Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression

Sooryanarayana Varambally,[1,6,9] Jianjun Yu,[1,2,9] Bharathi Laxman,[1] Daniel R. Rhodes,[1,2] Rohit Mehra,[1] Scott A. Tomlins,[1] Rajal B. Shah,[1,4,6] Uma Chandran,[7] Federico A. Monzon,[7] Michael J. Becich,[7] John T. Wei,[4,6] Kenneth J. Pienta,[4,5,6] Debashis Ghosh,[2,3] Mark A. Rubin,[8] and Arul M. Chinnaiyan[1,2,4,6,*]

[1]Department of Pathology, University of Michigan Medical School, Ann Arbor, Michigan 48109
[2]Department of Bioinformatics, University of Michigan Medical School, Ann Arbor, Michigan 48109
[3]Department of Biostatistics, University of Michigan Medical School, Ann Arbor, Michigan 48109
[4]Department of Urology, University of Michigan Medical School, Ann Arbor, Michigan 48109
[5]Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, Michigan 48109
[6]Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, Michigan 48109
[7]Department of Pathology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261
[8]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115
[9]These authors contributed equally to this work.
*Correspondence: arul@umich.edu

Summary

Molecular profiling of cancer at the transcript level has become routine. Large-scale analysis of proteomic alterations during cancer progression has been a more daunting task. Here, we employed high-throughput immunoblotting in order to interrogate tissue extracts derived from prostate cancer. We identified 64 proteins that were altered in prostate cancer relative to benign prostate and 156 additional proteins that were altered in metastatic disease. An integrative analysis of this compendium of proteomic alterations and transcriptomic data was performed, revealing only 48%–64% concordance between protein and transcript levels. Importantly, differential proteomic alterations between metastatic and clinically localized prostate cancer that mapped concordantly to gene transcripts served as predictors of clinical outcome in prostate cancer as well as other solid tumors.

Introduction

Prostate cancer is a highly prevalent disease in older men of the Western world (Chan et al., 2004; Linton and Hamdy, 2003). Unlike other cancers, more men die with prostate cancer than from the disease (Albertsen et al., 1998; Johansson et al., 1997). Deciphering the molecular networks that distinguish progressive disease from nonprogressive disease will shed light into the biology of aggressive prostate cancer as well as lead to the identification of biomarkers that will aid in the selection of patients that should be treated (Kumar-Sinha and Chinnaiyan, 2003). To begin to understand prostate cancer progression with a systems perspective, we need to characterize and integrate the molecular components involved (Grubb et al., 2003; Hood et al., 2004; Paweletz et al., 2001; Petricoin et al., 2002). A number of groups have employed gene expression microarrays to profile prostate cancer tissues (Dhanasekaran et al., 2001; Lapointe et al., 2004; LaTulippe et al., 2002; Luo et al., 2001, 2002b; Magee et al., 2001; Singh et al., 2002; Welsh et al., 2001; Yu et al., 2004) as well as other tumors (Alizadeh et al., 2000; Golub et al., 1999; Hedenfalk et al., 2001; Perou et al., 2000) at the transcriptome level, but much less work has been done at the protein level. Proteins, as opposed to nucleic acids, represent the functional effectors of cancer progression and thus serve as therapeutic targets as well as markers of disease.

In the present study, we utilized a high-throughput immunoblot approach to characterize proteomic alterations in human prostate cancer progression, focusing on the transition from clinically localized prostate cancer to metastatic disease. Using an integrative approach, we were able to analyze proteomic profiles with mRNA transcript data from several laboratories. Our analyses also indicated that the proteins that were

SIGNIFICANCE

Multiple molecular alterations occur during cancer development. To begin to understand these processes, we need to characterize and integrate these components. The present study is an attempt to integrate such disparate data as RNA and protein expression. We employed an immunoblot approach to characterize proteomic alterations in prostate cancer. This approach revealed over 100 proteomic alterations in prostate cancer progression, some of which may have the potential to be developed as biomarkers or therapeutic targets. Furthermore, we were able to integrate these proteomic profiles with transcript data from expression profiling data sets. We found that the proteins that were qualitatively concordant with gene expression could be used as a predictor of clinical outcome. Thus, this approach revealed the presence of a "progression signature" in prostate tumors.

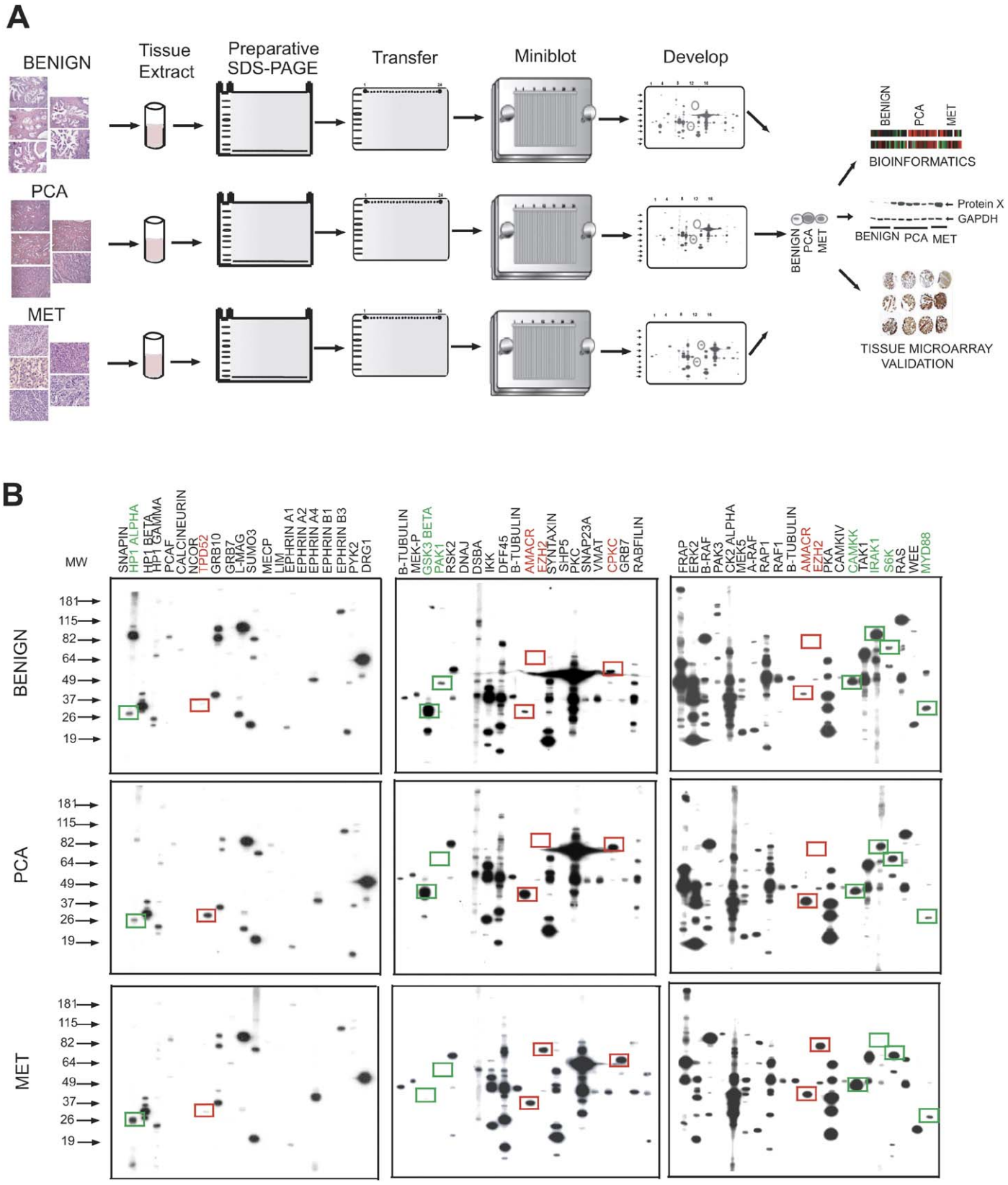


Figure 1. High-throughput immunoblot analysis to define proteomic alterations in prostate cancer progression

A: A flowchart of the general methodology employed to profile proteomic alterations in tissue extracts. Pooled tissue extracts (n = 5 each) from clinically localized prostate cancer, hormone-refractory metastatic prostate cancer, and benign prostate tissues were separated on preparative SDS-PAGE gels and transferred to PVDF membranes. The membranes were incubated with commercial antibodies using a miniblotter system. PCA, clinically localized prostate cancer; MET, metastatic prostate cancer.

qualitatively concordant with gene expression could be used to define a multiplex gene predictor of clinical outcome.

Results and discussion

In order to derive a first approximation of the prostate cancer proteome, we employed high-throughput immunoblot analysis. This method, while not feasible for use on many individual samples, allowed us to screen pooled tissue extracts for qualitative levels of hundreds of proteins (and posttranslational modifications) using commercially available antibodies. The basic approach is illustrated in Figure 1A. Extracts from five tissue specimens of benign prostate, clinically localized prostate cancer, and metastatic prostate cancer from distinct patients were pooled. Each of the three pools of tissue extracts were run on preparative SDS-PAGE gels, transferred to PVDF, and incubated with different antibodies using a miniblot apparatus. Figure 1B displays representative data using the high-throughput immunoblot approach. Known proteomic alterations in prostate cancer progression such as EZH2 (Varambally et al., 2002) and AMACR (Jiang et al., 2001; Luo et al., 2002a; Rubin et al., 2002) are highlighted in red, while novel associations such as GSK-3β and IRAK1 are highlighted in green. To further increase the number of proteins analyzed, we used an analogous high-throughput immunoblot methodology provided by commercial services (see Experimental Procedures). Thus, in total we assessed 1484 antibodies against 1354 distinct proteins or posttranslational modifications. Of these antibodies, 521 detected a band of the expected molecular weight in at least one of the pooled extracts. Antibodies that did not detect the correct molecular weight protein product may represent lack of antibody sensitivity (or poor quality antibody) or absence of protein expression in prostate tissues.

To validate the proteomic alterations identified by this screen in individual tissue extracts (as opposed to pooled extracts), we analyzed 86 proteins and two posttranslational modifications by conventional immunoblot analysis using four to five tissue extracts per class (see Figure S1). As with most gene expression studies done in prostate, our proteomic screen employed grossly dissected tumor specimens. Thus, the proteomic alterations that we detected could be due to differences in the stromal-epithelial ratio of the tissues in addition to actual alterations in the epithelial cells. In order to evaluate the proteomic alterations in situ, we employed high-density tissue microarrays (Kononen et al., 1998). As only a subset of the identified proteins have antibodies that are compatible with immunohistochemical analysis, a single tissue microarray containing 216 specimens from 51 cases was stained using 20 of these IHC-compatible antibodies. Representative tissue microarray elements are shown in Figure 2A. Each tissue microarray element was evaluated by a pathologist and scored for staining (scale of 1 to 4) as per cell type considered (e.g., epithelial, stromal, etc. . . .). Details of cell compartments used as the readout for each element are shown in Table S11. Using an in situ technique such as evaluation by immunohistochemistry allowed us to distinguish stromally expressed versus epithelially expressed proteins. In general, proteins that demonstrated a decrease in expression in the metastatic tumors most often were stromally expressed proteins. As the amount of stroma per unit area decreased with tumor progression, metastatic samples demonstrated a parallel significant decrease in protein expression of paxillin and ABP-280, among others (see Table S11). In order to visualize and cluster the tissue microarray data (Nielsen et al., 2003), the qualitative evaluations were normalized (see Experimental Procedures). Similar to gene expression analyses (Eisen et al., 1998; Perou et al., 2000), unsupervised hierarchical clustering of the data revealed that the in situ protein levels could be used to accurately classify prostate samples as benign, clinically localized prostate cancer, or metastatic disease (Figure 2B).

This high-throughput immunoblotting of prostate extracts led to the identification of several known and previously unknown proteomic alterations in prostate cancer. For example, previous studies have shown that the antiapoptosis protein XIAP (Krajewska et al., 2003), the racemase AMACR (Jiang et al., 2001; Luo et al., 2002a; Rubin et al., 2002), and the polycomb group protein EZH2 (Varambally et al., 2002) are dysregulated in prostate cancer progression. Novel associations (increases or decreases in protein expression) with prostate cancer progression identified by this screen include the E2 ubiquitin ligase UBC9, the cytosolic phosphoprotein stathmin, the death receptor DR3, and the Aurora-A kinase (STK15), among others. Descriptions and potential functional context of the various proteomic alterations identified by this screen are found in the Supplemental Data.

Having amassed this compendium of proteomic alterations in prostate cancer progression, we next examined the general concordance with the prostate cancer transcriptome. To this end, we developed an integrative model to incorporate qualitative proteomic alterations as assessed by high-throughput immunoblotting (but applicable to other proteomic technologies), with transcriptomic data derived from eight prostate cancer gene expression studies (Figure 3). As both the genomic and proteomic approaches involve analysis of grossly dissected tissues, this facilitates molecular comparisons to be made. The high-throughput immunoblot analysis of benign prostate, clinically localized prostate cancer and metastatic disease yielded 521 proteins of the expected molecular weight. Immunoreactive bands in each of the three tissue extracts were assessed, and comparisons were made between benign tissue and clinically localized prostate cancer (Figure 3A) and between clinically localized prostate cancer and metastatic disease (Figure 3B). Qualified proteins that were overexpressed were coded red, underexpressed proteins were coded blue, and unchanged proteins were coded white. Based on this analysis, 64 proteins were dysregulated in clinically localized prostate cancer relative to benign prostate tissue, while 156 proteins were dysregulated between metastatic disease relative to clinically localized prostate cancer. As might be expected, most of the proteins analyzed were unchanged in the context of prostate cancer

B: Representative high-throughput immunoblots performed for pooled benign, clinically localized prostate cancer, and metastatic prostate cancer tissues. Each lane represents analysis of an individual protein. Three representative blots are displayed for each tissue extract. Selected proteins altered in prostate cancer progression are highlighted. MW, molecular weight. See the Supplemental Data for complete data set.

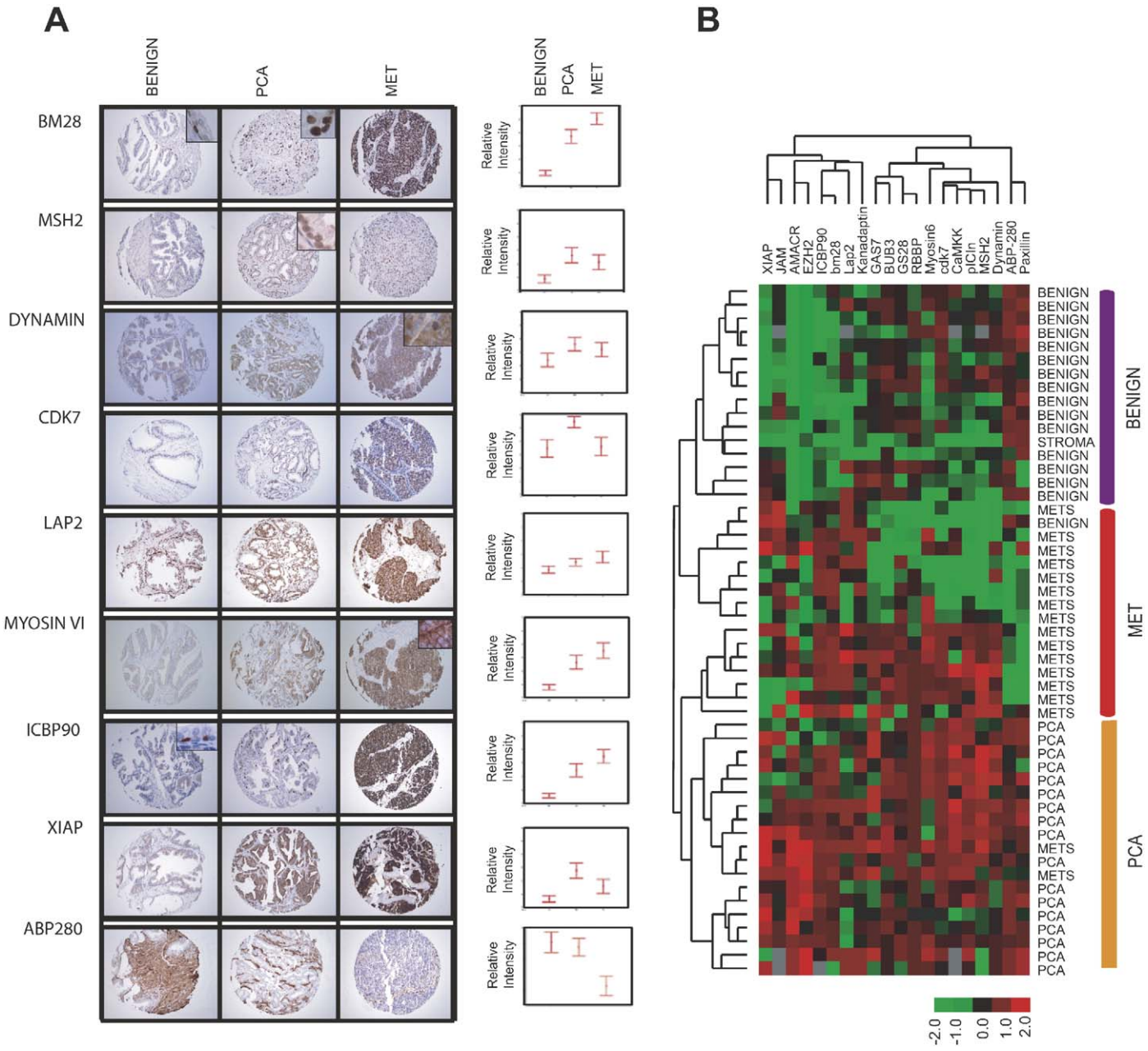


Figure 2. Tissue microarray analyses of protein markers deregulated in prostate cancer progression

A: Selected images of tissue microarray elements representing immunohistochemical analysis of proteins altered in prostate cancer progression. Relative levels of proteins as assessed by blinded pathology analysis of tissue microarrays (n = 216 specimens) are provided to the right. See the Supplemental Data for details regarding 20 markers evaluated by immunohistochemistry.

B: Cluster analysis of 20 proteins dysregulated in prostate cancer progression evaluated for in situ protein levels by tissue microarrays. Unsupervised hierarchical clustering of protein levels (columns) and samples (rows) was performed, and a heat map was generated. Red color represents high protein levels, while black refers to intermediate levels, and green represents low or absent protein levels.

progression (i.e., 87.7% [457/521] of the proteins were unchanged between clinically localized prostate cancer and benign and 70.1% [365/521] of the proteins were unchanged between clinically localized and metastatic disease).

The set of qualifiable proteins (n = 521) was then mapped to the NCBI LocusLink and UniGene databases to identify each corresponding gene. Data for mRNA were extracted for these genes using eight publicly available prostate cancer gene expression data sets (see Experimental Procedures). Over 90% of the genes were represented in at least one microarray study, allowing for integrative analysis to be performed. All eight of the prostate profiling studies made a comparison between clinically localized prostate cancer and benign tissue, while only four of these studies made a comparison between clinically

localized disease and metastatic disease. Genes that could only be found in one-fourth of studies or less were excluded, leading to 481 genes involved in the former comparison and 492 involved in the latter comparison. Since we assessed over- and underexpressed genes separately, a one-sided Welch's t test was conducted per each gene per each profiling study (see Experimental Procedures). As with the proteomic approach, comparisons between benign and clinically localized prostate cancer (Figure 3A) and localized disease and metastatic disease (Figure 3B) were made. If an mRNA transcript was significantly overexpressed in a particular study, it was coded red; underexpressed transcripts were coded blue, and white was used for unchanged transcripts.

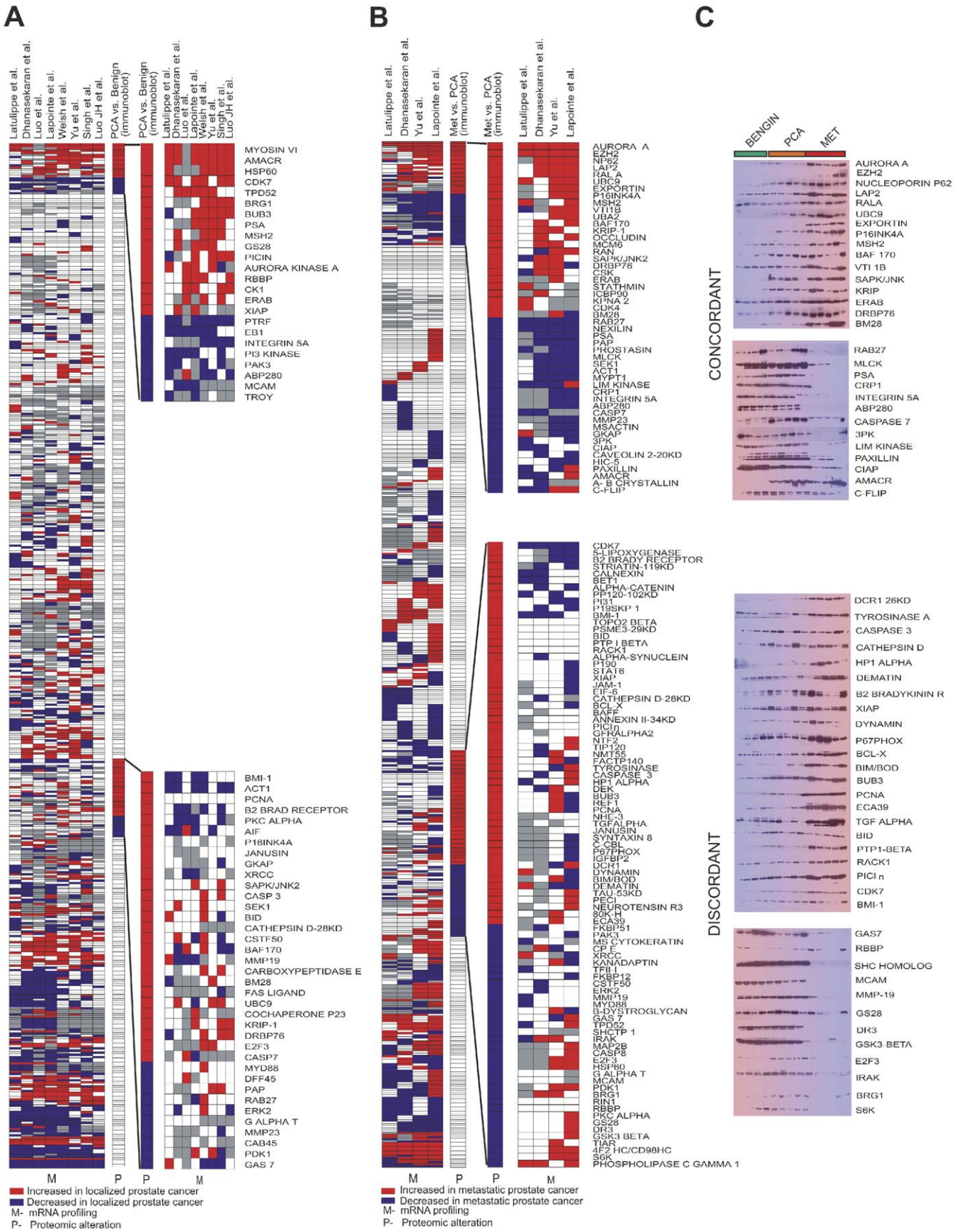
Figure 3 presents an integrative analysis of proteomic data with gene expression metadata in prostate cancer progression. An mRNA transcript alteration was considered "concordant" with a proteomic alteration if a majority of the microarray profiling studies (at least 50%) showed the same qualitative differential (increased, decreased, or unchanged) as the high-throughput immunoblot approach. According to these criteria, 289 (60.1%) out of 481 mRNA transcripts were concordant with protein levels in clinically localized prostate cancer relative to benign prostate tissue. Similarly, 291 (59.1%) out of 492 mRNA transcripts were concordant with protein levels in metastatic prostate cancer relative to clinically localized disease. Out of the 156 proteomic alterations identified between metastatic and localized prostate cancer, 50 were concordant with mRNA transcript and 90 were discordant with mRNA transcript, while the remaining alterations did not have mRNA measurements to map to (Figures 3B and 3C). Thus, similar to studies done in yeast (Griffin et al., 2002; Washburn et al., 2003), bacteria (Baliga et al., 2002), and cell lines (Tian et al., 2004), there was only weak concordance between protein and mRNA levels in prostate cancer progression.

To further explore the poor concordance we observed between protein and metadata from transcriptomic analyses, we profiled the pooled samples as well as the individual samples that comprised the pools on Affymetrix HG-U133 plus 2 microarrays. The same integrative analysis was carried out to examine the concordant relationship between the protein alterations observed in the pooled tissues by immunoblotting and transcript alterations observed in the corresponding pooled and individual tissues. The individual samples were included in order to calculate statistical significance for transcript alterations. Similar or even lower concordance was observed between protein and transcript (61.0% concordance in clinically localized prostate cancer relative to benign prostate tissue, and 48.2% for metastatic prostate cancer relative to clinically localized disease; Figure 4A; Figure S4A).

We also investigated the protein and mRNA concordance in individual samples. We focused on the 86 proteins identified as outliers in the larger high-throughput screen (see Figure S1). The immunoblot intensities were semiquantitated, and correlation coefficients were calculated for each protein (see Experimental Procedures). We found that a total 55 out of 86 proteins were observed to a have a positive correlation with mRNA, which led to 64.0% concordance between proteins and transcripts (Figure 4B). On subclassification, we observed a concordance of 54.7% and 66.3% in case of localized prostate cancer relative to benign prostate tissues and the metastatic disease relative to localized prostate cancer, respectively.

This proteomic screen identified proteins that are altered from benign prostate to clinically localized prostate cancer and a distinct set of alterations between clinically localized disease to metastatic disease. Since our group is interested in the transition from clinically localized to metastatic disease, we next focused on this comparison. As the metastatic tissues analyzed in this study are androgen independent (Shah et al., 2004), and by contrast the clinically localized tumors are generally androgen dependent, we evaluated whether there was an enrichment of androgen-regulated proteomic alterations discovered by our screening. Androgen-regulated genes (ARGs) are essential for the normal development of the prostate as well as the pathogenesis of prostate cancer (Culig et al., 1998; Koivisto et al., 1998; Mooradian et al., 1987). Pertinent to this analysis, Velasco et al. developed a meta-analysis of ARGs that represents a cross-comparison of four gene expression (DePrimo et al., 2002; Nelson et al., 2002; Segawa et al., 2002; Velasco et al., 2004) and two SAGE data sets (Waghray et al., 2001; Xu et al., 2001). ARGs were then defined as a union of these six data sets, all of which represented functional induction of mRNA transcript by androgen in vitro. Interestingly, 27 out of the 150 protein alterations (exclusive of posttranslational modifications) that we identified as being differential between metastatic and clinically localized disease were designated as androgen regulated by the Velasco et al. (2004) ARG compendium (see Supplemental Data). To demonstrate that this finding is statistically significant, we selected random sets of 150 genes from the Yu et al. (2004) or the Glinsky et al. (2004) prostate cancer profiling studies and found that the chance of selecting 27 ARGs was minimal ($p < 0.001$ for both the Yu et al. and Glinsky et al. data). Thus, androgen-regulated proteins are significantly enriched in the differential comparison between androgen-dependent and independent prostate cancer.

While examining concordant proteomic alterations, interestingly, we found that EZH2, a polycomb group protein that we and others have previously characterized as being overexpressed in aggressive prostate and breast cancer (Kleer et al., 2003; Varambally et al., 2002) was one of the 50 proteins identified as being concordantly overexpressed in metastatic tissues at the mRNA and protein level (Figures 3B and 3C). As *EZH2* was a member of this 50-gene concordant signature, we hypothesized that proteomic alterations that distinguish metastatic prostate cancer from clinically localized disease may serve as a multiplex signature of prostate cancer progression when applied to clinically localized disease (i.e., "more aggressive" genes would be expressed in progressive prostate cancer). While antibodies have yet to be developed to test all of these proteomic alterations in situ by immunohistochemistry, we postulated that mRNA transcript levels could be used instead due to their concordance with protein levels in this signature. To test this hypothesis, we selected prostate cancer gene expression data sets that monitored over 85% of the genes in the concordant genomic/proteomic signature and included biochemical recurrence information (time to PSA recurrence), as well as reported on at least 50 clinically localized specimens. According to Dobbin and Simon (2005), the number of samples required for developing prognostic markers was approximately 51 or above for a general human gene expression data set with the variance of a gene over samples as 0.5, type I error as 0.001, and type II error as 0.05. Thus, we chose $n = 50$ as our minimal sample size requirement in this analysis.


A
Latulippe et al.
Dhanasekaran et al.
Luo et al.
Lapointe et al.
Welsh et al.
Yu et al.
Singh et al.
Luo JH et al.
PCA vs. Benign (immunoblot)
PCA vs. Benign (immunoblot)
MYOSIN VI
AMACR
HSP60
CDK7
TPD52
BRG1
BUB3
PSA
MSH2
GS28
PICIN
AURORA KINASE A
RBBP
CK1
ERAB
XIAP
PTRF
EB1
INTEGRIN 5A
PI3 KINASE
PAK3
ABP280
MCAM
TROY
BMI-1
ACT1
PCNA
B2 BRAD RECEPTOR
PKC ALPHA
AIF
P16INK4A
JANUSIN
GKAP
XRCC
SAPK/JNK2
CASP 3
SEK1
BID
CATHEPSIN D-28KD
CSTF50
BAF170
MMP19
CARBOXYPEPTIDASE E
BM28
FAS LIGAND
UBC9
COCHAPERONE P23
KRIP-1
DRBP76
E2F3
CASP7
MYD88
DFF45
PAP
RAB27
ERK2
G ALPHA T
MMP23
CAB45
PDK1
GAS 7
M P P M
Increased in localized prostate cancer
Decreased in localized prostate cancer
M- mRNA profiling
P- Proteomic alteration
B
Latulippe et al.
Dhanasekaran et al.
Yu et al.
Lapointe et al.
Met vs. PCA (immunoblot)
Met vs. PCA (immunoblot)
AURORA A
EZH2
NP62
LAP2
RAL A
UBC9
EXPORTIN
P16INK4A
MSH2
VTI1B
UBA2
BAF170
KRIP-1
OCCLUDIN
MCM6
RAN
SAPK/JNK2
DRBP76
CSK
ERAB
STATHMIN
ICBP90
KPNA 2
CDK4
BM28
RAB27
NEXILIN
PSA
PAP
PROSTASIN
MLCK
SEK1
ACT1
MYPT1
LIM KINASE
CRP1
INTEGRIN 5A
ABP280
CASP7
MMP23
MSACTIN
GKAP
3PK
CIAP
CAVEOLIN 2-20KD
HIC-5
PAXILLIN
AMACR
A- B CRYSTALLIN
C-FLIP
CDK7
5-LIPOXYGENASE
B2 BRADY RECEPTOR
STRIATIN-119KD
CALNEXIN
BET1
ALPHA-CATENIN
PP120-102KD
PI31
P19SKP 1
BMI-1
TOPO2 BETA
PSME3-29KD
BID
PTP 1-BETA
RACK1
ALPHA-SYNUCLEIN
P190
STAT6
XIAP
JAM-1
EIF-6
CATHEPSIN D-28KD
BCL-X
BAFF
ANNEXIN II-34KD
PICIn
GFRALPHA2
NTF2
TIP120
NMT55
FACTP140
TYROSINASE
CASPASE 3
HP1 ALPHA
DEK
BUB3
REF1
PCNA
NHE-3
TGFALPHA
JANUSIN
SYNTAXIN 8
C-CBL
P67PHOX
IGFBP2
DCR1
DYNAMIN
BIM/BOD
DEMATIN
TAU-53KD
PECI
NEUROTENSIN R3
80K-H
ECA39
FKBP51
PAK3
MS CYTOKERATIN
CP E
XRCC
KANADAPTIN
TFII I
FKBP12
CSTF50
ERK2
MMP19
MYD88
B-DYSTROGLYCAN
GAS 7
TPD52
SHCTP 1
IRAK
MAP2B
CASP8
E2F3
HSP60
G ALPHA T
MCAM
PDK1
BRG1
RIN1
RBBP
PKC ALPHA
GS28
DR3
GSK3 BETA
TIAR
4F2 HC/CD98HC
S6K
PHOSPHOLIPASE C GAMMA 1
M P P M
Increased in metastatic prostate cancer
Decreased in metastatic prostate cancer
M- mRNA profiling
P- Proteomic alteration
C
BENGIN
PCA
MET
CONCORDANT
AURORA A
EZH2
NUCLEOPORIN P62
LAP2
RALA
UBC9
EXPORTIN
P16INK4A
MSH2
BAF 170
VTI 1B
SAPK/JNK
KRIP
ERAB
DRBP76
BM28
RAB27
MLCK
PSA
CRP1
INTEGRIN 5A
ABP280
CASPASE 7
3PK
LIM KINASE
PAXILLIN
CIAP
AMACR
C-FLIP
DISCORDANT
DCR1 26KD
TYROSINASE A
CASPASE 3
CATHEPSIN D
HP1 ALPHA
DEMATIN
B2 BRADYKININ R
XIAP
DYNAMIN
P67PHOX
BCL-X
BIM/BOD
BUB3
PCNA
ECA39
TGF ALPHA
BID
PTP1-BETA
RACK1
PICI n
CDK7
BMI-1
GAS7
RBBP
SHC HOMOLOG
MCAM
MMP-19
GS28
DR3
GSK3 BETA
E2F3
IRAK
BRG1
S6K

The prostate cancer gene expression data sets that fulfilled these criteria were carried out by Yu et al. (2004) and Glinsky et al. (2004), both of which represent Affymetrix oligonucleotide data sets and each of which measured at least 44 out of the 50 genes in the concordant signature. Although the Singh et al. and LaPointe et al. studies reported over 50 samples in their studies, the number of samples for which we have available follow-up information was less than 30 (29 and 20 samples for the LaPointe and Singh data sets, respectively). In addition, the average follow-up time for the samples in LaPointe study was only 10.7 months. Thus, we excluded both data sets in the analysis. We then chose to build our prediction models with the Yu et al. data set and test the performance on the Glinsky et al. data set. Utilizing an approach described earlier (Ramaswamy et al., 2003), unsupervised hierarchical clustering in the space of this 44-gene concordant signature resulted in two main clusters of individuals in the Yu et al. study (Figure 5A). Kaplan-Meier (KM) survival analysis of the clusters indicated that the two groups of individuals are significantly different based on time to recurrence status ($p = 0.035$; Figure 5A). Notably, when we use the 90 discordant genes (mRNA transcripts that are not qualitatively concordant with protein levels) we found that these signatures did not generate a clinical outcome distinction ($p = 0.238$; Supplemental Data). Moreover, by permutation test, we also observed that random sets of 44 genes did not generate such prognostic distinctions (see Experimental Procedures), indicating that our concordant signature was not likely due to chance. To assess the validity of this concordant signature, we utilized the Glinsky et al. study as an independent test set (Figure 5B). Each of the samples in the Glinsky data set was classified as high- or low-risk based on a *k*-nearest neighbor (*k*-NN) model developed using the Yu et al. study as a training set ($k = 3$). Based on the class predictions derived from the concordant signature, KM survival analysis revealed a significant difference in survival based on the risk stratification ($p = 0.001$; Figure 5B). As expected, this was not the case with the discordant signature when applied to the Glinsky et al. sample set ($p = 0.556$; Supplemental Data). A similar result was observed when a predictive model built on the Glinsky et al. data was applied to the Yu et al. data ($p < 0.001$ and $p = 0.02$ for the Glinsky et al. and Yu et al. data, respectively). We then carried out multivariate Cox proportional hazards regression analysis of the risk of recurrence on the Glinsky et al. validation set. Table 1 shows that the concordant signature predicted recurrence independently of the other clinical parameters such as surgical margin status, Gleason sum, and preoperative PSA. With an overall hazard ratio of 3.66 (95% CI: 1.36–7.02; $p < 0.001$), it was by far the strongest predictor of prostate cancer recurrence in the model.

Next, we sought to refine the concordant signature of prostate cancer progression by reducing the number of genes required. By using the Yu et al. study as a training set, we ranked the 44 concordant genes by a univariate Cox model. The same clustering procedure was employed to identify two clusters based on the top number of genes ranging from a minimum of 5 to a maximum of 44. Based on this iterative analysis, we identified nine genes that demarcated two main clusters that differed most significantly by KM survival analysis (Figure 5A; Experimental Procedures). The Glinsky et al. study was again used as an independent validation set confirming that the 9-gene concordant signature identified two groups of individuals that differed significantly based on recurrence (Figure 5B; Figure S5). Taken together, this integrative analysis suggests that mRNA transcripts that correlate with protein levels in metastatic prostate cancer can be used as gene predictors of progression in clinically localized disease.

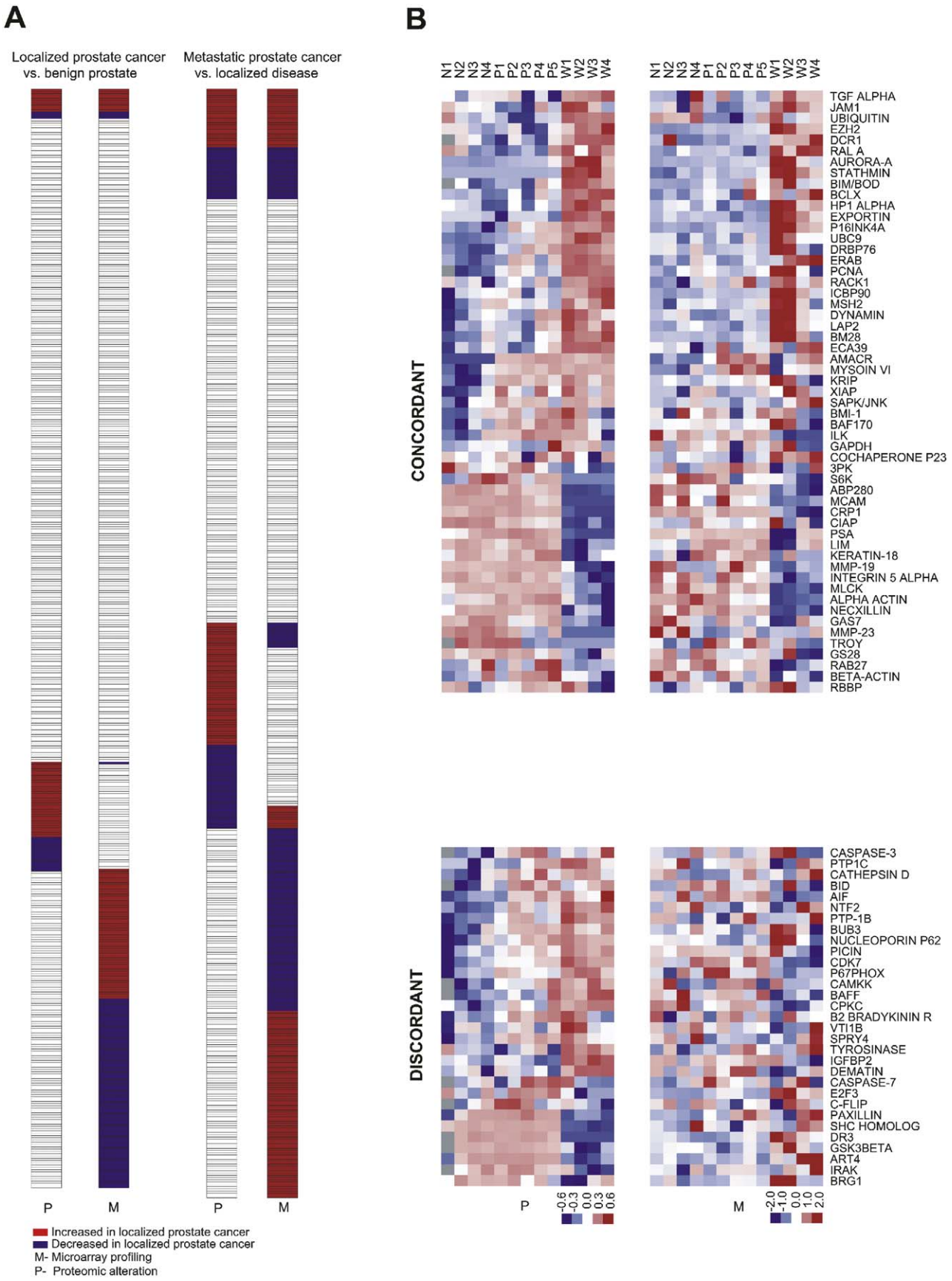
Next, we sought to explore the generality of the concordant progression signature in other solid tumors. We identified four tumor profiling data sets from the Oncomine compendium (Rhodes et al., 2004) that fulfilled the same criteria that we used in the prostate cancer analyses (see above). In 95 primary breast adenocarcinomas (van 't Veer et al., 2002), tumors bearing the 50-gene concordant progression signature were more likely to progress to metastasis than those lacking this signature ($p = 0.0025$; Supplemental Data). We observed a similar result in 80 primary breast infiltrating ductal carcinomas (Huang et al., 2003) ($p = 0.002$; Figure 5C). Moreover, this result was also observed in a series of 84 primary lung adenocarcinomas (Bhattacharjee et al., 2001) ($p = 0.03$; Figure 5C) and 56 gliomas (Freije et al., 2004) ($p = 0.01$; Figure 5C). Furthermore, we used two common gene expression prediction models (diagonal linear discriminant analysis and *k*-NN analysis) and conducted direct comparisons of the performances of the progression signature and the "study-specific" signature in each individual study where such a specific signature was available (see Table S12). The result indicated that the progression signature was able to retrieve similar or even superior prediction performance in most of the studies, especially when employing

Figure 3. Integrative analysis of proteomic and transcriptomic metadata in prostate cancer progression

A: Color map of integrative analysis relating protein alterations to gene expression in clinically localized prostate cancer relative to benign prostate tissue. For gene expression meta-analysis (transcript analysis), the first author of each prostate cancer gene expression study is indicated in columns, while individual genes are represented as rows. Red color indicates significantly increased expression at the $p = 0.05$ threshold level for prostate cancer relative to benign tissue, while blue indicates downregulation at the same threshold, and white indicates unchanged expression. Protein levels (protein) in pooled clinically localized prostate cancer extracts (as described in Figure 1) were qualified by high-throughput immunoblot analysis as overexpressed (red), underexpressed (blue), or unchanged (white) and mapped to the corresponding mRNA transcript. Proteins that were not expressed (or corresponding antibodies that did not produce an immunoreactive band of the correct molecular weight) or proteins whose corresponding mRNA transcript level was not present in over one-fourth of the profiling studies were excluded from the integrative analysis. Proteomic alterations in prostate cancer that were concordant or discordant with the meta-analysis of gene expression were expanded to the right. See the Experimental Procedures and Supplemental Data for details and complete data set.

B: As in **A**, except the integrative analysis was carried out between metastatic prostate cancer relative to clinically localized prostate cancer.

C: Conventional immunoblot validation of selected proteins differentially expressed between metastatic prostate cancer and clinically localized prostate cancer. Individual tissue extracts from three to four benign, five clinically localized prostate cancer, and five metastatic prostate cancer samples are shown. Validation of additional markers that distinguish clinically localized prostate cancer and benign tissue can be found in the Supplemental Data.


A
Localized prostate cancer vs. benign prostate
Metastatic prostate cancer vs. localized disease
P
M
P
M
Increased in localized prostate cancer
Decreased in localized prostate cancer
M- Microarray profiling
P- Proteomic alteration
B
CONCORDANT
N1 N2 N3 N4 P1 P2 P3 P4 P5 W1 W2 W3 W4
TGF ALPHA
JAM1
UBIQUITIN
EZH2
DCR1
RAL A
AURORA-A
STATHMIN
BIM/BOD
BCLX
HP1 ALPHA
EXPORTIN
P16INK4A
UBC9
DRBP76
ERAB
PCNA
RACK1
ICBP90
MSH2
DYNAMIN
LAP2
BM28
ECA39
AMACR
MYSOIN VI
KRIP
XIAP
SAPK/JNK
BMI-1
BAF170
ILK
GAPDH
COCHAPERONE P23
3PK
S6K
ABP280
MCAM
CRP1
CIAP
PSA
LIM
KERATIN-18
MMP-19
INTEGRIN 5 ALPHA
MLCK
ALPHA ACTIN
NECXILLIN
GAS7
MMP-23
TROY
GS28
RAB27
BETA-ACTIN
RBBP
DISCORDANT
CASPASE-3
PTP1C
CATHEPSIN D
BID
AIF
NTF2
PTP-1B
BUB3
NUCLEOPORIN P62
PICIN
CDK7
P67PHOX
CAMKK
BAFF
CPKC
B2 BRADYKININ R
VTI1B
SPRY4
TYROSINASE
IGFBP2
DEMATIN
CASPASE-7
E2F3
C-FLIP
PAXILLIN
SHC HOMOLOG
DR3
GSK3BETA
ART4
IRAK
BRG1
P
-0.6 -0.3 0.0 0.3 0.6
M
-2.0 -1.0 0.0 1.0 2.0

the *k*-NN prediction model. This is remarkable, as this signature was derived exclusively from prostate samples but had utility not only in prostate cancer data sets but also in breast cancer, lung cancer, and glioma data sets. Again, this suggests that there is likely biology inherent in the integrated predictor. Of note, we found that the smaller 9-gene model was only effective in discriminating prognostic classes in the Freije et al. glioma study ($p = 0.016$) but not in the other solid tumor data sets (Supplemental Data). This suggests that the 9-gene model may be relatively specific for prostate cancer while the 50-gene model has more universal applicability. Taken together, our observations suggest that the progression proteomic/genomic signature identified by the integrative analysis of metastatic prostate cancer may have utility in the prognostication of clinically localized solid tumors in general. Biologically, this suggests that aggressive tumors of different tissue origin begin to share the molecular machinery of a dedifferentiated state.

While these proteomic alterations have potential to serve as a multiplex biomarker of cancer aggressiveness, they may also shed light into the biology of neoplastic progression. As proteins, rather than RNA transcripts, are the primary effectors of the cell, they play the central and most distal role in the functional pathways to cancer. Interestingly, EZH2, which we previously have shown to have a role in prostate cancer progression (Varambally et al., 2002), is a member of this concordant genomic/proteomic signature, suggesting that other members of this signature may have utility as biomarkers as well as could have a role in the biology of progression. For example, this screen identified Aurora-A kinase (STK15) as being overexpressed in metastatic prostate cancer as well as being a member of the 50-gene concordant signature. This serine-threonine kinase has been shown to be amplified in a number of human cancers (Jeng et al., 2004; Neben et al., 2004), play a key role in G2/M cell cycle progression (Hirota et al., 2003), and inhibit p53 (Katayama et al., 2004), among other functions. Another candidate cancer regulatory molecule in the 50-gene concordant signature was KRIP1 (KAP-1), which is known to repress transcription via binding the methyltransferase SETDB1 (Schultz et al., 2002). Refer to the Supplemental Data for additional markers identified by this screen as well as pathway enrichment analysis.

In this study, we initially used a pooling strategy to perform high-throughput immunoblot analysis. While it would be more ideal to involve replicate protein measurements across multiple prostate tissues and then make comparisons to mRNA, the difficulty in monitoring thousands of antibodies on many individual samples and the cost of running multiple samples across thousands of antibodies required us to adopt the pooling approach. Further, analyses of concordance with mRNA expression on individual samples that comprised the pool confirmed the general feasibility of this strategy. We also noticed that there were recognized problems with annotations for microarrays. A recent study (Mecham et al., 2004) reported that up to 50% Affymetrix probes do not have a matching sequence in the Reference Sequence database (Refseq), questioning the reliability of such probes. As this study represents an initial foray in the area of integrative analyses, we used basic gene identifier-based matching for cross-platform annotations. Another potential limitation in the present study is that many immunoblots exhibit reactivity at multiple sizes potentially representing multiple protein isoforms. Thus, measuring the protein intensity for one "expected" band may not be adequate for determining a correlation with transcripts. However, most of the reported changes here are the result of alterations in the reported or predicted molecular weight isoform. In future studies, we will investigate the various isoforms and proteolytically cleaved products.

Taken together, the present study provides a general framework for the integrative analysis of proteomic and transcriptomic data from human tumors (Figure 6). Proteomic profiling of prostate cancer progression identified over 100 altered proteins in the transition from clinically localized to metastatic disease (a significant fraction of which were androgen regulated). While this approach was useful to integrate high-throughput immunoblot data, the general paradigm can also be applied to mass spectrometry or protein microarray-based technologies as they mature in the future. Differential proteins were then mapped to mRNA transcript levels to assess mRNA/protein concordance levels in a human disease state. Importantly, gene expression alterations that matched protein alterations qualitatively could be used as predictors of prostate cancer progression in clinically confined disease. Thus, this would suggest that clinically aggressive prostate cancer bears a "signature" set of genes/proteins that is characteristic of metastatic disease. The observation that the concordant proteomic/genomic signature can be applied to other solid tumors suggests commonalities in the undifferentiated state of advanced tumors.

Experimental procedures

High-throughput immunoblot analysis

Tissues utilized were from the radical prostatectomy series at the University of Michigan and from the Rapid Autopsy Program, which are both part of University of Michigan Prostate Cancer Specialized Program of Research Excellence (SPORE) Tissue Core. Institutional Review Board approval was obtained to procure and analyze the tissues used in this study. To develop the tissue extract pools, the following frozen tissue blocks were identified: five each of benign prostate tissues, clinically localized prostate cancer, and hormone-refractory metastatic tissues (Shah et al., 2004). Based on examination of the frozen sections of each tissue block, specimens were

Figure 4. Integrative genomic and proteomic analysis of pooled and individual prostate tissue extracts

A: Color maps of integrative analyses relating protein alterations observed in pooled tissues by immunoblotting and transcript alterations observed in the pooled and individual tissues by gene expression analyses. Please refer to Figure 3A for color scheme.

B: Color maps depicting integrative genomic and proteomic analysis of individual prostate tissue samples. Proteins in each tissue extract were assessed based on intensities derived from conventional immunoblot analysis. We focused on 86 distinct proteins identified as differential based on high-throughput analysis (Figures 3 and 4A). Transcriptomic profiles from the same samples were derived from Affymetrix microarrays. The immunoblot intensities were semiquantitated, and a correlation was calculated for each protein. Concordance was defined based on positive correlation between proteins and transcripts (see the Experimental Procedures and Supplemental Experimental Procedures).

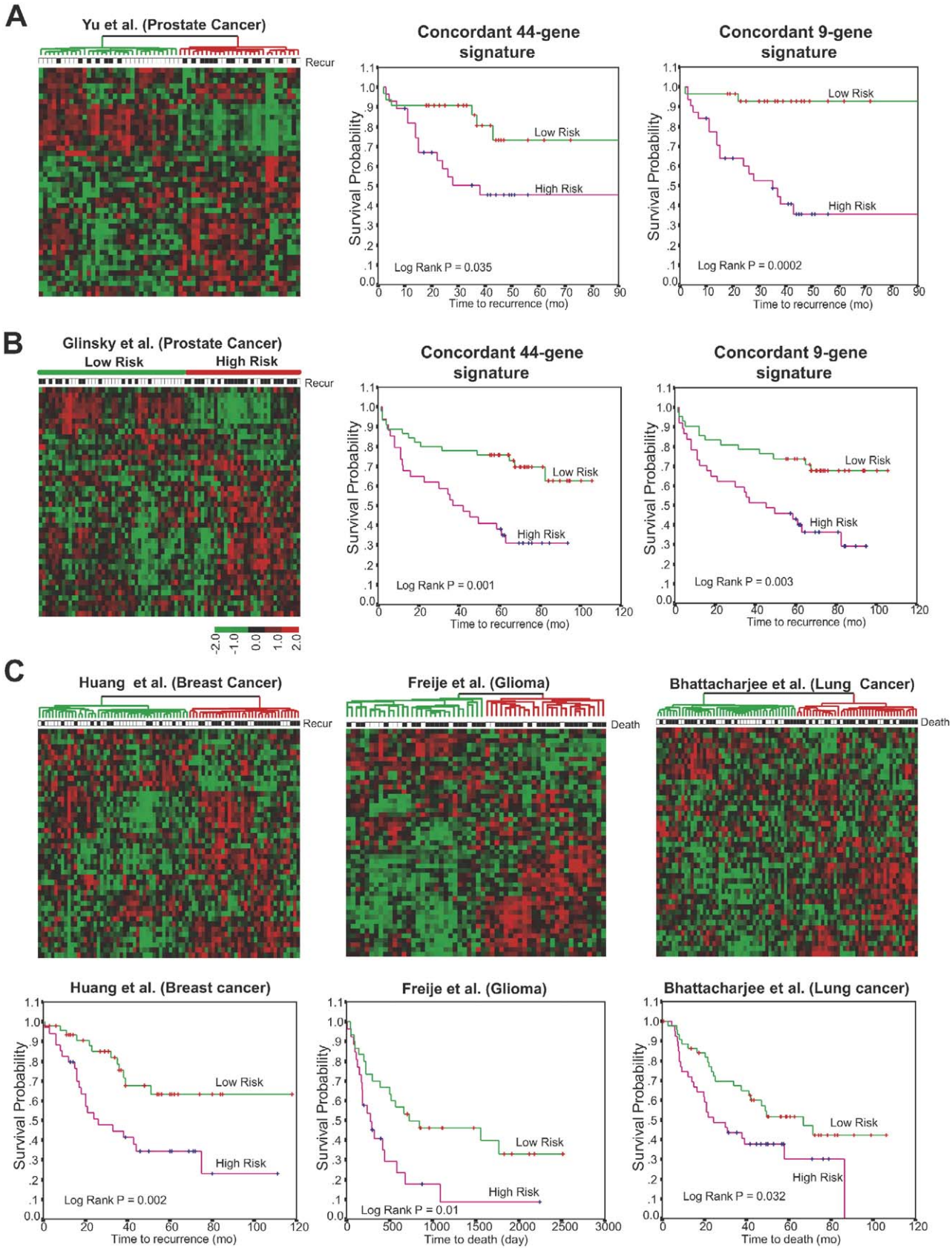

A
Yu et al. (Prostate Cancer)
Recur
Concordant 44-gene signature
Concordant 9-gene signature
Survival Probability
Low Risk
High Risk
Log Rank P = 0.035
Log Rank P = 0.0002
Time to recurrence (mo)
B
Glinsky et al. (Prostate Cancer)
Low Risk
High Risk
Recur
Concordant 44-gene signature
Concordant 9-gene signature
Log Rank P = 0.001
Log Rank P = 0.003
-2.0
-1.0
0.0
1.0
2.0
C
Huang et al. (Breast Cancer)
Freije et al. (Glioma)
Bhattacharjee et al. (Lung Cancer)
Death
Huang et al. (Breast cancer)
Freije et al. (Glioma)
Bhattacharjee et al. (Lung cancer)
Log Rank P = 0.002
Log Rank P = 0.01
Log Rank P = 0.032
Time to death (day)
Time to death (mo)

Table 1. Multivariable proportional hazards analysis of the risk of recurrence as a first event on the Glinsky et al. validation set

Variable	Hazard ratio (95% CI)	p value
High-risk signature (versus low-risk signature)	3.66 (1.77–7.59)	<0.001
PSA	1.04 (1.00–1.09)	0.043
Gleason sum score: score > 7 (versus score ≤ 7)	1.73 (0.79–3.76)	0.17
Tumor stage: stage T2 (versus stage T1)	0.85 (0.42–1.75)	0.67
Age	1.06 (1.00–1.13)	0.06
Surgical margins: positive (versus negative)	2.18 (0.92–5.18)	0.08

grossly dissected maintaining at least 90% of the tissue of interest. Total proteins were extracted from each tissue by homogenizing samples in boiling lysis buffer. One hundred micrograms of protein from each tissue extract pool was boiled in sample buffer and subjected to 4%–15% preparative SDS-PAGE and transferred to PVDF and probed with different antibodies. To supplement the number of proteins analyzed, the same extracts were analyzed using two commercial service providers, BD Biosciences and Kinexus. Detailed methods for high-throughput immunoblot analyses are provided in the Supplemental Data. Validation immunoblots for selected proteins in different functional classes were carried out using 4%–15% linear gradient SDS-PAGE for protein separation. The signal intensities were semiquantitated using Scion Image software.

Microarray analyses

Total RNA from the individual and pooled samples were analyzed on Affymetrix U133 2.0 Plus arrays by the University of Michigan Comprehensive Cancer Center Affymetrix Core. The amount and integrity of RNA were analyzed by spectrophotometry and the Agilent Bioanalyzer (Agilent Technologies). Biotin-labeled cRNA synthesis, hybridization, washing, staining, and scanning were done following the manufacturer's protocols (Affymetrix). All RNA samples and arrays met standard quality control metrics. The expression profiling data were deposited into GEO under accession number GSE3325 (available at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE3325).

Tissue microarray analysis

A prostate cancer progression tissue microarray analysis composed of benign prostate tissue, clinically localized prostate cancer, and hormone-refractory metastatic prostate cancer was developed. These cases came from well-fixed radical prostatectomy specimens as described previously (Rubin et al., 2002). Protein expression was determined using a validated scoring method (Dhanasekaran et al., 2001; Rubin et al., 2002; Varambally et al., 2002) where staining was evaluated for intensity and the percentage of cells staining positive. Detailed methods for tissue microarray analyses can be found in the Supplemental Data.

Integrative molecular analysis

To map the antibodies and their respective protein targets, we retrieved the official gene names from the NCBI LocusLink for our antibody/protein lists. To complement protein levels, transcriptome data were assembled from eight publicly available prostate cancer gene expression data sets (Dhanasekaran et al., 2001; Lapointe et al., 2004; LaTulippe et al., 2002; Luo et al., 2001, 2002b; Singh et al., 2002; Welsh et al., 2001; Yu et al., 2004), and each probe was mapped to Unigene Build #173 (Table S3). Expression values from multiple clones or probe sets mapping to the same Unigene Cluster ID were averaged. Each gene in each study was normalized across samples so that the mean equaled zero and the standard deviation equaled 1. Missing data were imputed by the k-NN (k = 5) imputation approach (Troyanskaya et al., 2001).

Eight prostate cancer profiling studies were included in the analysis of clinically localized prostate cancer relative to benign prostate tissue, while only four studies were included in the analysis of metastatic prostate cancer versus localized prostate cancer due to the availability of metastatic samples in those studies (Table S4). Detailed methods for integrative molecular analysis can be found in the Supplemental Data.

Integrative genomic and proteomic analysis of individual prostate cancer samples

We carried out profiling of mRNA expression analysis in 13 of the 14 individual samples used for the individual protein measurements (one was excluded due to an insufficient amount of tissue). We examined the concordance between proteins and transcripts for individual samples, focusing on the 86 proteins identified as outliers in the larger high-throughput screen (see Figure S1). The immunoblot intensities were semiquantitated using Scion Image software, and the Spearman's rank correlation was calculated for each protein. An mRNA transcript alteration was considered "concordant" with a proteomic alteration if a positive correlation was found.

Clinical outcomes analysis

Six different cancer profiling studies (Bhattacharjee et al., 2001; Freije et al., 2004; Glinsky et al., 2004; Huang et al., 2003; van 't Veer et al., 2002; Yu et al., 2004) were used for evaluation of prognostic value of these concordant genes. Detailed study information is shown in Table S5, and gene expression data utilized for the study are described in Table S6. Average linkage hierarchical clustering using an uncentered correlation similarity metric was used to identify two main clusters of clinically localized prostate cancer samples based on the 44 concordant mRNA transcripts that were qualita-

Figure 5. Proteomic alterations in metastatic prostate cancer identify gene predictors of cancer aggressiveness

A: A concordant 44 (out of 50)-gene predictor was developed based on proteomic alterations that were concordant with gene expression (Figure 3B) and subsequently evaluated for prognostic utility on a prostate cancer gene expression data set (Yu et al.). Hierarchical clustering of the tumor samples (columns) and genes (rows) is provided (left panel). Red indicates high relative levels of gene expression, while green represents low relative levels of gene expression. Horizontal bars above the heat maps indicate the recurrence status of each patient (black box, biochemical or tumor recurrence; white box, recurrence-free). Patients were categorized into two major clusters defined by the 44-gene signature. The prediction model was further refined to a 9-gene signature (see Supplemental Data). Kaplan-Meier survival analysis based on the groups defined by the 44-gene concordant cluster (middle panel) and the 9-gene concordant cluster (right panel). Similar analyses performed with discordant genes can be found in the Supplemental Data.

B: The concordant 44-gene predictor and the refined concordant 9-gene predictor were evaluated in an independent prostate cancer profiling data set. Each sample was assigned to a low-risk or high-risk group by k-nearest neighbor classification using cluster-defined low-/high-risk groups in the Yu et al. training data set (left panel; Supplemental Experimental Procedures). Kaplan-Meier plot of the predicted high-/low-risk groups in the space of the concordant 44 genes (middle panel) or the concordant 9 genes (right panel).

C: Same as **A**, except the concordant predictor was evaluated in other solid tumors. Huang et al. (2003) breast adenocarcinoma (left panel), Freije et al. (2004) glioma (middle panel), and Bhattacharjee et al. (2001) lung adenocarcinoma (right). Analysis using the van 't Veer breast cancer profiling data set (van 't Veer et al., 2002) can be found in the Supplemental Data.

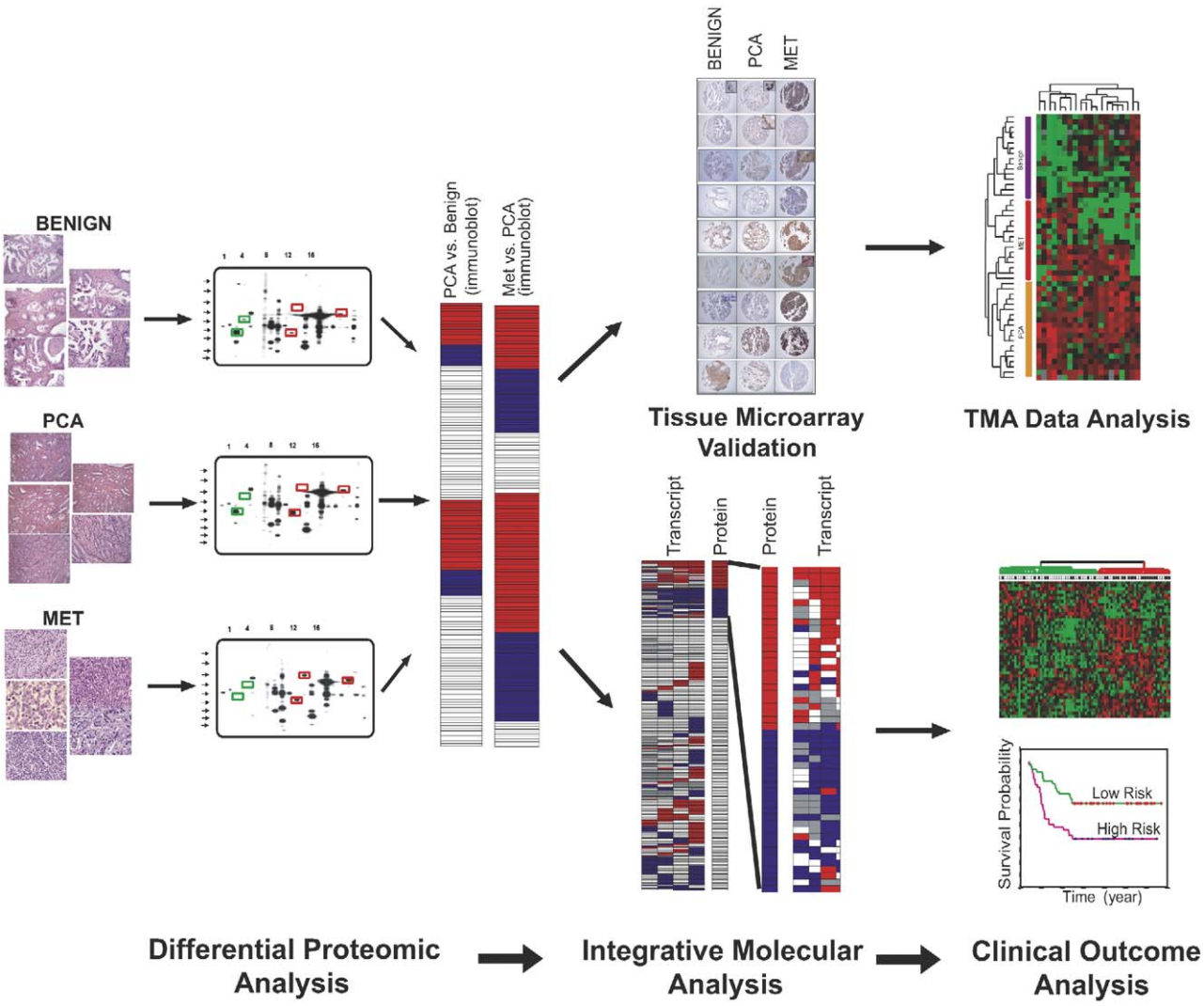


Figure 6. Integrative molecular analysis of cancer to identify gene predictors of clinical outcome

Proteomic profiles comparing metastatic prostate cancer to clinically localized prostate cancer were used to identify a composite gene predictor of clinical outcome in localized disease. This integrated proteomic-transcriptomic signature represents a prostate cancer progression signature and can be extended to other solid tumors.

tively concordant with protein expression in the Yu et al. (2004) study (only 44 out of 50 of the concordant signature were assessed on these arrays).

To validate the prognostic association of the 44-gene concordant signature, an independent (clinically localized) prostate cancer gene expression data set from Glinsky et al. (Glinsky et al., 2004) was used. The Yu et al. clustering functioned as the "training set" to define high-/low-risk groups. This "progression" signature was then refined by reducing the number of genes involved. By using Yu et al. study as a training set, we ranked the concordant genes by univariate Cox model. The generality of this "progression" signature was evaluated by using other solid tumor data sets. The signature was applied to two breast cancer (Huang et al., 2003; van 't Veer et al., 2002), one lung cancer (Bhattacharjee et al., 2001), and one glioma (Freije et al., 2004) gene expression study. Detailed methods for clinical outcomes analysis can be found in the Supplemental Data.

Multivariable analysis

We used a Cox proportional hazards regression model to carry out the multivariate analysis. The dichotomized values of the concordant "progression" signature, preoperative PSA, Gleason sum score from prostatectomy specimens, preoperative clinical stage, age, and status of surgical margins were included as covariates. The calculation was performed with the R 2.0 statistical package.

Pathway analysis

To better understand the biological pathways at work in the concordant and discordant signature, we searched for association of these genes with gene sets defined by Gene Ontology and Transfac analysis (Rhodes et al., 2005). The overlap of the signature with each gene set was counted, and the significance of the overlap was evaluated with Fisher's exact test.

Supplemental data

The Supplemental Data include supplemental results, discussion, and experimental procedures; six supplemental figures; and 12 supplemental tables. The Supplemental Data can be found with this article online at http://www.cancercell.org/cgi/content/full/8/5/393/DC1/.

Acknowledgments

The authors would like to thank Dr. William Gerald (Memorial Sloan-Kettering Cancer Center, NY) for providing the primary gene expression data files

from MSKCC as well as the other authors who made their microarray data sets publicly available. We would also like to thank Jill Macoska and Joe Washburn of the UMCCC microarray core for timely microarray analysis; James Montie and Francesca Demichelis for suggestions; Vasudeva Mahavisno, Shanker Kalyana-Sundaram, Terrence Barrette, and Doug Gibbs for bioinformatics support; and Jason Harwood, Prakash Chinnaiyan, Mukesh Nyati, and Qi Cao for technical support. We would like to thank the BD power blot team and Kinexus for technical assistance. We would like to thank the staff of the Microscopy and Image Analyses laboratory at the University of Michigan for their help in the various microscopic analyses employed in this study. We would like to thank StressGen Biotechnologies for their kind gifts of antibodies. This research was supported in part by the Department of Defense (PC051081 to A.M.C. and S.V.; PC040517 to R.M.), the American Cancer Society (RSG-02-179-MGO to A.M.C., M.A.R., and D.G.), the National Institutes of Health (Prostate SPORE P50CA69568 to K.J.P., S.V., A.M.C., J.T.W., D.G., and R.B.S.; R01AG21404 to M.A.R.; R01GM72007-01 to D.G.; and Early Detection Research Network UO1 CA111275-01 to A.M.C., J.T.W., and M.A.R.), and the UM Cancer Center Support Grant (5P30 CA46592) funding of the UMCC Bioinformatics Core. A.M.C. is a Pew Biomedical Scholar. D.R.R. and S.A.T. are Fellows of the Medical Scientist Training Program, and D.R.R. is supported by the Cancer Biology Training Program.

Received: April 8, 2005
Revised: July 25, 2005
Accepted: October 5, 2005
Published: November 14, 2005

References

Albertsen, P.C., Hanley, J.A., Gleason, D.F., and Barry, M.J. (1998). Competing risk analysis of men aged 55 to 74 years at diagnosis managed conservatively for clinically localized prostate cancer. JAMA *280*, 975–980.

Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., et al. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature *403*, 503–511.

Baliga, N.S., Pan, M., Goo, Y.A., Yi, E.C., Goodlett, D.R., Dimitrov, K., Shannon, P., Aebersold, R., Ng, W.V., and Hood, L. (2002). Coordinate regulation of energy transduction modules in Halobacterium sp. analyzed by a global systems approach. Proc. Natl. Acad. Sci. USA *99*, 14913–14918.

Bhattacharjee, A., Richards, W.G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M., et al. (2001). Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc. Natl. Acad. Sci. USA *98*, 13790–13795.

Chan, J.M., Jou, R.M., and Carroll, P.R. (2004). The relative impact and future burden of prostate cancer in the United States. J. Urol. *172*, S13–S16.

Culig, Z., Hobisch, A., Hittmair, A., Peterziel, H., Cato, A.C., Bartsch, G., and Klocker, H. (1998). Expression, structure, and function of androgen receptor in advanced prostatic carcinoma. Prostate *35*, 63–70.

DePrimo, S.E., Diehn, M., Nelson, J.B., Reiter, R.E., Matese, J., Fero, M., Tibshirani, R., Brown, P.O., and Brooks, J.D. (2002). Transcriptional programs activated by exposure of human prostate cancer cells to androgen. Genome Biol. *3*, RESEARCH0032.

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.

Dobbin, K., and Simon, R. (2005). Sample size determination in microarray experiments for class comparison and prognostic classification. Biostatistics *6*, 27–38.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA *95*, 14863–14868.

Freije, W.A., Castro-Vargas, F.E., Fang, Z., Horvath, S., Cloughesy, T., Liau, L.M., Mischel, P.S., and Nelson, S.F. (2004). Gene expression profiling of gliomas strongly predicts survival. Cancer Res. *64*, 6503–6510.

Glinsky, G.V., Glinskii, A.B., Stephenson, A.J., Hoffman, R.M., and Gerald, W.L. (2004). Gene expression profiling predicts clinical outcome of prostate cancer. J. Clin. Invest. *113*, 913–923.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., et al. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science *286*, 531–537.

Griffin, T.J., Gygi, S.P., Ideker, T., Rist, B., Eng, J., Hood, L., and Aebersold, R. (2002). Complementary profiling of gene expression at the transcriptome and proteome levels in Saccharomyces cerevisiae. Mol. Cell. Proteomics *1*, 323–333.

Grubb, R.L., Calvert, V.S., Wulkuhle, J.D., Paweletz, C.P., Linehan, W.M., Phillips, J.L., Chuaqui, R., Valasco, A., Gillespie, J., Emmert-Buck, M., et al. (2003). Signal pathway profiling of prostate cancer using reverse phase protein arrays. Proteomics *3*, 2142–2146.

Hedenfalk, I., Duggan, D., Chen, Y., Radmacher, M., Bittner, M., Simon, R., Meltzer, P., Gusterson, B., Esteller, M., Kallioniemi, O.P., et al. (2001). Gene-expression profiles in hereditary breast cancer. N. Engl. J. Med. *344*, 539–548.

Hirota, T., Kunitoku, N., Sasayama, T., Marumoto, T., Zhang, D., Nitta, M., Hatakeyama, K., and Saya, H. (2003). Aurora-A and an interacting activator, the LIM protein Ajuba, are required for mitotic commitment in human cells. Cell *114*, 585–598.

Hood, L., Heath, J.R., Phelps, M.E., and Lin, B. (2004). Systems biology and new technologies enable predictive and preventative medicine. Science *306*, 640–643.

Huang, E., Cheng, S.H., Dressman, H., Pittman, J., Tsou, M.H., Horng, C.F., Bild, A., Iversen, E.S., Liao, M., Chen, C.M., et al. (2003). Gene expression predictors of breast cancer outcomes. Lancet *361*, 1590–1596.

Jeng, Y.M., Peng, S.Y., Lin, C.Y., and Hsu, H.C. (2004). Overexpression and amplification of Aurora-A in hepatocellular carcinoma. Clin. Cancer Res. *10*, 2065–2071.

Jiang, Z., Woda, B.A., Rock, K.L., Xu, Y., Savas, L., Khan, A., Pihan, G., Cai, F., Babcook, J.S., Rathanaswami, P., et al. (2001). P504S: a new molecular marker for the detection of prostate carcinoma. Am. J. Surg. Pathol. *25*, 1397–1404.

Johansson, J.E., Holmberg, L., Johansson, S., Bergstrom, R., and Adami, H.O. (1997). Fifteen-year survival in prostate cancer. A prospective, population-based study in Sweden. JAMA *277*, 467–471.

Katayama, H., Sasai, K., Kawai, H., Yuan, Z.M., Bondaruk, J., Suzuki, F., Fujii, S., Arlinghaus, R.B., Czerniak, B.A., and Sen, S. (2004). Phosphorylation by aurora kinase A induces Mdm2-mediated destabilization and inhibition of p53. Nat. Genet. *36*, 55–62.

Kleer, C.G., Cao, Q., Varambally, S., Shen, R., Ota, I., Tomlins, S.A., Ghosh, D., Sewalt, R.G., Otte, A.P., Hayes, D.F., et al. (2003). EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. Proc. Natl. Acad. Sci. USA *100*, 11606–11611.

Koivisto, P., Kolmer, M., Visakorpi, T., and Kallioniemi, O.P. (1998). Androgen receptor gene and hormonal therapy failure of prostate cancer. Am. J. Pathol. *152*, 1–9.

Kononen, J., Bubendorf, L., Kallioniemi, A., Barlund, M., Schraml, P., Leighton, S., Torhorst, J., Mihatsch, M.J., Sauter, G., and Kallioniemi, O.P. (1998). Tissue microarrays for high-throughput molecular profiling of tumor specimens. Nat. Med. *4*, 844–847.

Krajewska, M., Krajewski, S., Banares, S., Huang, X., Turner, B., Bubendorf, L., Kallioniemi, O.P., Shabaik, A., Vitiello, A., Peehl, D., et al. (2003). Elevated expression of inhibitor of apoptosis proteins in prostate cancer. Clin. Cancer Res. *9*, 4914–4925.

Kumar-Sinha, C., and Chinnaiyan, A.M. (2003). Molecular markers to identify patients at risk for recurrence after primary treatment for prostate cancer. Urology Suppl. *62*, 19–35.

Lapointe, J., Li, C., Higgins, J.P., van de Rijn, M., Bair, E., Montgomery, K., 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.

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.

Linton, K.D., and Hamdy, F.C. (2003). Early diagnosis and surgical management of prostate cancer. Cancer Treat. Rev. *29*, 151–160.

Luo, J., Duggan, D.J., Chen, Y., Sauvageot, J., Ewing, C.M., Bittner, M.L., Trent, J.M., and Isaacs, W.B. (2001). Human prostate cancer and benign prostatic hyperplasia: molecular dissection by gene expression profiling. Cancer Res. *61*, 4683–4688.

Luo, J., Zha, S., Gage, W.R., Dunn, T.A., Hicks, J.L., Bennett, C.J., Ewing, C.M., Platz, E.A., Ferdinandusse, S., Wanders, R.J., et al. (2002a). α-methylacyl-CoA racemase: a new molecular marker for prostate cancer. Cancer Res. *62*, 2220–2226.

Luo, J.H., Yu, Y.P., Cieply, K., Lin, F., Deflavia, P., Dhir, R., Finkelstein, S., Michalopoulos, G., and Becich, M. (2002b). Gene expression analysis of prostate cancers. Mol. Carcinog. *33*, 25–35.

Magee, J.A., Araki, T., Patil, S., Ehrig, T., True, L., Humphrey, P.A., Catalona, W.J., Watson, M.A., and Milbrandt, J. (2001). Expression profiling reveals hepsin overexpression in prostate cancer. Cancer Res. *61*, 5692–5696.

Mecham, B.H., Klus, G.T., Strovel, J., Augustus, M., Byrne, D., Bozso, P., Wetmore, D.Z., Mariani, T.J., Kohane, I.S., and Szallasi, Z. (2004). Sequence-matched probes produce increased cross-platform consistency and more reproducible biological results in microarray-based gene expression measurements. Nucleic Acids Res. *32*, e74.

Mooradian, A.D., Morley, J.E., and Korenman, S.G. (1987). Biological actions of androgens. Endocr. Rev. *8*, 1–28.

Neben, K., Korshunov, A., Benner, A., Wrobel, G., Hahn, M., Kokocinski, F., Golanov, A., Joos, S., and Lichter, P. (2004). Microarray-based screening for molecular markers in medulloblastoma revealed STK15 as independent predictor for survival. Cancer Res. *64*, 3103–3111.

Nelson, P.S., Clegg, N., Arnold, H., Ferguson, C., Bonham, M., White, J., Hood, L., and Lin, B. (2002). The program of androgen-responsive genes in neoplastic prostate epithelium. Proc. Natl. Acad. Sci. USA *99*, 11890–11895.

Nielsen, T.O., Hsu, F.D., O'Connell, J.X., Gilks, C.B., Sorensen, P.H., Linn, S., West, R.B., Liu, C.L., Botstein, D., Brown, P.O., and van de Rijn, M. (2003). Tissue microarray validation of epidermal growth factor receptor and SALL2 in synovial sarcoma with comparison to tumors of similar histology. Am. J. Pathol. *163*, 1449–1456.

Paweletz, C.P., Charboneau, L., Bichsel, V.E., Simone, N.L., Chen, T., Gillespie, J.W., Emmert-Buck, M.R., Roth, M.J., Petricoin, I.E., and Liotta, L.A. (2001). Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. Oncogene *20*, 1981–1989.

Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., et al. (2000). Molecular portraits of human breast tumours. Nature *406*, 747–752.

Petricoin, E.F., III, Ornstein, D.K., Paweletz, C.P., Ardekani, A., Hackett, P.S., Hitt, B.A., Velassco, A., Trucco, C., Wiegand, L., Wood, K., et al. (2002). Serum proteomic patterns for detection of prostate cancer. J. Natl. Cancer Inst. *94*, 1576–1578.

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.

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 integrated data-mining platform. Neoplasia *6*, 1–6.

Rhodes, D.R., Kalyana-Sundaram, S., Mahavisno, V., Barrette, T.R., Ghosh, D., and Chinnaiyan, A.M. (2005). Mining for regulatory programs in the cancer transcriptome. Nat. Genet. *37*, 579–583.

Rubin, M.A., Zhou, M., Dhanasekaran, S.M., Varambally, S., Barrette, T.R., Sanda, M.G., Pienta, K.J., Ghosh, D., and Chinnaiyan, A.M. (2002). α-methylacyl coenzyme A racemase as a tissue biomarker for prostate cancer. JAMA *287*, 1662–1670.

Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G., and Rauscher, F.J., III. (2002). SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. Genes Dev. *16*, 919–932.

Segawa, T., Nau, M.E., Xu, L.L., Chilukuri, R.N., Makarem, M., Zhang, W., Petrovics, G., Sesterhenn, I.A., McLeod, D.G., Moul, J.W., et al. (2002). Androgen-induced expression of endoplasmic reticulum (ER) stress response genes in prostate cancer cells. Oncogene *21*, 8749–8758.

Shah, R.B., Mehra, R., Chinnaiyan, A.M., Shen, R., Ghosh, D., Zhou, M., Macvicar, G.R., Varambally, S., Harwood, J., Bismar, T.A., et al. (2004). Androgen-independent prostate cancer is a heterogeneous group of diseases: lessons from a rapid autopsy program. Cancer Res. *64*, 9209–9216.

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.

Tian, Q., Stepaniants, S.B., Mao, M., Weng, L., Feetham, M.C., Doyle, M.J., Yi, E.C., Dai, H., Thorsson, V., Eng, J., et al. (2004). Integrated genomic and proteomic analyses of gene expression in mammalian cells. Mol. Cell. Proteomics *3*, 960–969.

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.

van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature *415*, 530–536.

Varambally, S., Dhanasekaran, S.M., Zhou, M., Barrette, T.R., Kumar-Sinha, C., Sanda, M.G., Ghosh, D., Pienta, K.J., Sewalt, R.G., Otte, A.P., et al. (2002). The polycomb group protein EZH2 is involved in progression of prostate cancer. Nature *419*, 624–629.

Velasco, A.M., Gillis, K.A., Li, Y., Brown, E.L., Sadler, T.M., Achilleos, M., Greenberger, L.M., Frost, P., Bai, W., and Zhang, Y. (2004). Identification and validation of novel androgen-regulated genes in prostate cancer. Endocrinology *145*, 3913–3924.

Waghray, A., Feroze, F., Schober, M.S., Yao, F., Wood, C., Puravs, E., Krause, M., Hanash, S., and Chen, Y.Q. (2001). Identification of androgen-regulated genes in the prostate cancer cell line LNCaP by serial analysis of gene expression and proteomic analysis. Proteomics *1*, 1327–1338.

Washburn, M.P., Koller, A., Oshiro, G., Ulaszek, R.R., Plouffe, D., Deciu, C., Winzeler, E., and Yates, J.R., III. (2003). Protein pathway and complex clustering of correlated mRNA and protein expression analyses in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA *100*, 3107–3112.

Welsh, J.B., Sapinoso, L.M., Su, A.I., Kern, S.G., Wang-Rodriguez, J., Moskaluk, C.A., Frierson, H.F., Jr., and Hampton, G.M. (2001). Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. Cancer Res. *61*, 5974–5978.

Xu, L.L., Su, Y.P., Labiche, R., Segawa, T., Shanmugam, N., McLeod, D.G., Moul, J.W., and Srivastava, S. (2001). Quantitative expression profile of androgen-regulated genes in prostate cancer cells and identification of prostate-specific genes. Int. J. Cancer *92*, 322–328.

Yu, Y.P., Landsittel, D., Jing, L., Nelson, J., Ren, B., Liu, L., McDonald, C., Thomas, R., Dhir, R., Finkelstein, S., et al. (2004). Gene expression alterations in prostate cancer predicting tumor aggression and preceding development of malignancy. J. Clin. Oncol. *22*, 2790–2799.