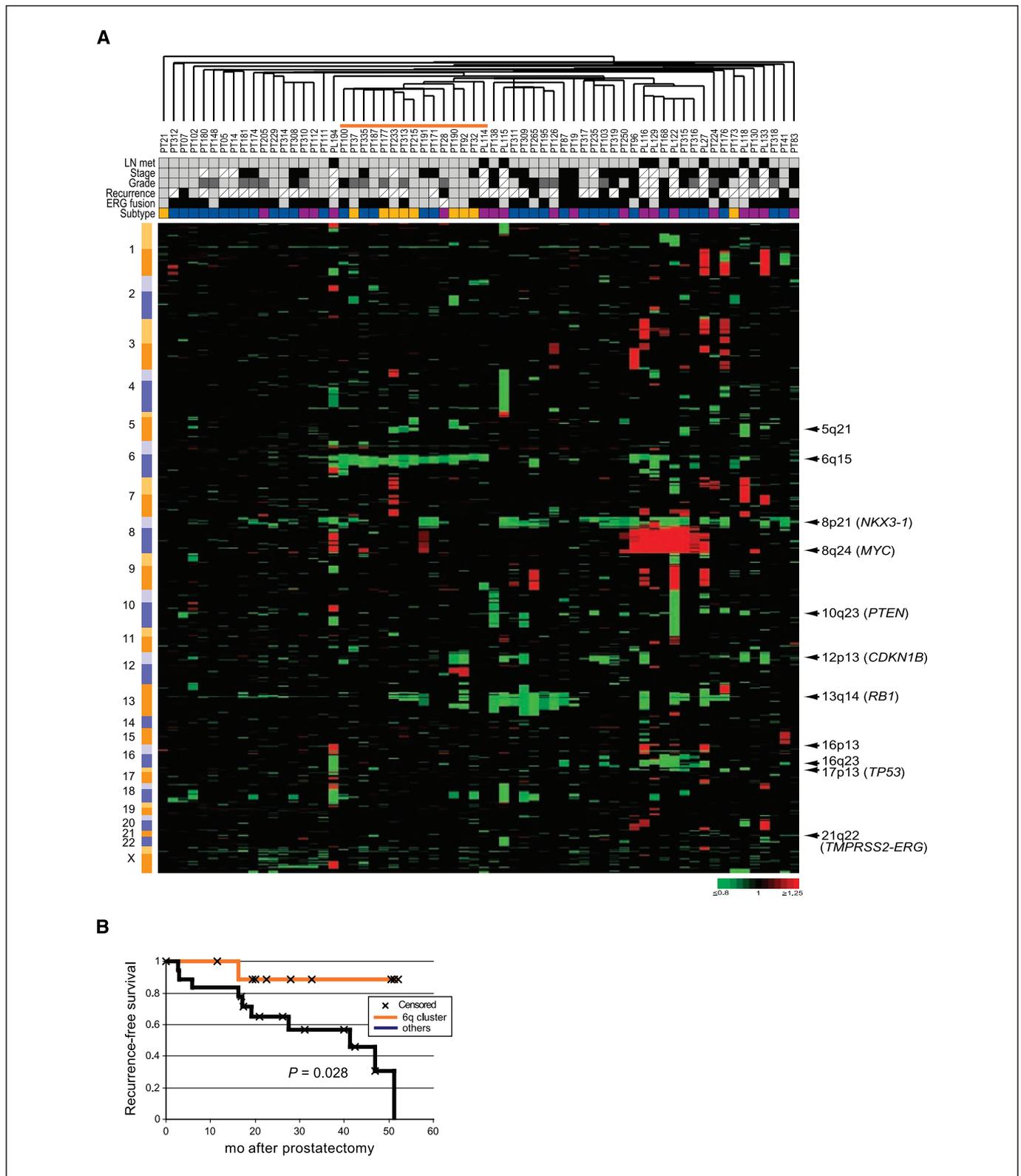
To more directly assess the relation between CNAs and gene-expression subtypes, we applied supervised analysis using the two-class (i.e., one subtype versus all others) SAM method (16). Supervised analysis affirmed that expression subtypes exhibited distinct patterns of CNAs (Fig. 2A and Table 1). In particular,

<sup>7</sup> <http://genome-www.stanford.edu/microarray>

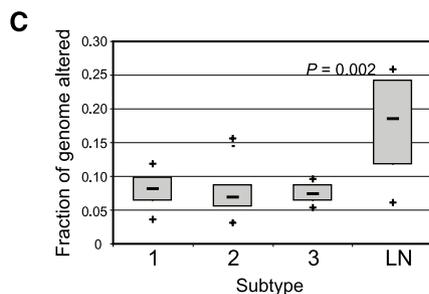
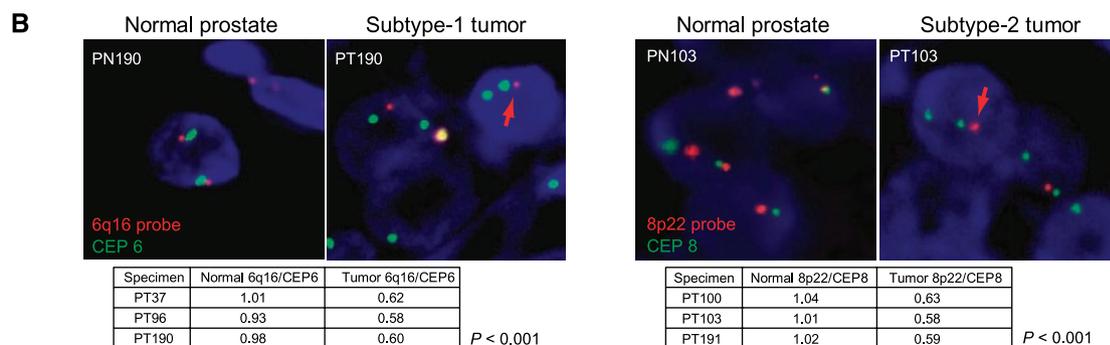
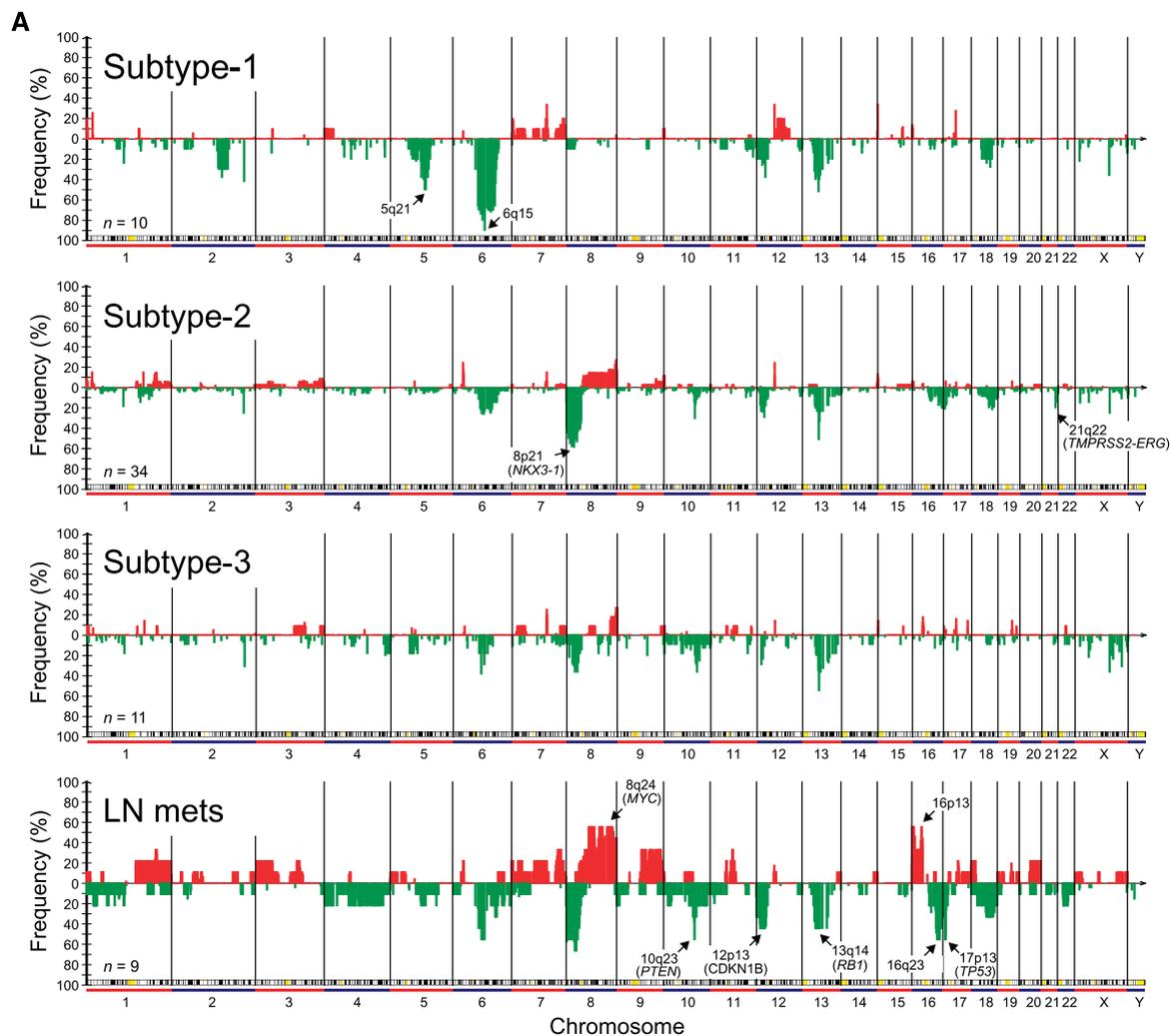
<sup>8</sup> <http://www.ncbi.nlm.nih.gov/geo>

<sup>9</sup> <http://rana.lbl.gov/EisenSoftware.htm>

<sup>10</sup> <http://genome.ucsc.edu>



**Figure 1.** Hierarchical cluster analysis of prostate cancer genomic copy number profiles. **A**, heatmap representation of the 64 prostate cancer specimens (*columns*) clustered across the 692 cytobands harboring CNAs ordered by genome position (*rows*). Mean  $\log_2$  array CGH ratios for cytobands are depicted by ratio-fold colorimetric scale (shown); *red* and *green*, gain and loss, respectively. *Shaded boxes below the sample dendrogram*, clinicopathologic features of tumor samples. Lymph node metastasis: primary tumor (*gray*); lymph node met (*black*). Pathologic stage: T2 (*gray*); T3 (*black*). Gleason grade: 6 (*gray*); 7 (*dark gray*);  $\geq 8$  (*black*). PSA-recurrence: nonrecurrent (*gray*); recurrent (*black*). *TMPRSS2-ERG* fusion by RT-PCR (28): no fusion expressed (*gray*); fusion expressed (*black*). Gene-expression subtype: subtype-1 (*yellow*), subtype-2 (*blue*), subtype-3 (*purple*). Missing data or not applicable (*white*). Selected recurrent CNAs (cytobands and cancer genes) are indicated. Orange bar, 6q15-deletion tumor cluster. **B**, Kaplan-Meier analysis comparing recurrence-free survival associated with primary tumors either within or outside the 6q15-deletion cluster (log-rank test; *P* value indicated).



**Figure 2.** Expression subtypes exhibit distinct patterns of CNA. **A**, frequency plots summarizing the distribution of CNAs in subtype-1, subtype-2, subtype-3 (primaries only), and therapy-naïve pelvic lymph node metastases. *Red* and *green*, gain and loss, respectively. Selected CNAs harboring known or novel cancer genes are identified. Number (*n*) of specimens represented in each group is indicated. **B**, FISH analysis verifying subtype-selective deletions at 6q16 (subtype-1) and 8p22 (subtype-2). Representative FISH images are shown; *red*, locus-specific FISH probes (*arrow*, tumor nuclei with single signal); *green*, control centromere probes. Counting FISH signals from 50 cells per specimen from three tumor-normal pairs confirms significantly decreased locus/centromere ratios (*P* values indicated; Student's *t* test) in tumor cases exhibiting deletion by array CGH. **C**, fraction of genome (cytobands) harboring CNA for each of the subgroups. *Box-and-whisker* plots, median, lower and upper quartiles, and the smallest and largest values.

**Table 1.** Subtype-characteristic CNAs identified by SAM

Subtype-1 (FDR, 4%)	Subtype-2 (FDR, 5%)	Lymph node metastases (FDR, 2%)*
Positive associations:		
Loss	Loss	Loss
2q22.3-q23.1 (2) <sup>†</sup>	8p11.21-p23.3	1p36.12-p36.13 (2)
2q23.3	(10; <i>NKX3-1</i> )	2q37.1
2q24.2	16q23.1	4q31.1
5q21.1-q21.3 (2)	17p13.2	10q23.31 ( <i>PTEN</i> )
6q13	17q21.2	16q22.2-q23.3 (5)
6q14.3-q22.31(8)	18q23	
	21q22.2	
	( <i>TMPRSS2-ERG</i> )	
	Gain	Gain
	6p22.1	8p11.23-p12 (2)
		8q12.3-q21.13 (7)
		8q22.1-q23.1 (4)
		8q23.3-q24.22 (6; <i>MYC</i> )
		9q21.13-q21.33 (5)
		9q22.32-q31.3 (5)
Negative associations:		11q12.3-q13.1 (2)
Loss		16p11.2-p12.1 (2)
8p12-p23.1 (6; <i>NKX3-1</i> )		16p12.3-p13.3 (6)
8p11.21		20q13.2

\*Only the top 50 of 91 significant cytobands are shown.

<sup>†</sup> For significant cytoband intervals, the number of cytobands is indicated in parentheses.

subtype-1 was characterized by frequent DNA loss at 5q21 (proximal to *APC* at 5q22) and 6q13-q22 (peak loss at 6q15), without loss from 8p, whereas subtype-2 tumors exhibited frequent loss at 8p11-p23 (peak loss at 8p21; *NKX3-1*), and among other loci, at 21q22, an intrachromosomal rearrangement resulting in the expression of the *TMPRSS2-ERG* fusion. Reverse transcription-PCR (RT-PCR) analysis of the *TMPRSS2-ERG* fusion in this sample set (28) indicated a 21q22 deletion in 34% of cases (15 of 44) expressing the fusion, somewhat lower than a prior report of ~50% (29). Like 21q22 deletion, we found expression of the *TMPRSS2-ERG* fusion also to be significantly associated with subtype-2 cases in our sample set ( $P = 0.001$ ,  $\chi^2$  test). Consistent with this finding, GSEA identified significant enrichment of genes with putative ETS binding sites (see Materials and Methods) in subtype-2 ( $P = 0.002$ ; Fig. 3A). For two of the subtype-selective deletions, 6q13-q22 (subtype-1) and 8p11-p23 (subtype-2), we also carried out FISH analysis on a subset of the same tumor specimens as a technical validation of findings (Fig. 2B).

Gene-expression subtype-3 included most lymph node metastases (8 of 9 cases), as well as primary tumors (11 cases). Despite shared expression patterns, preliminary array CGH analysis suggested the lymph node metastases comprised a distinct and homogeneous group, so they were considered separately. Although subtype-3 primary tumors as a group showed no characteristic CNAs, the lymph node metastases were characterized by an overall higher frequency of CNAs ( $P = 0.002$ , Student's  $t$  test; Fig. 2C) and specifically and most notably by frequent gains at 8q12-q24

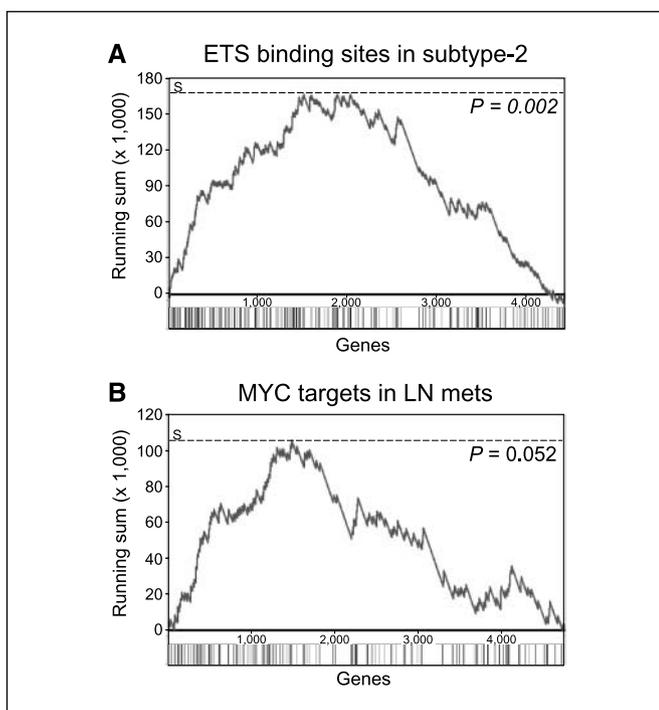
(including *MYC*) and 16p11-p13 (peak at 16p13) and loss at 10q23 (*PTEN*) and 16q22-q23 (peak loss at 16q23, distal to *CDHI* at 16q22), among others (Fig. 2A). Consistent with the finding of gain spanning 8q24 (*MYC*), GSEA analysis disclosed in the group of lymph node metastases significant enrichment of a curated set of *MYC*-up-regulated target genes (ref. 21;  $P = 0.052$ ; Fig. 3B), as well as an empirically derived set of cell-cycle-regulated genes (ref. 22;  $P = 0.034$ ; data not shown). Array CGH analysis also identified other aberrations, like deletions centered at 12p13 (*CDKN1B*) and 13q14 (*RBI*), which were frequent but shared among subtypes.

## Discussion

Our findings are consistent with the existence of distinct molecular subtypes of prostate cancer that are distinguished by their unique gene expression signatures, DNA CNAs, biological features, and clinical behavior. Previously, we had shown that expression of AZGP1 protein, a gene overexpressed in subtype-1 cancers, predicts favorable outcome after prostatectomy, a finding recently validated independently (30). Here, we identify the loss of chromosomal material from 5q and 6q, without loss from 8p, to be common genetic features of these clinically favorable subtype-1 cancers. The relationship between these specific deletions and the gene-expression patterns that characterize subtype-1 cases remains to be investigated. In general, a subset of gene-expression patterns is likely directly attributable to underlying CNA at the same loci (ref. 12 and Supplementary Fig. S2), whereas for others (e.g., *AZGP1* at 7q22), any link would be indirect.

Subtype-2 cancers display DNA CNAs that are distinct from subtype-1 cancers. Deletion at 8p21 (*NKX3-1*) and 21q22 (*TMPRSS2-ERG* fusion) is common in this subtype, whereas loss of material from chromosomes 5q and 6q is rare. Consistent with finding 21q22 deletion and *TMPRSS2-ERG* fusion in subtype-2 cancers, GSEA analysis identified the expression of putative ETS target genes to be a feature characterizing subtype-2 cases. Recent studies have found *TMPRSS2-ERG* gene fusions to be associated with aggressive tumor features (6, 29) and prostate cancer death in a watchful waiting cohort (31), although we did not find similar associations in the present cohort (ref. 28; possibly due to patient selection and limited follow-up). Nevertheless, the association of *TMPRSS2-ERG* fusion with adverse clinical features in other studies is consistent with our finding that tumors with subtype-2 expression features (*MUC1+*, *AZGP1-*) show higher rates of tumor recurrence after prostatectomy (9). Additional studies should clarify the relationship of tumor subtype-2 and *TMPRSS2-ERG* fusion with clinical outcome. Of note, Tomlins et al. (32) recently described an inverse relation between the expression of ETS genes and genes at 6q21. While distinct from the peak deletion we found at 6q15, their observation may well reflect the distinct CNAs we report here that characterize subtype-1 and subtype-2 tumors.

Therapy-naïve pelvic lymph node metastases, which classify predominantly with subtype-3, exhibit overall higher frequencies of CNA, suggestive of increased genomic instability, and specific CNAs including gain at 8q (*MYC*) and 16p13 and loss at 10q23 (*PTEN*) and 16q23. However, because no matched primary tumors are available from these cases (the prostatectomy procedure is aborted once pelvic lymph node metastases are discovered), we cannot distinguish whether this pattern of CNAs is characteristic of a distinct class of highly aggressive primary tumors with a strong propensity to metastasize early in their course, whether it reflects genetic changes acquired by subtype-3 primaries when they



**Figure 3.** GSEA links CNAs to subtype gene-expression patterns. GSEA identifies enrichment of genes (A) with putative ETS binding sites in subtype-2 versus others, and (B) of MYC-up-regulated targets in lymph node metastases versus others. Enrichment is evidenced by the early positive deflection of the Kolmogorov-Smirnov running sum. The significance of the maximum running sum (S) is evaluated by comparison with 500 trials with randomly permuted subtype class labels; the P value is the frequency that S in the actual data is equaled or exceeded in the permuted data.

become metastatic, or whether it encompasses one or more currently undefined subclasses. Regardless, frequent 8q24 (*MYC*) amplification with enrichment of MYC-up-regulated target transcripts in the pelvic lymph node metastases underscores the likely pathogenic relevance of deregulated *MYC* expression in this subgroup. Deregulated *MYC* expression drives cell-cycle progres-

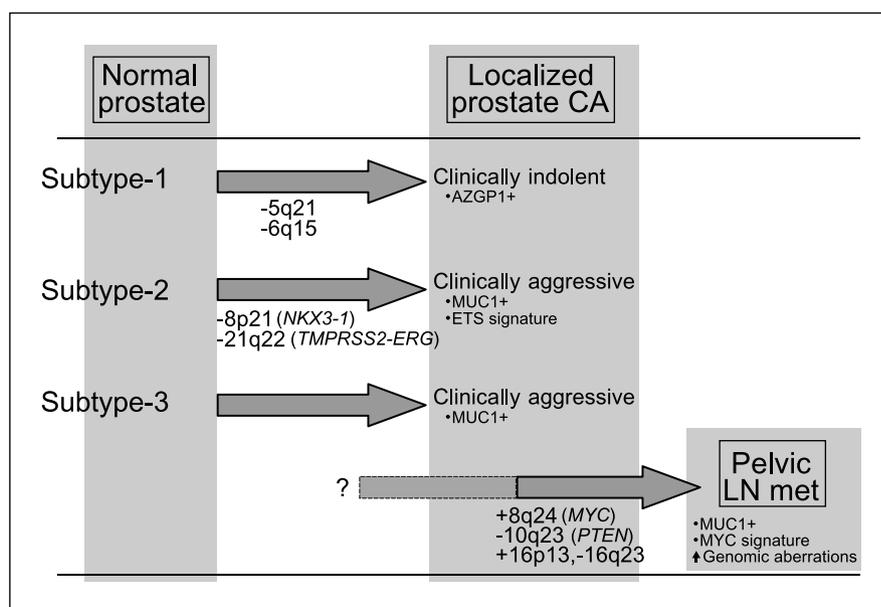
sion and has been linked to genomic instability (33), both prominent features of the tumors in the pelvic lymph node metastasis subset.

Our finding that the more aggressive subtypes-2 and -3 exhibit significantly lower frequencies of 5q21 and 6q15 deletion challenges the prevailing view of a linear pathway of prostate cancer progression. Rather, our data support the existence of alternative parallel pathways of tumorigenesis (Fig. 4), where loss of TSGs at 5q21 and 6q15 generally leads to the development of what seem to be more clinically favorable prostate cancers, whereas deletion at 8p21 (*NKX3-1*) and 21q22 (*TMPRSS2-ERG* fusion), among others, result in more aggressive cancers. Because some lymph node metastases do show loss on chromosomes 5q and 6q, it is quite possible that some subtype-1 cancers can progress to metastasis; however, given the relatively better prognosis of subtype-1 cancers, such progression would be expected to occur less frequently.

We found DNA deletions to be overall more frequent than DNA gains, consistent with most prior chromosome-based CGH studies (7) and indicating a central role of TSGs in prostate cancer initiation. Our data also suggest possible synergies between specific prostate cancer genes or loci. For example, subtype-2 tumors are characterized by concomitant deletion at 8p21 and 21q22, suggesting that the *NKX3-1* tumor suppressor (or another TSG at 8p21) and *TMPRSS2-ERG* fusion might cooperate to affect prostate tumorigenesis. Likewise, subtype-1 tumors exhibit frequent co-deletion of the 5q21 and 6q15 loci, implying cooperating TSGs at these loci. Although no known TSGs map to these latter loci, candidates with known functions relevant to tumor suppressor activity include, at 5q21, *RGMB* (cell adhesion), *CHD1* (chromatin remodeling), and *EFNA5* (cell-cell signaling); and at 6q15, *PNRC1* (nuclear receptor coactivation), *ANKRD6* (WNT signaling and cell polarity), *CASP8AP2* (apoptosis), and *MAP3K7* (transforming growth factor  $\beta$  signal transduction). Genomic profiling of additional subtype-1 specimens should enable further refinement of deletion boundaries to facilitate the identification of the TSGs.

Our discovery of prostate cancer subtypes associated with distinct patterns of both gene-expression profiles and genomic

**Figure 4.** Alternative genetic pathways of prostate tumorigenesis. Schematic model synthesizing major findings from the current study. See text for details; aberrations shared among subtypes, like deletions at 12p13 (*CDKN1B*) and 13q14 (*RB1*), are not shown.



CNAs are reminiscent of recent findings in other tumor types. In particular, breast cancer has been subdivided into five different gene-expression subtypes (Luminal-A, Luminal-B, Basal-like, ERBB2-associated, and normal-like), associated with distinct prognoses (34) and, more recently, different underlying patterns of genomic CNA (17, 35). Our findings indicate that prostate cancer can similarly be usefully considered as comprising distinct gene expression and genetic subtypes, with implications for diagnostic classification and potentially for patient management.

The CNAs we have identified also represent novel candidate biomarkers of prognosis that could be tested clinically. In particular, studies of biopsy material from patients followed but not treated (i.e., watchful waiting cohorts), or from patients who have been treated when their tumors were localized, are needed to determine if subtype markers might effectively predict which tumors progress. Discriminatory subtype markers might include gene-expression signatures, selected IHC stains, genetic markers (e.g., 5q21/6q15 deletion), or combinations of these. Of particular importance is the identification of markers (AZGP1 expression, 5q21/6q15 deletion) that define what seem to be less aggressive cancers. These markers offer promise to better identify tumors of

low risk that can be watched when used in combination with markers of poor prognosis (e.g., MUC1 or EZH2 protein expression, *TMPRSS2-ERG* fusion, and chromosome 8p loss; refs. 9, 31, 36, 37). Although additional investigations are needed, the current study establishes the existence of clinically relevant molecular genetic subtypes of prostate cancer arising via alternative genetic pathways. These subtypes provide a new framework for understanding prostate cancer biology and may explain in part the observed heterogeneity of the disease with regard to both tumor progression and response to therapies.

## Acknowledgments

Received 2/19/2007; revised 6/23/2007; accepted 7/5/2007.

**Grant support:** NIH grants CA97139 (J.R. Pollack), CA111782 (J.D. Brooks), CA85129 (J.R. Pollack and J.D. Brooks), and GM07365 (K. Salari); Swedish Cancer Society, Radiumhemmet's Research Foundations and Cancer and Allergy Foundation (C. Li).

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.

We thank the Stanford Functional Genomic Facility for microarray manufacture, the Stanford Microarray Database for database support, and Ilana Galperin (Stanford Cytogenetics Laboratory) for assistance with FISH analysis. We also thank the members of the Pollack lab for helpful discussion.

## References

- Jemal A, Tiwari RC, Murray T, et al. Cancer statistics, 2004. *CA Cancer J Clin* 2004;54:8–29.
- Klotz L. Active surveillance versus radical treatment for favorable-risk localized prostate cancer. *Curr Treat Options Oncol* 2006;7:355–62.
- DeMarzo AM, Nelson WG, Isaacs WB, Epstein JI. Pathological and molecular aspects of prostate cancer. *Lancet* 2003;361:955–64.
- Lee WH, Morton RA, Epstein JI, et al. Cytidine methylation of regulatory sequences near the pi-class glutathione *S*-transferase gene accompanies human prostatic carcinogenesis. *Proc Natl Acad Sci U S A* 1994;91:11733–7.
- Tomlins SA, Rhodes DR, Perner S, et al. Recurrent fusion of *TMPRSS2* and *ETS* transcription factor genes in prostate cancer. *Science* 2005;310:644–8.
- Wang J, Cai Y, Ren C, Iltmann M. Expression of variant *TMPRSS2/ERG* fusion messenger RNAs is associated with aggressive prostate cancer. *Cancer Res* 2006;66:8347–51.
- Karan D, Lin MF, Johansson SL, Batra SK. Current status of the molecular genetics of human prostatic adenocarcinomas. *Int J Cancer* 2003;103:285–93.
- Grossmann ME, Huang H, Tindall DJ. Androgen receptor signaling in androgen-refractory prostate cancer. *J Natl Cancer Inst* 2001;93:1687–97.
- Lapointe J, Li C, Higgins JP, et al. Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc Natl Acad Sci U S A* 2004;101:811–6.
- Schuler GD. Pieces of the puzzle: expressed sequence tags and the catalog of human genes. *J Mol Med* 1997;75:694–8.
- Pollack JR, Perou CM, Alizadeh AA, et al. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet* 1999;23:41–6.
- Pollack JR, Sorlie T, Perou CM, et al. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc Natl Acad Sci U S A* 2002;99:12963–8.
- Gollub J, Ball CA, Binkley G, et al. The Stanford Microarray Database: data access and quality assessment tools. *Nucleic Acids Res* 2003;31:94–6.
- Wang P, Kim Y, Pollack J, Narasimhan B, Tibshirani R. A method for calling gains and losses in array CGH data. *Biostatistics* 2005;6:45–58.
- Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 1998;95:14863–8.
- Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 2001;98:5116–21.
- Bergamaschi A, Kim YH, Wang P, et al. Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer. *Genes Chromosomes Cancer* 2006;45:1033–40.
- Lingjaerde OC, Baumbusch LO, Liestol K, Glad IK, Borresen-Dale AL. CGH-Explorer: a program for analysis of array-CGH data. *Bioinformatics* 2005;21:821–2.
- Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545–50.
- Kim YH, Girard L, Giacomini CP, et al. Combined microarray analysis of small cell lung cancer reveals altered apoptotic balance and distinct expression signatures of MYC family gene amplification. *Oncogene* 2006;25:130–8.
- Zeller KI, Jegga AG, Aronow BJ, O'Donnell KA, Dang CV. An integrated database of genes responsive to the Myc oncogenic transcription factor: identification of direct genomic targets. *Genome Biol* 2003;4:R69.
- Whitfield ML, Sherlock G, Saldanha AJ, et al. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell* 2002;13:1977–2000.
- Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Barrette TR, Ghosh D, Chinnaiyan AM. Mining for regulatory programs in the cancer transcriptome. *Nat Genet* 2005;37:579–83.
- Kel AE, Gossling E, Reuter I, Cheremushkin E, Kel-Margoulis OV, Wingender E. MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res* 2003;31:3576–9.
- Matys V, Fricke E, Geffers R, et al. TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* 2003;31:374–8.
- Graves BJ, Petersen JM. Specificity within the ets family of transcription factors. *Adv Cancer Res* 1998;75:1–55.
- Kim H, Lapointe J, Kaygusuz G, et al. The retinoic acid synthesis gene *ALDH1a2* is a candidate tumor suppressor in prostate cancer. *Cancer Res* 2005;65:8118–24.
- Lapointe J, Kim YH, Miller MA, et al. A variant *TMPRSS2* isoform and *ERG* fusion product in prostate cancer with implications for molecular diagnosis. *Mod Pathol* 2007;20:467–73.
- Perner S, Demichelis F, Beroukhim R, et al. *TMPRSS2:ERG* fusion-associated deletions provide insight into the heterogeneity of prostate cancer. *Cancer Res* 2006;66:8337–41.
- Henshall SM, Horvath LG, Quinn DI, et al. Zinc- $\alpha$ -glycoprotein expression as a predictor of metastatic prostate cancer following radical prostatectomy. *J Natl Cancer Inst* 2006;98:1420–4.
- Demichelis F, Fall K, Perner S, et al. *TMPRSS2:ERG* gene fusion associated with lethal prostate cancer in a watchful waiting cohort. *Oncogene* 2007;26:4596–9.
- Tomlins SA, Mehra R, Rhodes DR, et al. Integrative molecular concept modeling of prostate cancer progression. *Nat Genet* 2007;39:41–51.
- Felsner DW, Bishop JM. Transient excess of MYC activity can elicit genomic instability and tumorigenesis. *Proc Natl Acad Sci U S A* 1999;96:3940–4.
- Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001;98:10869–74.
- Chin K, DeVries S, Fridlyand J, et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiology. *Cancer Cell* 2006;10:529–41.
- Varambally S, Dhanasekaran SM, Zhou M, et al. The polcomb group protein EZH2 is involved in progression of prostate cancer. *Nature* 2002;419:624–9.
- Paris PL, Andaya A, Fridlyand J, et al. Whole genome scanning identifies genotypes associated with recurrence and metastasis in prostate tumors. *Hum Mol Genet* 2004;13:1303–13.