Human Molecular Genetics, 2004, Vol. 13, No. 13 **1303–1313**
DOI: 10.1093/hmg/ddh155
Advance Access published on May 11, 2004

Whole genome scanning identifies genotypes associated with recurrence and metastasis in prostate tumors

Pamela L. Paris[1], Armann Andaya[1], Jane Fridlyand[1], Ajay N. Jain[1], Vivian Weinberg[1], David Kowbel[1], John H. Brebner[1], Jeff Simko[1,2], J.E. Vivienne Watson[1], Stas Volik[1], Donna G. Albertson[1], Daniel Pinkel[1], Janneke C. Alers[3], Theodorus H. van der Kwast[3], Kees J. Vissers[3], Fritz H. Schroder[4], Mark F. Wildhagen[4], Phillip G. Febbo[5,8], Arul M. Chinnaiyan[6], Kenneth J. Pienta[6], Peter R. Carroll[7], Mark A. Rubin[8], Colin Collins[1,*] and Herman van Dekken[3]

[1]Comprehensive Cancer Center and [2]Department of Anatomic Pathology, University of California at San Francisco, San Francisco, CA 94115, USA, [3]Department of Pathology and [4]Department of Urology, Erasmus Medical Center, University Medical Center Rotterdam, Rotterdam, The Netherlands, [5]Dana Farber Cancer Institute, Harvard Medical School, Boston, MA 02115, USA, [6]Department of Pathology, University of Michigan Medical School, Ann Arbor, MI 48109, USA, [7]Department of Urology, Comprehensive Cancer Center, University of California at San Francisco, San Francisco, CA 94115, USA and [8]Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA

Received February 4, 2004; Revised and Accepted April 28, 2004

Prostate cancer is the most commonly diagnosed non-cutaneous neoplasm among American males and is the second leading cause of cancer-related death. Prostate specific antigen screening has resulted in earlier disease detection, yet ~30% of men will die of metastatic disease. Slow disease progression, an aging population and associated morbidity and mortality underscore the need for improved disease classification and therapies. To address these issues, we analyzed a cohort of patients using array comparative genomic hybridization (aCGH). The cohort comprises 64 patients, half of whom recurred postoperatively. Analysis of the aCGH profiles revealed numerous recurrent genomic copy number aberrations. Specific loss at 8p23.2 was associated with advanced stage disease, and gain at 11q13.1 was found to be predictive of postoperative recurrence independent of stage and grade. Moreover, comparison with an independent set of metastases revealed approximately 40 candidate markers associated with metastatic potential. Copy number aberrations at these loci may define metastatic genotypes.

INTRODUCTION

Prostate cancer is the most commonly diagnosed non-cutaneous neoplasm among males in Western countries and is estimated to result in 28 900 deaths this year in the US alone (1). The advent of widespread prostate specific antigen (PSA) screening has resulted in increased detection of prostate cancer at earlier stages. A persistent and recalcitrant problem is that men with similar stage tumors often exhibit markedly different clinical outcomes following therapy (i.e. surgery or radiation). Early detection combined with slowly progressing tumors means a significant subset of men may be candidates for watchful waiting or active surveillance rather than treatment, and this will become increasingly important as the population ages (2). Thus, it is imperative that new methods be developed for patient stratification based on risk of recurrence to enable appropriate patient management.

*To whom correspondence should be addressed at: PO Box 808, University of California at San Francisco, San Francisco, CA 94143, USA. Tel: +1 4155027067; Fax: +1 4155025665; Email: collins@cc.ucsf.edu

The development of array comparative genomic hybridization (aCGH) has important implications for analysis of tumor genomes as well as for development of predictive biomarkers and identification of genes involved in tumor progression. aCGH allows very high-resolution quantitative detection of copy number aberrations in tumor genomes (3–8); moreover, associations with clinical outcome can be made (9). Recurrent copy number changes reveal loci encoding tumor suppressors and oncogenes, the identification of which is now facilitated by completion of the human genome sequence and an impressive repertoire of genome annotation tools (10,11). The arrays used in this study contain approximately 2400 bacterial artificial chromosome (BAC) clones and have an average genome-wide resolution of 1.4 Mb (6,12). To maximize the clinical utility of data collected from aCGH experiments, it is necessary to use clinical specimens obtained from patients with substantial follow-up. Thus, we developed a methodology for performing aCGH with DNA extracted from archived prostate tumors that were formalin-fixed and paraffin-embedded (13). To limit the impact of tumor heterogeneity on the sensitivity of aberration detection, we used a novel tumor microdissection method (13) and tumor specific signal thresholding (see Materials and Methods).

In the current study, aCGH was used to analyze 64 tumors from men at intermediate to high risk of recurrence following radical prostatectomy. This cohort comprises 32 patients who biochemically progressed following prostatectomy and 32 who did not. Rising PSA following prostatectomy was used as a biochemical marker of disease recurrence. This unique cohort has a median clinical follow-up of 11 years for non-progressors (ranging from 8 to 15 years), which is longer than the time to recurrence for all progressors (ranging from <1 to 8 years), thereby increasing our confidence in outcome classification.

Our previous work demonstrated that analysis of archived prostate tissue by aCGH is capable of detecting single copy changes (13). In the present study, aCGH was performed to identify genomic profiles capable of distinguishing indolent from aggressive tumors and loci linked to tumor progression. Because aCGH is performed on arrayed BAC clones with known genome 'addresses', it is theoretically possible to identify multiple BAC-based markers of disease progression (9,14–17). Moreover, because the array data can be integrated with underlying genome annotations (10,11), it is straightforward to identify candidate tumor suppressors and oncogenes encoded at loci associated with tumor progression and/or clinical outcome such as metastasis and response to therapy.

To extend this study, we included an independent set of metastases in an exploratory exercise to determine whether markers present in primary tumors might be predictive of occult metastasis or proclivity for metastasis. In addition, their inclusion allows identification of known cancer related genes and/or novel genes that may play a direct role in metastasis, and may aid in defining new therapeutic approaches. Finally, an ability to compare patterns of recurrent copy number changes in non-recurring primary tumors, primary tumors that metastasize and metastatic tumors may provide important insights into the evolution of prostate cancer.

RESULTS

Recurrent copy number aberrations

Tumor based thresholds were calculated for all samples. Theoretically a log2ratio of 0.5 represents a single copy gain and a log2ratio of -1 corresponds to loss of one copy. However, a number of factors impact on this theoretical value, which include the amount of contaminating normal tissue and stroma. The log2ratio thresholds ranged from an absolute value cut-off of 0.19–0.52, with an average of 0.34. A subset of 10 samples was also analyzed with CGH to metaphase chromosomes, and the two techniques were concordant (13).

The overall frequency of copy number changes in the cohort of 64 primary tumors is shown in Figure 1A. The most frequent gains ($>40\%$) in both groups include 11p15.4 (66%), 2p25.1 (60%), 13q34 (60%), 11q13.1 (52%) and 2p22.1 (45%). Frequently ($>40\%$) lost loci were 8p21.2 (46%) and 8p23.2 (45%). Based on the July 2003 freeze of the UCSC human genome browser (http://genome.ucsc.edu/cgi-bin/hgGateway), the genomic position of these copy number aberrations correspond to: 2p22 (37891432–39128299), 2p25 (9619095–11073793), 11p15 (9238525–11610583), 11q13 (63810754–66394362), 13q34 (109071421–111904343), 8p21.2 (19764266–25211627) and 8p23 (2080710–4357590).

As expected, tumors that progressed had significantly more aberrations (Wilcoxon rank sum P-value, $P = 0.006$) than those that did not. The median value for the aberrations for the non-progressors was 10.5 (range 1–56) and 20.5 for progressors (range 1–90). Figure 1B and C show BAC clones that differed by $\geq 10\%$ in their frequency of copy number gain or loss between the progressors and non-progressors, respectively. The list of BAC clones is supplied in the supplementary material.

Deletion of 8p23 is associated with advanced stage

Deletion of 8p23 was more common in progressors than in non-progressors (50 versus 31%). An association was found between pathological advanced stage disease ($pT \geq 3$) and loss of 8p ($P = 0.0015$). A possible homozygous loss was seen for BAC RP11-112F7 on 8p23.2 (UCSC July 2003 freeze: 3284324–3324954). The deletion with the greatest magnitude corresponded to a log2ratio of -0.670. The minimal region of loss is $\sim$1 Mb and overlaps exons 3–11 of the CUB and Sushi multiple domains 1 (*CSMD1*) gene. A TaqMan primer-probe set was designed for *CSMD1*. On a panel of eight RNAs (six pT2, two pT3) from a separate cohort of prostatectomy patients, *CSMD1* showed a marked decrease in expression for the patients of higher stage ($pT \geq 3$) disease (Fig. 2).

Gain at 11q13.1 predicts recurrence independent of stage and grade

Univariate analysis indicated a statistically significant association with BAC CTD-2220I9 on 11q13.1 (UCSC July 2003 freeze: 64313688–64470546) and biochemical failure status, $P < 0.002$. This association was even stronger in the subgroup of 39 samples with negative surgical resection margins.

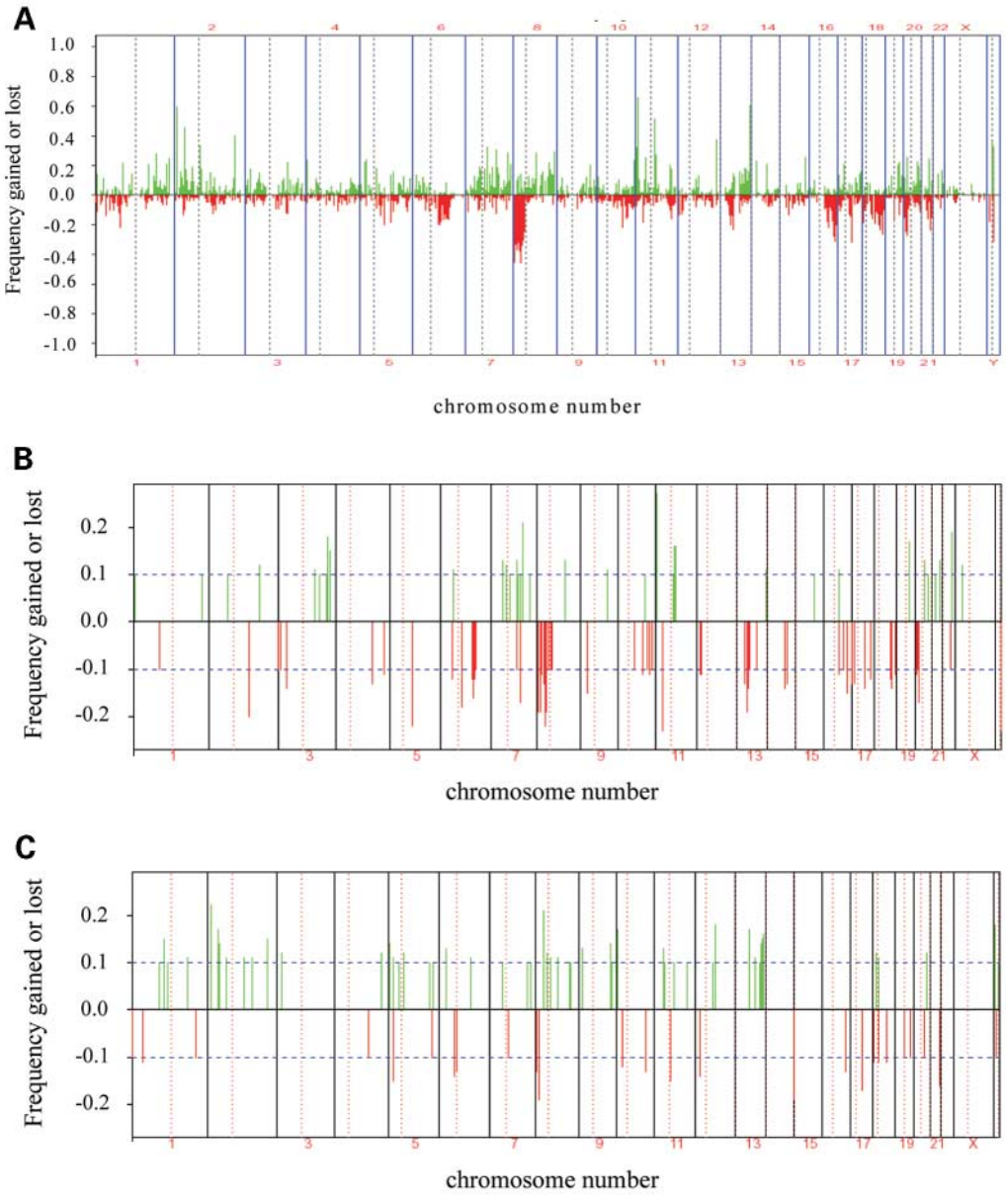

Figure 1. aCGH copy number frequency plots for 32 non-progressing primary tumors and 32 primary tumors that progressed. (**A**) Combined copy number changes for all 64 primary tumors. Gains are shown in red and losses in green. The vertical dashed lines represent the centromeres. The frequency of amplifications and deletions unique to tumors that progress are shown in (**B**) and those unique to non-progressors are shown in (**C**). The blue dashed horizontal lines represent the 10% frequency cut-off. The list of BAC clones is supplied in the Supplementary Material.

Importantly, the 11q13.1 biomarker retained its significance when adjusted for the clinical parameters (grade, stage, age at operation, margin and preoperative PSA). Distribution of the log2ratios for that clone in the negative margin cases is shown for progressors and non-progressors in Figure 3.

The minimal region of the 11q13 amplicon is ~600 kb. This region of the genome is gene rich (17 genes and 4790 ESTs represented in 53 Unigene clusters). The candidate genes that overlap with BAC CTD-2220I9 are *MAP4K2* (mitogen-activated protein kinase kinase kinase kinase 2), *MEN1* (multiple endocrine neoplasia I), *SF1* (splicing factor 1), *PPP2R5B* [protein phosphatase 2, regulatory subunit B (B56), beta isoform], *NAALADASEL* (*N*-acetylated alpha-linked acidic dipeptidase-like) and *EHD1* (EH-domain containing 1). The newly available Oncomine cancer expression database was queried for each of these six genes to prioritize candidate genes (18). There was no difference in *SF1* and *EHD1* expression levels for radical prostatectomy patients based on PSA recurrence (19). In primary and metastatic tumors, there was no difference in expression for *NAALADASEL*, and a decrease in *PPP2R5B* expression for metastatic tumors was observed (20,21). Only *MEN1* and *MAP4K2* showed a trend towards an increase in expression for progressors versus non-progressors (22). Real-time expression

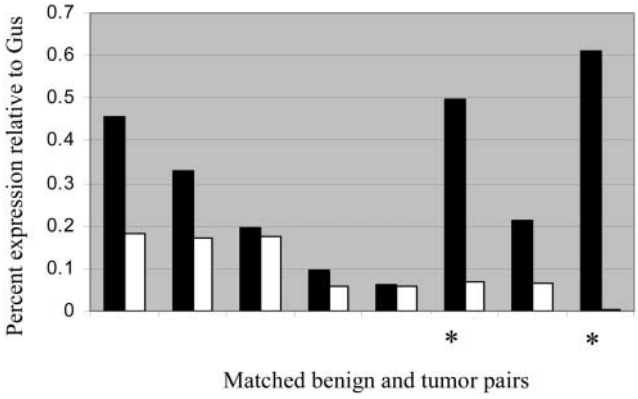

Figure 2. Real-time RT–PCR expression results for the *CSMD1* gene at 8p23. Matched benign and tumor pairs were used in this study and are indicated by black and white bars, respectively. Each tissue sample was run in triplicate. The SD for the cycle threshold values of all three replicates was less than 0.3. Results are displayed relative to a GUS reference gene. An asterisk denotes samples that are stage pT $\geq$ 3.

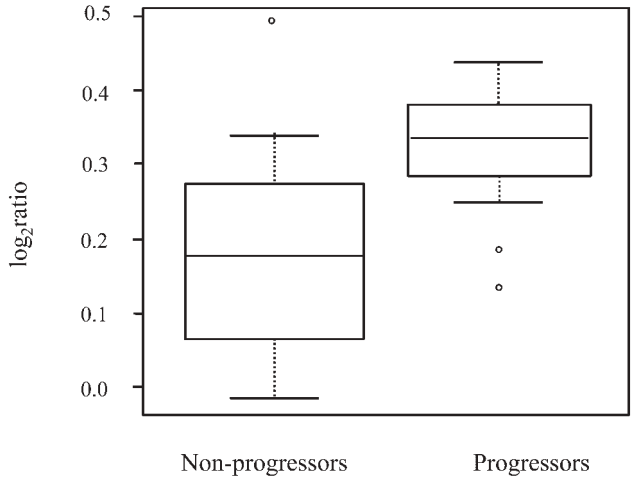

Figure 3. Box plot for BAC CTD-2220I9 in negative margin cases. The boxplot compares the distribution of the log2ratios of CTD-2220I9 between progressors and non-progressors for those with negative surgical margins. The solid horizontal lines are first, second (median) and third quartiles, respectively. The whiskers extend to 1.5 SD away from the median where SD is a distribution of the value in a subgroup for this clone. The outlying points indicate outliers (>1.5 SD away from the median).

analysis was performed for both MEN1 and MAP4K2. On a panel of 10 RNAs from a separate set of radical prostatectomy patients, only *MEN1* showed increased expression (Fig. 4). Four of the five cases where *MEN1* was upregulated also showed an increase in copy number at 11q13 by aCGH. In one tumor MEN1 was overexpressed despite normal copy number.

Identification of candidate markers of metastasis

The clinical information for the progressors consisted of recurrence type, which led us to look for BAC based predictors of distant metastases. A set of organ metastases was used to identify copy number changes that confer a more aggressive phenotype on primary tumors. Figure 5 shows the result of the analysis of copy number changes in primary tumors that ultimately metastasized and organ metastases versus non-progressors. Only those BAC clones with a $P \leq 0.05$ between primary tumors that metastasized, organ metastases and non-progressing primary tumors are shown. Approximately 40 loci were identified that were infrequently (0–20%) altered in primary tumors that did not progress. In contrast, aberrations at these loci were frequent in primary tumors that metastasized (20–45%) and organ metastases (20–90%). It is noteworthy that six BAC clones were never aberrant in the non-progressor cohort ($N_{max} = 32$) but were (20–30%) in both metastatic cohorts.

DISCUSSION

To maximize both biological information and clinical correlates, tumor based thresholds were used for the determination of aCGH gains and losses. Using spot checks, we confirmed that samples with the smallest threshold corresponded to aCGH data with extremely good signal-to-noise and vice versa for the sample with the largest threshold value. The tumor based threshold method allowed data with varying signal-to-noise ratios to be compared with one another. It should be noted that when a BAC was gained in one cohort it was rarely seen to also be deleted in the same cohort. This demonstrates the good signal to noise obtained with our whole genome aCGH technique.

Previously reported and frequently changed loci in prostate cancer +2p22.1 (23), +11q13.1 (24), −8p21.2 (25), −8p21.3 (26) and −8p23.2 (27) were also identified in this study, and often at higher resolution. Generally, the progressors exhibited a higher frequency of change for these loci. Newly defined amplicons in prostate tumors include 2p25.1, 11p15.4 and 13q34. We observed the expected wide range of inter-tumor heterogeneity in copy number aberration size at these loci. This phenomenon is well known and may reflect utilization of different fragile sites, independent mechanisms of aberration formation and/or biological selection. In this study BAC clones at the 8p23 deletion and 11q13 gain were identified computationally as having associations with the clinical phenotypes of tumor stage and recurrence, respectively. Individual chromosome specific profiles were then used to define minimum recurrent aberrations identifying MEN1 and CSMD1. Future work with the new sub-megabase resolution contig arrays is expected to further narrow these regions (28). Many retroviruses induce tumors by insertion of their viral DNA adjacent to oncogenes, resulting in altered expression. Known retroviral integration sites in murine tumors can be used as a genetic screen for the identification of candidate cancer genes (29). A search was performed for common retroviral insertion sites using the retroviral tagged cancer gene database (http://genome2.ncifcrf.gov/RTCGD). The *MRVI1* gene (murine retrovirus integration site 1 homolog) mapping to the 11p15 amplicon contains three common retroviral insertion sites. The 11q13 locus contains several known viral insertion sites that occur within *MAP4K2*, *MEN1* and *RASGRP2* (RAS guanyl releasing protein 2). Quantitative RT–PCR studies and expression

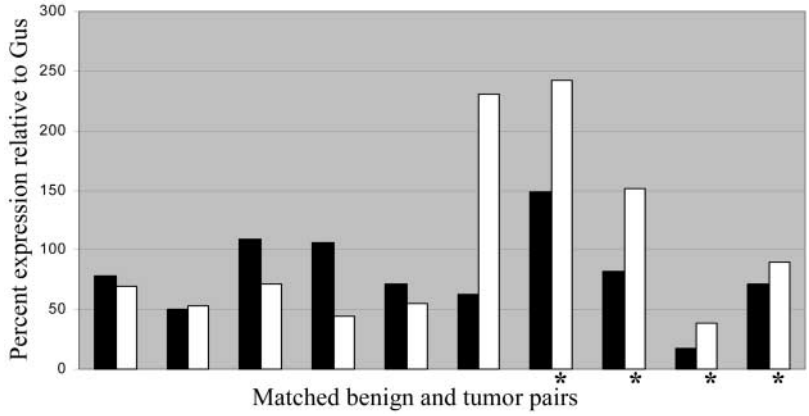

Figure 4. Real-time RT–PCR expression results for the *MEN1* gene at 11q13. Matched benign and tumor pairs were used in this study and are indicated by black and white bars, respectively. Each tissue sample was run in triplicate. The SD for the cycle threshold values of all three replicates was less than 0.3. Results are displayed relative to a GUS reference gene. An asterisk denotes samples that showed genomic gain at 11q13 by aCGH.

database mining are underway to evaluate the expression levels of genes mapping to these frequently altered loci in order to obtain information regarding the etiology of prostate cancer.

Deletions along 8p are common in prostate cancer (25–27). Whole arm deletion of 8p strongly associated with higher pathologic stage disease in our study. In a recent prostate study, 8p was found to be the most valuable predictor of stage (30). Our findings confirm and extend these previous studies. The identified 8p biomarkers may aid in therapy determination at the time of a biopsy. The deleted 8p23 BAC clones on the genomic array overlap with a single gene called *CSMD1* (31,32). This is the first report of CSMD1 undergoing deletion in prostate cancer. This finding is supported by prostate cancer expression microarray experiments that found CSMD1 decreased expression to be associated with relapse and survival (33). The function of this protein is currently unknown. Sushi domains exist in adhesion proteins, and therefore make *CSMD1* a likely target for deletion by an aggressive tumor. In a recent aCGH study of 14 fallopian tumors, 12 tumors showed deletion involving the *CSMD1* region (8). The minimal region of recurrent loss in their study directly overlapped ours (RP11-82K8 to RP11-140K14). TaqMan results, in a separate cohort of patients, provided further evidence for a decrease in *CSMD1* expression in higher stage ($pT \geq 3$) prostate tumors. In addition to being a marker of advanced stage disease, deletion of 8p23.2 may be a marker of disease recurrence and therefore warrants future studies. There is considerable evidence implicating the *NKX3.1* gene at 8p21 as a tumor suppressor gene in prostate cancer (34–36). *CSMD1* and *NKX3.1* map ~18 Mb apart in non-contiguous deletions, and thus may represent independent tumor suppressor genes.

Previous studies have identified loci that associate with aggressive behavior for prostate cancer (37–43); however, very few genes have been identified to date. Amplifications on 11q13 have been reported in other cancers (44–47), but rarely in prostate cancer (24,48). We have significantly narrowed the region on 11q13 identified by El Gedaily *et al.* (24) and Kasahara *et al.* (48) in advanced prostate cancer cases. In our intermediate and high risk of recurrence cohort, we identified a BAC clone mapping to 11q13.1 that showed a statistically significant increase in copy number in tumors from patients who failed following radical prostatectomy as compared with those who did not recur and that was an independent predictor of recurrence. MEN1 maps to this locus and recent prostate cancer gene expression profiling experiments identified elevated expression of MEN1 to be associated with recurrence (49). Our real-time expression analysis on RNAs from prostatectomy patients showed an increase in expression for *MEN1* for several cases, all but one exhibited corresponding genomic gain for the 11q13 locus. The exception is interesting because it implies mechanisms independent of amplification can result in increased MEN1 expression in prostate cancer. There was no correlation between increased expression and stage or grade, possibly indicating that this marker acts independent of those clinical parameters. Future work will evaluate *MEN1* as a biomarker of disease recurrence after radical prostatectomy. *MEN1* may be a surrogate biomarker or act as an oncogene.

Genes that have been implicated in pathways involving prostate cancer were also found to lie in regions of genomic gain and loss in this study. The oncogene *MYC* showed genomic gain in the progressors, and at an even higher frequency in the metastatic cohort, as compared with the non-progressors. The tumor suppressor *RB1* was shown to be more frequently deleted in the progressors, and at an even a higher frequency in the metastatic tumors than the non-progressors. Apoptosis of prostate cancer cells wild-type for Rb has been shown to occur by means of an intracellular pathway that involves the activation of Rb and repression of MYC transcription (50). We observed that the combination of +8q24.21/−13q14.2 occurred in 40% of the metastases suggesting future work is warranted regarding this pathway in prostate cancer.

Recent work in the gene expression field reported that a subset of primary solid tumors share the gene expression signature of their corresponding organ metastases (51). We propose that genomic changes in metastatic tumors can

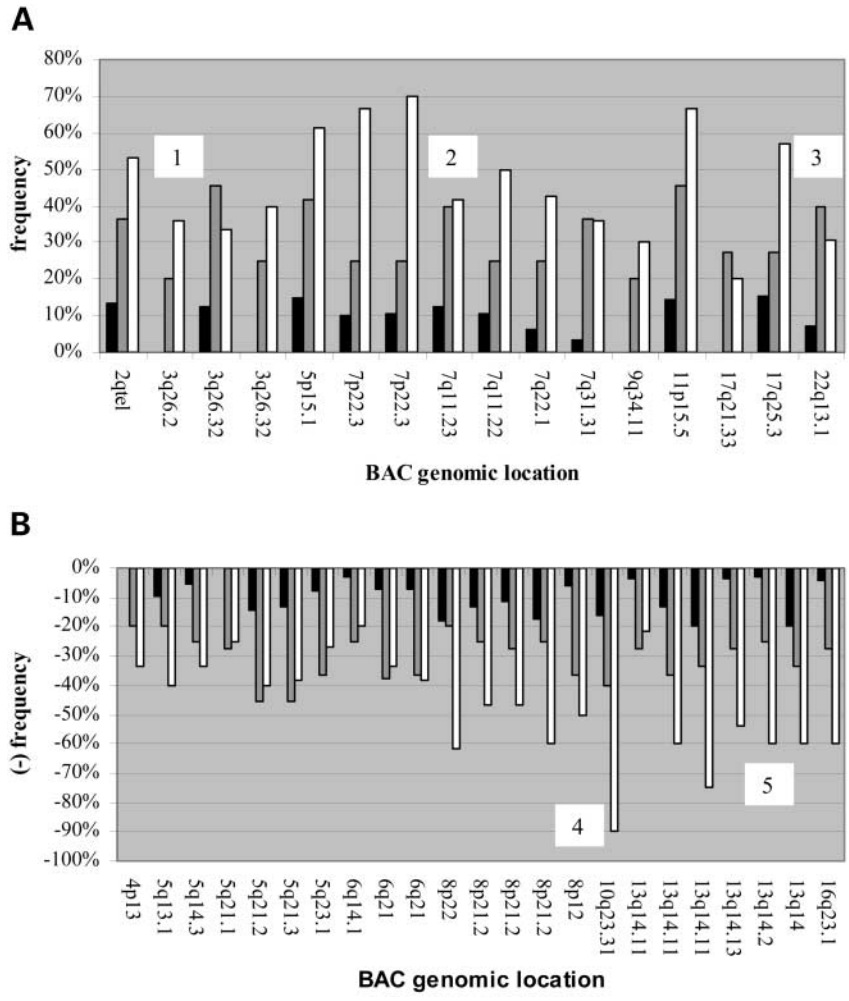

Figure 5. A set of 39 candidate BAC biomarkers associated with metastasis. Black bars represent tumors from patients ($N_{max} = 32$) who did not progress. Gray bars correspond to primary tumors from patients ($N_{max} = 12$) who progressed to metastasis. Tumors represented by the black and gray bars are from a single cohort, whereas white bars represent an independent cohort of metastatic tumors ($N_{max} = 15$). Copy number changes [(**A**) gains, (**B** losses)] are reported only if they occur in each metastatic cohort at a frequency of $\geq$20% and in the non-progressor cohort $<$20% and were statistically significant ($P \leq 0.05$). 1, *EVI1* locus; 2, *LIMK1* locus; 3, *PDGFB* locus; 4, *PTEN* locus; 5, *RB* locus. The list of BAC clones is supplied in Supplementary Material.

guide identification of the most important genomic changes in primary tumors. This should be especially useful in slow growing, highly heterogeneous tumors, such as prostate cancer. In addition, prostate cancer cells exhibit highly heterogeneous genomic profiles. Metastatic prostate tumors that have evolved further and are more homogenous can help elucidate which genetic changes confer aggressive phenotypes in primary tumors. Pattern recognition analysis identified a combination of BAC clones that may be utilized as biomarkers for predicting metastasis at the time of biopsy or surgery, and therefore assist in the identification of patients who would benefit from the use of adjuvant therapy. For example, the BAC gained at 22q13.1 in Figure 5A maps to platelet derived growth factor beta, *PDGFB*. This is intriguing since the receptors for *PDGF* have been shown to be expressed in advanced prostate cancer (52). The beta receptor in particular has been shown recently to serve as a recurrence predictor in a five-gene model (19). *LIMK1* (LIM domain kinase 1), mapping to 7q11.23 in Figure 5A, has been shown recently to be overexpressed in prostate tumors and metastatic cell lines (53). In the same report, partial reduction in *LIMK1* was shown to abolish the metastatic invasiveness of prostate cells *in vitro* (53). Additionally, the tumor suppressor *PTEN* maps to the BAC identified at 10q23.1 in the exploratory biomarker analysis shown in Figure 5B. Prostate specific deletion of the murine *Pten* tumor suppressor gene has been shown to lead to metastatic prostate cancer (54). At this stage the clinical implications of these markers is purely speculative; however, given their potential impact on patient care, further study clearly is warranted.

A significant strength of aCGH is its ability to map expeditiously and quantitatively both copy number gains and single copy losses at multiple independent loci in clinical specimens. This study has significantly expanded the catalog of recurrent aberrations found in prostate cancer and this may enable a deepened understanding of its etiology and progression. Amplification at 11q13.1 is predictive of postoperative disease recurrence, deletion at 8p23 is strongly associated

with advanced disease stage and candidate genes have been identified at each locus and these may form the basis for future diagnostics and therapies pending independent clinical validation. Genomic profiles were obtained from organ metastases and used to interrogate the genomic profiles of primary tumors for genome aberrations associated with metastasis. This exploratory analysis yielded a large number (~40) of biomarkers that may define metastatic genotypes. An array containing these BAC clones is being tested currently in a blinded fashion to assess their predictive value using prostate tumors from patients with known outcome. Our work suggests that studies that examine only primary tumors may be missing the key players in metastatic disease. Studies that combine primary tumors with metastatic tumors may lead to more effective therapeutics and more accurate diagnostics. The clinical significance of the putative biomarkers described in this study will require confirmation in independent cohorts and ultimately, prospective studies.

MATERIALS AND METHODS

Patients

Prostatectomy patients were selected retrospectively from Erasmus University Medical Center in The Netherlands. The cohort consists of 64 prostate cancer patients who were either at intermediate or at high risk of recurrence at diagnosis (55). Following surgery, PSAs were monitored every 3 months during the first year, bi-annually in the second year, followed by yearly monitoring. Of these patients 32 never had a biochemical failure following surgery (PSA < 0.2 ng/ml) or any other evidence of disease recurrence. The median follow-up for the non-progressors was 11 years. The other 32 patients failed biochemically. In this study, a biochemical relapse was defined as: (a) two consecutive PSA serum levels ≥ 0.2 ng/ml with an interval of at least 3 months followed by an elevated PSA (≥ 0.2 ng/ml); or (b) a single observation of PSA > 1 ng/ml followed by an elevated PSA (≥ 0.2 ng/ml). PSA levels ≥ 0.2 ng/ml occurring in the first 3 months after radical prostatectomy were not considered a biochemical relapse if followed by undetectable (< 0.1 ng/ml) PSA levels. The progression-free survival was defined as the interval between the time of surgery and the first elevated PSA serum level (≥ 0.2 ng/ml). Other clinical parameters, such as Gleason score, pathological stage, age at operation, pre-operative PSA and surgical margin status are listed in Table 1 for the 64 patients (32 progressors, 32 non-progressors).

In order to study metastatic tumors, 15 hormone refractory, metastatic tumors from the Rapid Autopsy Program from the University of Michigan Prostate SPORE were evaluated (56). Fifteen tissue slices at 15 μm were extracted with a Wizard Genomic DNA Isolation kit (Promega, Madison, WI, USA) according to the manufacturer's protocol. The DNA was further purified by phenol/chloroform extraction, followed by ethanol precipitation.

Tissue processing

All paraffin-embedded formalin fixed prostate tissue blocks were stained with DAPI to outline tumor areas. A bore (1 mm–1 cm in diameter) attached to a microscope was used to punch a few millimeters deep into the selected tumor region. Hematoxylin and eosin (H&E) staining was performed for the first and the last slice corresponding to the punch to ensure the tumor region was consistent from top to bottom.

DNA was extracted using the Puregene DNA isolation kit (Gentra Systems, Minneapolis, MN, USA).

aCGH

The human version 2.0 BAC arrays were provided by the UCSF Array Core. Each array consists of 2460 BAC clones spotted in triplicate on chromium slides. The resolution is ~1.4 Mb. The aCGH protocol that was followed is detailed in our recent aCGH archived tissue technique paper (13). Also included in this reference are details regarding the in-house imaging system and software that was used to process the arrays.

Statistical analysis

The tumor/reference fluorescence intensity ratios were converted to the $\log_2$ domain. The observed log2ratios were not included if there were fewer than two replicate spots (out of three) or if the standard deviation of the replicates was above 0.2. Each array was normalized to have a median log2ratio of 0. The clones that were present in $<75\%$ of the samples (or 48 samples) were removed from the dataset (348 or 13% of the clones); 2127 clones remained in the dataset.

To identify the gained and lost clones in individual samples, we constructed sample specific thresholds (57). The clones with log2ratios above or below a tumor's threshold were considered gained or lost, respectively. To calculate thresholds, we used discrete-time hidden Markov model (58) to segment clones on individual chromosomes into the states corresponding to underlying copy numbers. The number of states was determined with the BIC (62) criterion. In order to increase robustness of the procedure, we used only those chromosomes that contained less than three different states and only those states that contained at least 20 clones. We assumed that experimental measurement error is independent of the underlying copy number. Thus, for each state and chromosome that met the above criteria, we calculated the median absolute deviation (MAD) of the log2ratios of the clones on that chromosome belonging to a given state. The final estimate of the standard deviation of the experimental noise, SD, was then calculated as the median of the above MAD values across all used states and chromosomes. Finally, the thresholds were calculated conservatively as $2.5 \times$ SD for a given tumor. The *ad hoc* justification for using this threshold lies in considering the standard normal distribution (1.2% of the standard normals are expected to exceed the absolute cut-off of 2.5). The frequency of gains and losses for a given clone in a group of interest was calculated as the proportion of samples in which a clone was gained or lost in that group.

We imputed the missing values (8.8% of the observations) using the *K*-nearest neighbors (KNN) algorithm (63). We computed all pairwise correlations among the clones and assign five closest neighbors to each clone (i.e. the clones most highly correlated with a given clone). Then, if a

Table 1. Clinical characteristics of the cohort

Patient	OperYr	AgeOper	PreOpPSA	Stage	Grade	ResMargin	ProgFreeSurvival	ProgressType
Progressors								
1	1991	61	11.2	2a	5	0	91	PSA
2	1992	63	23.6	2c	7	0	73	PSA
3	1990	67	64.3	3c	7	1	2	M1b
4	1994	61	23.7	4a	7	1	26	M1a
5	1988	68	17.1	3c	7	1	3	PSA
6	1991	71	29.6	4a	7	0	44	LR
7	1994	72	7.1	2c	7	1	23	PSA
8	1988	70	1	4a	7	0	10	LR
9	1988	69	29	3c	7	1	33	M1b
10	1993	65	11.8	3a	7	0	52	PSA
11	1986	65	No data	3b	7	1	38	M1b
12	1987	64	No data	3c	7	1	83	PSA
13	1989	60	6.5	3a	6	0	13	PSA
14	1989	65	5.5	3c	7	1	37	PSA
15	1990	61	21	3b	7	0	3	LR
16	1991	61	7.8	3a	7	1	71	LR
17	1992	66	6.4	3a	7	1	41	PSA
18	1992	70	25.1	4a	7	0	1	PSA
19	1991	69	0.7	2a	6	0	54	LR + M1b
20	1992	63	18.7	4a	7	1	8	LR + M1b
21	1992	53	16.5	3a	7	0	2	LR + M1b
22	1990	55	13.2	4a	7	1	38	LR
23	1990	51	2.8	4a	6	1	16	PSA
24	1992	58	108	4a	5	0	5	LR
25	1992	47	11.5	3a	7	0	2	M1b + M1c
26	1992	59	3.2	3a	7	1	75	M1b
27	1989	59	16.9	4a	7	0	13	M1b
28	1989	74	19.4	3a	7	1	17	PSA
29	1990	65	25.2	3a	7	1	83	PSA
30	1990	71	16.1	3a	7	0	98	PSA
31	1991	69	17.8	4a	7	0	4	M1a + M1b + M1c
32	1991	62	73.7	4a	7	1	2	LR + M1b
Non-progressors								
33	1990	59	9.2	2b	7	0		
34	1990	64	7.4	2b	6	0		
35	1991	65	11.1	2c	5	0		
36	1989	60	19.4	3a	7	0		
37	1992	59	17.2	2c	7	0		
38	1992	51	16.1	2a	7	1		
39	1993	72	12.2	3b	7	1		
40	1994	66	9.7	2c	6	1		
41	1992	67	4.8	3a	5	1		
42	1993	52	2.2	3a	7	0		
43	1993	58	5.8	2c	6	0		
44	1987	59	No data	4a	7	1		
45	1993	67	2.2	2c	6	0		
46	1994	65	10.6	2c	6	0		
47	1993	55	16.4	2c	6	0		
48	1994	62	21.5	2c	6	0		
49	1989	62	5.2	3b	7	1		
50	1990	59	18.6	3a	7	0		
51	1992	65	23.1	4a	7	0		
52	1992	67	2.2	3a	6	0		
53	1992	63	4.5	3a	5	0		
54	1991	49	21.8	3a	6	0		
55	1991	59	8	3a	6	0		
56	1991	70	11.7	2c	7	0		
57	1990	69	2.6	4a	5	0		
58	1990	61	2.5	4a	5	0		
59	1991	44	17.3	3a	6	0		
60	1991	68	13.5	3a	6	0		
61	1991	72	15.3	3a	7	0		
62	1991	51	7.5	3a	7	1		
63	1991	65	27.8	3b	7	1		
64	1991	52	5.6	3a	7	0		

OperYr, year of prostatectomy; AgeOper, age of patient at time of surgery; PreOpPSA, preoperative PSA; Res Margin, resection margin status, 0 (negative) and 1 (positive); ProgFreeSurvival, length of time (months) to biochemical relapse; ProgressType, PSA = no data on local or distance recurrence, LR = local recurrence, M1a = metastasis in non-regional lymph node, M1b = bone metastasis, M1c = other site metastasis. The Gleason grading system (59) and the TNM classification (60) were used. The usage of elevated (≥0.2 ng/ml) PSA as a first indicator for imminent local or distant recurrent disease has been reported by several authors (61).

particular clone was missing in a given sample, the missing value was replaced by the average of the values of that clone's five closest neighbors in that sample. To impute all missing values, we iterated the above procedure twice using 10 neighbors in the second iteration.

For a given phenotype, we looked for the clones with significantly different underlying copy number between the subtypes. Since the clones that rarely show an abnormal copy number are not a priori likely to contribute to the difference among subgroups, we reduce the multiplicity of the comparisons by only considering the clones that show the gain or loss separately in at least 20% of the samples. There were 122 such clones, with 35% located on chromosome 8 (almost all on 8p).

To test univariately for association between the copy number and a phenotype, for each clone we tested the null hypothesis that the distribution of the copy was the same in each of the subgroups, by using a *t*-statistic with pooled variance when two groups were being compared and an *F*-statistic for more than two groups (64). The *P*-value for a clone was computed by considering the distribution of the *t*-statistic under the null hypothesis of no difference between the two groups. An adjustment for multiple comparisons was made so as to limit the probability of finding at least one false positive result. In practice, this was implemented using the maxT method (65) by randomly permuting the subgroup labels, recomputing the statistic for each clone and recording the maximum absolute value of the statistic over all clones. We repeated the procedure 10 000 times. The observed statistic for a given clone was compared with the distribution of the recorded maxima and the adjusted *P*-value equal to the proportion of the recorded maximum values exceeding the absolute observed *t*-statistic for a given clone.

This patient sample was selected to have a 50% probability of recurrence. To determine independent predictors of recurrence, we fit the multivariate Cox-proportional hazards model using all of the clinical variables and the significant loci identified in the univariate analysis. All of the statistical analyses were done in the environment of the freely available statistical package R (66).

We looked for BAC clones that could serve as a group to identify tumors with metastatic potential. The frequency of change for each BAC was computed and copy number aberrations that occurred in $\geq$20% of the progressors that later metastasized and organ metastases, but in $<$20% of the non-progressors and vice versa were considered a differentiating BAC. For each clone the proportion of cases either above or below the defined threshold between non-progressors and those with metastases was compared using Fisher's one--sided exact test.

Validation of candidate genes with TaqMan

For validation of candidate genes, RNA was extracted from tumors and benign tissue obtained from 10 radical prostatectomy specimens and analyzed using TaqMan RNA quantitation. For each case, this was performed as follows. Twenty 13 μm slices were obtained from UCSF Comprehensive Cancer Center fresh-frozen tissue bank, with the first, 10th and 20th sections H&E stained for evaluation of tissue types and amounts. Areas of interest on these slides were then marked to act as guides for microdissecting the areas for analysis from the remaining unstained sections; only areas with $>$70% of the tissue of interest (benign prostate epithelium or tumor) were marked. The unstained slides were dehydrated using steps of 70, 90 and 100% ethanol for 1 min each followed by immersion in xylene for 5 min. After the slides dried, outlined areas were microdissected from the slides using a scalpel blade, and the dissected tissue suspended into a lysis buffer containing 1% β-mercaptoethanol (RNeasy kit, Qiagen). The tissue was homogenized using a Qiashredder spin column (Qiagen) and the RNA extracted according to manufacturer's suggestions. The RNAs were run on an Agilent 2100 BioAnalyzer (Palo Alto, CA, USA) to assess RNA quality. Tissue samples were retained in the study if there was not any significant RNA degradation.

TaqMan assays were carried out by the UCSF Genome Core. TaqMan primer–probe sets were available as Assays on Demand kits through ABI (Foster City, CA, USA) for MAP4K2 and MEN1. The CSMD1 TaqMan primer–probe set was designed in-house and synthesized by IDT (Coralville, IA, USA). The forward primer (5′ TTTCCAGATTTTTATCC AAACTCTCTAA 3′) and probe (5′ FAM-CACGTGGACC ATTGAAGTGTCTCATGG-BHQ1 3′) (BHQ = black hole quencher 1 from Biosearch Technologies, Novato, CA, USA) lie within exon 19 and the reverse primer (5′ GTGTGAAA GATCATTTGAACTCCTTT 3′) spans exons 19 and 20 of CMSD1. Each tissue sample was run in triplicate. Good methodology was demonstrated by the SD for the cycle threshold values of all three replicates being less than 0.3. The prostate tissue sample was retained in the study if the direction of change in expression for the candidate gene was the same as the two reference genes (18S and GUS). Results are displayed as percentage of expression relative to GUS, since similar changes were seen for 18S.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

ACKNOWLEDGEMENTS

We would like to thank David Ginzinger and Mamie Yu of the UCSF Genome Core for help with the TaqMan experiments, Lars Schmidt and Karen Chew of the Tissue Core for facilitating UCSF tissue acquisition and Rosemary Akhurst and Allan Balmain for reviewing the manuscript. This work was supported by the UCSF Prostate Cancer SPORE, NIH Grant P50CA89520, UCSF Department of Urology Prostate Cancer Fellowship (P.L.P.), the Dutch Cancer Society Grants EUR 97-1404 and DDHK 2002-2700, and the University of Michigan Prostate Cancer SPORE, NIH Grant P50CA69568 (M.A.R., A.M.C., K.J.P.).

REFERENCES

1. Jemal, A., Murray, T., Samuels, A., Ghafoor, A., Ward, E. and Thun, M.J. (2003) Cancer statistics, 2003. *CA Cancer J. Clin.*, **53**, 5–26.

2. Steineck, G., Helgesen, F., Adolfsson, J., Dickman, P.W., Johansson, J.E., Norlen, B.J. and Holmberg, L. (2002) Quality of life after radical prostatectomy or watchful waiting. *N. Engl. J. Med.*, **347**, 790–796.
3. Bruder, C.E., Hirvela, C., Tapia-Paez, I., Fransson, I., Segraves, R., Hamilton, G., Zhang, X.X., Evans, D.G., Wallace, A.J., Baser, M.E. *et al.* (2001) High resolution deletion analysis of constitutional DNA from neurofibromatosis type 2 (NF2) patients using microarray-CGH. *Hum. Mol. Genet.*, **10**, 271–282.
4. Hui, A.B., Lo, K.W., Teo, P.M., To, K.F. and Huang, D.P. (2002) Genome wide detection of oncogene amplifications in nasopharyngeal carcinoma by array based comparative genomic hybridization. *Int. J. Oncol.*, **20**, 467–473.
5. Hui, A.B., Lo, K.W., Yin, X.L., Poon, W.S. and Ng, H.K. (2001) Detection of multiple gene amplifications in glioblastoma multiforme using array-based comparative genomic hybridization. *Lab. Invest.*, **81**, 717–723.
6. Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W.L., Chen, C., Zhai, Y. *et al.* (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.*, **20**, 207–211.
7. Veltman, J.A., Fridlyand, J., Pejavar, S., Olshen, A.B., Korkola, J.E., DeVries, S., Carroll, P., Kuo, W.L., Pinkel, D., Albertson, D. *et al.* (2003) Array-based comparative genomic hybridization for genome-wide screening of DNA copy number in bladder tumors. *Cancer Res.*, **63**, 2872–2880.
8. Snijders, A.M., Nowee, M.E., Fridlyand, J., Piek, J.M., Dorsman, J.C., Jain, A.N., Pinkel, D., van Diest, P.J., Verheijen, R.H. and Albertson, D.G. (2003) Genome-wide-array-based comparative genomic hybridization reveals genetic homogeneity and frequent copy number increases encompassing CCNE1 in fallopian tube carcinoma. *Oncogene*, **22**, 4281–4286.
9. Wilhelm, M., Veltman, J.A., Olshen, A.B., Jain, A.N., Moore, D.H., Presti, J.C., Jr., Kovacs, G. and Waldman, F.M. (2002) Array-based comparative genomic hybridization for the differential diagnosis of renal cell cancer. *Cancer Res.*, **62**, 957–960.
10. Volik, S., Zhao, S., Chin, K., Brebner, J.H., Herndon, D.R., Tao, Q., Kowbel, D., Huang, G., Lapuk, A., Kuo, W.L. *et al.* (2003) End-sequence profiling: sequence-based analysis of aberrant genomes. *Proc. Natl Acad. Sci. USA*, **100**, 7696–7701.
11. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
12. Snijders, A.M., Nowak, N., Segraves, R., Blackwood, S., Brown, N., Conroy, J., Hamilton, G., Hindle, A.K., Huey, B., Kimura, K. *et al.* (2001) Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat. Genet.*, **29**, 263–264.
13. Paris, P.L., Albertson, D.G., Alers, J.C., Andaya, A., Carroll, P., Fridlyand, J., Jain, A.N., Kamkar, S., Kowbel, D., Krijtenburg, P.J. *et al.* (2003) High-resolution analysis of paraffin-embedded and formalin-fixed prostate tumors using comparative genomic hybridization to genomic microarrays. *Am. J. Pathol.*, **162**, 763–770.
14. O'Hagan, R.C., Brennan, C.W., Strahs, A., Zhang, X., Kannan, K., Donovan, M., Cauwels, C., Sharpless, N.E., Wong, W.H. and Chin, L. (2003) Array comparative genome hybridization for tumor classification and gene discovery in mouse models of malignant melanoma. *Cancer Res.*, **63**, 5352–5356.
15. Albertson, D.G. and Pinkel, D. (2003) Genomic microarrays in human genetic disease and cancer. *Hum. Mol. Genet.*, **12** (Suppl. 2), R145–152.
16. Massion, P.P., Kuo, W.L., Stokoe, D., Olshen, A.B., Treseler, P.A., Chin, K., Chen, C., Polikoff, D., Jain, A.N., Pinkel, D. *et al.* (2002) Genomic copy number analysis of non-small cell lung cancer using array comparative genomic hybridization: implications of the phosphatidylinositol 3-kinase pathway. *Cancer Res.*, **62**, 3636–3640.
17. Orntoft, T.F., Thykjaer, T., Waldman, F.M., Wolf, H. and Celis, J.E. (2002) Genome-wide study of gene copy numbers, transcripts, and protein levels in pairs of non-invasive and invasive human transitional cell carcinomas. *Mol. Cell Proteomics*, **1**, 37–45.
18. Rhodes, D.R., Yu, J., Shanker, K., Deshpande, N., Varambally, R., Ghosh, D., Barrette, T., Pandey, A. and Chinnaiyan, A.M. (2004) ONCOMINE: a cancer microarray database and data-mining platform. *Neoplasia* **6**, 1–6.
19. Singh, D., Febbo, P.G., Ross, K., Jackson, D.G., Manola, J., Ladd, C., Tamayo, P., Renshaw, A.A., D'Amico, A.V., Richie, J.P. *et al.* (2002) Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell*, **1**, 203–209.
20. Ramaswamy, S., Tamayo, P., Rifkin, R., Mukherjee, S., Yeang, C.H., Angelo, M., Ladd, C., Reich, M., Latulippe, E., Mesirov, J.P. *et al.* (2001) Multiclass cancer diagnosis using tumor gene expression signatures. *Proc. Natl Acad. Sci. USA*, **98**, 15149–15154.
21. LaTulippe, E., Satagopan, J., Smith, A., Scher, H., Scardino, P., Reuter, V. and Gerald, W.L. (2002) Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastatic disease. *Cancer Res.*, **62**, 4499–4506.
22. Dhanasekaran, S.M., Barrette, T.R., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K.J., Rubin, M.A. and Chinnaiyan, A.M. (2001) Delineation of prognostic biomarkers in prostate cancer. *Nature*, **412**, 822–826.
23. Cher, M.L., Bova, G.S., Moore, D.H., Small, E.J., Carroll, P.R., Pin, S.S., Epstein, J.I., Isaacs, W.B. and Jensen, R.H. (1996) Genetic alterations in untreated metastases and androgen-independent prostate cancer detected by comparative genomic hybridization and allelotyping. *Cancer Res.*, **56**, 3091–3102.
24. El Gedaily, A., Bubendorf, L., Willi, N., Fu, W., Richter, J., Moch, H., Mihatsch, M.J., Sauter, G. and Gasser, T.C. (2001) Discovery of new DNA amplification loci in prostate cancer by comparative genomic hybridization. *Prostate*, **46**, 184–190.
25. Swalwell, J.I., Vocke, C.D., Yang, Y., Walker, J.R., Grouse, L., Myers, S.H., Gillespie, J.W., Bostwick, D.G., Duray, P.H., Linehan, W.M. *et al.* (2002) Determination of a minimal deletion interval on chromosome band 8p21 in sporadic prostate cancer. *Genes Chromosomes Cancer*, **33**, 201–205.
26. Oba, K., Matsuyama, H., Yoshihiro, S., Kishi, F., Takahashi, M., Tsukamoto, M., Kinjo, M., Sagiyama, K. and Naito, K. (2001) Two putative tumor suppressor genes on chromosome arm 8p may play different roles in prostate cancer. *Cancer Genet. Cytogenet.*, **124**, 20–26.
27. Washburn, J.G., Wojno, K.J., Dey, J., Powell, I.J. and Macoska, J.A. (2000) 8pter-p23 deletion is associated with racial differences in prostate cancer outcome. *Clin. Cancer Res.*, **6**, 4647–4652.
28. Ishkanian, A.S., Malloff, C.A., Watson, S.K., DeLeeuw, R.J., Chi, B., Coe, B.P., Snijders, A., Albertson, D.G., Pinkel, D., Marra, M.A. *et al.* (2004) A tiling resolution DNA microarray with complete coverage of the human genome. *Nat. Genet.*, **36**, 299–303.
29. Kim, R., Trubetskoy, A., Suzuki, T., Jenkins, N.A., Copeland, N.G. and Lenz, J. (2003) Genome-based identification of cancer genes by proviral tagging in mouse retrovirus-induced T-cell lymphomas. *J. Virol.*, **77**, 2056–2062.
30. Chu, L.W., Troncoso, P., Johnston, D.A. and Liang, J.C. (2003) Genetic markers useful for distinguishing between organ-confined and locally advanced prostate cancer. *Genes Chromosomes Cancer*, **36**, 303–312.
31. Sun, P.C., Uppaluri, R., Schmidt, A.P., Pashia, M.E., Quant, E.C., Sunwoo, J.B., Gollin, S.M. and Scholnick, S.B. (2001) Transcript map of the 8p23 putative tumor suppressor region. *Genomics*, **75**, 17–25.
32. Toomes, C., Jackson, A., Maguire, K., Wood, J., Gollin, S., Ishwad, C., Paterson, I., Prime, S., Parkinson, K., Bell, S. *et al.* (2003) The presence of multiple regions of homozygous deletion at the CSMD1 locus in oral squamous cell carcinoma question the role of CSMD1 in head and neck carcinogenesis. *Genes Chromosomes Cancer*, **37**, 132–140.
33. Henshall, S.M., Afar, D.E., Hiller, J., Horvath, L.G., Quinn, D.I., Rasiah, K.K., Gish, K., Willhite, D., Kench, J.G., Gardiner-Garden, M. *et al.* (2003) Survival analysis of genome-wide gene expression profiles of prostate cancers identifies new prognostic targets of disease relapse. *Cancer Res.*, **63**, 4196–4203.
34. Xu, L.L., Srikantan, V., Sesterhenn, I.A., Augustus, M., Dean, R., Moul, J.W., Carter, K.C. and Srivastava, S. (2000) Expression profile of an androgen regulated prostate specific homeobox gene *NKX3.1* in primary prostate cancer. *J. Urol.*, **163**, 972–979.
35. Bhatia-Gaur, R., Donjacour, A.A., Sciavolino, P.J., Kim, M., Desai, N., Young, P., Norton, C.R., Gridley, T., Cardiff, R.D., Cunha, G.R. *et al.* (1999) Roles for Nkx3.1 in prostate development and cancer. *Genes Dev.*, **13**, 966–977.
36. He, W.W., Sciavolino, P.J., Wing, J., Augustus, M., Hudson, P., Meissner, P.S., Curtis, R.T., Shell, B.K., Bostwick, D.G., Tindall, D.J. *et al.* (1997) A novel human prostate-specific, androgen-regulated homeobox gene (*NKX3.1*) that maps to 8p21, a region frequently deleted in prostate cancer. *Genomics*, **43**, 69–77.

37. Neville, P.J., Conti, D.V., Krumroy, L.M., Catalona, W.J., Suarez, B.K., Witte, J.S. and Casey, G. (2003) Prostate cancer aggressiveness locus on chromosome segment 19q12-q13.1 identified by linkage and allelic imbalance studies. *Genes Chromosomes Cancer*, **36**, 332–339.
38. Neville, P.J., Conti, D.V., Paris, P.L., Levin, H., Catalona, W.J., Suarez, B.K., Witte, J.S. and Casey, G. (2002) Prostate cancer aggressiveness locus on chromosome 7q32–q33 identified by linkage and allelic imbalance studies. *Neoplasia*, **4**, 424–431.
39. Witte, J.S., Goddard, K.A., Conti, D.V., Elston, R.C., Lin, J., Suarez, B.K., Broman, K.W., Burmester, J.K., Weber, J.L. and Catalona, W.J. (2000) Genomewide scan for prostate cancer-aggressiveness loci. *Am. J. Hum. Genet.*, **67**, 92–99.
40. Takahashi, S., Shan, A.L., Ritland, S.R., Delacey, K.A., Bostwick, D.G., Lieber, M.M., Thibodeau, S.N. and Jenkins, R.B. (1995) Frequent loss of heterozygosity at 7q31.1 in primary prostate cancer is associated with tumor aggressiveness and progression. *Cancer Res.*, **55**, 4114–4119.
41. Elo, J.P., Harkonen, P., Kyllonen, A.P., Lukkarinen, O., Poutanen, M., Vihko, R. and Vihko, P. (1997) Loss of heterozygosity at 16q24.1–q24.2 is significantly associated with metastatic and aggressive behavior of prostate cancer. *Cancer Res.*, **57**, 3356–3359.
42. Alers, J.C., Rochat, J., Krijtenburg, P.J., Hop, W.C., Kranse, R., Rosenberg, C., Tanke, H.J., Schroder, F.H. and van Dekken, H. (2000) Identification of genetic markers for prostatic cancer progression. *Lab. Invest.*, **80**, 931–942.
43. Alers, J.C., Krijtenburg, P.J., Vis, A.N., Hoedemaeker, R.F., Wildhagen, M.F., Hop, W.C., van Der Kwast, T.T., Schroder, F.H., Tanke, H.J. and van Dekken, H. (2001) Molecular cytogenetic analysis of prostatic adenocarcinomas from screening studies: early cancers may contain aggressive genetic features. *Am. J. Pathol.*, **158**, 399–406.
44. Kusano, N., Okita, K., Shirahashi, H., Harada, T., Shiraishi, K., Oga, A., Kawauchi, S., Furuya, T. and Sasaki, K. (2002) Chromosomal imbalances detected by comparative genomic hybridization are associated with outcome of patients with hepatocellular carcinoma. *Cancer*, **94**, 746–751.
45. Brookes, S., Lammie, G.A., Schuuring, E., de Boer, C., Michalides, R., Dickson, C. and Peters, G. (1993) Amplified region of chromosome band 11q13 in breast and squamous cell carcinomas encompasses three CpG islands telomeric of FGF3, including the expressed gene *EMS1*. *Genes Chromosomes Cancer*, **6**, 222–231.
46. Fantl, V., Richards, M.A., Smith, R., Lammie, G.A., Johnstone, G., Allen, D., Gregory, W., Peters, G., Dickson, C. and Barnes, D.M. (1990) Gene amplification on chromosome band 11q13 and oestrogen receptor status in breast cancer. *Eur. J. Cancer*, **26**, 423–429.
47. Lammie, G.A., Fantl, V., Smith, R., Schuuring, E., Brookes, S., Michalides, R., Dickson, C., Arnold, A. and Peters, G. (1991) D11S287, a putative oncogene on chromosome 11q13, is amplified and expressed in squamous cell and mammary carcinomas and linked to BCL-1. *Oncogene*, **6**, 439–444.
48. Kasahara, K., Taguchi, T., Yamasaki, I., Kamada, M., Yuri, K. and Shuin, T. (2002) Detection of genetic alterations in advanced prostate cancer by comparative genomic hybridization. *Cancer Genet. Cytogenet.*, **137**, 59–63.
49. Lapointe, J., Li, C., Higgins, J.P., van de Rijn, M., Bair, E., Montgomery, M., Ferrari, M., Egevad, L., Rayford, W., Bergerheim, U. *et al.* (2004) Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc. Natl Acad. Sci. USA*, **101**, 811–816.
50. Zhao, X. and Day, M.L. (2001) RB activation and repression of C-MYC transcription precede apoptosis of human prostate epithelial cells. *Urology*, **57**, 860–865.
51. Ramaswamy, S., Ross, K.N., Lander, E.S. and Golub, T.R. (2003) A molecular signature of metastasis in primary solid tumors. *Nat. Genet.*, **33**, 49–54.
52. Chott, A., Sun, Z., Morganstern, D., Pan, J., Li, T., Susani, M., Mosberger, I., Upton, M.P., Bubley, G.J. and Balk, S.P. (1999) Tyrosine kinases expressed *in vivo* by human prostate cancer bone marrow metastases and loss of the type 1 insulin-like growth factor receptor. *Am. J. Pathol.*, **155**, 1271–1279.
53. Davila, M., Frost, A.R., Grizzle, W.E. and Chakrabarti, R. (2003) LIM kinase 1 is essential for the invasive growth of prostate epithelial cells: implications in prostate cancer. *J. Biol. Chem.*, **278**, 36868–36875.
54. Wang, S., Gao, J., Lei, Q., Rozengurt, N., Pritchard, C., Jiao, J., Thomas, G.V., Li, G., Roy-Burman, P., Nelson, P.S. *et al.* (2003) Prostate-specific deletion of the murine Pten tumor suppressor gene leads to metastatic prostate cancer. *Cancer Cell*, **4**, 209–221.
55. D'Amico, A.V., Whittington, R., Malkowicz, S.B., Fondurulia, J., Chen, M.H., Kaplan, I., Beard, C.J., Tomaszewski, J.E., Renshaw, A.A., Wein, A. *et al.* (1999) Pretreatment nomogram for prostate-specific antigen recurrence after radical prostatectomy or external-beam radiation therapy for clinically localized prostate cancer. *J. Clin. Oncol.*, **17**, 168–172.
56. Rubin, M.A., Putzi, M., Mucci, N., Smith, D.C., Wojno, K., Korenchuk, S. and Pienta, K.J. (2000) Rapid ('warm') autopsy study for procurement of metastatic prostate cancer. *Clin. Cancer Res.*, **6**, 1038–1045.
57. Fridlyand, J., Snijders, A., Pinkel, D., Albertson, D.G. and Jain, A.N. (2004) Application of hidden Markov models to the analysis of the array CGH data. *J. Multivar. Anal.* (Special Genomic Issue), in press.
58. Rabiner, L.R. (1989) A tutorial on hidden Markov models and selected applications in sppech recognition. *Proc. IEEE*, **77**, 257–286.
59. Gleason, D.F. (1992) Histologic grading of prostate cancer: a perspective. *Hum. Pathol.*, **23**, 273–279.
60. Hermanek, P., Hutter, R.V.P. and Sobin, L.H. (1997) *TNM Atlas IUAC*, 4th edn. Springer, New York.
61. Kupelian, P., Katcher, J., Levin, H., Zippe, C. and Klein, E. (1996) Correlation of clinical and pathologic factors with rising prostate-specific antigen profiles after radical prostatectomy alone for clinically localized prostate cancer. *Urology*, **48**, 249–260.
62. Schwarz, G. (1978) Estimating the dimension of a model. *Ann. Stat.*, **6**, 461–464.
63. Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D. and Altman, R.B. (2001) Missing value estimation methods for DNA microarrays. *Bioinformatics*, **17**, 520–525.
64. Snedecor, G.W. and Cochran, W.G. (1989) *Statistical Methods*, 8th edn. Iowa State University Press, Iowa.
65. Westfall, P.H. and Young, S.S. (1993) *Resampling-Based Multiple Testing: Examples and Methods for P-value Adjustment*. Wiley, New York.
66. Ihaka, R. and Gentleman, R. (1996) R: a language for data analysis and graphics. *J. Comput. Graph. Stat.*, **5**, 299–314.