Oncogene (2004) 23, 3487–3494

www.nature.com/onc

Integration of high-resolution array comparative genomic hybridization analysis of chromosome 16q with expression array data refines common regions of loss at 16q23–qter and identifies underlying candidate tumor suppressor genes in prostate cancer

JE Vivienne Watson*[,1], Norman A Doggett[2], Donna G Albertson[1], Armann Andaya[1], Arul Chinnaiyan[3], Herman van Dekken[4], David Ginzinger[1], Christopher Haqq[1], Karen James[1], Sherwin Kamkar[1], David Kowbel[1], Daniel Pinkel[1], Lars Schmitt[1], Jeffry P Simko[1], Stanislav Volik[1], Vivian K Weinberg[1], Pamela L Paris[1] and Colin Collins[1]

[1]Collins Lab, UCSF Comprehensive Cancer Center, University of California, 2340 Sutter Street, San Francisco, USA; [2]Los Alamos National Laboratory, Bioscience Division, Los Alamos, NM 87545, USA; [3]Department of Pathology, University of Michigan Medical School, Ann Arbor, MI 48109-0602, USA; [4]Erasmus Medical Center, University Medical Center, Rotterdam, Netherlands

We have constructed a high-resolution genomic microarray of human chromosome 16q, and used it for comparative genomic hybridization analysis of 16 prostate tumors. We demarcated 10 regions of genomic loss between 16q23.1 and 16qter that occurred in five or more samples. Mining expression array data from four independent studies allowed us to identify 11 genes that were frequently underexpressed in prostate cancer and that co-localized with a region of genomic loss. Quantitative expression analyses of these genes in matched tumor and benign tissue from 13 patients showed that six of these 11 (WWOX, WFDC1, MAF, FOXF1, MVD and the predicted novel transcript Q9H0B8 (NM_031476)) had significant and consistent downregulation in the tumors relative to normal prostate tissue expression making them candidate tumor suppressor genes.
Oncogene (2004) **23,** 3487–3494. doi:10.1038/sj.onc.1207474
Published online 12 April 2004

Keywords: prostate cancer; chromosome 16q; array CGH; tumor suppressor genes

Introduction

Chromosome 16q is a region of the genome associated with copy number abnormality and loss of heterozygosity in prostate cancer as well as carcinomas of many other tissue types. Several regions at distal 16q have been implicated as having a role in prostate cancer, these include 16q23.2 and 16q24.3 through allelic loss (Strup *et al*., 1999; Verhage *et al*., 2003) and 16q23 through familial association studies (Paris *et al*., 2000). In addition, both chromosomal comparative genomic hybridization (CGH) and genomic microarray CGH have detected loss of distal 16q at a high frequency in prostate tumors (Chu *et al*., 2003; Paris *et al*., 2003). More importantly, loss of 16q24 has been seen to be associated with metastatic prostate cancer, implying a specific function for the gene or genes driving genomic loss in this region (Elo *et al*., 1999; Matsuyama *et al*., 2003). Candidate tumor suppressor genes mapping within distal 16q include WWOX at 16q23.1 (Bednarek *et al*., 2001; Paige *et al*., 2001), WFDC1 at 16q24.1 (Rowley *et al*., 1995), CDH13 at 16q24.2 (Toyooka *et al*., 2001), and CBFA2T3 at 16q24.3 (Kochetkova *et al*., 2002); however, none of these have been shown to have altered structure or expression in prostate cancer.

Prostate tumors are morphologically heterogeneous, and areas of different tumor grade are commonly seen to be intimately associated with the areas of normal stroma. An accompanying genetic heterogeneity has been demonstrated, which goes some way to arguing for a multifocal origin of prostate tumors and explains to a certain extent why identifying tumor-associated genetic change has been a difficult task in this organ (Ruijter *et al*., 1999).

We have taken a two-step approach to elucidating the identity of the genes on 16q with potential involvement in prostate cancer. First, we constructed an array comprised of 326 BAC, PAC and cosmid clones mapped to chromosome 16q by DNA fingerprinting and mapping of end sequences (Han *et al*., 2000; Paige *et al*., 2000). The position of each clone relative to the human genome assembly (Ensembl Sept 2002) is known so that analysis of the underlying sequence is straightforward. We have used this array for CGH with DNA from 14 microdissected prostate tumors: nine high grade (Gleason score 8–10), two intermediate grade (Gleason score 7) and three low grade (Gleason score 6 or less) and two prostate tumor metastases, all known to have

*Correspondence: Dr JE Vivienne Watson;
E-mail: vwatson@cc.ucsf.edu
Received 5 September 2003; revised 18 December 2003; accepted 7 January 2004; Published online 12 April 2004

some 16q loss by whole genome array CGH. Multiple regions of recurrent deletion at 16q23–qter were defined within the tumors, and all known genes and ESTs that lay within these regions were identified. Second, we interrogated expression microarray data from four independent studies of prostate tumors (Dhanasekaran *et al*., 2001; Luo *et al*., 2001; Magee *et al*., 2001; Welsh *et al*., 2001), to obtain expression profiles for the genes that localized to recurrent chromosomal deletions. In total, 10 genes that co-localized with regions of genomic loss were found to have reduced expression in prostate tumors. The expression of each of these genes was measured in matched normal and malignant prostate tissue from 13 patients using quantitative RT–PCR. We also measured the relative expression of WFDC1, already known to be a prostate cancer growth suppressor (Rowley *et al*., 1995), which maps within a common deleted region of 16q24 but for which there was no available expression data. Using this combined approach we have identified six candidate tumor suppressor genes in this region whose expression is consistently and significantly reduced in prostate cancer: WWOX, MAF, WFDC1, FOXF1, MVD, and the predicted novel transcript Q9H0B8.

Results

CGH analysis

In order to refine the regions of loss observed on chromosome 16q in prostate cancer we have used a high-resolution contig array of human chromosome 16q incorporating 326 BAC, PAC, and cosmid clones precisely mapped to the human genome assembly. The contig spans 45.3 Mb, with 78% coverage. Regions of the contig with a paucity of clones are indicated by brackets in Figure 1a. The array was optimized using the average log 2 ratio of four control hybridizations with normal XX and normal XY lymphocyte DNA (Figure 1a). Clones for which the average log 2 ratio plus 2 standard deviations over four XX vs XY hybridizations was greater than 0.3 were eliminated from the analysis (81 clones). The high level of noise observed in the log 2 ratios of these clones is due, in part, to the high level of intra- and interchromosomal duplication that occurs on 16q. Regions of the genome with high repeat content or with strong sequence similarity to other loci have been shown by others to be intractable by high-resolution chromosomal CGH (Buckley *et al*., 2002). Using the sequence visualization tool Genome Cryptographer (Collins *et al*., 2001) we observed that the region of poor hybridization at 78.93–80.79 Mb (Figure 1a, region 4) contains a 287 kb intrachromosomal duplication, and the second region of poor hybridization at 71.95–73.41 Mb (Figure 1a region 3) contains three highly represented pseudogenes (25–41 copies). Genome Cryptographer analysis is presented in the supplemental data (http://shark.ucsf.edu/~stas/Viv/watson_et.al.2003.html). This is compared to two regions that hybridize consistently well, which contain one small interchromosomal duplicon of 45 kb (Figure 1a, region 2) and three low copy number pseudogenes (4–7 copies) (Figure 1a, region 1).

To test the capability of the array to detect homozygous loss we performed CGH with DNA from the cell line HCT116, which has a well-defined 210 kb homozygous deletion at 16q23 (Paige *et al*., 2000). Log 2 ratios of −1.31 and −1.32 were seen with two PAC clones that lie within the homozygous deletion. That any HCT116 DNA hybridized to these PAC clones is likely due to incomplete suppression of the high level of Alu repetitive elements found in this region, under our hybridization conditions (Ried *et al*., 2000).

DNA isolated from fresh-frozen or paraffin-embedded tumors was hybridized onto the array in competition with normal XY DNA. We analysed 16 tumors: 11 high grade with a combined Gleason score of 8 or more, including two hormone refractory tumors and two metastases; two intermediate grade (Gleason 7); and three low-grade tumors (Gleason 6). The two metastases were the only samples that were formalin fixed and paraffin embedded. DNA extracted from these samples had been demonstrated previously to give good results with CGH to metaphases (van Dekken *et al*., 2000) and to a whole genome scanning microarray (Paris *et al* 2003). The fresh-frozen tumors were dissected without using a microscope, and therefore some attenuation of tumor-specific signal with that from normal tissue was expected in these cases. However, through a combination of conservative designation of areas of pure tumor tissue and careful use of the scalpel, we were confidant that the level of contamination was within the detection limits of CGH array hybridization, previously set at 30% (Hodgson *et al*., 2001).

All the tumors had been preselected for analysis as having genomic loss at 16q23qter through hybridization to a whole genome scanning array (manuscript in preparation). Tumors with decreased copy number of the whole arm of 16q relative to the rest of the genome detected by the scanning array (2/20) were not included in this study.

A consistently high degree of concordance between the high-resolution 16q array data and the genome scanning array was seen and is illustrated in Figure 1c where data from the two CGH arrays for one sample are superimposed. Thresholds for significant gain or loss of a particular clone (log 2 ratio of <-0.26 or >0.26) are shown with a dashed line. From this example it can be seen that the high-resolution array of 16q was able to define much more precisely the regions of loss in these tumors compared to the 2400 element whole genome array. We found that the quality of high-resolution array CGH data from formalin-fixed, paraffin-embedded samples was just as good as that obtained using DNA from frozen samples, indicating that DNA from specimens preserved in this way is not inherently difficult to work with.

Figure 2 shows the frequency of gains and losses across 16q for all 16 tumors. In total, 15 separate regions of copy number loss seen in $>30\%$ of tumors were delineated (Table 1 and Figure 2). Three regions of copy

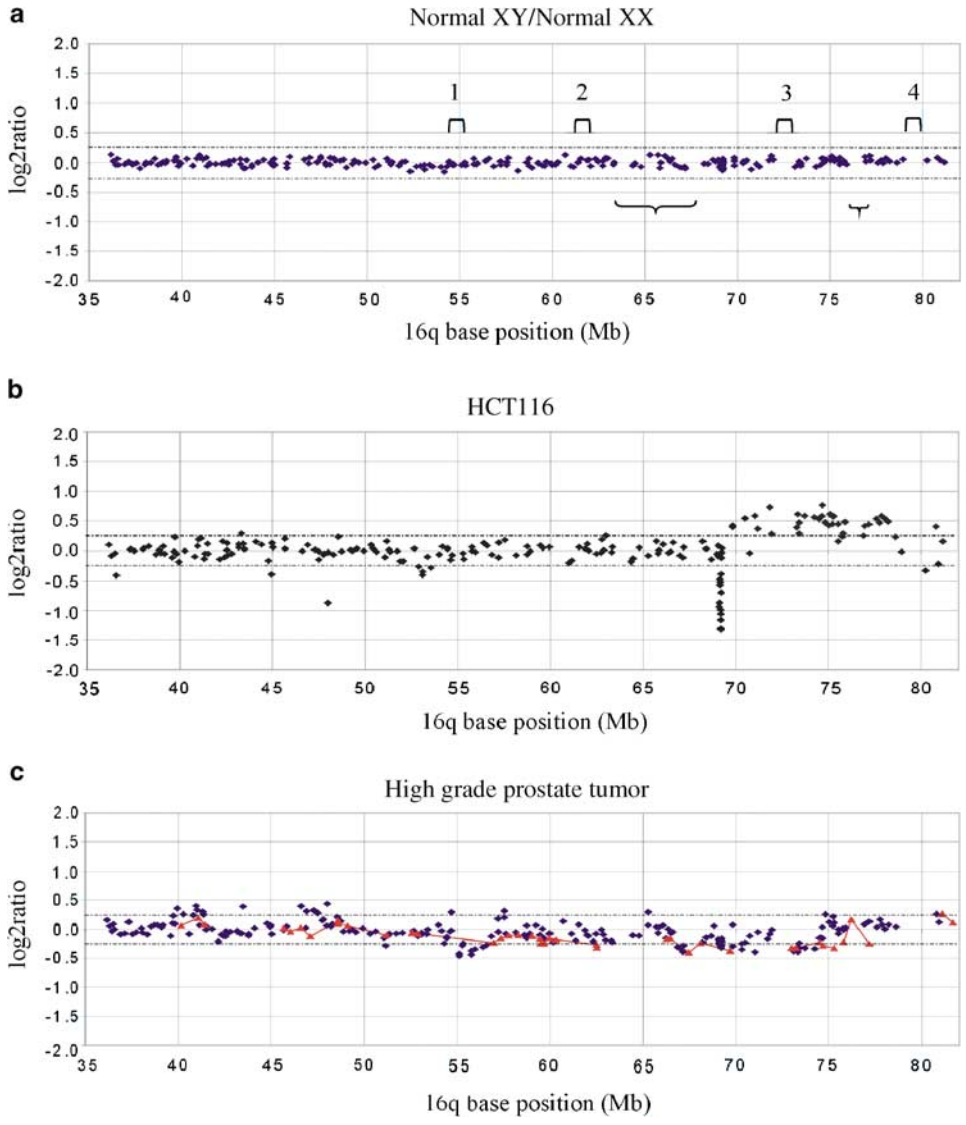

Figure 1 Chromosome 16q array CGH. Examples of 16q contig array CGH data. The map position of each genomic clone is represented along the *x*-axis. A value for the log 2 ratio of cy3/cy5 intensity for test/reference DNA is given for each clone. (**a**) Relative signal intensities for the average of four normal XY/XX hybridizations. Regions indicated 1–4 were selected as representatives of good hybridization (1 and 2), or poor hybridization (3 and 4), for sequence analysis by Genome Cryptographer. The region indicated by the bracket had a paucity of clones in the original contig. (**b**) Data for colon cancer cell line HCT116. (**c**) An example of CGH data for a single high-grade tumor (black). Data from the same tumor analysed on a 2400-element whole genome array are also shown for comparison (red)

number gain were also observed in >30% of tumors but these were not pursued further. We analysed the sequence of the BAC clones at the regions of loss at 16q23.1–qter to identify all known and predicted genes.

Expression array data mining

In an analysis of published data from four independent studies of gene expression in prostate tumors (Dhanasekaran *et al*., 2001; Luo *et al*., 2001; Magee *et al*., 2001; Welsh *et al*., 2001), we obtained all available expression data for genes and ESTs mapping to chromosome 16q. The four studies utilized two different expression array platforms, and between them had data on 273 unique targets on 16q (Table 2). Through a meta-analysis, these datasets had previously been shown to share significantly similar results independent of the method and technology used (Rhodes *et al*., 2002). A total of 68 tumors with combined Gleason scores ranging from 5 to 9 were analysed in these expression studies. Within each study the average expression (a median centered log base 2 value) for each gene was calculated across all of the tumors. We identified a subset of 10 genes that showed downregulation of expression in at least one of the studies, and that co-localized to a region of recurrent genomic loss at 16q23.1–qter. These were: WWOX, MAF, PLCG2, CDH13, NOC4, ICSBP1, FOXF1, MVD, FANCA, and Q9H0B8 (RefSeq gene

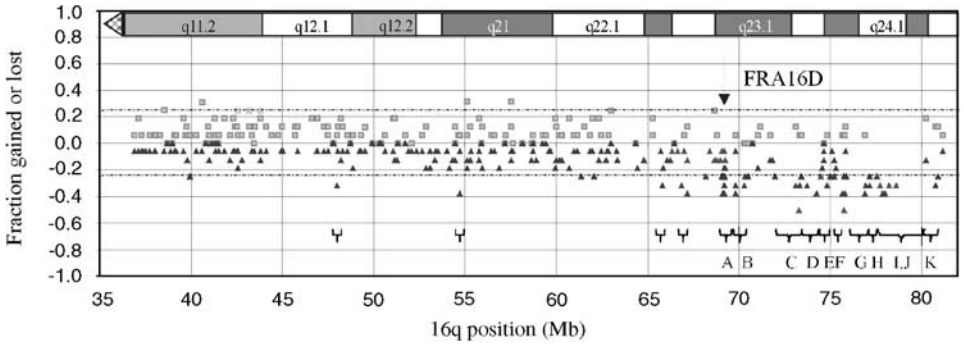

Figure 2 Frequency of gains and losses on chromosome 16q. Array elements (BAC, PAC or cosmid clones) are placed along the *x*-axis according to the map position of the start of the clone (Ensembl Sept 2002 freeze, UCSC June 2002 freeze). The number of times a particular clone showed a significant copy number change in 16 CGH experiments is represented as a fraction and plotted either above the *x*-axis for gains (gray squares) or below the *x*-axis for losses (black triangles). Regions of loss in five or more tumors lie beneath the dashed line and are indicated by brackets. Those chosen for analysis and the candidate genes they contain are as follows: A WWOX; B MAF; C PLGC2; D CDH13; E WFDC1; F Q9H0B8; G NOC4; H ICSBP1; I FOXF1; J MVD; K FANCA. The arrow indicates approximate position of the common fragile site FRA16D

Table 1 Results of 16qCGH analysis and expression array data mining of prostate tumors: deleted regions and co-localized genes with reduced expression in prostate cancer

	16q position (Mb) UCSC June 2002			*Genes selected for analysis*			*Expression data: median log2 ratio*			
Deleted region	*Start*	*End*	*Size (Mb)*	*Name*	*Ref Seq ID*	*Accession*	*Dhanasekaran et al.*	*Luo et al.*	*Magee et al.*	*Welsh et al.*
1	47.808	48.016	0.209	GPR114	NM_153837	AY140956	No data	No data	No data	No data
2	54.694	55.117	0.423	None predicted						
3	65.702	65.850	0.148	Q96PW5	Not avail.	AB067510	No data	No data	No data	No data
4	66.622	66.992	0.369	Q9BQ00	NM_030970	BC001809	No data	No data	No data	No data
5	66.992	67.186	0.194	CASPR4	NM_033401	AF463518	No data	No data	No data	No data
6	69.041	69.828	0.786	WWOX	NM_016373	AF325432	No data	−0.017	0.350	0.835
7	69.828	70.330	0.502	MAF	NM 005360	AI091435	−0.908	No data	No data	2.962
8	71.945	73.415	1.469	PLCG2	NM 002661	H57180	−0.256	−0.540	−0.079	0.420
9	73.415	74.486	1.072	CDH13	NM 001257	BC030653	−0.336	0.051	No data	−2.776
10	74.668	75.024	0.356	WFDC1	NM_021197	AF169631	No data	No data	No data	No data
11	75.291	75.842	0.551	Q9H0B8	NM 031476	AA460304	−0.765	No data	No data	No data
12	75.899	77.142	1.243	NOC4	NM 006067	AA186413	No data	−0.036	No data	No data
13	77.142	77.567	0.425	ICSBP1	NM 002163	AA865141	−0.248	−0.191	−2.996	−1.804
14	77.567	80.235	2.668	FOXF1	NM 001451	AA521040	No data	No data	−0.613	1.788
14	77.567	80.235	2.668	MVD	NM 002461	N50834	−0.312	−0.027	−1.643	No data
15	80.235	80.918	0.683	FANCA	NM 000135	AA644129	No data	−0.302	No data	No data

Table 2 Details of expression array studies used in meta-analysis

Study	*Number of tumors screened*	*Combined Gleason score of tumors*	*Array platform*	*Number of target elements on 16q*
Dhanasekaran *et al.*	17	6–7	cDNA (Research Genetics)	222
Luo *et al.*	16	5–8	cDNA (Research Genetics)	151
Magee *et al.*	11	6–9, 2 met	Oligonucleotide (Affymetrix 6800 chip)	97
Welsh *et al.*	24	5–9, 1 met	Oligonucleotide (Affymetrix U95a chip)	213

NM_031476 encoding the hypothetical protein DKFZp434B044) (Table 1). An additional gene, WFDC1, was selected that mapped to a region of consistent loss but for which no expression data were available. WFDC1 protein has previously been shown to suppress the growth of the prostate cancer cell line PC3 (Rowley *et al*., 1995), and was therefore considered to be a good candidate.

Quantitative expression analysis

Quantitative expression analysis for 11 candidate genes was carried out using an RT–PCR TaqMan assay. For this study, RNA was isolated from microdissected matched tumor and benign tissue from an independent cohort of 10 prostate cancer patients. In addition, we analysed RNA from matched pairs of tissue from three patients from which DNA had been concurrently extracted and analysed by 16q array CGH. The total sample set comprised five high-grade tumors (combined Gleason score of 8–10); one intermediate-grade tumor (Gleason 7) and seven low-grade tumors (Gleason 6 or less). Only those samples that showed consistent patterns of expression relative to two reference genes were used. Expression levels were calculated relative to the reference gene GUS. The expression of each gene in each tumor and matched normal sample is plotted in Figure 3. Using both a matched pair *t*-test and a Wilcoxon matched pairs test, six out of the 11 genes analysed were shown to be significantly downregulated in the tumor samples relative to the adjacent benign tissue. These were WWOX, MAF, WFDC1, FOXF1, MVD and predicted gene Q9H0B8. Although there were only three cases for which there was CGH data, each gene that showed decreased genomic copy number in the tumor had reduced expression in the tumor in each case. Interestingly, in these three cases, some genes (MAF, WFDC1 and CDH13) showed reduced expression even when normal copy number was observed, suggesting an alternative mechanism of transcriptional attenuation in these cases. The number of cases studied here was too small to correlate expression differences with tumor grade.

Discussion

New technologies have enabled a large amount of data to be generated very quickly when studying the expression profiles of cancer cells. However, when searching for those genes that are specifically involved in the disease process, one can be overwhelmed with candidates. It has been known for many years that 16q23–ter loss is associated with prostate cancer and other carcinomas, but this region of the genome is gene rich and contains a common fragile site, both of which compound the complexity of identifying the specific genes associated with cancer through positional cloning methods. In an attempt to overcome these problems we adopted the combined approach described here. We created a high-resolution genomic microarray and used it to define more precisely regions of 16q genomic loss in prostate tumors. We then interrogated previously published expression array data to find genes in these regions that show downregulation across a broad spectrum of prostate tumors. Finally, we tested these candidates using quantitative RT–PCR to measure their expression in a separate cohort of matched tumor and benign prostatic tissue.

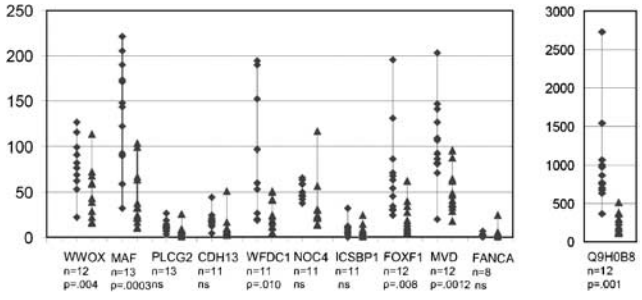

Figure 3 Quantitative expression analysis of candidate genes. Expression relative to the reference gene GUS is shown for each candidate gene in '*n*' paired samples of benign and malignant prostatic tissue (indicated by diamonds and triangles respectively). Expression of the candidate gene Q9H0B8 is shown on separate axes due to the greater level of expression of this gene. Significance values were calculated using a matched pair *t*-test

High-resolution genomic microarray analysis of 16 prostate tumors known to have loss at 16q defined a region of loss of at least 250 kb within the distal 22 Mb of the chromosome in all the tumors. Six out of 11 candidate genes mapping within this interval tested by quantitative RT–PCR were significantly downregulated in 13 prostate tumors compared to their adjacent benign tissue. Two of these genes, WWOX and WFDC1, have been previously demonstrated to have a tumor growth suppression effect; WWOX on the anchorage independent growth and *in vivo* tumorigenicity of breast cancer cell lines (Bednarek *et al*., 2001) and WFDC1 on the growth of a prostate cancer cell line (Larsen *et al*., 1998). A further one of the candidate genes, MAF, is a transcription factor which binds the promoter of p53 and regulates its expression (Hale *et al*., 2000). A role for the other candidate genes has yet to be established in cancer, although we can speculate how they might be involved. FOXF1 is a forkhead transcription factor shown to be important in the hedgehog signaling pathway in embryonic development (Mahlapuu *et al*., 2001), and MVD (mevalonate pyrophosphate decarboxylase) is part of the cholesterol biosynthesis pathway required for prenylation of many proteins including Ras, a modification that is essential for protein function (Wadhwa *et al*., 2003). Deregulation of either of these genes will likely affect these important cellular pathways that may, in turn, impact upon the maintenance of normal prostate cell growth.

Q9H0B8 is an anonymous EST that we have identified as a novel candidate tumor suppressor gene in prostate cancer. It shows very high levels of expression in benign prostatic tissue accompanied by significantly decreased expression in matched tumor tissue, making it a potentially powerful biomarker for malignant disease or a target for novel therapeutics. Q9H0B8 encodes a predicted protein of 497 amino acids that shows 59% identity over 400 amino acids with the CocoaCrisp protein (NP_113649). Both these proteins are cysteine-rich and contain an SCP domain, making them part of the large CRISP (cysteine-rich secretory protein) family that function in some vertebrates as venoms and toxins (Smith *et al*., 2001). CRISP proteins

1–3 are expressed primarily in the male genital tract and are thought to mediate cell–cell interactions with other cells during sperm maturation or during fertilization (Giese *et al*., 2002). Downregulation of Q9H0B8 expression may therefore be a marker for de-differentiation of prostate cells. Mutation analysis and transfection studies will help to elucidate the function and growth suppression capabilities of all these candidate genes.

The common fragile site FRA16D lies at (16q23.2), which corresponds approximately with the point along the chromosome arm at which a transition to frequent high level losses occurs, suggesting that the fragile site maybe involved in the mechanism of loss and rearrangements seen at 16q in many tumors (Ried *et al*., 2000). In particular, translocations that involve the nearby MAF gene have been observed in multiple myeloma (Chesi *et al*., 1998) and homozygous loss at the WWOX locus, lying right at the fragile site, has been seen in multiple tumor types (Paige *et al*., 2000). Indeed, it has been argued that rearrangement of 16q is a consequence of generalized genomic instability that accompanies cancer progression. However, we have observed discrete regions of loss distal to FRA16D in many of the prostate tumors, indicating independent breakpoints and selection for loss of multiple genes. Also, in the three cases for which we have CGH analysis and quantitative expression data from the same tumor, we have observed that downregulation of our candidate genes is not necessarily accompanied by genomic loss, suggesting an independent selection pressure for a reduction in expression of tumor suppressor genes in prostate cancer. In addition, observations from whole genome CGH analysis of 10 high-grade and 10 low-grade tumors did not indicate an association of 16q loss with a general increase in genomic rearrangement (manuscript in preparation), strengthening the argument that loss of part or all of chromosome 16q is a cause and not simply a consequence of cancer progression.

Overall, we have demonstrated that determination of regions of genomic deletion in tumors through high-resolution array CGH in combination with large-scale expression microarray data can be a powerful tool for the identification of candidate tumor suppressor genes. By applying this method for 'structural genomic filtering' of expression microarray data we have extracted specific and relevant data that may otherwise have taken much longer to identify or potentially even have been missed, and have pinpointed novel candidate genes that may be of relevance in many carcinomas.

Materials and methods

Tissues and patient resources

Prostate tissue (primary tumor and benign) was collected from 22 patients during radical prostatectomy and from a further two patients with local recurrence via transurethral resection of the prostate (UCSF, CHR Approval number H5664-06255-13). DNA from formalin-fixed paraffin-embedded tissues of a bone metastasis and a regional lymph node metastasis were obtained from Erasmus Medical Center, Rotterdam, Netherlands (Paris *et al*., 2003). All tissue specimens, except those that were paraffin embedded, were frozen at −20°C in OCT mounting medium (TissueTek, Sakura, Japan) immediately after retrieval from the operating room and stored at −80°C until used. For each frozen patient specimen, the tissue of interest (tumor or benign prostate) was enriched for by microdissecting the tissue from 24–36 sequential 14 μm frozen sections mounted onto glass slides, using a scalpel blade. Every eighth slide of this sequence was stained by standard hematoxylin and eosin and marked by a pathologist to act as a guide slide, to increase the accuracy of the dissection and enrichment.

The frozen sections were dehydrated in an ethanol series, and those for RNA extraction were further processed by immersion in xylene for 5 min. Tissue for DNA extraction was scraped off the frozen sections directly into proteinase K lysis buffer (10 mM Tris pH 8.3, 50 mM KCl_2, 1.5 mM $MgCl_2$, 0.5% Tween 20, 100 μg/ml proteinase K) and then incubated overnight at 55°C. DNA was cleaned by phenol : chloroform : isoamylalcohol (50 : 49 : 1) extraction and then purified through a Phaselock gel column (Brinkmann Eppendorf Westbury, NY, USA) before ethanol precipitation and resuspension in 1 × TE pH 7.0. RNA extraction was performed using a Qiashredder kit (Qiagen Inc., Valencia, CA, USA) to homogenize the tissue and then by using an RNeasy RNA extraction kit (Qiagen Inc.). RNA was quantified on a spectrophotometer and its quality evaluated using RNA 6000 Pico LabChips and a Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) when possible. HCT116 cell line DNA was a gift from C Chen (UCSF Comprehensive Cancer Center).

Preparation of array targets The 16q contig has been previously described (Han *et al*., 2000). Additional clones from around the WWOX locus (Paige *et al*., 2000) and from the RP11 BAC library were also added to the array. BAC, PAC and cosmid clones were obtained as agar stabs. A control set of 180 BAC clones from the RP11 library that map to 17 different chromosomes was also included on the array. In all, 220 μl of an overnight culture of each clone was used to inoculate 3 ml Luria broth containing an appropriate antibiotic. After incubation for 7 h at 37°C with shaking at 220 r.p.m., DNA was extracted using the AutoGenprep 740 machine, following the manufacturer's protocol (AutoGen, Holliston, MA, USA). Approximately 80 μl of DNA was collected, containing 0.5–1 μg DNA. DNA quality and quantity was assessed by restriction endonuclease digestion of 16 μl of each DNA with *Hind*III followed by gel electrophoresis. In total, 1 μl of each DNA containing between 50 and 200 ng/μl was amplified by degenerate oligo primer (DOP) PCR in a 96-well format as previously described (Hodgson *et al*., 2001). PCR reactions checked on 1% agarose gels typically showed DNA fragments sizes ranging from 0.2 to 5 kb. The concentration of DNA following PCR was 100–120 ng/μl. Products were ethanol precipitated, the pellets were rinsed with 70% ethanol and then resuspended overnight in 12 μl 20% DMSO.

Printing of arrays DNA was spotted at high density onto chromium-coated slides (Nanofilm, Westlake Village, CA, USA) using a custom DNA arraying device developed at the UCSF Comprehensive Cancer Center. Each clone was represented four times at two well-separated locations on the array.

Hybridization of arrays Prostate tumor DNA, normal female DNA (Sigma-Aldrich Corp., St Louis, MO, USA) or HCT116 cell line DNA was labeled as test DNA, and normal human male DNA (Sigma-Aldrich Corp., St Louis, MO, USA) was labeled as reference DNA by random priming using a BioPrime kit (Invitrogen, Carlsbad, CA, USA). Briefly, 10 μl 2.5 $\times$ random primers (750 μg/ml) was mixed with 400 ng DNA brought to a total volume of 20.5 μl with H_2O. The reaction was heated at 100°C for 10 min, then cooled on ice for 5 min. A dinucleotide mix (final concentration of 200 μM dTTP, 400 μM each dATP, dGTP, dCTP) was added to the DNA with Cy3-dUTP (test) or Cy5-dUTP (reference) to a final concentration of 80 μM (Amersham Biosciences, Piscataway, NJ, USA) and 40U Klenow (Invitrogen). After overnight incubation at 37°C, unincorporated nucleotides were removed from the pooled reference and test labeling reactions using a Microspin G50 column (Amersham). The probes were then ethanol-precipitated with approximately 60 μg cot1 DNA (Roche Applied Science, Indianapolis, IL, USA). The precipitated pellet was resuspended in 50 μl hybridization buffer (50% formamide, 10% dextran sulfate, 4% SDS, 2 $\times$ SSC pH 7.0) with 500 μg yeast tRNA before the probe was denatured at 74°C for 10 min. The probe was allowed to preanneal at 37°C for up to an hour before application to the array.

The array was incubated with 50 μl prehybridization solution (50% formamide, 10% dextran sulfate, 2 $\times$ SSC pH 7.0, 4% SDS) for 20 min, which was then removed before the probe was applied. The arrays were hybridized for 40 h at 37°C. Washing was as follows: 2 $\times$ SSC 2 min, room temperature; 50% formamide 2 $\times$ SSC 15 min 50°C; 2 $\times$ SSC/0.1% SDS, 20 min, 50°C; PN buffer (0.1 M NaH_2PO_4, pH 8.0, 0.1% (w/v) NP40). The slides were briefly immersed in 2 $\times$ SSC before mounting in glycerol containing 1 μg/μl DAPI, buffered in 10% PBS. Slides were left in the dark for at least 1 h before imaging.

Imaging and analysis We acquired 16-bit 1024 $\times$ 1024 pixel DAPI, Cy3 and Cy5 images using a custom built CCD camera system (UCSF Comprehensive Cancer Center). We used 'UCSF SPOT' software (Jain *et al*., 2002) to automatically segment the spots and to calculate $\log_2$ ratios of the total integrated Cy3 and Cy5 intensities for each spot. A ratio of Cy3 and Cy5 intensities, averaged for the quadruplicate spots, was obtained per clone. Clones were excluded for analysis if the standard deviation of the replicate spots exceeded 0.2 or if none or only a single spot had contributed to the mean of the replicate spots. Clone identities and relative positions were assigned to each spot using a custom program 'SPROC'. All data from 16q were normalized to the average log 2 ratio of the control set of clones mapping to the rest of the genome to enable detection of whole arm copy number change of 16q. Data from a particular BAC clone were excluded from further analysis if the average log 2 ratio of the clone across four normal male vs normal female hybridizations of the array ± 2 s.d. was $> +0.3$ or < -0.3. The threshold for significant gains or losses was calculated as four times the average standard deviation of the four normal vs normal hybridizations of the array. This was equal to 0.26.

Whole genome scanning array Details of CGH to a 2400 element whole genome array have been described elsewhere (Paris *et al*., 2003).

Analysis of underlying sequence All genes and ESTs that lay in regions of loss were determined using Sept 2002 freeze of the human genome sequence at Ensembl (http://www.ensembl.org/Homo_sapiens/) and the June 2002 freeze at University of California Santa Cruz (http://genome.ucsc.edu/).

Expression array data mining Data were extracted from collected expression array experiments supplied by four different laboratories. The data for each array were log base 2 transformed and median centered as previously described (Rhodes *et al*., 2002). Target elements were determined as belonging to 16q using Unigene build 157 and Netaffx software (Affymetrix Inc., Santa Clara, CA, USA). The mean expression ratios of the candidate genes were calculated across all the prostatic tumors analysed in each study.

Quantitative RT–PCR Primers and probes were designed for WWOX, WFDC1 and ICSBP1 using Primer Express (Applied Biosystems Foster City, CA, USA). 5′FAM/3′TAMRA probes were obtained from Integrated DNA Technologies (Coralville, IA, USA). Primers and probes for MAF, PLCG2, CDH13, NOC4, MVD, Q9H0B8, FOXF1 and FANCA were obtained as 'Assays on Demand' (Applied Biosystems). Reactions were optimized and run on ABI 7700 (Applied Biosystems). Expression was measured relative to two reference genes: glucoronidate synthase (GUS) and ribosomal subunit 18S. Data for a particular gene on a particular patient tumor/benign pair that showed discordance in the tumor : benign expression ratio relative to GUS and 18S were not included in the analysis (13/139 tumor/benign pair expression ratios). The difference in expression between the sample and the reference gene GUS was calculated for tumor and benign tissue. Values are given as a proportion of GUS expression.

Acknowledgements
We thank Graeme Hodgson and Eric Gagne for advice, Maime Yu, Marcella Fasso, Ajay Jain and Jane Fridlyand for technical assistance, Karen Taylor for supplying PAC clones, Terrence Barrette for expression array data, and Peter Carroll and Karen Chew for patient material. This work was supported by the National Cancer Institute (Prostate Cancer SPORE Grant CA89520).

References

Bednarek AK, Keck-Waggoner CL, Daniel RL, Laflin KJ, Bergsagel PL, Kiguchi K, Brenner AJ and Aldaz CM. (2001). *Cancer Res.*, **61,** 8068–8073.

Buckley PG, Mantripragada KK, Benetkiewicz M, Tapia-Paez I, Diaz De Stahl T, Rosenquist M, Ali H, Jarbo C, De Bustos C, Hirvela C, Sinder Wilen B, Fransson I, Thyr C, Johnsson BI, Bruder CE, Menzel U, Hergersberg M, Mandahl N, Blennow E, Wedell A, Beare DM, Collins JE, Dunham I, Albertson D, Pinkel D, Bastian BC, Faruqi AF, Lasken RS, Ichimura K, Collins VP and Dumanski JP. (2002). *Hum. Mol. Genet.*, **11,** 3221–3229.

Chesi M, Bergsagel PL, Shonukan OO, Martelli ML, Brents LA, Chen T, Schrock E, Ried T and Kuehl WM. (1998). *Blood*, **91,** 4457–4463.

Chu LW, Troncoso P, Johnston DA and Liang JC. (2003). *Genes Chromosomes Cancer*, **36,** 303–312.

Collins C, Volik S, Kowbel D, Ginzinger D, Ylstra B, Cloutier T, Hawkins T, Predki P, Martin C, Wernick M, Kuo WL,

Alberts A and Gray JW. (2001). *Genome Res.*, **11,** 1034–1042.
Dhanasekaran SM, Barrette TR, Ghosh D, Shah R, Varambally S, Kurachi K, Pienta KJ, Rubin MA and Chinnaiyan AM. (2001). *Nature*, **412,** 822–826.
Elo JP, Harkonen P, Kyllonen AP, Lukkarinen O and Vihko P. (1999). *Br. J. Cancer*, **79,** 156–160.
Giese A, Jude R, Kuiper H, Raudsepp T, Piumi F, Schambony A, Guerin G, Chowdhary BP, Distl O, Topfer-Petersen E and Leeb T. (2002). *Gene*, **299,** 101–109.
Hale TK, Myers C, Maitra R, Kolzau T, Nishizawa M and Braithwaite AW. (2000). *J. Biol. Chem.*, **275,** 17991–17999.
Han CS, Sutherland RD, Jewett PB, Campbell ML, Meincke LJ, Tesmer JG, Mundt MO, Fawcett JJ, Kim UJ, Deaven LL and Doggett NA. (2000). *Genome Res.*, **10,** 714–721.
Hodgson G, Hager JH, Volik S, Hariono S, Wernick M, Moore D, Albertson DG, Pinkel D, Collins C, Hanahan D and Gray JW. (2001). *Nat. Genet.*, **29,** 459–464.
Jain AN, Tokuyasu TA, Snijders AM, Segraves R, Albertson DG and Pinkel D. (2002). *Genome Res.*, **12,** 325–332.
Kochetkova M, McKenzie OLD, Bais AJ, Martin JM, Secker GA, Seshadri R, Powell JA, Hinze SJ, Gardner AE, Spendlove HE, O'Callaghan NJ, Cleton-Jansen A-M, Cornelisse C, Whitmore SA, Crawford J, Kremmidiotis G, Sutherland GR and Callen DF. (2002). *Cancer Res.*, **62,** 4599–4604.
Larsen M, Ressler SJ, Lu B, Gerdes MJ, McBride L, Dang TD and Rowley DR. (1998). *J. Biol. Chem.*, **273,** 4574–4584.
Luo J, Duggan DJ, Chen Y, Sauvageot J, Ewing CM, Bittner ML, Trent JM and Isaacs WB. (2001). *Cancer Res.*, **61,** 4683–4688.
Magee JA, Araki T, Patil S, Ehrig T, True L, Humphrey PA, Catalona WJ, Watson MA and Milbrandt J. (2001). *Cancer Res.*, **61,** 5692–5696.
Mahlapuu M, Enerback S, Carlsson P, Wadhwa R, Yaguchi T, Hasan MK, Taira K and Kaul SC. (2001). *Development*, **128,** 2397–2406.
Matsuyama H, Pan Y, Yoshihiro S, Kudren D, Naito K, Bergerheim US and Ekman P. (2003). *Prostate*, **54,** 103–111.
Paige AJ, Taylor KJ, Stewart A, Sgouros JG, Gabra H, Sellar GC, Smyth JF, Porteous DJ and Watson JE. (2000). *Cancer Res.*, **60,** 1690–1697.
Paige AJ, Taylor KJ, Taylor C, Hillier SG, Farrington S, Scott D, Porteous DJ, Smyth JF, Gabra H and Watson JE. (2001). *Proc. Natl. Acad. Sci. USA*, **98,** 11417–11422.
Paris PL, Albertson DG, Alers JC, Andaya A, Carroll P, Fridlyand J, Jain AN, Kamkar S, Kowbel D, Krijtenburg PJ, Pinkel D, Schroder FH, Vissers KJ, Watson VJ, Wildhagen MF, Collins C and Van Dekken H. (2003). *Am. J. Pathol.*, **162,** 763–770.
Paris PL, Witte JS, Kupelian PA, Levin H, Klein EA, Catalona WJ and Casey G. (2000). *Cancer Res.*, **60,** 3645–3649.
Rhodes DR, Barrette TR, Rubin MA, Ghosh D and Chinnaiyan AM. (2002). *Cancer Res.*, **62,** 4427–4433.
Ried K, Finnis M, Hobson L, Mangelsdorf M, Dayan S, Nancarrow JK, Woollatt E, Kremmidiotis G, Gardner A, Venter D, Baker E and Richards RI. (2000). *Hum. Mol. Genet.*, **9,** 1651–1663.
Rowley DR, Dang TD, Larsen M, Gerdes MJ, McBride L and Lu B. (1995). *J. Biol. Chem.*, **270,** 22058–22065.
Ruijter ET, Miller GJ, van de Kaa CA, van Bokhoven A, Bussemakers MJ, Debruyne FM, Ruiter DJ and Schalken JA. (1999). *J. Pathol.*, **188,** 271–277.
Smith DM, Collins-Racie LA, Marigo VA, Roberts DJ, Davis NM, Hartmann C, Schweitzer R, LaVallie ER, Gamer L, McCoy J and Tabin CJ. (2001). *Mech. Dev.*, **102,** 223–226.
Strup SE, Pozzatti RO, Florence CD, Emmert-Buck MR, Duray PH, Liotta LA, Bostwick DG, Linehan WM and Vocke CD. (1999). *J. Urol.*, **162,** 590–594.
Toyooka KO, Toyooka S, Virmani AK, Sathyanarayana UG, Euhus DM, Gilcrease M, Minna JD and Gazdar AF. (2001). *Cancer Res.*, **61,** 4556–4560.
Van Dekken H, Krijtenburg PJ and Alers JC. (2000). *Acta Histochem*, **102,** 85–94.
Verhage BA, van Houwelingen K, Ruijter TE, Kiemeney LA and Schalken JA. (2003). *Prostate*, **54,** 50–57.
Wadhwa R, Yaguchi T, Hasan MK, Taira K and Kaul SC. (2003). *Biochem. Biophys. Res. Commun.*, **302,** 735–742.
Welsh JB, Sapinoso LM, Su AI, Kern SG, Wang-Rodriguez J, Moskaluk CA, Frierson Jr HF and Hampton GM. (2001). *Cancer Res.*, **61,** 5974–5978.