nature genetics

Integrative molecular concept modeling of prostate cancer progression

Scott A Tomlins[1,8], Rohit Mehra[1,2,8], Daniel R Rhodes[1–3,8], Xuhong Cao[1], Lei Wang[1], Saravana M Dhanasekaran[1], Shanker Kalyana-Sundaram[1], John T Wei[2,4], Mark A Rubin[5,6], Kenneth J Pienta[2,4,7], Rajal B Shah[1,2,4] & Arul M Chinnaiyan[1–4]

Despite efforts to profile prostate cancer, the genetic alterations and biological processes that correlate with the observed histological progression are unclear. Using laser-capture microdissection to isolate 101 cell populations, we have profiled prostate cancer progression from benign epithelium to metastatic disease. By analyzing expression signatures in the context of over 14,000 'molecular concepts', or sets of biologically connected genes, we generated an integrative model of progression. Molecular concepts that demarcate critical transitions in progression include protein biosynthesis, E26 transformation-specific (ETS) family transcriptional targets, androgen signaling and cell proliferation. Of note, relative to low-grade prostate cancer (Gleason pattern 3), high-grade cancer (Gleason pattern 4) shows an attenuated androgen signaling signature, similar to metastatic prostate cancer, which may reflect dedifferentiation and explain the clinical association of grade with prognosis. Taken together, these data show that analyzing gene expression signatures in the context of a compendium of molecular concepts is useful in understanding cancer biology.

Prostate cancer is the most common noncutaneous malignancy in American men[1]. Although numerous groups have profiled prostate cancer using DNA microarrays (reviewed in ref. 2), genetic changes and biological processes mediating important transitions in progression remain undefined. For example, because of the difficulty in profiling small lesions, little is known about gene expression in the putative precursor lesions prostatic intraepithelial neoplasia (PIN) and proliferative inflammatory atrophy (PIA)[3–6]. Prostate cancer is most commonly graded using the Gleason grading system[7], which relies entirely on the architectural pattern of cancerous glands (with 1 as the most differentiated and 5 as the least differentiated). As prostate cancer is often multifocal, the overall Gleason score is the sum of the two most prevalent patterns, and affected individuals with a higher Gleason score tend to have more aggressive cancer[8]. Despite attempts to identify genetic signatures distinguishing low– and high–Gleason grade cancer, different signatures show little overlap between individual genes, and the processes driving the different architectural patterns are unknown[9–14]. Furthermore, the relationship between different Gleason grades of clinically localized and metastatic prostate cancer is unclear.

Traditionally, expression profiling analysis has focused on identifying individual genes dysregulated during the disease process. More recently, several techniques, including gene set enrichment analysis and other 'modular' approaches, have been developed to identify sets of dysregulated genes that share a biological function[15–19]. Here, we analyzed prostate cancer progression using an alternative resource, the Molecular Concept Map (MCM) (D.R.R. *et al.*, unpublished data), an analytical framework for exploring the network of relationships among a growing collection of 'molecular concepts', or biologically related gene sets. The MCM is the largest collection of gene sets for association analysis, and it is unique in that it computes pairwise associations among all gene sets in the database, allowing for the identification and visualization of 'enrichment networks' of linked concepts. This is especially useful for complex gene expression signatures. Integration of these signatures with the MCM allowed us to systematically link them to over 14,000 molecular concepts.

RESULTS

Profiling prostate cancer progression at the cellular level

We used laser-capture microdissection (LCM) to isolate 101 specific cell populations from 44 individuals, representing prostate cancer progression as described in Methods. Isolated total RNA was amplified by TransPlex whole-transcriptome amplification, before expression profiling on 20,000-element cDNA microarrays. We have recently validated TransPlex whole-transcriptome amplification for expression profiling[20], and several lines of evidence support the validity of our signatures, as described below and in the **Supplementary Discussion** online.

[1]Department of Pathology, [2]Comprehensive Cancer Center, [3]Bioinformatics Program and [4]Department of Urology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. [5]Department of Pathology, Brigham and Women's Hospital and [6]Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts, USA. [7]Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. [8]These authors contributed equally to this manuscript. Correspondence should be addressed to A.M.C. (arul@umich.edu).

Received 27 June; accepted 31 October; published online 17 December 2006; doi:10.1038/ng1935

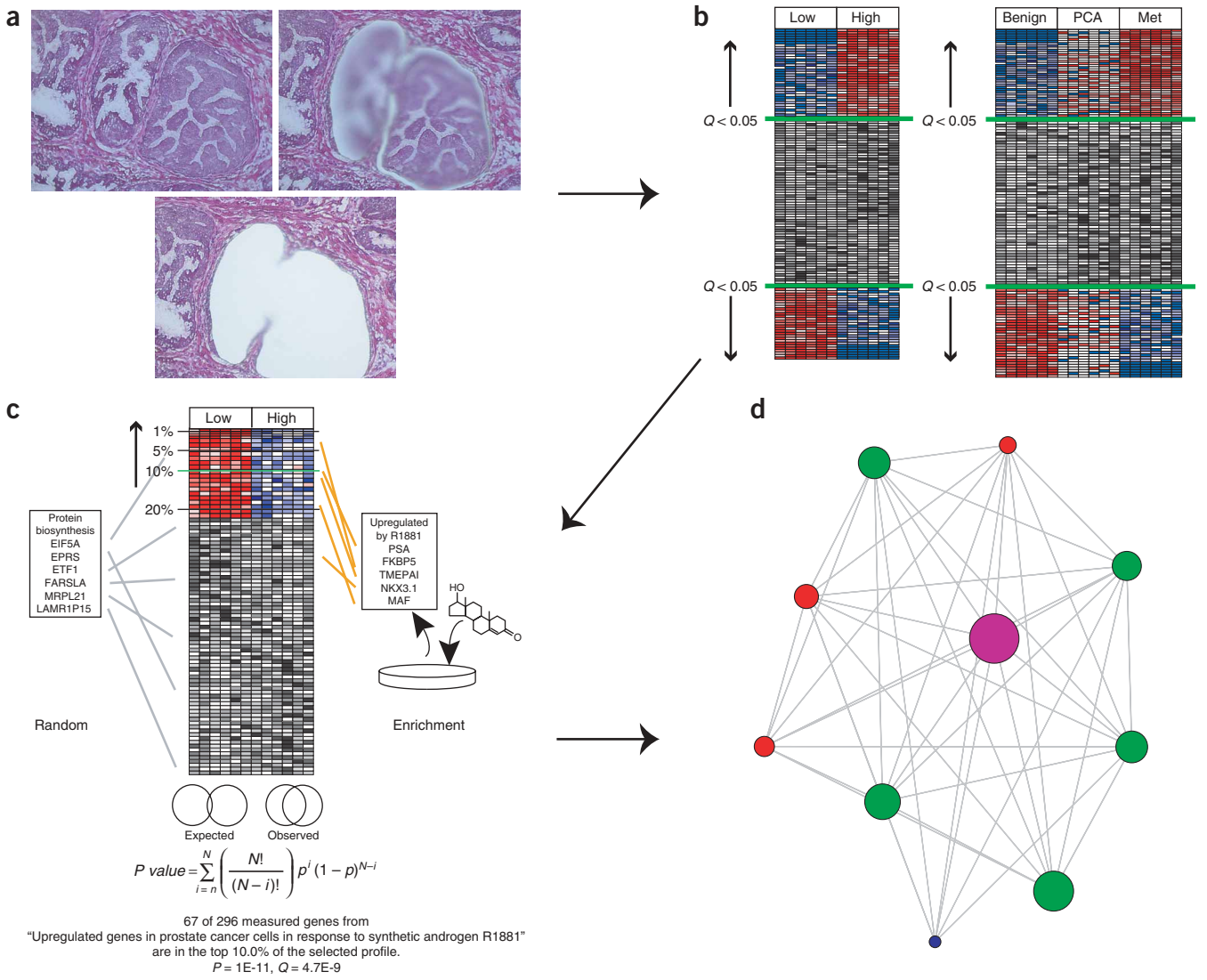

Figure 1 Integrative analysis of molecular concepts in prostate cancer progression. (**a**) Samples representing different aspects of progression, such as prostatic intraepithelial neoplasia (shown here), are obtained by LCM and hybridized to cDNA microarrays. Normalized data are loaded into the Oncomine database for analysis. (**b**) Using the tools available in Oncomine, gene signatures are identified for pairwise comparisons, such as low–Gleason grade prostate cancer versus high–Gleason grade prostate cancer, or correlation analyses, such as genes correlating with progression from benign epithelium to prostate cancer (PCA) to metastatic prostate cancer (Met). (**c**) Expression signatures are automatically compared with all concepts in the Molecular Concept Map (MCM), a resource containing approximately 15,000 molecular concepts (that is, biologically related genes), for enrichment by disproportionate overlap. (**d**) Enrichment networks are then visualized using the MCM for significant links between concepts.

To identify expression signatures, we loaded our data set into Oncomine, a bioinformatics resource developed by our group to catalog and analyze microarray studies[21]. To identify molecular correlates of prostate cancer progression, we analyzed our expression signatures using the MCM. In total, data from 12 databases and 340 high-throughput data sets were collected and analyzed, yielding over 14,000 molecular concepts (D.R.R. *et al.*, unpublished data). The types of molecular concepts are described in the **Supplementary Discussion**. Our experimental approach is shown in **Figure 1**.

Prostate cancer progression signatures

We used LCM to interrogate epithelial cells and minimize the bias of stroma. By profiling 12 stromal and 89 epithelial cell populations, we defined a stromal signature, determined the extent of stromal bias in studies using grossly dissected tissues and assessed the epithelial purity of our specific LCM-isolated cell populations (**Supplementary Fig. 1** and **Supplementary Discussion** online). To define a comprehensive progression signature, we also sought to incorporate putative precursor lesions. As described below, through several analyses, our results demonstrated that high-grade PIN and prostate cancer shared markedly similar expression signatures. However, these same analyses indicated that atrophic lesions, including PIA, shared few genetic changes with prostate cancer, suggesting that PIA may be only a very early precursor of, or unrelated to, cancer progression. Although enrichment analysis indicated that our PIA samples were biased by contaminating stroma, the epithelial PIA signature was still more similar to benign epithelium than to prostate cancer. Further profiling studies will be needed to define the role of PIA in progression.

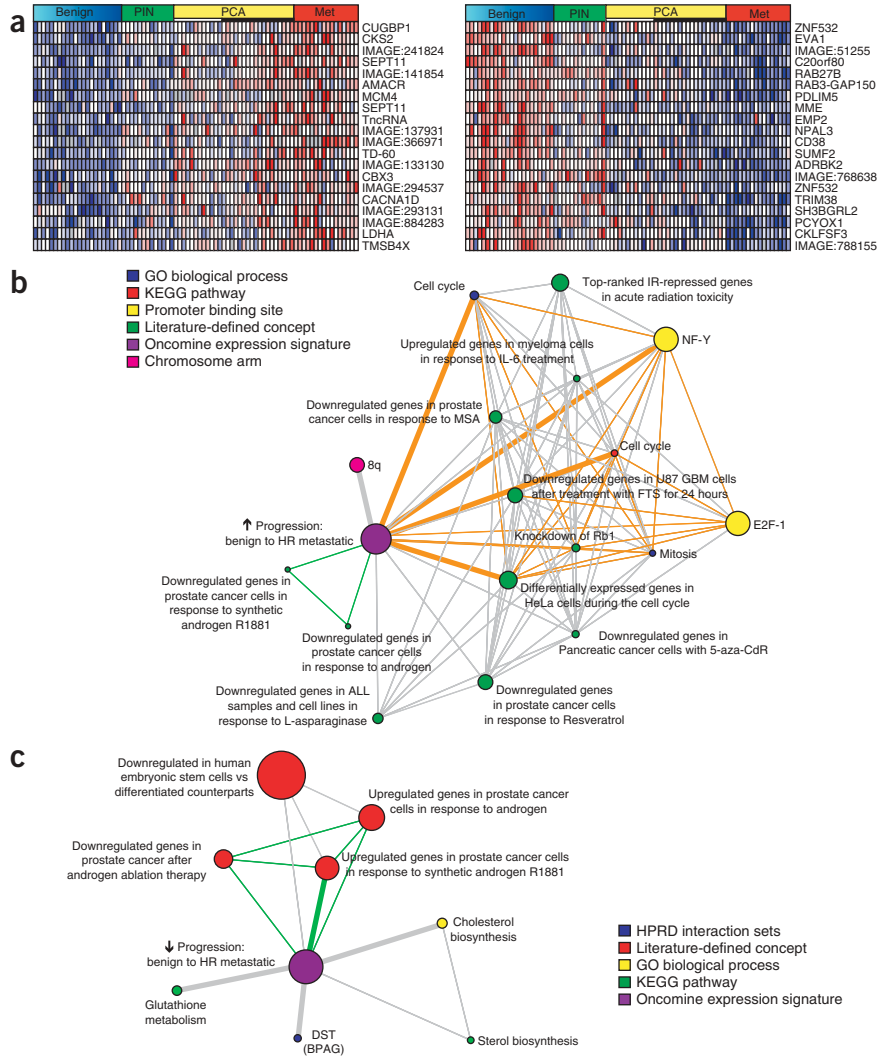

Figure 2 Expression signatures and molecular concept analysis of cancer progression in microdissected prostatic epithelia. (**a**) Robust prostate cancer progression signatures identified from microdissected material. Genes correlating with progression from benign epithelium to PIN to prostate cancer (PCA) to hormone-refractory metastatic prostate cancer (Met) were identified. The top 20 overexpressed (left) and underexpressed (right) features are shown. Columns represent individual arrays, and rows represent the indicated feature. Red and blue indicate relative overexpression and underexpression, respectively, and gray indicates features excluded during normalization. (**b**,**c**) Network view of the molecular concept analysis of our 'overexpressed during progression' (**b**) or 'underexpressed during progression' (**c**) signatures (purple nodes) defined in **a**. Each node represents a molecular concept or set of biologically related genes. The node size is proportional to the number of genes in the concept (as examples, the 'NF-Y' and 'mitosis' concepts contain 1,347 and 85 genes, respectively). Each edge represents a statistically significant enrichment ($P < 5 \times 10^{-4}$). The most enriched concept of each type in the progression signature is indicated by a thick edge. Enrichments with 'androgen concepts', indicating decreased androgen signaling during prostate cancer progression, are indicated by green edges. Enrichments with 'proliferation' concepts are indicated by orange edges. HPRD, Human Protein Reference Database.

We identified robust progression signatures comprising genes whose expression increased or decreased during the progression from benign epithelium to PIN to prostate cancer to metastatic prostate cancer (see Methods). The 'overexpressed in progression' and 'underexpressed in progression' signatures had 661 and 862 features at $Q < 0.05$, respectively (**Fig. 2a**). Several lines of evidence support the accuracy of our progression signatures. For example, the most enriched Oncomine signatures in our 'overexpressed in progression' and 'underexpressed in progression' signatures were the overexpressed ($P = 1.7 \times 10^{-68}$) and underexpressed ($P = 2.7 \times 10^{-82}$) prostate cancer progression signatures, respectively, from our group's previous study on an independent set of grossly dissected samples[22]. The most enriched chromosome arm in our 'overexpressed in progression' signature is 8q ($P = 4.5 \times 10^{-4}$), which shows one of the most frequent gains during progression[23]. Located at 8q24, the *MYC* oncogene, which ranks 24th in our 'overexpressed in progression' signature, is thought to be one of the targets of this amplification and has been shown to be amplified during progression[23,24].

Notably, stroma confounds the identification of truly underexpressed genes (such as tumor suppressors) in studies using grossly dissected tumors, owing to masking by large numbers of stromal transcripts whose decreased expression during progression reflect the decrease in the percentage of stroma[25–27]. For example, *MME*, which ranks 8th in our 'underexpressed in progression' signature, is known to be lost during prostate cancer progression and is thought to function as a tumor suppressor[28]. Although *MME* is significantly underexpressed in progression signatures from previous studies using grossly dissected tissue (ref. 11, $P = 2.5 \times 10^{-10}$; ref. 29, $P = 1.5 \times 10^{-5}$; ref. 14, $P = 0.005$), it ranks no higher than 639th in these studies because of stromal masking. Together, these results indicate that although signatures derived from grossly dissected and LCM-isolated samples share substantial overlap, our LCM-based progression signature can serve as a more specific resource for identifying genes underexpressed in epithelial cells during progression.

We also validated the differential expression of genes not previously implicated in progression using quantitative real-time-PCR on an independent sample set (**Supplementary Fig. 2** online). We validated the overexpression of two transcripts, *ZIC2* (which is involved in the hedgehog signaling pathway[30]) and a noncoding transcript at 11q13.1 (a region in which amplification predicts progression to metastatic disease[31]), as described in the **Supplementary Discussion**. We also validated the underexpression of *NPAL3*, which ranks tenth in our signature and maps to 1p36, a region lost in metastatic prostate cancer[32].

To move beyond the single-gene or single-concept approach, we analyzed our progression signatures and enriched concepts using the MCM, which allows for the identification of enrichment networks. Analyzing our 'overexpressed in progression' signature uncovered an enrichment network containing proliferation-related concepts (including the most enriched Gene Ontology (GO) biological process, 'cell cycle' ($P = 1.6 \times 10^{-6}$), and the most enriched literature concept,

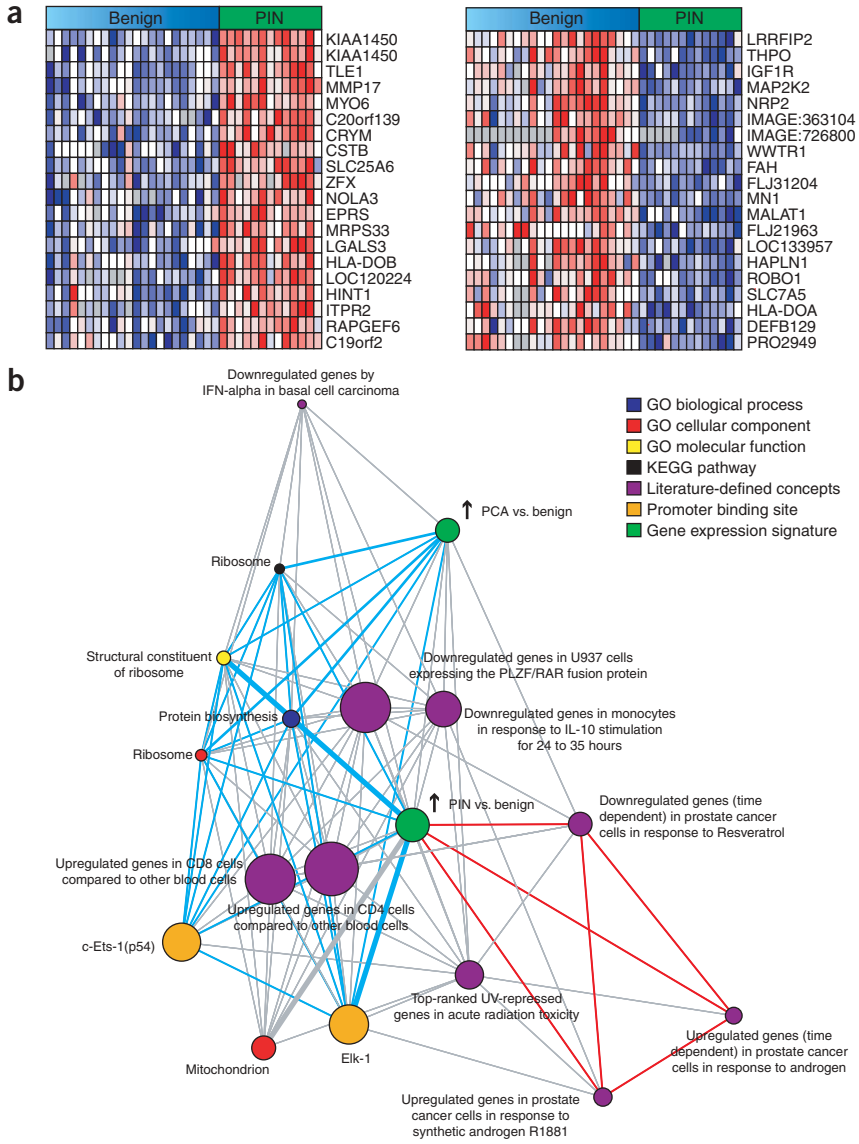

Figure 3 Molecular concept analysis comparing benign epithelium with prostatic intraepithelial neoplasia (PIN). The most differentially expressed genes between benign prostate epithelium and PIN were identified by Oncomine. (**a**) The top 20 overexpressed (left) and underexpressed (right) features in PIN compared with benign epithelium are indicated. Heat map color scale is as in **Figure 2**. (**b**) Network view of the molecular concepts enriched in our 'overexpressed in PIN versus benign epithelium' signature, including our 'overexpressed in prostate cancer (PCA) versus benign epithelium' signature. The node size is proportional to the number of genes in the concept (as examples, the 'upregulated genes in CD4 cells compared with other blood cells' and the KEGG pathway 'ribosome' concepts contain 2,588 and 85 genes, respectively). Each edge represents a statistically significant enrichment ($P < 5 \times 10^{-4}$). The most enriched concept of each type in our signatures is indicated by a thick edge. Enrichments between expression signatures are indicated by black edges. Enrichments with interconnected 'androgen concepts', indicating increased androgen signaling in PIN versus benign epithelium, are indicated by red edges. Enrichments with interconnected 'protein biosynthesis concepts', indicating increased protein biosynthesis in PIN versus benign epithelium, are indicated by blue edges.

'differentially expressed genes in HeLa cells during the cell cycle' ($P = 1.7 \times 10^{-10}$)) (**Fig. 2b**). Other concepts in this set have strong biological relationships to proliferation (**Supplementary Discussion**).

MCM analysis of our 'underexpressed in progression' signature demonstrated an interconnected androgen signaling concept set (**Fig. 2c**). These concepts represent both *in vitro* measures (including the most enriched literature concept, 'upregulated genes in prostate cancer cells in response to synthetic androgen R1881' ($P = 2 \times 10^{-14}$)) and *in vivo* measures of androgen signaling ('downregulated genes in prostate cancer after androgen ablation therapy' ($P = 1.7 \times 10^{-4}$)). The enrichment of genes normally upregulated by androgen in the 'underexpressed in progression' signature is consistent with decreased androgen signaling activity in the androgen-ablated state, as our metastatic prostate cancers have recurred after androgen ablation therapy. Other enriched concepts in our 'underexpressed in progression' signature are described in the **Supplementary Discussion**.

The transition from benign epithelium to PIN

We also attempted to identify molecular correlates of individual histological transitions in prostate cancer progression. For individual transitions, we observed far fewer genes differentially expressed ($Q < 0.05$) between PIN and prostate cancer (1.2% of measured features) than between benign epithelium and PIN (13.4%) or between localized and metastatic prostate cancer (15.7%), suggesting that PIN and prostate cancer share similar expression signatures. Additional analyses, including clustering and prediction analysis of microarrays (**Supplementary Fig. 3** online) and MCM analysis (described below), support this observation, suggesting that key processes distinguishing prostate cancer from benign epithelium also occur in PIN.

We identified robust expression signatures (with 1,213 and 1,335 features, respectively, at $Q < 0.05$) for genes over- and underexpressed in PIN versus benign epithelium (**Fig. 3a**). MCM analysis of our 'overexpressed in PIN relative to benign epithelium' signature identified an enrichment network of protein biosynthesis concepts (**Fig. 3b**), including the most enriched GO biological process ('protein biosynthesis', $P = 1.3 \times 10^{-7}$) and KEGG pathway ('ribosome', $P = 4.9 \times 10^{-5}$). This network also contained the most enriched Transfac promoter binding site concept, the *ETS* transcription factor *ELK1* ($P = 2.1 \times 10^{-9}$), representing genes with defined *ELK1* binding sites in their proximal promoters. Transcriptional targets of other *ETS* family members, including *ETS1* ($P = 1.3 \times 10^{-8}$) and *GABPA* (also known as *NRF-2*) ($P = 7.6 \times 10^{-8}$), also show strong enrichment due to overlapping transcription factor matrices. These results suggest that a major process distinguishing PIN from benign epithelium is increased protein biosynthesis, probably through *ETS* target genes. MCM analysis also supports the genetic similarity of PIN and prostate cancer, because in addition to marked overlap ($P < 1 \times 10^{-100}$), the signatures shared enrichment of the 'protein biosynthesis' and '*ETS*

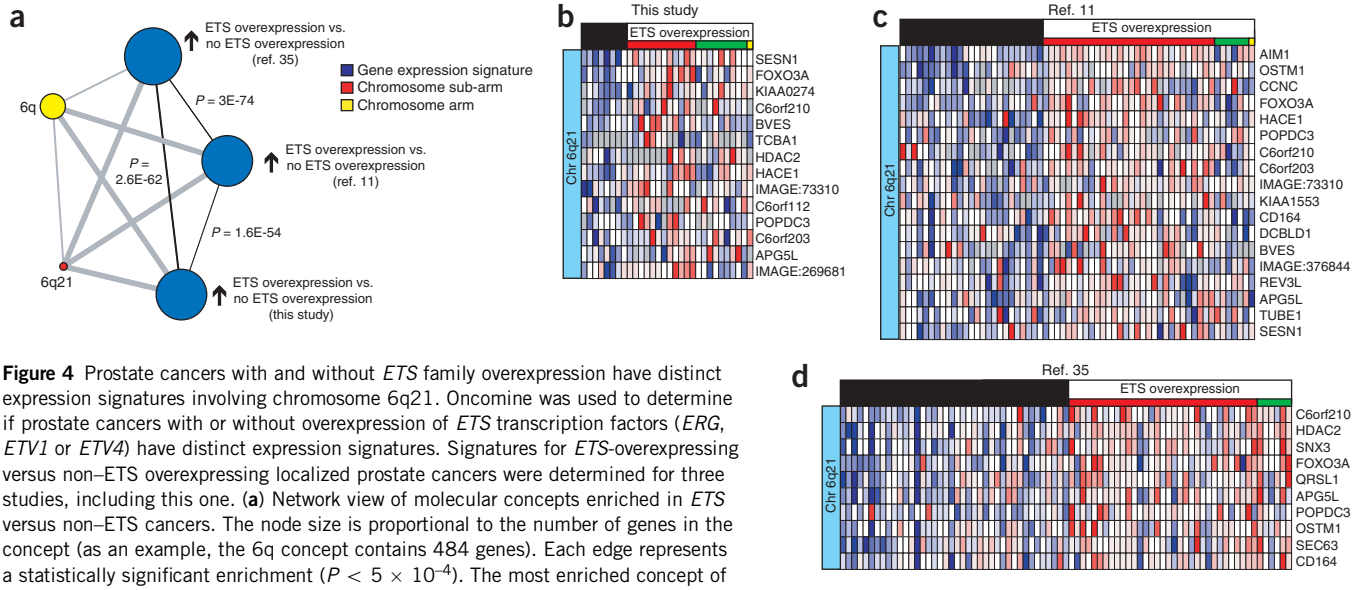

Figure 4 Prostate cancers with and without *ETS* family overexpression have distinct expression signatures involving chromosome 6q21. Oncomine was used to determine if prostate cancers with or without overexpression of *ETS* transcription factors (*ERG*, *ETV1* or *ETV4*) have distinct expression signatures. Signatures for *ETS*-overexpressing versus non–ETS overexpressing localized prostate cancers were determined for three studies, including this one. (**a**) Network view of molecular concepts enriched in *ETS* versus non–ETS cancers. The node size is proportional to the number of genes in the concept (as an example, the 6q concept contains 484 genes). Each edge represents a statistically significant enrichment ($P < 5 \times 10^{-4}$). The most enriched concept of each type in the expression signatures is indicated by a thick edge. Enrichments between expression signatures are indicated by black edges, and *P* values are shown for these enrichments. (**b**–**d**) Heat map of genes from the 6q21 concept enriched in ETS-overexpressing tumors in (**b**) this study, (**c**) ref. 11 and (**d**) ref. 35. In each study, cancers overexpressing *ERG* (red), *ETV1* (green) or *ETV4* (yellow) were compared with cancers without *ETS* family overexpression (black).

target gene' concepts. Furthermore, genes overexpressed in PIN relative to benign epithelium also showed enrichment of concepts indicating increased androgen signaling (**Fig. 3b**), suggesting a link between androgen signaling, *ETS* transcription factors and protein biosynthesis, as described below.

The transition from PIN to prostate cancer

We identified a limited number of transcripts differentially expressed in localized prostate cancer relative to high-grade PIN (199 overexpressed and 23 underexpressed features at $Q < 0.05$). This was intriguing in the context of our work demonstrating that the *ETS* family members *ERG*, *ETV1* and *ETV4* are markedly overexpressed in prostate cancer through fusions with the androgen-regulated gene *TMPRSS2* (refs. 33,34) and our observation that *ETS* target genes involved in protein biosynthesis are overexpressed in PIN (and prostate cancer) relative to benign prostatic epithelium. We observed mutually exclusive overexpression of *ERG*, *ETV1* or *ETV4* (>3.5 normalized *z* score units) in 20 of 30 localized prostate cancer samples (from 12 of 19 cases) but did not observe overexpression of these genes in any of the 22 benign or 13 PIN samples (**Supplementary Fig. 4** online). Notably, our data set contains three individuals in which we captured both PIN and prostate cancer lesions, and an ETS family member (*ERG*) was markedly overexpressed in the prostate cancer samples. There was a marked upregulation of *ERG* in prostate cancer relative to PIN in each of these three individuals: *ERG* was the most upregulated feature in individuals 5 and 7 and the second most upregulated in individual 17 (**Supplementary Fig. 4**). In addition, in individuals in which multiple prostate cancer foci were profiled, either all or no foci overexpressed *ERG*, *ETV1* or *ETV4*, suggesting that this is a clonally selected event occurring early in prostate cancer development.

It is unclear if expression signatures differ between tumors with *TMPRSS2-ETS* fusions and those without. Thus, in this study and in two others[11,35], we divided localized prostate cancers into those that overexpressed *ETS* (*ERG*, *ETV1* or *ETV4*) and those that did not, and we attempted to identify expression signatures. In each study, we identified molecular signatures differentiating *ETS* and non-*ETS* tumors (157 features in this study, 524 features in ref. 11, 3,328 features in ref. 35; $Q < 0.05$). Notably, the three signatures were highly overlapping (**Fig. 4a**). For example, the 'overexpressed in *ETS* versus non-*ETS*' signatures from ref. 35 and this study were the two most enriched Oncomine signatures in the 'overexpressed in *ETS* versus non-*ETS*' signature in ref. 11 ($P = 3 \times 10^{-74}$ and $P = 1.6 \times 10^{-54}$, respectively). Although few concepts were significantly enriched across all three studies, 6q21 was the most enriched chromosome sub-arm in all three 'overexpressed in *ETS* versus non-*ETS*' signatures (this study, $P = 1.20 \times 10^{-4}$; ref. 11, $P = 1.80 \times 10^{-8}$; ref. 35, $P = 8.30 \times 10^{-6}$) (**Fig. 4b–d**). This suggests a cooperating amplification at 6q21 in *ETS* tumors or loss of 6q21 in non-*ETS* tumors; notably, multiple studies have identified loss of 6q21 in approximately half of localized prostate cancers[36]. Thus, down-regulation of genes at 6q21 may be important to tumor development in non-*ETS* prostate cancers, providing an important direction for future studies. For example, *FOXO3A*, which showed reduced expression in non-*ETS* tumors in all three studies, has been proposed as a prostate cancer tumor suppressor through promoting the expression of *CDKN1B* (also known as $p27^{kip1}$) and *BCL2L11* (also known as *BIM*)[37,38].

Clinically localized to metastatic prostate cancer

Metastatic prostate cancer is generally considered incurable, and treatment becomes palliative in nature. Treatment with anti-androgens usually results in regression, but unfortunately, the cancer almost invariably progresses with a hormone-refractory phenotype. Thus, identifying signatures and concepts differentiating clinically localized from untreated, or hormone-naive, and hormone-refractory metastatic prostate cancer is essential to understanding progression. As our study set did not contain enough hormone-naive metastatic samples ($n = 3$) for comprehensive analysis, we used two Oncomine data sets (from refs. 11 and 14, respectively) containing benign epithelium, localized

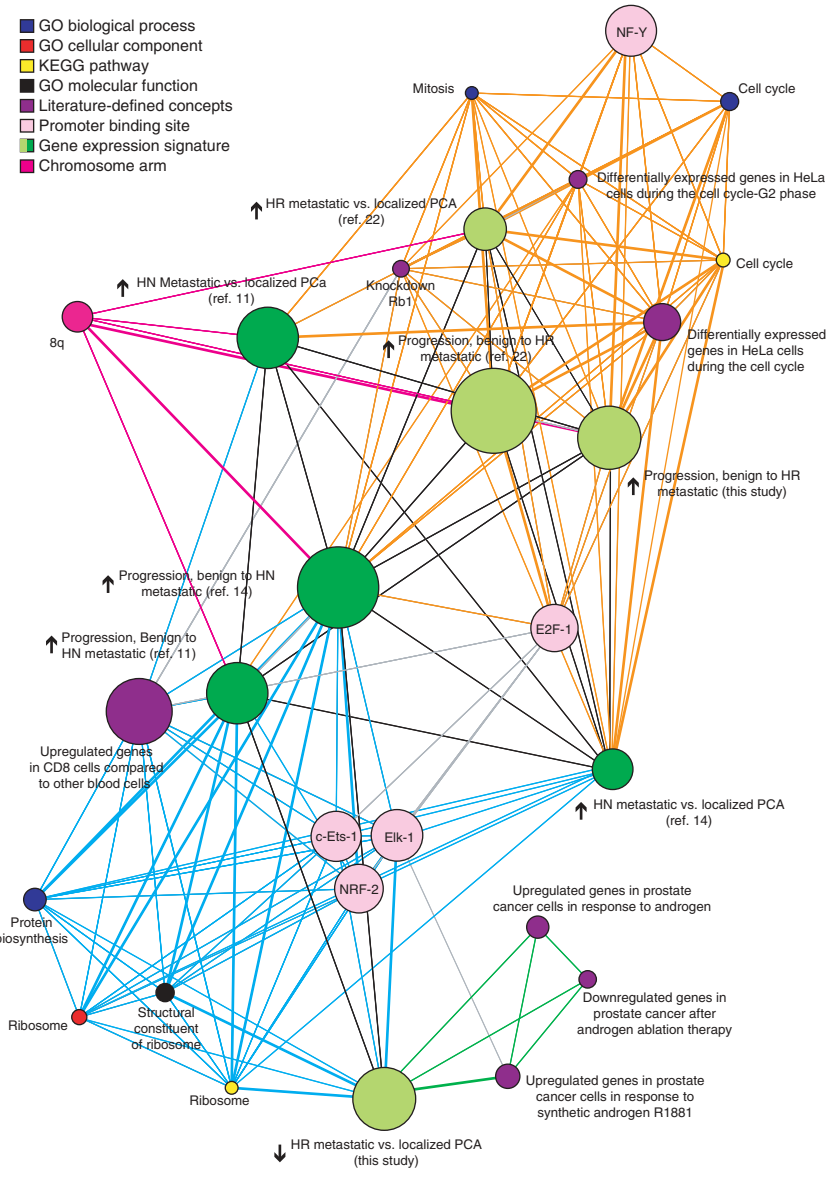

Figure 5 Differential expression of proliferation, protein biosynthesis and androgen signaling concepts in clinically localized, hormone-naive metastatic and hormone-refractory metastatic prostate cancer. To identify concepts differentially expressed in localized, hormone-naive (HN) and hormone-refractory (HR) metastatic prostate cancer, we identified enrichment networks for progression signatures (as defined in the text) and 'metastatic cancer versus localized prostate cancer' signatures from our study and three others[11,14,22]. HN and HR signatures are indicated by light and dark green nodes, respectively. The node size is proportional to the number of genes in the concept (as an example, the 'upregulated genes in prostate cancer cells in response to synthetic androgen R1881' concept contains 301 genes). Each edge represents a statistically significant enrichment ($P < 5 \times 10^{-4}$). The most enriched concept of each type in the expression signatures is indicated by a thick edge. Enrichments between expression signatures are indicated by black edges. Enrichments with protein biosynthesis, proliferation and androgen signaling concepts are indicated by blue, orange and green edges, respectively. Enrichments with chromosome arm 8q are shown in magenta.

prostate cancer and hormone-naive metastatic prostate cancer samples. MCM analysis of these studies uncovered two distinct interaction networks (centered on protein biosynthesis and proliferation concepts) enriched in the 'overexpressed in progression from benign epithelium to localized cancer to hormone-naive metastatic prostate cancer' signature and the 'overexpressed in hormone-naive metastatic cancer versus localized prostate cancer' signatures (**Fig. 5**). For example, 'protein biosynthesis' was the most enriched GO biological process in the 'overexpressed in progression' signatures from both ref. 14 and ref. 11 ($P = 2.29 \times 10^{-29}$ and 7.0×10^{-13}, respectively). Similarly, 'differentially expressed genes in HeLa cells during the cell cycle' was the most enriched literature concept in the 'overexpressed in hormone-naive metastatic cancer versus localized prostate cancer' signatures from both studies (ref. 14, $P = 1.3 \times 10^{-19}$; ref. 11, $P = 3.0 \times 10^{-15}$) (**Fig. 5**). Notably, in both studies, progression signatures are more strongly linked to protein biosynthesis, whereas 'hormone-naive metastatic versus localized prostate cancer' signatures are more strongly linked to proliferation concepts. Increased protein biosynthesis defined our 'PIN versus benign epithelium' and 'localized prostate cancer versus benign epithelium' signatures, suggesting that protein biosynthesis concepts are similarly overexpressed in hormone-naive metastatic prostate cancer, whereas increased proliferation distinguishes hormone-naive metastatic cancer from localized prostate cancer.

To understand the transition from hormone-naive to hormone-refractory metastatic prostate cancer, we compared the hormone-naive metastatic enrichment networks just described with corresponding hormone-refractory metastatic signatures from our study and ref. 22. The 'progression to hormone-refractory metastatic prostate cancer' and 'hormone-refractory metastatic versus localized prostate cancer' signatures from both studies were highly enriched with the proliferation network (**Fig. 5**). For example, 'differentially expressed genes in HeLa cells during the cell cycle' was the most enriched literature concept in both 'overexpressed in progression' signatures (this study, $P = 1.7 \times 10^{-10}$; ref. 22, $P = 2.4 \times 10^{-36}$) and in the 'hormone-refractory metastatic versus localized prostate cancer' signature in ref. 22 ($P = 3.2 \times 10^{-28}$). However, these signatures did not show significant enrichment of protein biosynthesis concepts. Rather, our 'underexpressed in hormone-refractory metastatic cancer versus localized prostate cancer' signature showed strong enrichment with protein biosynthesis concepts (**Fig. 5**). Additionally, the strongest enrichment in this signature was for decreased androgen signaling concepts, consistent with the castrated hormone-refractory state. For example, 'upregulated genes in prostate cancer cells in response to synthetic androgen R1881' was the most enriched literature concept ($P = 2.1 \times 10^{-31}$). Other decreased androgen signaling concepts in our 'hormone-refractory metastatic cancer versus localized prostate cancer' signatures are described in the **Supplementary Discussion** and **Supplementary Figure 5** online. These results suggest that although hormone-refractory and hormone-naive metastatic prostate

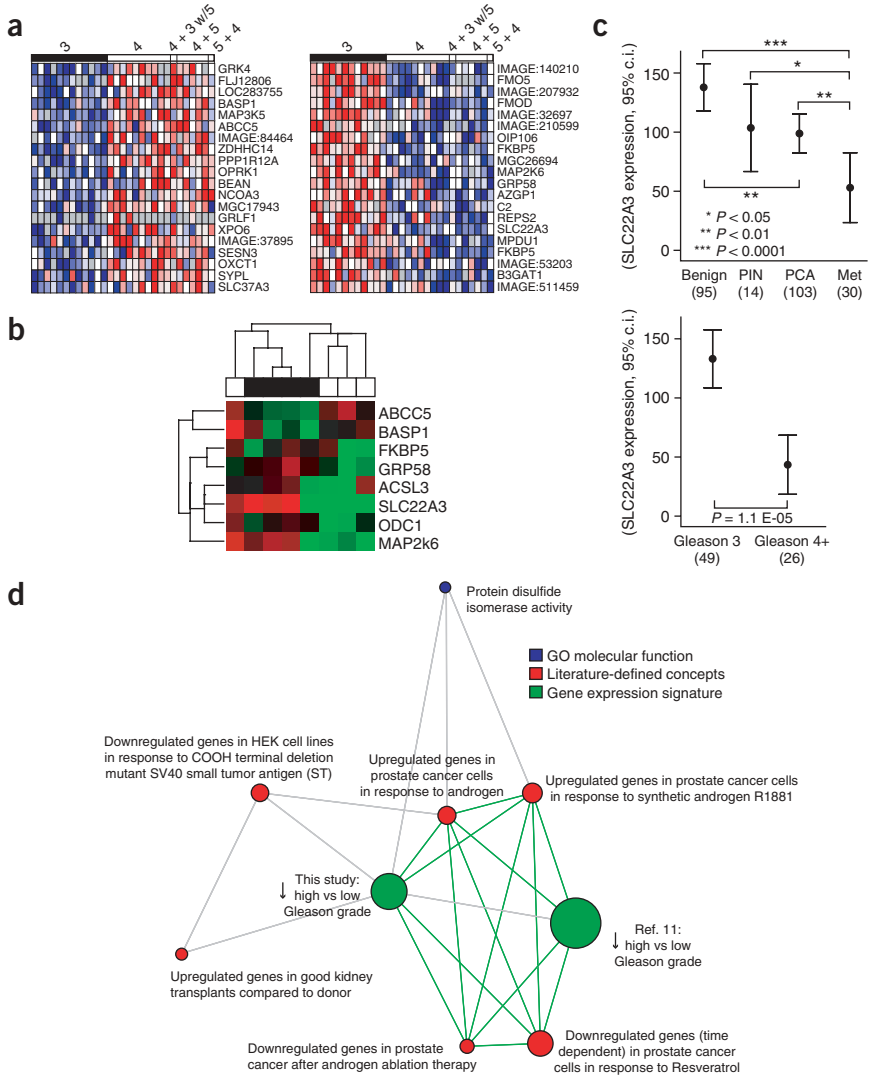

Figure 6 Molecular concept analysis comparing low–Gleason grade to high–Gleason grade prostate cancer. (**a**) Molecular Gleason signatures were identified from genes showing the greatest differential expression between low– (black, Gleason pattern 3) and high–Gleason grade (white, Gleason pattern > 3) prostate cancer samples. The top 20 overexpressed (left) and underexpressed (right) features are indicated. (**b**) Heat map of quantitative real-time PCR validation experiments in an independent panel of grossly dissected prostate cancer samples containing >90% Gleason pattern 3 (black) or 4 (white) cells. Rows represent genes, and columns represent samples. Red and green indicate relative overexpression or underexpression, respectively. (**c**) Validation of decreased SLC22A3 expression in progression and high–Gleason grade prostate cancer. Expression of SLC22A3 was measured using immunohistochemistry on a tissue microarray. Staining intensity of epithelial cells was scored as strong (3), moderate (2), weak (1) or negative (0) and multiplied by the percentage of epithelial cells stained per core. Error bars indicate 95% confidence intervals of the mean. Expression of SLC22A3 in prostate cancer progression was determined by analysis of benign, PIN, localized prostate cancer (PCA) and metastatic prostate cancer (Met) tissue cores (left). The number of cores per class is indicated. SLC22A3 expression was compared between the localized prostate cancer cores containing only Gleason pattern 3 (Gleason 3) and cores containing only Gleason patterns 4 or 5 (Gleason 4+) (right). (**d**) Network view of the molecular concepts enriched in 'underexpressed from low- to high–Gleason grade prostate cancer' signatures from this study and from ref. 11. Enrichments with interconnected 'androgen concepts', indicating decreased androgen signaling in high–Gleason grade prostate cancer, are indicated by green edges.

cancer share increased proliferation, only hormone-refractory metastatic cancers show a marked decrease of androgen signaling and protein biosynthesis concepts. As protein biosynthesis decreases with decreased androgen signaling, whereas protein biosynthesis and androgen signaling activity are increased in PIN compared with benign epithelium, this supports a link between androgen signaling, ETS transcriptional targets and protein biosynthesis. This is in agreement with studies showing that androgen ablation causes a reduction in nucleolar size and regression of PIN[39].

Decreased androgen signaling in high–Gleason grade cancer

Using LCM also allowed us to profile distinct Gleason patterns of cancerous epithelium. We divided our cancer samples into two classes, low-grade (only Gleason pattern 3) and high-grade (samples with Gleason patterns >3). We did not identify robust signatures distinguishing low- and high-grade samples (two features at $Q < 0.05$) (**Fig. 6a**). However, the 'overexpressed in high–Gleason grade' ($P = 6.4 \times 10^{-49}$) and 'underexpressed in high–Gleason grade' ($P = 1.5 \times 10^{-36}$) signatures of ref. 11 are the most enriched Oncomine signatures in our 'overexpressed in high–Gleason grade' and 'underexpressed in high–Gleason grade' signatures, respectively, supporting the existence of more subtle multigene signatures. We also validated several of the most differentially expressed features using quantitative real-time PCR on an independent set of grossly dissected tumors (**Fig. 6b**). We further validated the underexpression of *SLC22A3* in high–Gleason grade tumors by immunohistochemistry on tissue microarrays. *SLC22A3* ranked 15th in our 'underexpressed in high–Gleason grade' signature and 47th in our 'underexpressed in progression' signature. Immunohistochemistry confirmed the decrease in *SLC22A3* expression during progression (benign epithelium to prostate cancer, $P = 0.003$; localized prostate cancer to metastatic prostate cancer, $P = 0.009$) as well as lower expression in high–Gleason grade cancer compared with low–Gleason grade cancer ($P = 1.1 \times 10^{-5}$) (**Fig. 6c**). A recent report defining LCM-isolated signatures from low– and high–Gleason grade prostate cancer also found marked underexpression of *SLC22A3* (the third lowest expression as a multiple of the control) in high–Gleason grade cancer[9], further validating *SLC22A3* as a marker for Gleason grade.

Although few transcripts were differentially expressed between low–Gleason grade cancer and high–Gleason grade cancer, MCM analysis identified strong enrichment of decreased androgen signaling in high–Gleason grade cancers (**Fig. 6d**). For example, 'upregulated genes in prostate cancer cells in response to synthetic androgen R1881' (which includes *SLC22A3*) was the most enriched literature concept in

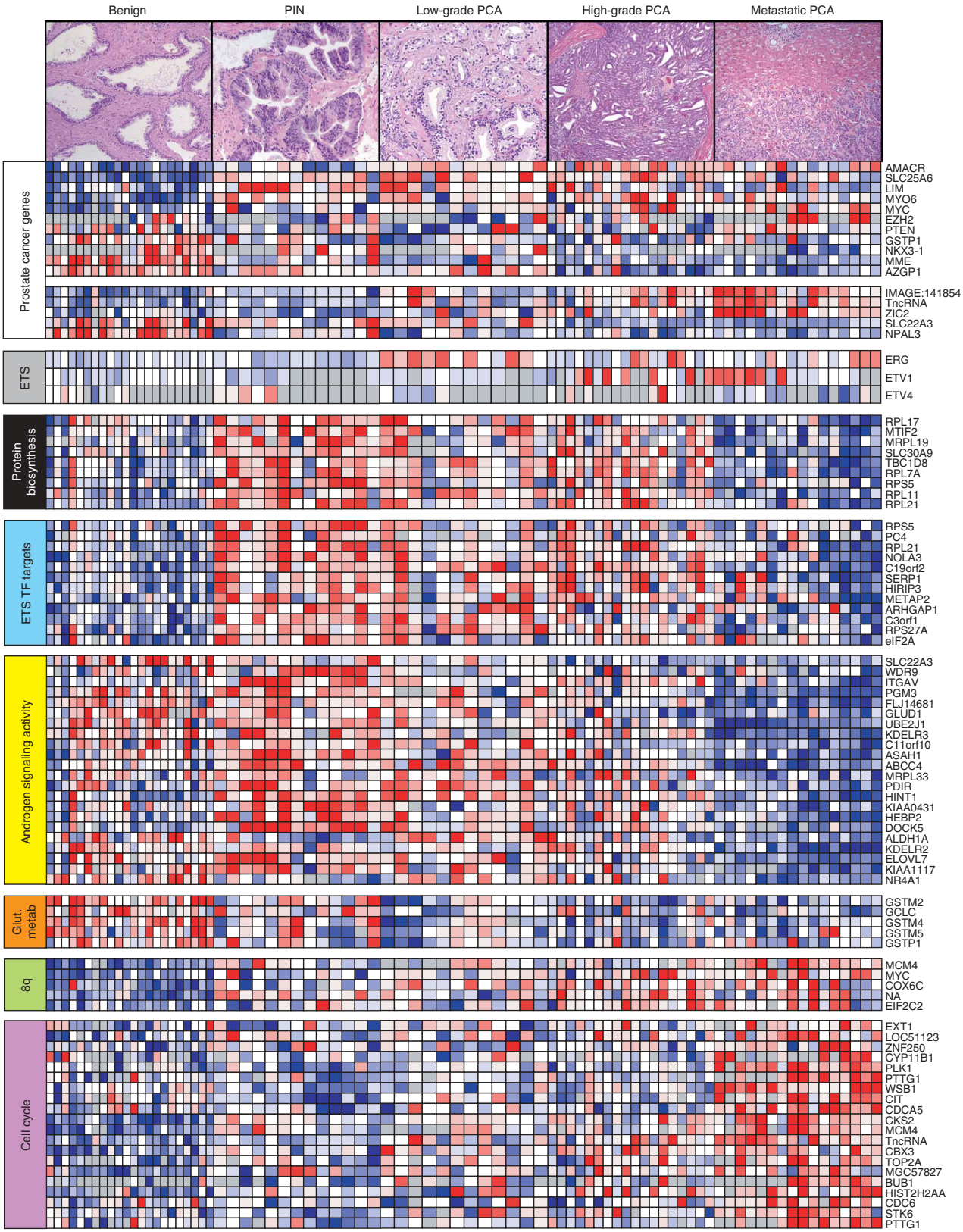

Figure 7 Molecular concept heat map of prostate cancer progression. LCM was used to generate specific expression profiles for epithelial cells from the histological transitions in prostate cancer (PCA) progression. Columns below the histological images represent individual samples from each class, and rows represent individual features. Heat map color scale is as in **Figure 2**. Genes previously implicated in prostate cancer progression and those validated in this study ('prostate cancer genes') as well as representative genes from enriched molecular concepts as described in the text are shown at right.

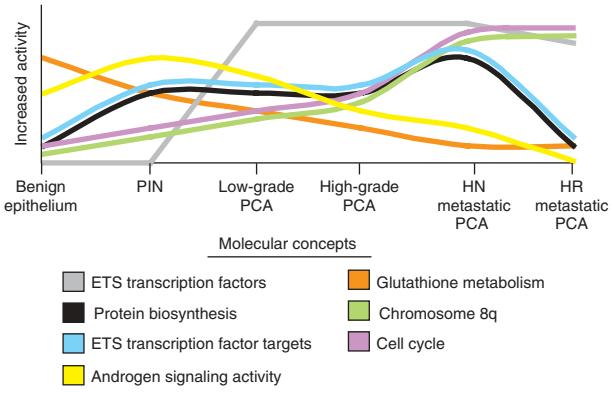

Figure 8 Molecular concept model of prostate cancer progression. The relative expression of enriched concepts identified by expression profiling of specific cell populations was used to develop a molecular concept model of prostate cancer (PCA) progression to hormone-naive (HN) and hormone-refractory (HR) metastatic PCA.

our 'underexpressed in high–Gleason grade versus low–Gleason grade cancer' signature and that of ref. 11 ($P = 1.8 \times 10^{-16}$ and 1×10^{-11}, respectively). Other enriched concepts related to androgen signaling are described in the **Supplementary Discussion**.

A molecular concept model of prostate cancer progression

By combining LCM-based profiling with an integrative molecular concept analysis, we identified genes, concepts and enrichment networks correlating with the histologic progression of prostate cancer, leading to a unified molecular model (**Figs. 7 and 8**). We confirmed the differential expression of several genes (including *TMPRSS2-ETS* gene fusions, *AMACR*, *MYC*, *EZH2*, *PTEN*, *GSTP1*, *NKX3-1*, *MME* and *AZGP1*) and concepts (including increased expression of genes on 8q and proliferation genes) previously linked to prostate cancer progression[40,41]. We also identified molecular concepts correlating with known histological features of prostate cancer progression. For example, the defining histological characteristic of PIN is an enlarged nucleolus[4,42] (which is the organelle responsible for controlling protein biosynthesis), consistent with our concept analyses.

DISCUSSION

Our concept-based analysis allows us to make several insights into prostate cancer progression. MCM analysis identifies strong enrichment of *ETS* transcription factor targets in genes involved in protein biosynthesis. Although we are unaware of direct evidence linking *ETS* target genes to protein biosynthesis regulation in prostate cancer, our results suggest that this pathway is upregulated during the transition from benign epithelium to PIN and is downregulated during the transition from localized to hormone-refractory metastatic prostate cancer.

This result is important in light of our recent discovery of *TMRPSS2-ETS* gene fusions in the majority of prostate cancers[33,34]. Our analysis suggests that this is one of the few expression changes distinguishing PIN and prostate cancer, consistent with the similar histological appearance of PIN and prostate cancer cells. As *ETS* targets are overexpressed in PIN, possibly through subtle, direct dysregulation of *ETS* family members or through a distinct genetic lesion with overlapping targets (such as *MYC* or *PTEN*, which can regulate protein biosynthesis[43]), these gene fusions may serve to lock in the deregulation of this pathway and allow for the development of overt carcinoma. Additionally, as *ETS* targets are already overexpressed in PIN, *TMPRSS2-ETS* fusions may bypass feedback mechanisms, resulting in the differential expression of a limited number of targets, or may result in modest expression changes.

Our study confirms the central role of androgen signaling in prostate cancer. We identified increased androgen signaling in PIN compared with benign epithelium and decreased androgen signaling in localized cancer compared with PIN, in high–Gleason grade cancer compared with low–Gleason grade cancer and in hormone refractory metastatic cancer compared with localized cancer. Furthermore, enrichment analysis suggests that androgen signaling is also decreased in hormone-naive metastatic prostate cancer compared with localized cancer, as 'upregulated genes (time-dependent) in prostate cancer cells in response to androgen' is the most enriched literature concept in the 'underexpressed in hormone-naive metastatic cancer versus localized prostate cancer' signature in ref. 11 ($P = 2.9 \times 10^{-14}$). A recent report[44] uses an experimentally derived set of androgen-regulated genes to analyze ref. 11, and the authors also observe lower androgen signaling in high–Gleason grade prostate cancer compared with low–Gleason grade prostate cancer and in hormone-naive metastatic prostate cancer compared with localized prostate cancer. The authors propose a model in which localized prostate cancer cells become more aggressive by selectively downregulating androgen-responsive genes, resulting in increased proliferation, dedifferentiation or reduced apoptosis. An alternative explanation is that the relative amount of androgen signaling during progression reflects the differentiation status of prostatic epithelium, or the contribution of distinct cells of origin, whereas separate lesions drive the increased proliferation seen late in progression. For example, 8q is the most enriched chromosome arm in the 'overexpressed in high–Gleason grade' signatures in both our study and ref. 11, and 8q is significantly enriched across progression signatures along with proliferation-related concepts (**Fig. 5**). As described above, *MYC* (8q24), a master regulator of cell-cycle control, has been shown to be amplified in progression, with amplification correlating with increased Gleason grade and progression to metastatic prostate cancer[23,24].

The marked decrease in androgen signaling concepts observed in hormone-refractory metastatic prostate cancer compared with localized cancer is consistent with a recent profiling study[45], in which the authors attribute it to the castrated state. They note that although androgen-regulated genes are markedly downregulated in hormone-refractory metastatic cancer, they are still major transcripts in the cancerous cell. Hormone-refractory metastatic cancers select for mechanisms to maintain androgen signaling (such as androgen receptor amplifications or mutations[46,47]), consistent with survival of cancerous prostate cells requiring minimal androgen signaling. This is consistent with *TMPRSS2-ETS* gene fusions driving prostate cancer development, as even minimal androgen signaling activity would result in inappropriate expression of *ETS* family members, owing to the strong androgen promoter-enhancer elements regulating *TMPRSS2* expression[48]. Our results support a model in which genetic changes resulting in increased proliferation drive the transition to hormone-naive metastatic prostate cancer, whereas androgen ablation forces the selection of lesions that restore a minimal level of androgen signaling to allow continued survival in the castrated state.

By combining specific profiling with an integrated analysis, we have identified concepts correlating with observed histological transitions in prostate cancer progression. By using the MCM, we identify enrichment networks of linked concepts differentially expressed during progression, such as *ETS* target genes and protein biosynthesis. As all concepts in the MCM are automatically tested for association, we have also been able to identify distinct enrichment networks by their

lack of association, such as the proliferation and protein biosynthesis networks, which are both enriched in the progression to hormone-naive metastatic disease without sharing significant overlap. Our work also demonstrates that enriched signatures and concepts can be identified across studies, microarray platforms and signatures from grossly dissected and LCM-isolated tissues. More broadly, our work demonstrates that integrative analysis of expression profiles with a compendium of molecular concepts provides insight into biological and disease processes.

METHODS

Samples. Tissues 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 Tissue Core. Tissues were also obtained from a radical prostatectomy series at the University Hospital Ulm. All samples were collected with informed consent of the patients and prior institutional review board approval at each institution. For the reference sample in all hybridizations, a commercially available pool of benign prostate tissue total RNA (CPP, Clontech) was used.

Sample classes, as shown in the MIAME checklist (**Supplementary Methods** online), include stroma from individuals with no history of prostate disease (STROMA_NOR), stromal nodules of benign prostatic hyperplasia (BPH) (STROMA_BPH), stroma adjacent to prostate cancer foci (STROMA_PCA), epithelium from individuals with no history of prostate disease (EPI_NOR), epithelium from nodules of BPH (EPI_BPH), epithelium from individuals with prostate cancer (EPI_ADJ_PCA), atrophic epithelium (EPI_ATR) including PIA (EPI_ATR_PIA), PIN, localized prostate cancer and hormone-naive prostate cancer (MET_HN) or hormone-refractory metastatic prostate cancer (MET_HR). For this study, we defined benign epithelium to include EPI_NOR, EPI_BPH and EPI_ADJ_PCA. As described in the text, owing to the low number of PIA and hormone-naive metastatic samples, we excluded them from all signatures except for the stromal versus epithelial signature.

Laser-capture microdissection. LCM was performed from frozen tissue sections with the SL Microtest device using μCUT software (MMI). Approximately 10,000 cells were captured for each sample. Serial sections were used if cells could not be obtained from a single section. Total RNA was isolated from captured cells with the RNAqueous Micro kit (Ambion) and treated with DNase I according to the manufacturer's instructions. RNA quantification was performed using Ribogreen (Molecular Probes). Hybridizations to assess biological and technical reproducibility are described in the MIAME checklist (**Supplementary Methods**).

RNA amplification, cDNA microarrays and data analysis. Exponential RNA amplification was performed using a TransPlex Whole-Transcriptome Amplification kit (Rubicon Genomics) as described[20], and complete details are provided in the MIAME checklist. Complete details of printing the 20,000 element–spotted cDNA microarrays, post-processing, labeling hybridization and normalization of the arrays are available in the MIAME checklist. A single commercially available pool of benign prostate tissue (Clontech) was used as the reference (Cy3) in all hybridizations and was amplified in parallel to LCM-isolated samples (Cy5). Arrays were autogridded by GenePix 4.0. Features flagged by GenePix as not found during grid alignment and areas of obvious defects were manually flagged, and both were excluded from further analysis. To create the data set for uploading into the Oncomine database, features were ranked based on the sum of the medians (Cy3 + Cy5 intensity for each feature), and the bottom 10% were excluded. The median of ratios (log2 of Cy5/Cy3) for each included feature was normalized using locally weighted regression (lowess) with a window of 0.6 using custom software written in Perl and R. To exclude unreliable features, features showing an average normalized median of ratios of >1.5 or <0.75 across a series of self-self hybridizations (including unamplified and whole transcriptome–amplified samples performed for both print runs used in this study) were removed from all arrays in the individual print run. Finally, to remove biases associated with the use of two print runs, all features were median centered per print run before compilation into the final data set. The pairs of individual hybridizations representing the replicate hybridizations (as described in the MIAME checklist) were averaged before we used the data.

Gene signature generation and MCM analysis. A complete description of the methods used to identify gene signatures in the Oncomine database and gene set enrichment in the context of the MCM is available (refs. 19,21,49 and D.R.R. *et al.*, unpublished data). All gene expression studies catalogued in Oncomine, including this study, are normalized in the same manner. All data were log transformed and median centered per array, and standard deviation was normalized to one per array. For differential expression, Student's *t*-test was used for two-class differential expression analyses (for example, high-grade versus low-grade), and Pearson's correlation was used for multiclass ordinal analyses (for example, benign epithelium, PIN, localized prostate cancer and metastatic prostate cancer) and genes were rank ordered by *P* values. *P* values for expression signatures are also corrected for multiple hypothesis testing (*Q* value) using the false discovery rate method[50]. The top 1%, 5%, 10% and 20% of each expression signature was used for enrichment analysis against all concepts in the MCM. Each pair of molecular concepts was also tested for association using Fisher's exact test. Each Oncomine-generated gene signature, including those described here, generated four molecular concepts based on the 1%, 5%, 10% and 20% cutoffs. Each concept was analyzed independently and the most significant of the four was reported. Results were stored if a given test had an odds ratio >1.25 and $P < 0.01$. *P* values $< 1 \times 10^{-100}$ were set to 1×10^{-100}. *Q* values were computed for all enrichment analyses. Concepts enriched in our signatures were identified in Oncomine, and enrichment networks were visualized using the MCM.

Quantitative real-time PCR. Quantitative real-time PCR was performed on an independent set of grossly dissected benign, localized prostate cancer and metastatic prostate cancer tissue samples essentially as described using SYBR Green dye on an Applied Biosystems 7300 Real-Time PCR system[33]. Standard curves of pooled cDNA were run for each primer pair, and the amount of target gene in each sample was normalized to the amount of *HMBS* or to the average of the amounts of *HMBS* and *GAPDH*, as indicated, in the corresponding sample. Tissues were homogenized in Trizol (Invitrogen), and total RNA was isolated using the standard Trizol protocol. All primer sequences are available in **Supplementary Table 1** online.

Immunohistochemistry. Immunohistochemistry was performed using a goat polyclonal antibody against SLC22A3 (sc-18516, Santa Cruz Biotechnologies) on a prostate cancer progression tissue microarray (TMA). The staining intensity of epithelial cells in each core was scored as strong (3), moderate (2), weak (1) or negative (0) and multiplied by the percentage of epithelial cells stained in the core. A complete description of the TMA is provided in the **Supplementary Methods**.

Accession codes. The complete microarray data set is available from the Gene Expression Omnibus (GSE6099) and Oncomine.

URLs. Oncomine can be found at http://www.oncomine.org.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank A. Menon for microarray production, B. Briggs, S. Varambally and B. Helgeson for technical assistance and R. Kuefer (University of Ulm) for tissue samples. Supported in part by Department of Defense (grants DAMD17-03-2-0033 to A.M.C. and M.A.R., PC040517 to R.M. and W81XWH-06-1-0224 to A.M.C.), the US National Institutes of Health (U54 DA021519-01A1 to A.M.C., R01 CA102872 to K.J.P and A.M.C and Prostate SPORE P50CA69568 to K.J.P., A.M.C. and R.B.S.), the Early Detection Research Network (UO1 CA111275-01 to A.M.C.) and the Cancer Center Bioinformatics Core (support grant 5P30 CA46592). K.J.P. is supported as an American Cancer Society Clinical Research Professor, D.R.R. is supported by the Cancer Biology Training Program, S.A.T. is supported by a Rackham Predoctoral Fellowship, A.M.C. is supported by a Clinical Translational Research Award from the Burroughs Welcome Foundation and S.A.T. and D.R.R. are Fellows of the Medical Scientist Training Program.

AUTHOR CONTRIBUTIONS

R.B.S. and A.M.C. share senior authorship.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Genetics* website for details).

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Jemal, A. *et al.* Cancer statistics, 2005. *CA Cancer J. Clin.* **55**, 10–30 (2005).
2. Nelson, P.S. Predicting prostate cancer behavior using transcript profiles. *J. Urol.* **172**, S28–S32 (2004).
3. Ashida, S. *et al.* Molecular features of the transition from prostatic intraepithelial neoplasia (PIN) to prostate cancer: genome-wide gene-expression profiles of prostate cancers and PINs. *Cancer Res.* **64**, 5963–5972 (2004).
4. Bostwick, D.G. & Qian, J. High-grade prostatic intraepithelial neoplasia. *Mod. Pathol.* **17**, 360–379 (2004).
5. De Marzo, A.M., Marchi, V.L., Epstein, J.I. & Nelson, W.G. Proliferative inflammatory atrophy of the prostate: implications for prostatic carcinogenesis. *Am. J. Pathol.* **155**, 1985–1992 (1999).
6. DeMarzo, A.M., Nelson, W.G., Isaacs, W.B. & Epstein, J.I. Pathological and molecular aspects of prostate cancer. *Lancet* **361**, 955–964 (2003).
7. Gleason, D.F. & Mellinger, G.T. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. *J. Urol.* **111**, 58–64 (1974).
8. Humphrey, P.A. Gleason grading and prognostic factors in carcinoma of the prostate. *Mod. Pathol.* **17**, 292–306 (2004).
9. True, L. *et al.* A molecular correlate to the Gleason grading system for prostate adenocarcinoma. *Proc. Natl. Acad. Sci. USA* **103**, 10991–10996 (2006).
10. Best, C.J. *et al.* Molecular differentiation of high- and moderate-grade human prostate cancer by cDNA microarray analysis. *Diagn. Mol. Pathol.* **12**, 63–70 (2003).
11. Lapointe, J. *et al.* Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc. Natl. Acad. Sci. USA* **101**, 811–816 (2004).
12. Singh, D. *et al.* Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* **1**, 203–209 (2002).
13. Welsh, J.B. *et al.* Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. *Cancer Res.* **61**, 5974–5978 (2001).
14. Vanaja, D.K., Cheville, J.C., Iturria, S.J. & Young, C.Y. Transcriptional silencing of zinc finger protein 185 identified by expression profiling is associated with prostate cancer progression. *Cancer Res.* **63**, 3877–3882 (2003).
15. Mootha, V.K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273 (2003).
16. Segal, E., Friedman, N., Koller, D. & Regev, A. A module map showing conditional activity of expression modules in cancer. *Nat. Genet.* **36**, 1090–1098 (2004).
17. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
18. Rhodes, D.R. & Chinnaiyan, A.M. Integrative analysis of the cancer transcriptome. *Nat. Genet.* **37** (Suppl.), S31–S37 (2005).
19. Rhodes, D.R. *et al.* Mining for regulatory programs in the cancer transcriptome. *Nat. Genet.* **37**, 579–583 (2005).
20. Tomlins, S.A. *et al.* Whole transcriptome amplification for gene expression profiling and development of molecular archives. *Neoplasia* **8**, 153–162 (2006).
21. Rhodes, D.R. *et al.* ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* **6**, 1–6 (2004).
22. Varambally, S. *et al.* Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. *Cancer Cell* **8**, 393–406 (2005).
23. Quinn, D.I., Henshall, S.M. & Sutherland, R.L. Molecular markers of prostate cancer outcome. *Eur. J. Cancer* **41**, 858–887 (2005).
24. Sato, K. *et al.* Clinical significance of alterations of chromosome 8 in high-grade, advanced, nonmetastatic prostate carcinoma. *J. Natl. Cancer Inst.* **91**, 1574–1580 (1999).
25. Kristiansen, G. *et al.* Expression profiling of microdissected matched prostate cancer samples reveals CD166/MEMD and CD24 as new prognostic markers for patient survival. *J. Pathol.* **205**, 359–376 (2005).
26. Moore, S. *et al.* Loss of stearoyl-CoA desaturase expression is a frequent event in prostate carcinoma. *Int. J. Cancer* **114**, 563–571 (2005).
27. Stuart, R.O. *et al. In silico* dissection of cell-type-associated patterns of gene expression in prostate cancer. *Proc. Natl. Acad. Sci. USA* **101**, 615–620 (2004).
28. Sumitomo, M. *et al.* Synergy in tumor suppression by direct interaction of neutral endopeptidase with PTEN. *Cancer Cell* **5**, 67–78 (2004).
29. Dhanasekaran, S.M. *et al.* Delineation of prognostic biomarkers in prostate cancer. *Nature* **412**, 822–826 (2001).
30. Koyabu, Y., Nakata, K., Mizugishi, K., Aruga, J. & Mikoshiba, K. Physical and functional interactions between Zic and Gli proteins. *J. Biol. Chem.* **276**, 6889–6892 (2001).
31. Paris, P.L. *et al.* Whole genome scanning identifies genotypes associated with recurrence and metastasis in prostate tumors. *Hum. Mol. Genet.* **13**, 1303–1313 (2004).
32. Nupponen, N.N., Kakkola, L., Koivisto, P. & Visakorpi, T. Genetic alterations in hormone-refractory recurrent prostate carcinomas. *Am. J. Pathol.* **153**, 141–148 (1998).
33. Tomlins, S.A. *et al.* Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, 644–648 (2005).
34. Tomlins, S.A. *et al.* TMPRSS2:ETV4 gene fusions define a third molecular subtype of prostate cancer. *Cancer Res.* **66**, 3396–3400 (2006).
35. Glinsky, G.V., Glinskii, A.B., Stephenson, A.J., Hoffman, R.M. & Gerald, W.L. Gene expression profiling predicts clinical outcome of prostate cancer. *J. Clin. Invest.* **113**, 913–923 (2004).
36. Konishi, N., Shimada, K., Ishida, E. & Nakamura, M. Molecular pathology of prostate cancer. *Pathol. Int.* **55**, 531–539 (2005).
37. Trotman, L.C. *et al.* Identification of a tumour suppressor network opposing nuclear Akt function. *Nature* **441**, 523–527 (2006).
38. Tran, H., Brunet, A., Griffith, E.C. & Greenberg, M.E. The many forks in FOXO's road. *Sci. STKE* **2003**, RE5 (2003).
39. Vailancourt, L. *et al.* Effect of neoadjuvant endocrine therapy (combined androgen blockade) on normal prostate and prostatic carcinoma. A randomized study. *Am. J. Surg. Pathol.* **20**, 86–93 (1996).
40. Tomlins, S.A., Rubin, M.A. & Chinnaiyan, A.M. Integrative biology of prostate cancer progression. *Ann. Rev. Pathol. Mech. Dis.* **1**, 243–271 (2006).
41. De Marzo, A.M. *et al.* Pathological and molecular mechanisms of prostate carcinogenesis: implications for diagnosis, detection, prevention, and treatment. *J. Cell. Biochem.* **91**, 459–477 (2004).
42. Fischer, A.H., Bardarov, S., Jr. & Jiang, Z. Molecular aspects of diagnostic nucleolar and nuclear envelope changes in prostate cancer. *J. Cell. Biochem.* **91**, 170–184 (2004).
43. Ruggero, D. & Pandolfi, P.P. Does the ribosome translate cancer? *Nat. Rev. Cancer* **3**, 179–192 (2003).
44. Hendriksen, P.J. *et al.* Evolution of the androgen receptor pathway during progression of prostate cancer. *Cancer Res.* **66**, 5012–5020 (2006).
45. Stanbrough, M. *et al.* Increased expression of genes converting adrenal androgens to testosterone in androgen-independent prostate cancer. *Cancer Res.* **66**, 2815–2825 (2006).
46. Chen, C.D. *et al.* Molecular determinants of resistance to antiandrogen therapy. *Nat. Med.* **10**, 33–39 (2004).
47. Feldman, B.J. & Feldman, D. The development of androgen-independent prostate cancer. *Nat. Rev. Cancer* **1**, 34–45 (2001).
48. Lin, B. *et al.* Prostate-localized and androgen-regulated expression of the membrane-bound serine protease TMPRSS2. *Cancer Res.* **59**, 4180–4184 (1999).
49. Rhodes, D.R. *et al.* Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proc. Natl. Acad. Sci. USA* **101**, 9309–9314 (2004).
50. Storey, J.D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **100**, 9440–9445 (2003).