DISCOVERY AND CHARACTERIZATION OF RECURRENT GENE FUSIONS IN PROSTATE CANCER

by

Scott A. Tomlins

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Molecular and Cellular Pathology)
in The University of Michigan
2007

Doctoral Committee:

Professor Kathleen R. Cho, Chair
Professor Arul M. Chinnaiyan
Associate Professor Thomas J. Giordano
Assistant Professor Anuj Kumar
Assistant Professor Andrew P. Lieberman

ACKNOWLEDGEMENTS

None of the work described here could have been completed without the help of a large number of people, to whom I am truly indebted. I would first like to thank my mentor, Arul Chinnaiyan, for enabling me to have a truly outstanding doctoral research experience. I would also like to thank my committee members and all the members of the Chinnaiyan lab, as much of the work described here was only possible due to the exceptional work of others. In particular, I would like to thank my colleague and friend Daniel Rhodes. I would also like to thank Saravana Dhanasekaran and Sooryanarayana Varambally, who have worked with me since my first day in the lab. I would also like to thank Rohit Mehra, Xuhong Cao and Anjana Menon, who have contributed to all of the work described here. Additionally, I would like to thank the students and residents who have worked on these projects. I would also like to thank the people from the University of Michigan and outside institutions who provided assistance, reagents, or comments on the projects described here, with information provided in the resulting manuscripts. Finally, I'd like to thank my friends and family for their support.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Prostate cancer displays considerable clinical, morphological and biological heterogeneity. The discrete genetic events that lead to prostate cancer development are unclear. Classical genetic techniques have provided only limited information about the pathogenesis of prostate cancer progression. Nevertheless, several candidate genes and pathways have been implicated in prostate cancer development, such as *GSTP1*, *PTEN*, *NKX3.1*, *AR*, *c-Myc* and the hedgehog pathway. High throughput techniques have expanded the number of candidate genes exponentially, including some whose role in prostate cancer pathogenesis has been studied, including *HPN, AMACR*, *FASN* and *EZH2*. However, the techniques used to study the prostate cancer genome, transcriptome and proteome generate massive amounts of data that have yet to be integrated and explored. In order to move beyond candidate gene identification and develop a comprehensive understanding of cancer pathogenesis, integrative approaches will be needed to analyze this data on a global level. Here we review candidate genes involved in prostate cancer pathogenesis in biological and clinical context and demonstrate how integrated analysis of high throughput data can augment our understanding of prostate cancer.

Although prostate cancer (PCA) has surpassed lung cancer as the most common non-cutaneous cancer in American men, its etiology remains largely unknown. Retrospective studies have suggested that by the age of 80, more than half of all American men have some cancer in their prostate (1), yet most men shown no symptoms and clinical treatment is not needed. However, some PCAs are very aggressive and quickly spread locally and metastasize throughout the body. The advent of the prostate specific antigen (PSA) screening test has led to increasing number of cancers being detected at an early stage. Together, these factors have complicated treatment decisions

and a more thorough understanding of PCA progression and pathogenesis is urgently needed.

Marked incidence differences between industrialized Western countries and East Asian countries suggest that lifestyle components contribute to PCA development. The higher incidence in Western countries is commonly attributed to dietary factors, such as increased meat and dairy intake along with reduced fruit and vegetable consumption, although this is a matter of some controversy. In addition, although there is evidence for a strong genetic component of PCA and several genetic loci have been implicated in hereditary PCA, the functional role of these genes in PCA pathogenesis is speculative (2). Classically, most human cancers are thought to develop through a series of sequentially accumulated genetic events, culminating in carcinoma and metastatic disease, and these results are reflected at the histologic, genetic, and clinical level. Unfortunately, many aspects of PCA biology do not fit with this model and it is unclear if this conceptual framework should be applied to understanding PCA pathogenesis.

Histologically, PCA demonstrates a progression to carcinoma somewhat reminiscent of other epithelial cancers. In the normal prostate, the glandular epithelium consists of two cell layers, the basal layer and the differentiated luminal secretory layer, with rare neuroendocrine (NE) cells distributed mostly in the basal layer. Prostatic intraepithelial neoplasia (PIN), characterized by nuclear and architectural changes in luminal epithelial cells with maintenance of the basal epithelium and basement membrane, is an accepted precursor lesion of frank adenocarcinoma. Proliferative inflammatory atrophy (PIA) is a lesion that has recently been proposed to be a candidate precursor lesion to PIN and/or PCA, based on the hypothesis that inflammation drives PCA development (3-6). This hypothesis is consistent with epidemiological data, as the Western diet is presumed to be high in potential carcinogens and low in protective antioxidants. Further, some candidate loci implicated in hereditary PCA encode viral and bacterial response genes and some genetic events in sporadic cancer (see GSTP1 below) are also consistent with this hypothesis (5).

PCA is also often multifocal, as several distinct foci of PIN and/or PCA can be found in a single prostate gland, each showing different degrees of cellular dysplasia and tissue disorganization. PCA is most commonly graded as described by the method of

Gleason (7), which assigns a score based on the glandular architecture (lower scores indicate a well differentiated glandular pattern). Reflecting the multifocal nature of PCA, the Gleason score is composed of the two most common patterns, as PCA is often multifocal with distinct patterns present in the same tumor. Even so, histologically identical tumors of the same stage can show remarkably different clinical courses, ranging from an indolent disease to rapidly invasive and metastatic. At present, it is unclear whether these histologic and clinical differences are driven by corresponding molecular differences and what the implications are for the pathogenesis of PCA.

The histological and clinical heterogeneity of PCA is also reflected at the molecular level. Little evidence exists for a clearly defined series of genetic events leading to PCA development. There are no obvious syndromes predisposing to PCA and while classical methods of identifying candidate genes have provided important candidates, very few PCAs share all or even many of the same alterations. However, molecular studies have identified several candidate genes and pathways that are likely to be important in PCA pathogenesis and progression that are consistent with important aspects of epidemiological and biological features of PCA. The candidate genes and pathways described below are shown in the context of the histological progression of PCA in **Figure 1.1**.

Decreased expression of the pi-class glutathione S-transferase through somatic hypermethylation of the *GSTP1* locus is one of the most frequently reported events in PCA. Glutathione S-transferases are enzymes that protect against cancer development by catalyzing the conjugation of the chemical scavenger glutathione to reactive chemical species. While GSTP1 is normally expressed in basal cells but not in luminal secretory cells in benign prostate epithelium, GSTP1 is transcriptionally silenced in almost all PCA cells (8). This silencing, mediated by hypermethylation of the CpG island sequences in the *GSTP1* promoter, occurs in more than 90% of PCA cases (9). However, *GSTP1* does not appear to act as a classical tumor suppressor, as PCA lines forced to express GSTP1 grow well in vitro and in vivo (10). Although the mechanism by which *GSTP1* is selectively targeted for methylation is not known, this phenomenon provides important information about PCA pathogenesis.

Importantly, up to 60% of PIN, the accepted precursor lesion of PCA, also display loss of GSTP1 expression through CpG island hypermethylation (8). Further, PIA lesions, the recently described candidate precursor to PIN and PCA, typically display high levels of GSTP1 expression (4), although approximately 6% show GSTP1 hypermethylation (11). These observations have provided support for the hypothesis that inflammation is important to the pathogenesis of PCA as described above. The consequence of GSTP1 loss in PIN and PCA is thought to be increased vulnerability to electrophilic or oxidant carcinogens, such as 2-amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine (PhIP), which is found in well done or charred meats. For example, PCA cell lines overexpressing GSTP1 form significantly fewer PhIP-DNA adducts upon exposure to activated PhIP than do control cells (12). GSTP1 expression in androgen-dependent LNCaP cells also significantly inhibits the oxidation of DNA bases upon exposure to oxidant stresses (5). While the mechanism of GSTP1 silencing is unknown, the frequency of occurrence in the context of known risk factors for PCA supports roles for *GSTP1* and inflammation in PCA development.

In 1997, *phosphatase and tensin homolog* (*PTEN*) was cloned and mapped to 10q23, a region undergoing frequent deletion in a variety of tumors, including PCA (13-15). Germline *PTEN* mutations in families with the related hamartoma syndromes, Cowden syndrome and Bannayan-Riley-Ruvalcaba syndrome, were shortly identified, confirming *PTEN* as a classic tumor suppressor gene (16, 17). The region containing *PTEN* is a frequent target for heterozygous deletion in primary and metastatic prostate tumors, where loss of heterozygosity (LOH) is found up to 60% of the time. Point mutations and deletions of the *PTEN* gene have also been reported in PCA cell lines and xenografts, as well as primary and metastatic tumors (18). Although the rate of second mutational events is generally less frequent than the LOH incidence, second *PTEN* mutations have been found in up to 50% of metastatic tumors (19). However, loss of PTEN protein expression also occurs in 20% of primary tumors, particularly those with advanced grade or stage (20). Epigenetic inactivation of PTEN through promoter hypermethylation has been described in PCA xenografts, however the same phenomenon has not been identified in primary human tumors (18).

The protein encoded by *PTEN* antagonizes the phosphoinositide 3-kinase (PI3K)/Akt pathway. A summary of the PI3K/Akt pathway in prostate cancer is depicted in **Figure 1.2**. PI3K phosphorylates phosphatidylinositol-4,5-biphosphate (PIP_2) generating phosphatidylinositol-3,4,5-triphosphate (PIP_3). PIP_3 acts as a second messenger that activates multiple downstream effectors including Akt (Protein kinase B), a serine threonine kinase which promotes cell growth, survival, invasiveness and angiogenesis (18). Thus, PTEN serves to suppress the PI3K/Akt signaling pathway and the loss of PTEN results in PI3K signaling activation.

PTEN null mice are embryonic lethal (21), and *PTEN* $^{+/-}$ mice develop PIN with varying penetrance, although development of carcinoma is rare (one report of a 65 week old mouse with invasive cancer showing LOH at the wild type allele) (21-23). Consistent with its role in PCA development, the age of prostate tumorigenesis is accelerated in different PIN/PCA prone backgrounds with *PTEN* haploinsufficiency, including *NKX3.1* $^{-/-}$ (24, 25), transgenic adenocarcinoma of mouse prostate (TRAMP, prostate specific expression of simian virus (SV) 40 early genes) (26) and *p27* $^{-/-}$ (*CDKN1B*) (27, 28). Further, prostate specific deletion of both *PTEN* alleles in two mouse models resulted in completely penetrant high grade PIN and carcinoma with varying rates of metastatic disease (29, 30). Recently, a hypomorphic *PTEN* mutant mouse series demonstrated that the extent of PTEN inactivation controls PCA incidence and progression in a dose dependent manner (31). Consistent with the importance of downstream PI3K signaling events to the development of PCA, expression of an activated allele of *Akt* in the mouse prostate results in a highly penetrant PIN phenotype (32). Treatment with the rapamycin derivative RAD001 eradicated the PIN, demonstrating that the mammalian target of rapamycin kinase (mTOR), which is known to mediate Akt signaling pathways, controls PIN development in this model (33). Other lines of evidence from mouse and xenograft models implicate PTEN as specifically controlling the progression to metastasis from localized PCA (34, 35).

Further mechanisms for the observed effects of PTEN specific to the prostate have been investigated. Recently, neutral endopeptidase (NEP), also known as MME, a protein frequently showing loss of expression in primary PCA (36), has been shown to directly bind PTEN (37). NEP recruits PTEN to the plasma membrane, which increases

the stability and phosphatase activity of PTEN, resulting in Akt inactivation. In four PCA cell lines that were characterized, NEP and PTEN were never expressed simultaneously and PC-3 cells, which express neither NEP nor PTEN, showed the strongest activation of Akt (37). Taken together, these results demonstrate the importance of PTEN loss to PCA pathogenesis and support dysregulation of PI3K signaling as a general feature of PCA.

A two megabase region surrounding 8p21 is one of the most frequent sites of LOH in PCA (up to 90%) as well as PIN (up to 60%) (38, 39). Although it does not act as a classic tumor suppressor, the androgen regulated *NKX3.1* homeobox gene was localized to this region (40) and has been demonstrated to play roles in both prostate development and PCA pathogenesis. NKX3.1 is the earliest known marker of prostate formation during embryogenesis and its expression in luminal epithelium continues through adulthood. Loss of *NKX3.1* causes defects in prostatic secretion and ductal morphogenesis in mice (41). Like other NKX homeobox proteins, NKX3.1 acts a transcriptional regulator and cell line experiments suggest that NKX3.1 can regulate PSA transcription (42). Although expression of NKX3.1 in prostatic epithelium precedes androgen receptor expression, subsequent NKX3.1 expression is dependent on androgen signaling (41). Depletion of androgen in cell culture medium down-regulates NKX3.1 expression in human PCA cells, and numerous studies have confirmed that NKX3.1 is androgen regulated (40).

In addition to developmental defects, *NKX3.1* null mice also display increasing prostatic epithelial hyperplasia and dysplasia with age. By 1 to 2 years of age, the mice develop features similar to PIN, however they do not develop carcinoma (41). NKX3.1's role in the promotion of PIN is independent of prostate development, as a prostate specific deletion of a *NKX3.1* allele in adult mice results in PIN formation (43). Consistent with the idea that LOH at *NKX3.1* promotes PCA pathogenesis, NKX3.1 has been shown to cooperate with other tumor suppressors in mouse models of cancer. As described above, compound *PTEN* $^{+/-}$;*NKX3.1* $^{-/-}$ mutant mice show increased development of high grade PIN with strong activation of Akt and older mice show invasive carcinoma and develop metastases (24, 25, 44).

Although 8p21 is deleted at high frequency in PCA as described above, the remaining *NKX3.1* allele has not been demonstrated to undergo mutational inactivation, as would be expected for a classical tumor suppressor (45). *PTEN* $^{+/-}$;*NKX3.1* $^{+/-}$ compound heterozygotes are consistent with this idea, as the wild type *PTEN* allele is lost, while the wild type *NKX3.1* allele remains intact (24). However, NKX3.1 protein expression is down-regulated in approximately 20% of PIN and localized PCA along with 78% of metastases (46), although a recent study did not confirm the loss of expression (47). However, heterozygous *NKX3.1* $^{+/-}$ mice also demonstrate dysplasia and PIN formation and display loss of NKX3.1 protein expression after the development of PIN (43). Recent evidence also suggests that NKX3.1 may be selectively methylated in PCA (48).

By utilizing *NKX3.1* $^{+/-}$ and *NKX3.1* $^{-/-}$ mice, Magee and colleagues clarified the mechanism of NKX3.1 dosage dependent tumor suppression (49). They demonstrated that *NKX3.1* controls the rate that proliferating luminal epithelial cells exit the cell cycle and that deletion of one or both alleles extends the proliferative phase during prostate regeneration. Through gene expression profiling, they identified discrete clusters of haploinsufficient and nonhaploinsufficient target genes. Interestingly, not all haploinsufficient genes exhibited the same response to loss of one *NKX3.1* allele, as some targets showed complete loss of expression, while others were less affected. Taken together, results from these experiments demonstrate a role for NKX3.1 as a haploinsufficient tumor suppressor during the early pathogenesis of PCA.

During the normal development of the prostate and the development of PCA, cell survival depends primarily on the androgen receptor (AR). The steroid receptor is normally bound to heat-shock proteins in the cytoplasm of prostate cells. Binding of the cognate ligand, dihyrotestosterone, causes dissociation of the AR which translocates to the nucleus. After translocating to the nucleus, the AR dimerizes, binds to regulatory sequences of specific target genes and activates their transcription. Most prostate tumors are initially responsive to androgen ablation, however almost all tumors eventually become androgen independent and proliferate despite androgen ablation. Determining how PCA cells become androgen independent is an intense area of investigation and many studies have focused on the role of the androgen receptor.

Recent evidence suggests that the mechanisms of androgen independence can either involve or bypass the androgen receptor, and that both types of mechanisms may co-exist in the same cancer. Pathways involving the androgen receptor include amplification or mutation of the receptor, dysregulation of growth factors or cytokines and alterations of coactivators. Several excellent reviews cover this subject in more detail than can be presented here, including those pathways that do not involve the androgen receptor gene directly (50, 51). A summary of the role of the androgen receptor in the development of hormone refractory PCA is presented in **Figure 1.3**.

Continued expression of the androgen receptor despite resistance to anti-androgen therapy occurs frequently in metastatic PCA. Independent studies have found increased androgen receptor mRNA in hormone refractory cancers compared to hormone dependent cancers and 20-30% of hormone refractory cancers have been shown to have an amplification of the *AR* gene (52-54). Androgen-refractory tumors have also been shown to contain a variety of mutations in the *AR*, increasing the number of ligands that can activate the receptor, although there is some controversy as to how often this occurs in vivo (55). Interestingly, prostate specific expression of a mutant androgen receptor in mice, *AR*-E231G, resulted in completely penetrant PCA and lung metastases, demonstrating the oncogenic effects of *AR* mutations (56). Recently, evidence from a DNA microarray study by Chen and colleagues elegantly demonstrates the importance of the *AR* to PCA progression, as summarized in **Figure 1.3**. By comparing microarray profiles of hormone-sensitive and hormone-refractory pairs of human prostate xenografts, the authors reported a two- to five-fold increase in the *AR* as the only gene expression change consistently associated with androgen independence (57). Further, this modest increase was shown to be necessary and sufficient for hormone independent progression in the animal models. Cells with the modest increased androgen receptor expression were also highly sensitive to androgens, and AR antagonists acted as agonists. Taken together, the importance of the AR to the development of hormone refractory disease suggests that strategies to down regulate the androgen receptor may also be important clinically. Studies on the role of the AR in PCA pathogenesis and progression will continue to be pursued with great interest.

Dysregulation of the *Myc* family of genes is one of the most common events in human cancer. In PCA, *c-Myc*, a strong positive regulator of cellular proliferation located at 8q24, appears to play a role in PCA pathogenesis. Gains in chromosome 8 are one of the most common amplifications in human prostate tumors, even at the early stage of PIN (58). However, entire copies of chromosome 8 are often gained and many genes localized to 8q24 are also amplified in PCA, suggesting that other target genes in the region may be important (59-61). However, high level amplification of *Myc* has been consistently identified during the progression to androgen independent metastatic disease (62-64). Cell culture work has shown that androgen-dependent LNCaP cells over-expressing Myc form colonies in soft agar independently of androgens and Myc appears to be an essential downstream target of the androgen receptor (65).

Two mouse models over-expressing Myc specifically in the prostate also suggest an earlier role in PCA progression. These models demonstrated that Myc over-expression induced PIN (66) or overt carcinoma (67) depending on the targeting promoter. In mice over-expressing Myc that develop carcinoma, microarray expression profiling studies defined a mouse Myc-PCA expression signature that shares features with human PCA as shown in **Figure 1.4**. *NKX3.1* gene expression was shown to be lost during the transition from PIN to carcinoma and this was confirmed at the protein level. Further, one of the most consistently regulated genes in the mouse Myc-PCA signature that was also coexpressed in human cancers showing a "Myc like" signature was *PIM-1*, which has been shown to cooperate with Myc in other cancers (68). Increased PIM-1 protein expression had been shown previously to predict recurrence in patients with localized PCA (69, 70). These results suggest that Myc controls aspects of PCA progression, similar to its role in other human cancers.

Recent evidence has also implicated the sonic hedgehog (SHH) pathway, known for its important role in prostate development, as being crucial to the development of PCA metastasis (**Figure 1.5**). Secreted SHH molecules bind to the patched receptor (PTCH), relieving PTCH-mediated inhibition of smoothened (SMO), a seven transmembrane G-protein coupled receptor. SMO signaling leads to activation of downstream target genes through *GLI*-dependent transcription. *PTCH* is also a target gene of *GLI*, forming a negative feedback mechanism (71). Hedgehog signaling is

required for proper growth, patterning and tissue polarity during prostate development, although hedgehog signaling has not been reported in normal adult prostate (72).

In 2004, four groups reported that the hedgehog pathway is activated in advanced PCA (73-76). Expression of *GLI-1*, *SHH*, and *PTCH* mRNA were reported with varying frequencies in benign prostate, localized PCA or metastatic PCA, however increased expression during the progression to advanced or metastatic cancer was consistently reported (73, 75, 76). Furthermore, treatment of PCA cell lines or xenografts with cyclopamide, a SHH pathway inhibitor, consistently inhibited growth, invasion and metastasis both in vitro and in vivo (73, 75, 76). Sheng and colleagues also demonstrated that suppressor of fused (SUFU), an inhibitor of SHH pathway activity, showed decreased expression in high grade tumors, and identified two somatic mutations suggesting a possible regulatory role for *SUFU* in PCA (76). Karhadkar and colleagues also demonstrated that constitutive activation of the SHH pathway can transform primitive prostate epithelial progenitor cells (75).

As this pathway has just begun to be studied, there are some discrepancies that will need to be resolved to completely understand the contribution of the SHH pathway to PCA pathogenesis. For example, Fan and colleagues demonstrated by in situ hybridization on human benign and cancerous tissue samples that *SHH* expression localized to the epithelium, while *GLI-1* expression localized to the surrounding stroma (74). This pattern was maintained in their xenograft models using LNCaP cells. However Sanchez and colleagues reported that GLI1, PTCH, and SHH are normally coexpressed in epithelial cells and not in the surrounding stroma (73). Karhadkar and colleagues also demonstrated high levels of *SHH* and *PTCH* in the DU145 cell line, and treatment of these cells with cyclopamide significantly inhibited cell growth (75), while Sanchez and colleagues reported that DU145 cells did not express *SHH* or *PTCH* and they found these cells to be insensitive to cyclopamide (73). It is interesting to note that although quantitative PCR revealed 10- to 100-fold changes between expression levels of SHH pathway members, no microarray studies have identified pathway members as being dysregulated between low and high grade tumors or localized and metastatic samples. The reasons for this discrepancy are unclear.

In addition to establishing a role for the SHH pathway in the development of metastatic PCA, Karhadkar and colleagues also demonstrated that the SHH pathway is required for epithelial cell regeneration after androgen ablation followed by androgen re-treatment. Taken together with their other observations, the authors suggest that PCA may develop from the trapping of normal prostate stem cells in a SHH pathway-dependent state of continuous renewal (75). The involvement of prostate stem cells in PCA has been proposed previously (77) and the recent identification of cells with stem cell like properties in other cancers has led to speculation that tissue-specific stem cells, through the SHH pathway, may drive many cancers, including PCA (78-80). The possibility of the SHH pathway mediating prostate stem cell behavior is consistent with several aspects of PCA. Consistent with the idea that inflammation drives PCA, the cycling between tissue damage and regeneration would promote the prolonged expansion of the stem cell pool. Subsequent genetic events could lock the stem cells into a state of continuous proliferation (78). A role for stem-cells in PCA is also consistent with the clinical course of androgen-ablation therapy, if the prostate tumor is thought to contain androgen independent cancer stem cells and differentiated tumor cells that have become androgen dependent. The initial clinical response would result from selective killing of differentiated tumor cells that are androgen-dependent, while the relapse results from the expansion of androgen-independent stem cells (77). While the hypothesis that stem cells are important to PCA is consistent with other observations, prospective markers to identify and isolate potential PCA stem cells will be needed to for experimental confirmation.

Although the candidate genes and pathways described above have provided profound insight into the pathogenesis of PCA, we are still far from understanding the key transitions in the progression and development of PCA. Due to the observed clinical and biological heterogeneity, many groups have attempted to globally characterize PCA at the molecular level, particularly using DNA microarrays. DNA microarrays have become the standard research tool for high-throughput examination of genome-wide expression changes in PCA, with over twenty five published studies examining various aspects of PCA biology (catalogued at www.oncomine.org). Most of these studies have focused on distinguishing PCA from benign prostate tissue and almost all studies report

distinct expression profiles separating cancerous from benign tissues. Unfortunately, due to differences in technology, experimental design, and the output of hundreds to thousands of candidate genes, translating these profiles to candidate genes and pathways involved in the pathogenesis of PCA has been much more difficult. Nevertheless, several attractive candidates, including *HPN*, *AMACR* and *EZH2*, have been identified and functional studies have provided evidence for their involvement in the pathogenesis of PCA progression. Further, recent studies have shown the existence of distinct classes of tumors based on their expression profiles (81-84), supporting the observed clinical heterogeneity of prostate tumors and suggesting that these profiles might represent underlying genetic changes during PCA pathogenesis and progression.

Marked over-expression of the type II serine protease *hepsin* (*HPN*) has been identified in nearly every DNA microarray study profiling PCA (85). *HPN* mRNA is up-regulated in ~90% of prostate tumors, with expression confined to epithelial cells, however the role of HPN in PCA progression is still unclear. In humans, HPN protein expression is the strongest at the PIN stage and decreases during the transition to metastatic cancer. Intensity of HPN protein expression in localized tumors has also been inversely correlated with PSA recurrence after surgical treatment (70). However other studies found that *HPN* mRNA is highest in high grade and stage tumors, although metastatic tissues were not examined (86, 87). Cell line experiments support a role for HPN early in PCA pathogenesis, as forced expression of HPN in PC-3 cells, a metastasis derived cell line that does not express endogenous HPN, decreased cell proliferation and invasion (88).

In an effort to understand the function of *HPN* in PCA, Klezovitch and colleagues created transgenic mice with forced expression of HPN specifically in the prostate (89). While these mice showed no obvious abnormalities in proliferation or apoptosis in the prostate, weakening of epithelial-stromal adhesion and disorganization and disruption of the basement membrane were observed. To determine the effect on cancer progression, these transgenic mice were crossed with the LPB-Tag mouse, which express the SV40 large T antigen specifically in the prostate. LPB-Tag mice develop high grade PIN and limited adenocarcinoma by 20 weeks of age, with no metastases. However, in addition to developing focal carcinoma, up to 55% of hepsin/LPB-Tag mice developed prominent

metastases. While these metastatic lesions showed NE differentiation, it is unclear if this is due to hepsin, as other metastases from LPB-Tag driven models also show NE differentiation. Crossing these transgenic hepsin mice with other mouse models that do not develop NE differentiation (*NKX3.1* $^{-/-}$ or *PTEN* $^{+/-}$) should clarify the role of hepsin in PCA progression. In addition to these experimental results, the magnitude and consistency of over expression in PCA suggest that *HPN* plays a role in PCA pathogenesis.

The enzyme α-methylacyl-CoA racemase (AMACR), which is involved in peroxisomal β-oxidation of dietary branched-chain fatty acids, has also been identified as an important contributor to PCA. The functionally active protein is specifically over-expressed in PCA epithelium compared to benign epithelium and DNA microarrays consistently reveal *AMACR* to be one of the most up-regulated transcripts from normal to cancerous tissue (85, 90-92). AMACR has also been found to be over-expressed, although not as frequently, in PIN (90, 91, 93). In addition to studies demonstrating it to be a possible biomarker (94-96), epidemiologic, genetic, and functional studies point to a role for AMACR and fatty acid metabolism in general in PCA pathogenesis.

The major dietary sources of branched chain fatty acids, meat and dairy products, have been implicated in the increased incidence of PCA in Western countries as described in the introduction. Recent evidence also suggests that PCA patients have higher levels of phytanic acid, which is primarily obtained from dietary intake of dairy and red meat and requires AMACR for its metabolism, than controls (97). Branched chain fatty acids themselves have also been shown to induce AMACR protein expression, but not transcript expression, in PCA cells in vitro (98). In addition, other enzymes involved in fatty acid metabolism, including D-bifunctional protein (99), fatty acid synthase (FASN) (100), and sterol-CoA desaturase (101) are also dysregulated in PCA. These observations suggest that general dysregulation of lipid metabolism, possibly though the generation of reactive oxygen species during beta-oxidation, as a feature of PCA. Genome scans for linkage in families with hereditary PCA suggest that the locus of *AMACR* (5p13) contains a PCA susceptibility gene (102, 103) and *AMACR* single nucleotide polymorphisms (SNP) cosegregate with PCA in hereditary PCA families (104). Finally, knockdown of AMACR expression in an androgen dependent PCA cell

line has also been shown to inhibit cellular proliferation (105). Future experiments, including the generation of transgenic mice which express AMACR in the prostate, should provide more information about the role of *AMACR* in PCA pathogenesis.

Although most microarray studies profiling PCA were designed to identify genes differentially expressed between normal and cancerous tissues, recent studies have also focused on identifying genes specifically dysregulated in metastatic cancer (83, 106, 107). *Enhancer of Zeste homolog 2* (*EZH2)*, a member of the polycomb group family of transcriptional repressors, has recently emerged as a candidate gene that is specifically over-expressed in metastatic and aggressive localized PCAs (69, 106, 107). In addition to being over-expressed in PCA, *EZH2* has been found to be over-expressed or amplified in other cancers, suggesting that *EZH2* may act as an oncogene (108). Recently, EZH2 expression in the content of a polycomb complex was examined during PCA progression in the *NKX3.1* $^{+/-}$; *PTEN* $^{+/-}$ mouse model described previously. PIN lesions displayed moderate elevation in the number of cells expressing EZH2 and cancerous tissues displayed further elevation as assayed at the transcript and protein level (109).

Besides acting as a histone methyltransferase responsible for gene silencing, recent evidence has demonstrated that EZH2 expression, controlled by the pRB/E2F pathway, is essential for proliferation in transformed and primary human cells, including PCA cells (107, 108, 110-112). EZH2 expression has also been shown to be down-regulated by activated p53 through p53/p21Waf1-mediated repression of the *EZH2* promoter, resulting in inhibition of cell growth (110). These results, along with DNA microarray array results demonstrating that EZH2 transcripts are often increased in PCAs with increased proliferation as assessed by Ki67 staining (83, 106) have prompted speculation that EZH2 expression may merely be a marker of proliferation.

However, several lines of evidence suggest that EZH2 has roles in cancer pathogenesis besides its role in controlling cell proliferation. In normal breast tissue sections examined with double immunoflourescence, EZH2 protein expression was only found in rare cells, which were always Ki67 positive. However, in poorly differentiated ductal carcinoma in situ and carcinoma, the majority of EZH2 positive cells were not Ki67 positive (113). EZH2 over-expression in immortalized normal breast cells had no effect on cell proliferation, however it caused increased invasion and promoted anchorage

independent growth (114). Finally, in the *NKX3.1* $^{+/-}$; *PTEN* $^{+/-}$ mouse model, increased EZH2 expression did not correlate with increased Ki67 levels (109). These results, along with the frequent amplification in various cancers, suggest that EZH2 has functional roles in cancer pathogenesis besides regulating proliferation, although further investigations are needed to determine the extent of its function in PCA pathogenesis and progression.

As described above, thousands of genes have been identified that are dysregulated during PCA progression. In addition to the marked difficulties in identifying which genes represent causal candidates rather than merely being markers of the neoplastic state, technological and experimental differences also complicate the comparison of multiple data sets (115, 116). Our group has recently developed a web site that catalogues expression profiling studies (www.oncomine.org) (117) and developed a meta-analysis method to compare expression profiling studies from different platforms (85). Analyzing four of the earliest profiling studies, a cohort of genes, including *AMACR*, *Myc*, *FASN* and *HPN*, were found to be significantly dysregulated across the data sets. Besides developing an improved way to identify candidate genes from profiling studies, bioinformatic approaches were also applied in an attempt to discover biologic pathways represented in the consistently dysregulated genes. Significant involvement of the polyamine biosynthesis pathway at a number of steps was identified, with enzymes directing substrates towards polyamines being over-expressed and the enzyme directing polyamine precursors toward an alternate pathway being under-expressed in PCA (85). Polyamines have been implicated in PCA biology and enzymes involved in their biosynthesis are candidates for chemotherapeutic agents (118). As an update to our previous work, we have performed a meta-analysis (85) of 8 studies profiling PCA vs. benign prostate tissue and identified 216 genes present in 6 of 8 studies that were dysregulated at a *Q* value (estimated false discovery rate) < 0.10. A heatmap of the 69 genes identified as being over-expressed in PCA compared to benign tissue are shown in **Figure 1.6**. Taken together, these results indicate that bioinformatic mining of the massive amounts of data generated by microarray experiments can identify candidate genes and pathways involved in PCA progression

Other techniques that measure genome or proteome wide changes during the development of PCA are also being used to identify candidate genes. Many of the

candidate genes described above were identified by their consistent loss or amplification during PCA progression. Recently developed techniques, such as array comparative genomic hybridization (CGH) (119) and SNP arrays (120), are allowing for global monitoring of gains and losses as well as LOH at increased resolutions. Although these studies have only recently begun in prostate, they have confirmed previous studies and are identifying novel chromosomal regions that may harbor important genes involved in PCA pathogenesis. A recent study using array CGH profiling 64 primary tumors identified 8p21.2, the location of *NKX3.1*, as the most frequently lost locus in the data set (121). Interestingly, amplifications at 11q13.1 were able to predict PSA recurrence independent of other clinical parameters, suggesting that this region may harbor a gene or genes controlling tumor aggressiveness. Using SNP arrays, Lieberfarb and colleagues confirmed that LOH at 8p21 as well as 10q23, the locus of *PTEN*, were amongst the most consistently observed events in 50 PCA samples (122). Further, they clustered the tumors based on their patterns of LOH and identified distinct genetic subtypes amongst the cancers. The lack of overlapping regions of LOH between the different tumor samples suggests that many sets of genetic lesions may lead to PCA development, rather than progression through a defined series of events. Although studies in this area are still in their infancy, these results support clinical and expression profiling results suggesting that PCA is a heterogeneous disease. Further technologic advances in these areas should also aid the search for candidate genes controlling various aspects of PCA progression.

As the ultimate effectors in most cellular processes are proteins, characterizing the PCA proteome should greatly facilitate understanding the disease pathogenesis. Technological developments, including tissue microarrays (123) and reverse phase protein microarrays (124), have allowed for high throughput assaying of a single protein's expression across hundreds of samples simultaneously and have been used extensively to validate PCA candidate genes. Alternatively, mass spectrometry (MS) based profiling has enormous potential to characterize and quantify large numbers of proteins. Although MS has recently been applied to various aspects of PCA, it has yet to provide much insight into PCA pathogenesis (125). Much MS work has focused on identifying protein changes in serum for diagnostic and prognostic applications, and while these studies have shown the ability to differentiate patients with cancer from

control patients, the identity of discriminatory peaks have not been determined (126, 127). Studies have also quantified the secreted proteins generated from androgen stimulation of human PCA cells in culture in attempt to discover biomarkers (128). Two recent studies profiling PCA tissue specimens demonstrate the potential of proteomics for characterizing PCA pathogenesis, similar to the role that DNA microarray technology has recently played (129, 130). Both groups generated protein profiles by surface-enhanced laser desorption/ionization-time of flight (SELDI-TOF) MS from benign and cancerous tissues. As this method does not identify the analyzed proteins, both groups used alternative techniques to identify and validate differentially expressed proteins between cancer and normal samples, although they did not characterize them functionally. Interestingly, the differential protein identified by Cheung and colleagues, growth differentiation factor 15, did not show differential expression at the transcript level using the same samples analyzed by MS (130). These results call attention to the fact that post translational regulation can result in protein expression changes during disease progression that cannot be measured by genomic or expression profiling studies, demonstrating the importance of characterizing the proteome for understanding PCA pathogenesis.

Other high throughput techniques with the potential to identify underlying causes of PCA are being developed at a rapid rate, including methods to analyze genome wide methylation (131-134) and genome wide RNA interference screens (135-137). Advances in these high throughput techniques have the potential to revolutionize the understanding of PCA pathogenesis, through cataloguing the PCA genome, transcriptome and proteome. Unfortunately, the problems inherent in comparing DNA microarray studies, differences in experimental conditions, platforms and output formats, are expanded when attempting to integrate data from these different fields (138). In an attempt to understand PCA progression, our group has recently conducted an integrative transcriptomic and proteomic analysis, as outlined in **Figure 1.7**. Using high throughput immunoblotting employing antibodies against over 1,300 proteins or post-translational modifications, we interrogated benign prostate, localized PCA and metastatic PCA tissue extracts. Sixty four proteins were altered in localized PCA compared to benign tissue and 156 additional proteins were dysregulated between localized and metastatic PCA. Candidate

progression markers were then evaluated using conventional Western blotting and tissue microarrays. Finally, in order to combine these proteomic alterations with the wealth of PCA expression data, we performed an integrative analysis with eight PCA profiling studies. Interestingly, this analysis demonstrated only 60% concordance between transcript and protein levels. Most importantly, differential proteomic alterations between localized and metastatic PCA that were concordant with expression data served as predictors of clinical outcome in PCA. Thus, in addition to serving as attractive biomarker candidates, these genes and their encoded proteins may represent important candidates for understanding the clinical heterogeneity of PCA and disease progression.

Classical genetic techniques have identified several candidate genes involved in the pathogenesis of PCA, however a clear picture of PCA development has not emerged. As described in this review, high throughput techniques have been used to expand our knowledge of the function of these genes, as well as identify thousands of other candidate genes, few of which have been demonstrated to have a functional role in PCA development. In order to identify the most promising candidates from the wealth of genomic, transcriptomic and proteomic data being generated, integrative methods must be developed. Clinical, histological and genetic observations reveal PCA to be a heterogeneous disease. Future work should be able to identify whether the development and progression of PCA has a single underlying series of genetic events or whether PCA is really a group of genetically distinct subtypes that cannot be distinguished at the histological level. The development of high throughput techniques and the careful integration of the data they produce should help to answer the unsolved etiology of PCA.

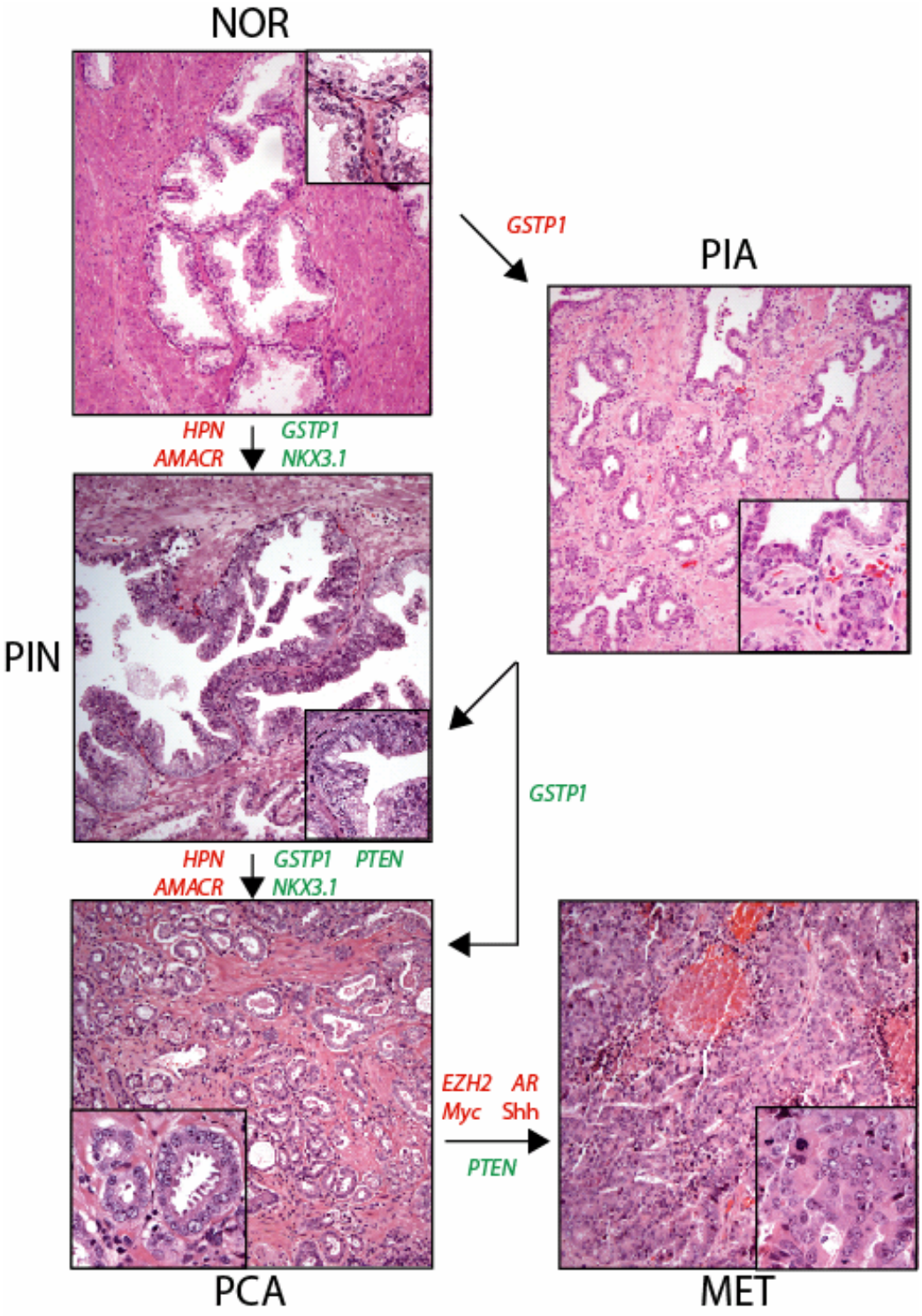

Figure 1.1. Genetic alterations during the histologic progression of prostate cancer. Photomicrographs of H&E stained sections representative of the indicated tissue or lesion (10x original magnification, 40x inserts). Arrows indicate proposed routes of progression from benign prostate glands (NOR) to precursor lesions (PIN or PIA) to localized cancer (PCA) to metastatic disease (MET). Pathway and candidate gene names in green or red represent genes and pathways under-expressed or over-expressed, respectively, during the indicated transition.

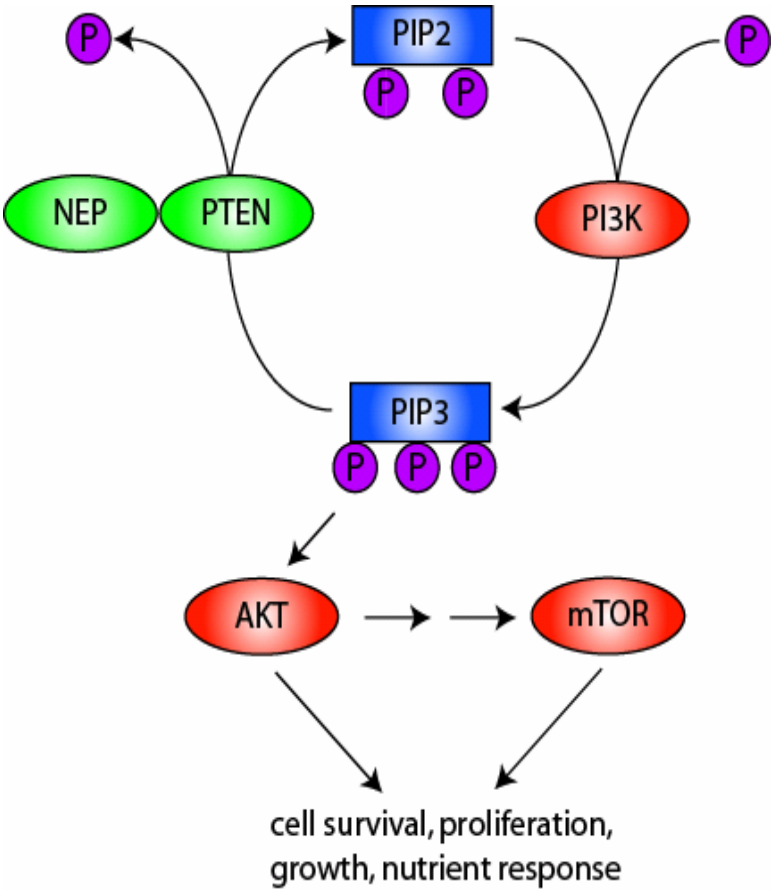

Figure 1.2. Components of the PI3K/PTEN pathway involved in prostate cancer development. Members of the phosphoinositide 3-kinase (PI3K) pathway are indicated, with genes/protein products represented in ellipses, lipid signaling molecules in rectangles and phosphate groups in purple circles. PI3K is recruited to the cell membrane by activated receptor tyrosine kinases, where it catalyzes the transfer of a phosphate group to phosphatidylinositol-4,5 bisphosphate (PIP2), yielding phosphatidylinositol-3,4,5 trisphosphate (PIP3). PIP3 transmits signals through the kinase Akt and downstream members such as mammalian target of rapamycin (mTOR). The phosphatase PTEN inhibits the activity of the PI3K pathway through inactivation of PIP3 to PIP2, and neutral endopeptidase directly binds and stabilizes PTEN. As described in the text, proteins indicated in red indicate increased activity in PCA development or mouse models predisposed to PCA development, while proteins in green show decreased activity in PCA through deletion or inactivation.

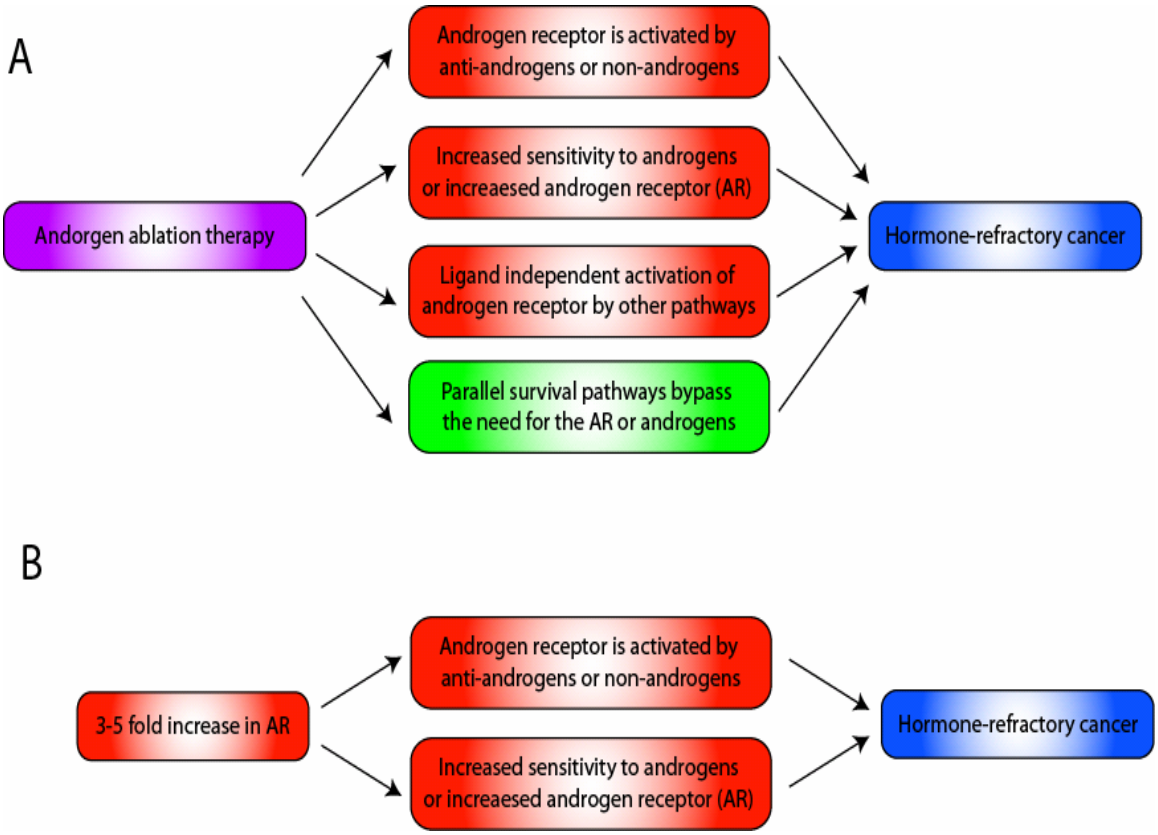

Figure 1.3. The role of the androgen receptor (AR) in the development of hormone refractory prostate cancer. A. Although PCA treated with hormone ablation initially responds clinically, almost invariably hormone refractory disease develops. The AR is thought to play a crucial role in this progression and possible mechanisms as described (51) are shown in red, while mechanisms not involving the AR are shown in green. **B.** As described in the text, Chen *et al*. identified a modest increase in the androgen receptor as the only consistent gene expression change through gene expression profiling of 7 isogenic pairs of hormone dependent/refractory xenografts. In animal models, the authors demonstrated that a modest increase in the level of AR expression resulted in hormone-refractory disease involving two of the mechanisms shown in **A**.

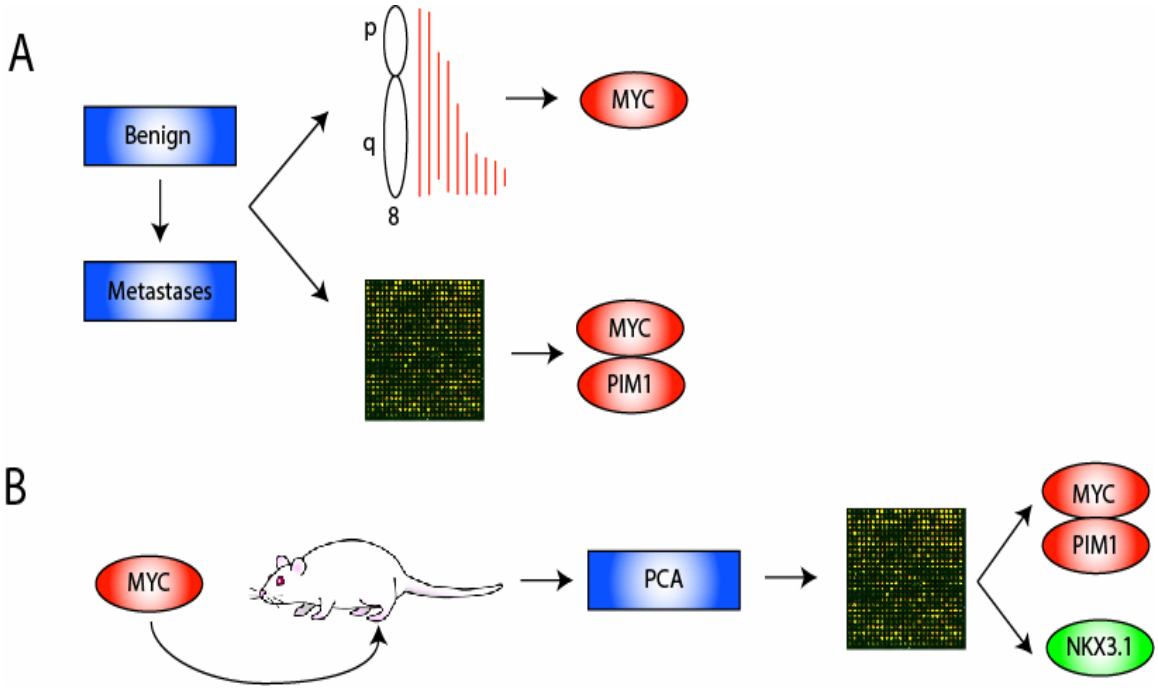

Figure 1.4. Integrative analyses reveal conserved MYC activation signatures in human prostate cancer and mouse models. **A**. As described in the text, during the transition from benign prostate to PCA and metastatic disease, multiple techniques including CGH have demonstrated that one of the most frequently gained regions is 8q24, the locus of *MYC*. Gene expression profiling of localized and metastatic PCA also demonstrate increased expression of *MYC*, often in parallel with increased expression of *PIM1*, a gene known to cooperate with *MYC* in other malignancies. **B**. Prostate specific expression of *MYC* in mouse models also resulted in PCA development, and gene expression profiling of these tumors revealed increased *PIM1* expression as seen in human tissue samples. Importantly, these tumors also showed decreased expression of *NKX3.1*, a prostate dosage dependent tumor suppressor as described in the text.

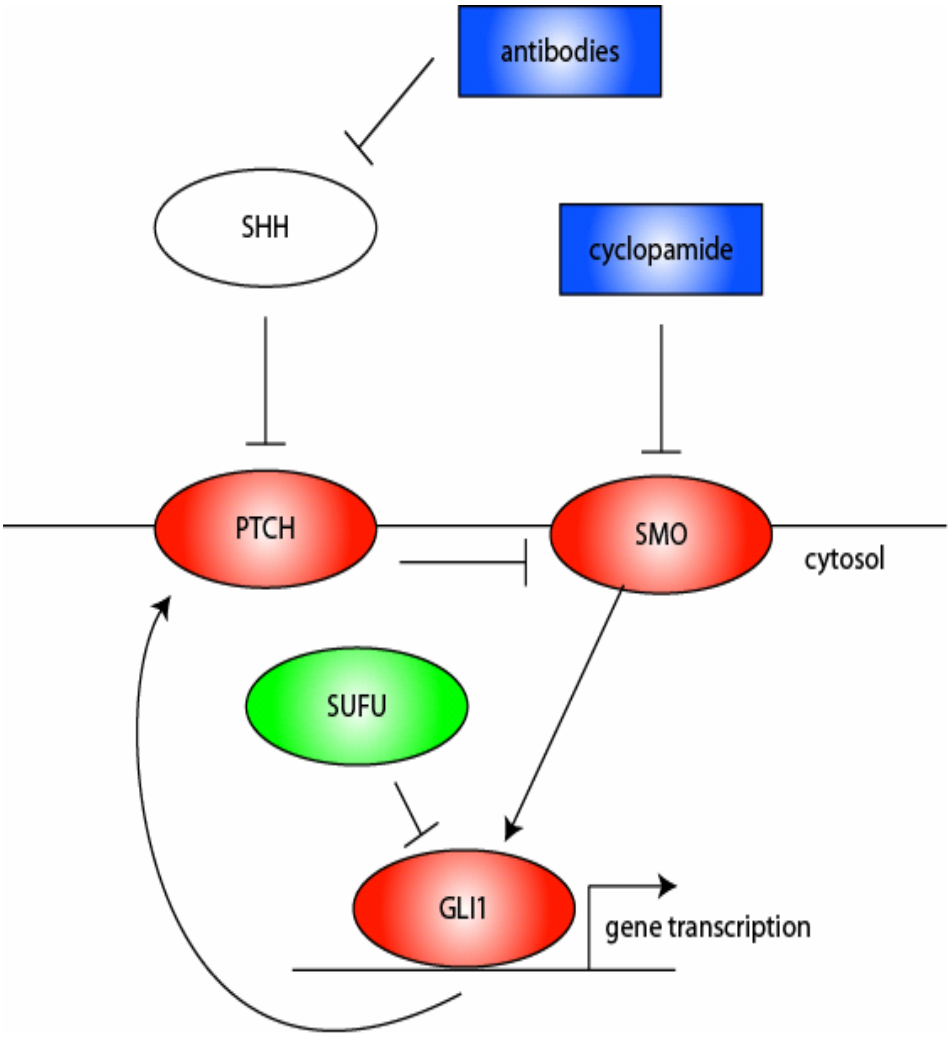

Figure 1.5. Activation and targeting of the hedgehog pathway in advanced prostate cancer. Components of the hedgehog pathway that have been implicated in advanced PCA are indicated. Protein components are indicated in ellipses and specific pathway inhibitors are indicated in rectangles. The color of the ellipse indicates transcript abundance in advanced PCA as described in the text, with red and green representing increased and decreased activity respectively. Positive (→) and negative (⊣) regulation between pathway members is indicated.

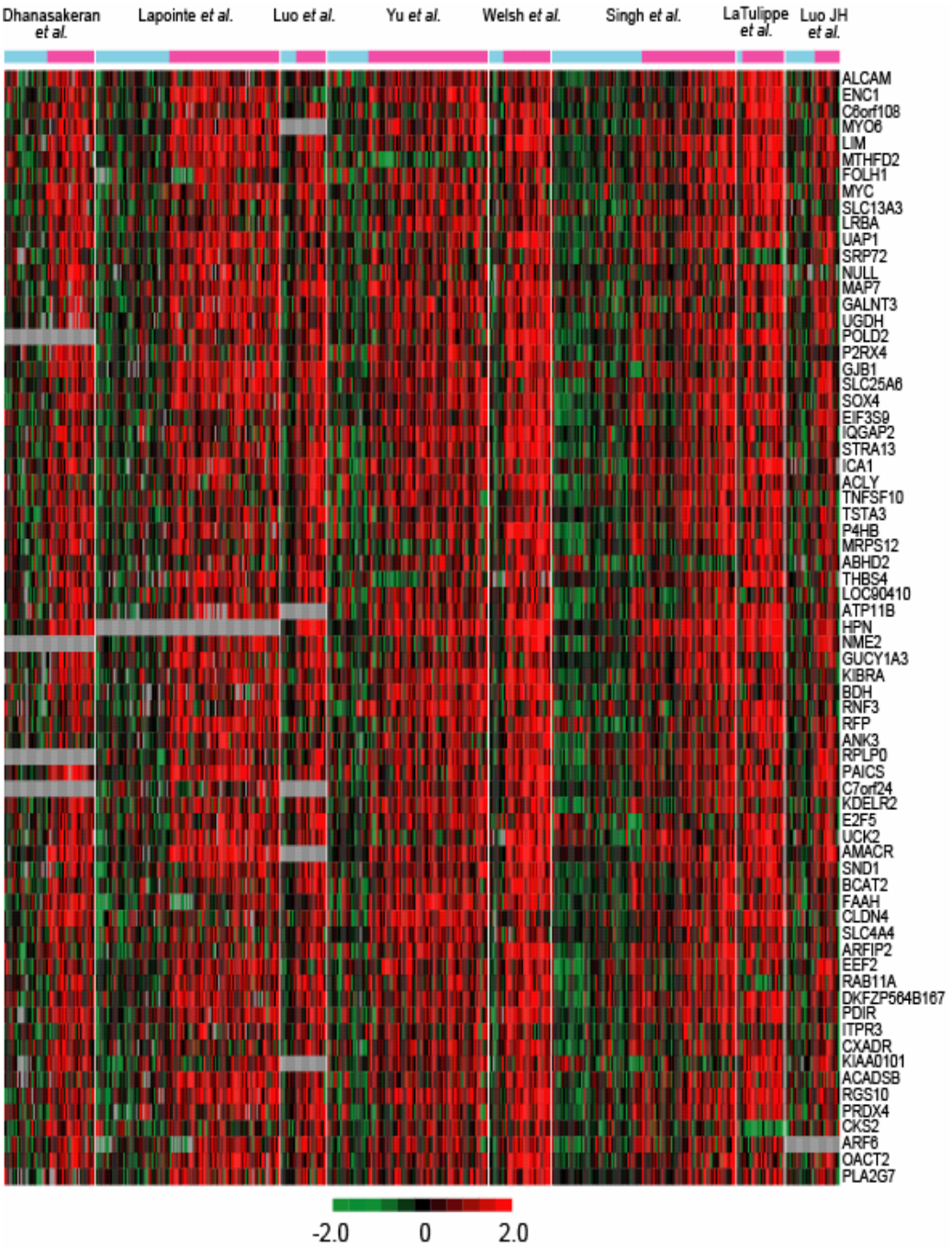

Figure 1.6. Meta-analysis of genes over-expressed in prostate cancer compared to benign prostate tissue. A meta-analysis as described (85) was performed on eight studies profiling PCA vs. benign prostate tissue (Dhanasekaran *et al.* (70), Lapointe *et al.* (83), Luo *et al.* (139), Yu *et al*. (81) Welsh *et al.* (140), Singh *et al.* (84), LaTulippe *et al.* (106) and Luo JH *et al*. (141)). 216 genes present in at least 6 of 8 studies were identified at a significance threshold of *Q (*estimated false discovery rate*)* < 0.10, and the 69 up-regulated genes are depicted. Rows represent the indicated genes and columns represent individual profiled samples, with benign or PCA samples indicated in blue or magenta respectively. Gray cells indicate missing values or genes that were not monitored in individual studies and expression values are indicated by the color bar.

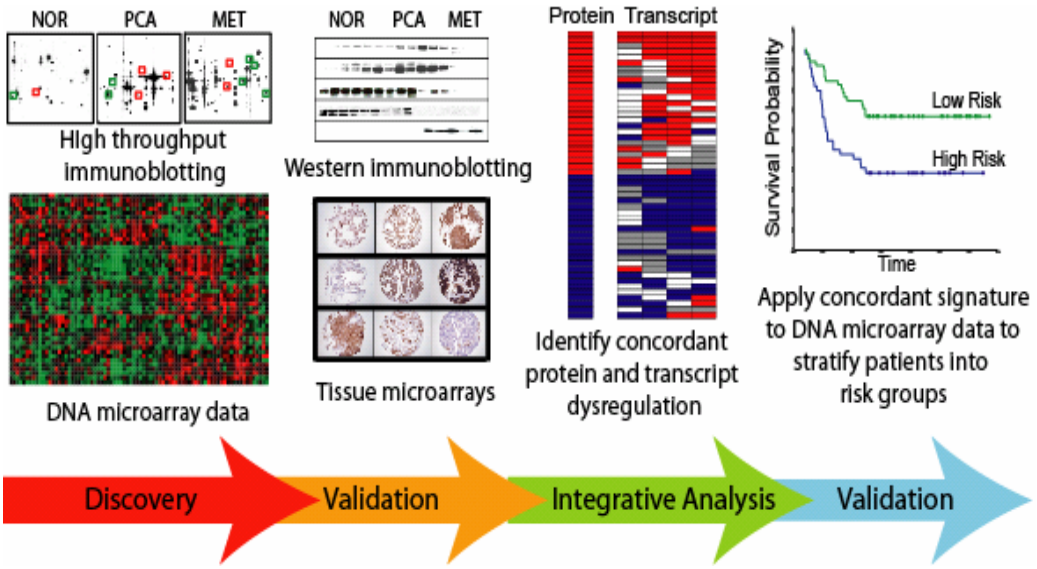

Figure 1.7. **Schema of an integrative transcriptome and proteome analysis of prostate cancer**. Steps in an integrative analysis combining high throughput proteome and expression data to characterize PCA progression are presented. We used high throughput immunoblotting to interrogate protein expression in benign prostate, localized PCA and metastatic PCA tissue extracts. Proteins showing altered expression were validated using conventional immunoblotting and tissue microarrays. Protein expression changes were then compared to transcript data from eight PCA profiling studies. We identified a signature of concordant transcript and protein changes between localized and metastatic PCA. This signature was able to identify localized tumors at increased risk for PSA recurrence after surgical resection.

REFERENCES

1. Sakr WA, Ward C, Grignon DJ, Haas GP. Epidemiology and molecular biology of early prostatic neoplasia. Mol Urol 2000;*4:*109-13;discussion 15.
2. Rubin MA, De Marzo AM. Molecular genetics of human prostate cancer. Mod Pathol 2004;*17:*380-8.
3. Franks LM. Atrophy and hyperplasia in the prostate proper. J Pathol Bacteriol 1954;*68:*617-21.
4. De Marzo AM, Marchi VL, Epstein JI, Nelson WG. Proliferative inflammatory atrophy of the prostate: implications for prostatic carcinogenesis. Am J Pathol 1999;*155:*1985-92.
5. Nelson WG, De Marzo AM, Isaacs WB. Prostate cancer. N Engl J Med 2003;*349:*366-81.
6. Smith CJ, Gardner WA, Jr. Inflammation-proliferation: possible relationships in the prostate. Prog Clin Biol Res 1987;*239:*317-25.
7. Gleason DF, Mellinger GT. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. J Urol 1974;*111:*58-64.
8. Nakayama M, Gonzalgo ML, Yegnasubramanian S, Lin X, De Marzo AM, Nelson WG. GSTP1 CpG island hypermethylation as a molecular biomarker for prostate cancer. J Cell Biochem 2004;*91:*540-52.
9. Lee WH, Morton RA, Epstein JI, et al. Cytidine methylation of regulatory sequences near the pi-class glutathione S-transferase gene accompanies human prostatic carcinogenesis. Proc Natl Acad Sci U S A 1994;*91:*11733-7.
10. Lin X, Tascilar M, Lee WH, et al. GSTP1 CpG island hypermethylation is responsible for the absence of GSTP1 expression in human prostate cancer cells. Am J Pathol 2001;*159:*1815-26.
11. Nakayama M, Bennett CJ, Hicks JL, et al. Hypermethylation of the human glutathione S-transferase-pi gene (GSTP1) CpG island is present in a subset of proliferative inflammatory atrophy lesions but not in normal or hyperplastic epithelium of the prostate: a detailed study using laser-capture microdissection. Am J Pathol 2003;*163:*923-33.
12. Nelson CP, Kidd LC, Sauvageot J, et al. Protection against 2-hydroxyamino-1-methyl-6-phenylimidazo[4,5-b]pyridine cytotoxicity and DNA adduct formation in human prostate by glutathione S-transferase P1. Cancer Res 2001;*61:*103-9.
13. Steck PA, Pershouse MA, Jasser SA, et al. Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers. Nat Genet 1997;*15:*356-62.
14. Li DM, Sun H. TEP1, encoded by a candidate tumor suppressor locus, is a novel protein tyrosine phosphatase regulated by transforming growth factor beta. Cancer Res 1997;*57:*2124-9.
15. Li J, Yen C, Liaw D, et al. PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. Science 1997;*275:*1943-7.
16. Marsh DJ, Dahia PL, Zheng Z, et al. Germline mutations in PTEN are present in Bannayan-Zonana syndrome. Nat Genet 1997;*16:*333-4.

17. Liaw D, Marsh DJ, Li J, et al. Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome. Nat Genet 1997;*16:*64-7.
18. Sansal I, Sellers WR. The biology and clinical relevance of the PTEN tumor suppressor pathway. J Clin Oncol 2004;*22:*2954-63.
19. Suzuki H, Freije D, Nusskern DR, et al. Interfocal heterogeneity of PTEN/MMAC1 gene alterations in multiple metastatic prostate cancer tissues. Cancer Res 1998;*58:*204-9.
20. McMenamin ME, Soung P, Perera S, Kaplan I, Loda M, Sellers WR. Loss of PTEN expression in paraffin-embedded primary prostate cancer correlates with high Gleason score and advanced stage. Cancer Res 1999;*59:*4291-6.
21. Di Cristofano A, Pesce B, Cordon-Cardo C, Pandolfi PP. Pten is essential for embryonic development and tumour suppression. Nat Genet 1998;*19:*348-55.
22. Podsypanina K, Ellenson LH, Nemes A, et al. Mutation of Pten/Mmac1 in mice causes neoplasia in multiple organ systems. Proc Natl Acad Sci U S A 1999;*96:*1563-8.
23. Suzuki A, de la Pompa JL, Stambolic V, et al. High cancer susceptibility and embryonic lethality associated with mutation of the PTEN tumor suppressor gene in mice. Curr Biol 1998;*8:*1169-78.
24. Kim MJ, Cardiff RD, Desai N, et al. Cooperativity of Nkx3.1 and Pten loss of function in a mouse model of prostate carcinogenesis. Proc Natl Acad Sci U S A 2002;*99:*2884-9.
25. Abate-Shen C, Banach-Petrosky WA, Sun X, et al. Nkx3.1; Pten mutant mice develop invasive prostate adenocarcinoma and lymph node metastases. Cancer Res 2003;*63:*3886-90.
26. Kwabi-Addo B, Giri D, Schmidt K, et al. Haploinsufficiency of the Pten tumor suppressor gene promotes prostate cancer progression. Proc Natl Acad Sci U S A 2001;*98:*11563-8.
27. Gao H, Ouyang X, Banach-Petrosky W, et al. A critical role for p27kip1 gene dosage in a mouse model of prostate carcinogenesis. Proc Natl Acad Sci U S A 2004;*101:*17204-9.
28. Di Cristofano A, De Acetis M, Koff A, Cordon-Cardo C, Pandolfi PP. Pten and p27KIP1 cooperate in prostate cancer tumor suppression in the mouse. Nat Genet 2001;*27:*222-4.
29. Backman SA, Ghazarian D, So K, et al. Early onset of neoplasia in the prostate and skin of mice with tissue-specific deletion of Pten. Proc Natl Acad Sci U S A 2004;*101:*1725-30.
30. Wang S, Gao J, Lei Q, et al. Prostate-specific deletion of the murine Pten tumor suppressor gene leads to metastatic prostate cancer. Cancer Cell 2003;*4:*209-21.
31. Trotman LC, Niki M, Dotan ZA, et al. Pten dose dictates cancer progression in the prostate. PLoS Biol 2003;*1:*E59.
32. Majumder PK, Yeh JJ, George DJ, et al. Prostate intraepithelial neoplasia induced by prostate restricted Akt activation: the MPAKT model. Proc Natl Acad Sci U S A 2003;*100:*7841-6.

33. Majumder PK, Febbo PG, Bikoff R, et al. mTOR inhibition reverses Akt-dependent prostate intraepithelial neoplasia through regulation of apoptotic and HIF-1-dependent pathways. Nat Med 2004;*10:*594-601.
34. Davies MA, Kim SJ, Parikh NU, Dong Z, Bucana CD, Gallick GE. Adenoviral-mediated expression of MMAC/PTEN inhibits proliferation and metastasis of human prostate cancer cells. Clin Cancer Res 2002;*8:*1904-14.
35. Bandyopadhyay S, Pai SK, Hirota S, et al. PTEN up-regulates the tumor metastasis suppressor gene Drg-1 in prostate and breast cancer. Cancer Res 2004;*64:*7655-60.
36. Freedland SJ, Seligson DB, Liu AY, et al. Loss of CD10 (neutral endopeptidase) is a frequent and early event in human prostate cancer. Prostate 2003;*55:*71-80.
37. Sumitomo M, Iwase A, Zheng R, et al. Synergy in tumor suppression by direct interaction of neutral endopeptidase with PTEN. Cancer Cell 2004;*5:*67-78.
38. Vocke CD, Pozzatti RO, Bostwick DG, et al. Analysis of 99 microdissected prostate carcinomas reveals a high frequency of allelic loss on chromosome 8p12-21. Cancer Res 1996;*56:*2411-6.
39. Emmert-Buck MR, Vocke CD, Pozzatti RO, et al. Allelic loss on chromosome 8p12-21 in microdissected prostatic intraepithelial neoplasia. Cancer Res 1995;*55:*2959-62.
40. He WW, Sciavolino PJ, Wing J, et al. A novel human prostate-specific, androgen-regulated homeobox gene (NKX3.1) that maps to 8p21, a region frequently deleted in prostate cancer. Genomics 1997;*43:*69-77.
41. Bhatia-Gaur R, Donjacour AA, Sciavolino PJ, et al. Roles for Nkx3.1 in prostate development and cancer. Genes Dev 1999;*13:*966-77.
42. Chen H, Nandi AK, Li X, Bieberich CJ. NKX-3.1 interacts with prostate-derived Ets factor and regulates the activity of the PSA promoter. Cancer Res 2002;*62:*338-40.
43. Abdulkadir SA, Magee JA, Peters TJ, et al. Conditional loss of Nkx3.1 in adult mice induces prostatic intraepithelial neoplasia. Mol Cell Biol 2002;*22:*1495-503.
44. Kim MJ, Bhatia-Gaur R, Banach-Petrosky WA, et al. Nkx3.1 mutant mice recapitulate early stages of prostate carcinogenesis. Cancer Res 2002;*62:*2999-3004.
45. Voeller HJ, Augustus M, Madike V, Bova GS, Carter KC, Gelmann EP. Coding region of NKX3.1, a prostate-specific homeobox gene on 8p21, is not mutated in human prostate cancers. Cancer Res 1997;*57:*4455-9.
46. Bowen C, Bubendorf L, Voeller HJ, et al. Loss of NKX3.1 expression in human prostate cancers correlates with tumor progression. Cancer Res 2000;*60:*6111-5.
47. Korkmaz CG, Korkmaz KS, Manola J, et al. Analysis of androgen regulated homeobox gene NKX3.1 during prostate carcinogenesis. J Urol 2004;*172:*1134-9.
48. Asatiani E, Huang WX, Wang A, et al. Deletion, methylation, and expression of the NKX3.1 suppressor gene in primary human prostate cancer. Cancer Res 2005;*65:*1164-73.
49. Magee JA, Abdulkadir SA, Milbrandt J. Haploinsufficiency at the Nkx3.1 locus. A paradigm for stochastic, dosage-sensitive gene regulation during tumor initiation. Cancer Cell 2003;*3:*273-83.

50. Koivisto P, Kolmer M, Visakorpi T, Kallioniemi OP. Androgen receptor gene and hormonal therapy failure of prostate cancer. Am J Pathol 1998;*152:*1-9.
51. Feldman BJ, Feldman D. The development of androgen-independent prostate cancer. Nat Rev Cancer 2001;*1:*34-45.
52. Koivisto P, Kononen J, Palmberg C, et al. Androgen receptor gene amplification: a possible molecular mechanism for androgen deprivation therapy failure in prostate cancer. Cancer Res 1997;*57:*314-9.
53. Linja MJ, Savinainen KJ, Saramaki OR, Tammela TL, Vessella RL, Visakorpi T. Amplification and overexpression of androgen receptor gene in hormone-refractory prostate cancer. Cancer Res 2001;*61:*3550-5.
54. Latil A, Bieche I, Vidaud D, et al. Evaluation of androgen, estrogen (ER alpha and ER beta), and progesterone receptor expression in human prostate cancer by real-time quantitative reverse transcription-polymerase chain reaction assays. Cancer Res 2001;*61:*1919-26.
55. Linja MJ, Visakorpi T. Alterations of androgen receptor in prostate cancer. J Steroid Biochem Mol Biol 2004;*92:*255-64.
56. Han G, Buchanan G, Ittmann M, et al. Mutation of the androgen receptor causes oncogenic transformation of the prostate. Proc Natl Acad Sci U S A 2005;*102:*1151-6.
57. Chen CD, Welsbie DS, Tran C, et al. Molecular determinants of resistance to antiandrogen therapy. Nat Med 2004;*10:*33-9.
58. Sakr WA, Partin AW. Histological markers of risk and the role of high-grade prostatic intraepithelial neoplasia. Urology 2001;*57:*115-20.
59. Wang R, Xu J, Saramaki O, et al. PrLZ, a novel prostate-specific and androgen-responsive gene of the TPD52 family, amplified in chromosome 8q21.1 and overexpressed in human prostate cancer. Cancer Res 2004;*64:*1589-94.
60. Rubin MA, Varambally S, Beroukhim R, et al. Overexpression, amplification, and androgen regulation of TPD52 in prostate cancer. Cancer Res 2004;*64:*3814-22.
61. Saramaki O, Willi N, Bratt O, et al. Amplification of EIF3S3 gene is associated with advanced stage in prostate cancer. Am J Pathol 2001;*159:*2089-94.
62. Edwards J, Krishna NS, Witton CJ, Bartlett JM. Gene amplifications associated with the development of hormone-resistant prostate cancer. Clin Cancer Res 2003;*9:*5271-81.
63. Nupponen NN, Kakkola L, Koivisto P, Visakorpi T. Genetic alterations in hormone-refractory recurrent prostate carcinomas. Am J Pathol 1998;*153:*141-8.
64. Visakorpi T, Kallioniemi AH, Syvanen AC, et al. Genetic changes in primary and recurrent prostate cancer by comparative genomic hybridization. Cancer Res 1995;*55:*342-7.
65. Bernard D, Pourtier-Manzanedo A, Gil J, Beach DH. Myc confers androgen-independent prostate cancer cell growth. J Clin Invest 2003;*112:*1724-31.
66. Zhang X, Lee C, Ng PY, Rubin M, Shabsigh A, Buttyan R. Prostatic neoplasia in transgenic mice with prostate-directed overexpression of the c-myc oncoprotein. Prostate 2000;*43:*278-85.
67. Ellwood-Yen K, Graeber TG, Wongvipat J, et al. Myc-driven murine prostate cancer shares molecular features with human prostate tumors. Cancer Cell 2003;*4:*223-38.

68. van Lohuizen M, Verbeek S, Krimpenfort P, et al. Predisposition to lymphomagenesis in pim-1 transgenic mice: cooperation with c-myc and N-myc in murine leukemia virus-induced tumors. Cell 1989;*56:*673-82.
69. Rhodes DR, Sanda MG, Otte AP, Chinnaiyan AM, Rubin MA. Multiplex biomarker approach for determining risk of prostate-specific antigen-defined recurrence of prostate cancer. J Natl Cancer Inst 2003;*95:*661-8.
70. Dhanasekaran SM, Barrette TR, Ghosh D, et al. Delineation of prognostic biomarkers in prostate cancer. Nature 2001;*412:*822-6.
71. Taipale J, Beachy PA. The Hedgehog and Wnt signalling pathways in cancer. Nature 2001;*411:*349-54.
72. Walterhouse DO, Lamm ML, Villavicencio E, Iannaccone PM. Emerging roles for hedgehog-patched-Gli signal transduction in reproduction. Biol Reprod 2003;*69:*8-14.
73. Sanchez P, Hernandez AM, Stecca B, et al. Inhibition of prostate cancer proliferation by interference with SONIC HEDGEHOG-GLI1 signaling. Proc Natl Acad Sci U S A 2004;*101:*12561-6.
74. Fan L, Pepicelli CV, Dibble CC, et al. Hedgehog signaling promotes prostate xenograft tumor growth. Endocrinology 2004;*145:*3961-70.
75. Karhadkar SS, Bova GS, Abdallah N, et al. Hedgehog signalling in prostate regeneration, neoplasia and metastasis. Nature 2004;*431:*707-12.
76. Sheng T, Li C, Zhang X, et al. Activation of the hedgehog pathway in advanced prostate cancer. Mol Cancer 2004;*3:*29.
77. Isaacs JT. The biology of hormone refractory prostate cancer. Why does it develop? Urol Clin North Am 1999;*26:*263-73.
78. Beachy PA, Karhadkar SS, Berman DM. Tissue repair and stem cell renewal in carcinogenesis. Nature 2004;*432:*324-31.
79. Pardal R, Clarke MF, Morrison SJ. Applying the principles of stem-cell biology to cancer. Nat Rev Cancer 2003;*3:*895-902.
80. Hudson DL. Epithelial stem cells in human prostate growth and disease. Prostate Cancer Prostatic Dis 2004;*7:*188-94.
81. Yu YP, Landsittel D, Jing L, et al. Gene expression alterations in prostate cancer predicting tumor aggression and preceding development of malignancy. J Clin Oncol 2004;*22:*2790-9.
82. Glinsky GV, Glinskii AB, Stephenson AJ, Hoffman RM, Gerald WL. Gene expression profiling predicts clinical outcome of prostate cancer. J Clin Invest 2004;*113:*913-23.
83. Lapointe J, Li C, Higgins JP, et al. Gene expression profiling identifies clinically relevant subtypes of prostate cancer. Proc Natl Acad Sci U S A 2004;*101:*811-6.
84. Singh D, Febbo PG, Ross K, et al. Gene expression correlates of clinical prostate cancer behavior. Cancer Cell 2002;*1:*203-9.
85. Rhodes DR, Barrette TR, Rubin MA, Ghosh D, Chinnaiyan AM. Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. Cancer Res 2002;*62:*4427-33.
86. Stephan C, Yousef GM, Scorilas A, et al. Hepsin is highly over expressed in and a new candidate for a prognostic indicator in prostate cancer. J Urol 2004;*171:*187-91.

87. Chen Z, Fan Z, McNeal JE, et al. Hepsin and maspin are inversely expressed in laser capture microdissectioned prostate cancer. J Urol 2003;*169:*1316-9.
88. Srikantan V, Valladares M, Rhim JS, Moul JW, Srivastava S. HEPSIN inhibits cell growth/invasion in prostate cancer cells. Cancer Res 2002;*62:*6812-6.
89. Klezovitch O, Chevillet J, Mirosevich J, Roberts RL, Matusik RJ, Vasioukhin V. Hepsin promotes prostate cancer progression and metastasis. Cancer Cell 2004;*6:*185-95.
90. Rubin MA, Zhou M, Dhanasekaran SM, et al. alpha-Methylacyl coenzyme A racemase as a tissue biomarker for prostate cancer. Jama 2002;*287:*1662-70.
91. Luo J, Zha S, Gage WR, et al. Alpha-methylacyl-CoA racemase: a new molecular marker for prostate cancer. Cancer Res 2002;*62:*2220-6.
92. Jiang Z, Woda BA, Rock KL, et al. P504S: a new molecular marker for the detection of prostate carcinoma. Am J Surg Pathol 2001;*25:*1397-404.
93. Wu CL, Yang XJ, Tretiakova M, et al. Analysis of alpha-methylacyl-CoA racemase (P504S) expression in high-grade prostatic intraepithelial neoplasia. Hum Pathol 2004;*35:*1008-13.
94. Sreekumar A, Laxman B, Rhodes DR, et al. Humoral immune response to alpha-methylacyl-CoA racemase and prostate cancer. J Natl Cancer Inst 2004;*96:*834-43.
95. Rogers CG, Yan G, Zha S, et al. Prostate cancer detection on urinalysis for alpha methylacyl coenzyme a racemase protein. J Urol 2004;*172:*1501-3.
96. Zielie PJ, Mobley JA, Ebb RG, Jiang Z, Blute RD, Ho SM. A novel diagnostic test for prostate cancer emerges from the determination of alpha-methylacyl-coenzyme a racemase in prostatic secretions. J Urol 2004;*172:*1130-3.
97. Xu J, Thornburg T, Turner AR, et al. Serum levels of phytanic acid are associated with prostate cancer risk. Prostate 2005.
98. Mobley JA, Leav I, Zielie P, et al. Branched fatty acids in dairy and beef products markedly enhance alpha-methylacyl-CoA racemase expression in prostate cancer cells in vitro. Cancer Epidemiol Biomarkers Prev 2003;*12:*775-83.
99. Zha S, Ferdinandusse S, Hicks JL, et al. Peroxisomal branched chain fatty acid beta-oxidation pathway is upregulated in prostate cancer. Prostate 2004.
100. Baron A, Migita T, Tang D, Loda M. Fatty acid synthase: a metabolic oncogene in prostate cancer? J Cell Biochem 2004;*91:*47-53.
101. Moore S, Knudsen B, True LD, et al. Loss of stearoyl-CoA desaturase expression is a frequent event in prostate carcinoma. Int J Cancer 2005;*114:*563-71.
102. Wiklund F, Gillanders EM, Albertus JA, et al. Genome-wide scan of Swedish families with hereditary prostate cancer: suggestive evidence of linkage at 5q11.2 and 19p13.3. Prostate 2003;*57:*290-7.
103. Hsieh CL, Oakley-Girvan I, Balise RR, et al. A genome screen of families with multiple cases of prostate cancer: evidence of genetic heterogeneity. Am J Hum Genet 2001;*69:*148-58.
104. Zheng SL, Chang BL, Faith DA, et al. Sequence variants of alpha-methylacyl-CoA racemase are associated with prostate cancer risk. Cancer Res 2002;*62:*6485-8.

105. Zha S, Ferdinandusse S, Denis S, et al. Alpha-methylacyl-CoA racemase as an androgen-independent growth modifier in prostate cancer. Cancer Res 2003;*63:*7365-76.
106. LaTulippe E, Satagopan J, Smith A, et al. Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastatic disease. Cancer Res 2002;*62:*4499-506.
107. Varambally S, Dhanasekaran SM, Zhou M, et al. The polycomb group protein EZH2 is involved in progression of prostate cancer. Nature 2002;*419:*624-9.
108. Bracken AP, Pasini D, Capra M, Prosperini E, Colli E, Helin K. EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. Embo J 2003;*22:*5323-35.
109. Kuzmichev A, Margueron R, Vaquero A, et al. Composition and histone substrates of polycomb repressive group complexes change during cellular differentiation. Proc Natl Acad Sci U S A 2005;*102:*1859-64.
110. Tang X, Milyavsky M, Shats I, Erez N, Goldfinger N, Rotter V. Activated p53 suppresses the histone methyltransferase EZH2 gene. Oncogene 2004;*23:*5759-69.
111. Attwooll C, Oddi S, Cartwright P, et al. A novel repressive E2F6 complex containing the polycomb group protein, EPC1, that interacts with EZH2 in a proliferation-specific manner. J Biol Chem 2005;*280:*1199-208.
112. Tonini T, Bagella L, D'Andrilli G, Claudio PP, Giordano A. Ezh2 reduces the ability of HDAC1-dependent pRb2/p130 transcriptional repression of cyclin A. Oncogene 2004;*23:*4930-7.
113. Raaphorst FM, Meijer CJ, Fieret E, et al. Poorly differentiated breast carcinoma is associated with increased expression of the human polycomb group EZH2 gene. Neoplasia 2003;*5:*481-8.
114. Kleer CG, Cao Q, Varambally S, et al. EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. Proc Natl Acad Sci U S A 2003;*100:*11606-11.
115. Rhodes DR, Chinnaiyan AM. Bioinformatics strategies for translating genome-wide expression analyses into clinically useful cancer markers. Ann N Y Acad Sci 2004;*1020:*32-40.
116. Ghosh D, Barette TR, Rhodes D, Chinnaiyan AM. Statistical issues and methods for meta-analysis of microarray data: a case study in prostate cancer. Funct Integr Genomics 2003;*3:*180-8.
117. Rhodes DR, Yu J, Shanker K, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia 2004;*6:*1-6.
118. Schipper RG, Romijn JC, Cuijpers VM, Verhofstad AA. Polyamines and prostatic cancer. Biochem Soc Trans 2003;*31:*375-80.
119. Pinkel D, Segraves R, Sudar D, et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet 1998;*20:*207-11.
120. Wang DG, Fan JB, Siao CJ, et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 1998;*280:*1077-82.

121. Paris PL, Andaya A, Fridlyand J, et al. Whole genome scanning identifies genotypes associated with recurrence and metastasis in prostate tumors. Hum Mol Genet 2004;*13:*1303-13.
122. Lieberfarb ME, Lin M, Lechpammer M, et al. Genome-wide loss of heterozygosity analysis from laser capture microdissected prostate cancer using single nucleotide polymorphic allele (SNP) arrays and a novel bioinformatics platform dChipSNP. Cancer Res 2003;*63:*4781-5.
123. Kononen J, Bubendorf L, Kallioniemi A, et al. Tissue microarrays for high-throughput molecular profiling of tumor specimens. Nat Med 1998;*4:*844-7.
124. Paweletz CP, Charboneau L, Bichsel VE, et al. Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. Oncogene 2001;*20:*1981-9.
125. Wright ME, Han DK, Aebersold R. Mass spectrometry based expression profiling of clinical prostate cncer. Mol Cell Proteomics 2005.
126. Semmes OJ, Feng Z, Adam BL, et al. Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. Assessment of platform reproducibility. Clin Chem 2005;*51:*102-12.
127. Ornstein DK, Rayford W, Fusaro VA, et al. Serum proteomic profiling can discriminate prostate cancer from benign prostates in men with total prostate specific antigen levels between 2.5 and 15.0 ng/ml. J Urol 2004;*172:*1302-5.
128. Martin DB, Gifford DR, Wright ME, et al. Quantitative proteomic analysis of proteins released by neoplastic prostate epithelium. Cancer Res 2004;*64:*347-55.
129. Liu AY, Zhang H, Sorensen CM, Diamond DL. Analysis of prostate cancer by proteomics using tissue specimens. J Urol 2005;*173:*73-8.
130. Cheung PK, Woolcock B, Adomat H, et al. Protein profiling of microdissected prostate tissue links growth differentiation factor 15 to prostate carcinogenesis. Cancer Res 2004;*64:*5929-33.
131. Yu YP, Paranjpe S, Nelson J, et al. High throughput screening of methylation status of genes in prostate cancer using an oligonucleotide methylation array. Carcinogenesis 2005;*26:*471-9.
132. Costello JF, Fruhwald MC, Smiraglia DJ, et al. Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. Nat Genet 2000;*24:*132-8.
133. Adorjan P, Distler J, Lipscher E, et al. Tumour class prediction and discovery by microarray-based DNA methylation analysis. Nucleic Acids Res 2002;*30:*e21.
134. Li LC, Carroll PR, Dahiya R. Epigenetic changes in prostate cancer: implication for diagnosis and treatment. J Natl Cancer Inst 2005;*97:*103-15.
135. Zheng L, Liu J, Batalov S, et al. An approach to genomewide screens of expressed small interfering RNAs in mammalian cells. Proc Natl Acad Sci U S A 2004;*101:*135-40.
136. Paddison PJ, Silva JM, Conklin DS, et al. A resource for large-scale RNA-interference-based screens in mammals. Nature 2004;*428:*427-31.
137. Berns K, Hijmans EM, Mullenders J, et al. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. Nature 2004;*428:*431-7.
138. Hood L, Heath JR, Phelps ME, Lin B. Systems biology and new technologies enable predictive and preventative medicine. Science 2004;*306:*640-3.

139. Luo J, Duggan DJ, Chen Y, et al. Human prostate cancer and benign prostatic hyperplasia: molecular dissection by gene expression profiling. Cancer Res 2001;*61:*4683-8.
140. Welsh JB, Sapinoso LM, Su AI, et al. Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. Cancer Res 2001;*61:*5974-8.
141. Luo JH, Yu YP, Cieply K, et al. Gene expression analysis of prostate cancers. Mol Carcinog 2002;*33:*25-35.

CHAPTER 2

RECURRENT FUSION OF *TMPRSS2* AND ETS TRANSCRIPTION FACTOR GENES IN PROSTATE CANCER

Recurrent chromosomal rearrangements have not been well characterized in common carcinomas. We used a bioinformatics approach to discover candidate oncogenic chromosomal aberrations based on outlier gene expression. Two ETS transcription factors, *ERG* and *ETV1*, were identified as outliers in prostate cancer. We identified recurrent gene fusions of the 5' untranslated region of *TMPRSS2* to *ERG* or *ETV1* in prostate cancer tissues with outlier expression. Using fluorescence in situ hybridization, we demonstrated that 23 of 29 prostate cancer samples harbor rearrangements in *ERG* or *ETV1*. Cell line experiments suggest that the androgen-responsive promoter elements of *TMPRSS2* mediate the over-expression of ETS family members in prostate cancer. These results have implications in the development of carcinomas and the molecular diagnosis and treatment of PCA.

A central aim in cancer research is to identify altered genes that play a casual role in cancer development. Many such genes have been identified through the analysis of recurrent chromosomal rearrangements that are characteristic of leukemias, lymphomas and sarcomas (1). These rearrangements are of two general types. In the first, the promoter/enhancer elements of one gene are aberrantly juxtaposed to a proto-oncogene, thus causing altered expression of an oncogenic protein. This type of rearrangement is exemplified by the apposition of immunoglobulin (*IG*) and T-cell receptor (*TCR*) genes to *MYC* leading to activation of this oncogene in B- and T-cell malignancies, respectively (2). In the second, the rearrangement fuses two genes, resulting in the production of a fusion protein that may have a new or altered activity. The prototypic example of this translocation is the *BCR-ABL* gene fusion in chronic myelogenous leukemia (CML) (3, 4). Importantly, this finding led to the development of the promising cancer drug imatinib

mesylate (Gleevec) (5). In contrast to leukemias, epithelial tumors (carcinomas) display many nonspecific, but few recurrent chromosomal rearrangements (6). This karyotypic complexity is thought to reflect secondary genomic alterations acquired during tumor progression.

We hypothesized that rearrangements and high-level copy number changes that result in marked over-expression of an oncogene should be evident in DNA microarray data, but not necessarily by traditional analytical approaches. In the majority of cancer types, heterogeneous patterns of oncogene activation have been observed; thus, traditional analytical methods that search for common activation of genes across a class of cancer samples (e.g., t-test or signal-to-noise ratio) will fail to find such oncogene expression profiles. Instead, a method that searches for marked over-expression in a subset of cases is needed. Toward this end, we developed a method termed Cancer Outlier Profile Analysis (COPA). COPA seeks to accentuate and identify outlier profiles by applying a simple numerical transformation based on the median and median absolute deviation of a gene expression profile (7).

We applied COPA to the Oncomine database (8), a compendium of 132 gene expression datasets representing 10,486 microarray experiments. COPA correctly identified several outlier profiles for genes in specific cancer types in which a recurrent rearrangement or high-level amplification is known to occur. We focused our analyses on outlier profiles of known causal cancer genes, as defined by the Cancer Gene Census (9), that ranked in the top 10 outlier profiles in an Oncomine dataset, as we felt these genes to be the most likely to participate in uncharacterized alterations. The general COPA methodology can be applied to any expression data and detailed results from the application of COPA to Oncomine datasets can be explored at www.oncomine.org.

Strikingly, in several independent datasets, COPA identified strong outlier profiles in prostate cancer for *ERG* (21q22.3) and *ETV1* (7p21.2), two genes that encode ETS family transcription factors and are involved in oncogenic translocations in Ewing's sarcoma and myeloid leukemias (10, 11). In total, COPA ranked *ERG* or *ETV1* within the top 10 outlier genes in six independent prostate cancer profiling studies.

Fusion of the 5' activation domain of the EWS gene to the highly conserved 3' DNA binding domain of an ETS family member, such as *ERG* (t(21;22)) or *ETV1*

(t(7;22)), is characteristic of Ewing's sarcoma (10, 12, 13). Because translocations involving ETS family members are functionally redundant in oncogenic transformation, only one type of translocation is typically observed in each case of Ewing's sarcoma. We hypothesized that if *ERG* and *ETV1* are similarly involved in the development of prostate cancer, their outlier profiles should be mutually exclusive—that is, each tumor should over-express only one of the two genes.

Thus, we examined the joint expression profiles of *ERG* and *ETV1* across several prostate cancer datasets and found that they invariably showed mutually exclusive outlier profiles, consistent with our hypothesis. Exclusive outlier expression of *ERG* and *ETV1* were identified in two large-scale transcriptome studies (14, 15), which profiled grossly dissected prostate tissues using different microarray platforms (**Figure 2.1)**. Similar results were obtained in prostate tissue samples obtained by laser capture microdissection (LCM). In addition to exclusive outlier expression of either *ERG* or *ETV1* in epithelial cells from prostate cancer or metastatic prostate cancer, *ETV1* and *ERG* were not over-expressed in the precursor lesion prostatic intraepithelial neoplasia (PIN) or adjacent benign epithelia. The observed exclusive outlier pattern is consistent with other translocations where an activating gene can fuse with multiple partners, such as the fusion of the immunoglobulin heavy chain promoter to *CCND1* or *FGFR3*, t(11,14) or t(4,14) respectively, in specific subsets of multiple myeloma (16).

To determine the mechanism responsible for *ERG* and *ETV1* over-expression, we identified prostate cancer cell lines and clinical specimens that over-expressed *ERG* or *ETV1* by quantitative PCR (QPCR) (**Figure 2.2**). The LNCaP prostate cancer cell line and two specimens obtained from a patient with hormone refractory metastatic disease (MET26-RP, residual primary carcinoma in the prostate and MET26-LN, a lymph node metastasis) over-expressed *ETV1*. A lymph node metastasis from a second patient (MET-28LN) and two prostate cancer cell lines, VCaP and DuCaP, over-expressed *ERG*. We did not find consistent amplification of *ERG* or *ETV1* in samples with respective transcript over-expression, so we considered the possibility of DNA rearrangements. We measured the expression level of *ETV1* exons by exon-walking QPCR in samples that displayed *ETV1* over-expression. We utilized five primer pairs spanning *ETV1* exons 2 through 7 and found that while LNCaP cells showed essentially uniform over-expression

of all measured *ETV1* exons, both MET26 specimens showed > 90% reduction in the expression of *ETV1* exons 2 and 3 compared with exons 4-7 (**Figure 2.2**).

To characterize the complete 5' *ETV1* transcript, we performed 5' RNA ligase-mediated rapid amplification of cDNA ends (RLM-RACE) on LNCaP cells and MET26-LN. In addition, we also performed RLM-RACE to obtain the complete 5' transcript of *ERG* in MET28-LN. Sequencing of the cloned products revealed fusions of the prostate-specific gene *TMPRSS2* (17) (21q22.2) with *ETV1* in MET26-LN and with *ERG* in MET28-LN (**Figure 2.2)**. In MET26-LN, two RLM-RACE PCR products were identified. The first product, *TMPRSS2:ETV1a*, resulted in a fusion of the complete exon 1 of *TMPRSS2* with the beginning of exon 4 of *ETV1* (**Figure 2.2**). The second product, *TMPRSS2:ETV1b*, resulted in a fusion of exons 1 and 2 of *TMPRSS2* with the beginning of exon 4 of *ETV1*. Both products are consistent with the exon-walking QPCR described above, where MET26-LN showed loss of over-expression in exons 2 and 3. In MET28-LN, a single RLM-RACE PCR product was identified and sequencing revealed a fusion of the complete exon 1 of *TMPRSS2* with the beginning of exon 4 of *ERG* (*TMPRSS2:ERGa*) (**Figure 2.2**).

Based on these results, we designed QPCR primer pairs with forward primers in *TMPRSS2* and reverse primers in exon 4 of *ERG* or *ETV1*. We performed SYBR Green QPCR using both primer pairs across a panel of samples from 42 cases of clinically localized prostate cancer and metastatic prostate cancer, with representative results depicted **(Figure 2.2)**. These 42 cases were selected based on previous cDNA microarray or QPCR results indicating over-expression of *ERG* or *ETV1*. We were limited to samples with remaining material and thus this cohort does not represent a random sampling. In addition to QPCR, we also performed standard reverse transcription PCR (RT-PCR) with the same primers used for QPCR, or with a different forward primer in *TMPRSS2* and reverse primers in exon 6 of *ERG* and exon 7 of *ETV1* on a subset of the samples with or without fusions as determined by QPCR**.** Electrophoresis of QPCR products and sequencing of cloned RT-PCR products from MET-26RP and MET-26LN revealed the presence of both *TMPRSS2:ETV1a* and *TMPRSS2:ETV1b*. By QPCR melt curve analysis and gel electrophoresis of QPCR and RT-PCR products, PCA4 produced a larger amplicon than *TMPRSS2:ERGa*. Subsequent RLM-RACE analysis and sequencing

of the RT-PCR product confirmed a fusion of the complete exon 1 of *TMPRSS2* with the beginning of exon 2 of *ERG* (*TMPRSS2:ERGb*). Evidence for the *TMPRSS2:ERG* and *TMPRSS2:ETV1* fusions were only found in cases that over-expressed *ERG* or *ETV1* respectively, by QPCR or DNA microarray. These results are also in agreement with the exclusive expression observed in our outlier analysis.

We used interphase fluorescence in situ hybridization (FISH) to validate the rearrangements at the chromosomal level on formalin fixed paraffin embedded (FFPE) specimens from the two cases initially used for RLM-RACE, MET26 and MET28 (**Figure 2.3**). Using probes for *TMPRSS2* and *ETV1*, normal peripheral lymphocytes (NPLs) demonstrated a pair of red and a pair of green signals (**Figure 2.3**). However, MET26 showed fusion of one pair of signals, indicative of probe overlap (**Figure 2.3**), consistent with the expression of the *TMPRSS2:ETV1*. Due to the proximity of *TMPRSS2* to *ERG* on chromosome 21, ~3 megabases, we utilized probes spanning the 5' and 3' region of the *ERG* locus to assay for gene rearrangements. Using these probes, we observed a pair of yellow signals in NPLs (**Figure 2.3**), however, in MET28, one pair of probes split into separate green and red signals, indicative of a rearrangement at the *ERG* locus (**Figure 2.3**), consistent with the expression of the *TMPRSS2:ERG*. We next performed both individual FISH analyses described above on serial tissue microarrays containing cores from 13 cases of localized prostate cancer and 16 cases of metastatic prostate cancer (**Figure 2.3**). Of 29 cases, 23 (79.3%) showed evidence of *TMPRSS2:ETV1* fusion (7 cases) or *ERG* rearrangement (16 cases).

As additional confirmation of the *ERG* rearrangement, we performed FISH on metaphase spreads of VCaP cells, which express the *TMPRSS2:ERGa* transcript. This assay revealed co-localization of 5' *TMPRSS2* and 3' *ERG* probes, with splitting of the 5' and 3' *ERG* signals, supporting the molecular results. In addition, Southern blotting using a probe in the intron between exons 1 and 2 of *TMPRSS2* revealed a unique band in VCaP cells, consistent with a rearrangement at this locus.

TMPRSS2 is expressed in normal and neoplastic prostate tissue and is strongly induced by androgen in androgen-sensitive prostate cell lines (17-19). To investigate whether the *TMPRSS2:ERG* fusion results in the androgen regulation of *ERG*, we assessed the expression of *ERG* by QPCR in androgen treated VCaP cells, which express

TMPRSS2:ERGa, and LNCaP cells, which do not express a fusion transcript. Both VCaP and LNCaP respond to androgen stimulation with increased expression of *PSA*, which is expressed at a similar level in both cells and is sensitive to the androgen receptor antagonists bicalutamide and flutamide (**Figure 2.4**). However, in addition to expressing ~2000 fold more *ERG* than LnCAP cells, only VCaP cells responded to androgen stimulation with increased *ERG* expression sensitive to bicalutamide and flutamide **(Figure 2.4)**. A similar increase in *ERG* expression upon androgen stimulation was observed in DuCaP cells, which express *TMPRSS2:ERGa*, while RWPE, PC3, and PC3 cells expressing the human androgen receptor express low levels of *ERG* that are not androgen responsive. These results suggest that the fusion with *TMPRSS2* may explain the aberrant expression of *ERG* or *ETV1* in specific subsets of prostate cancer.

The existence of recurring gene fusions of *TMPRSS2* to the oncogenic ETS family members *ERG* and *ETV1* may have important implications for understanding prostate cancer tumorigenesis and developing novel diagnostics and targeted therapeutics. Several lines of evidence suggest that these rearrangements occur in the majority of prostate cancer samples and drive the ETS family member expression. Across three independent microarray datasets, *ERG* or *ETV1* was markedly over-expressed in 95 of 167 (57%) prostate cancer cases, while over-expression was never observed across 54 benign prostate tissue samples. Furthermore, a recent study reported that *ERG* was the most commonly over-expressed oncogene by QPCR in prostate cancer, with 72.0% of cases over-expressing *ERG* (20). Using a combination of assays, we found evidence of fusion with *TMPRSS2* in 20 of 22 (>90%) cases that over-expressed *ERG* or *ETV1*, suggesting that the fusion is the most likely cause for the over-expression. By FISH analysis on a set 29 prostate cancer cases selected independently of any knowledge of *ERG* or *ETV1* expression, 23 of 29 (79%) had *TMPRSS2:ETV1* fusions or *ERG* rearrangement. It is possible that this cohort is not reflective of all prostate cancer samples and this may be an over-estimate of the prevalence of *TMPRSS2* fusions with ETS family members, as our split-signal approach can detect additional rearrangements involving *ERG*. However, the reported frequencies of *ERG* or *ETV1* over-expression in prostate cancer with our fusion transcript and FISH results suggest that *TMPRSS2* fusions with *ETV1* or *ERG* occur in the majority of prostate cancer cases. Coupled with the high incidence of prostate cancer

(an estimated 232,090 new cases will be diagnosed in the U.S. in 2005 (21)) the *TMPRSS2* fusion with ETS family members is likely to be the most common rearrangement yet identified in human malignancies and the only rearrangement present in the majority of one of the most prevalent carcinomas.

Future efforts will be directed at characterizing the expressed protein products, including the effects of N-terminal truncation of ERG and ETV1, identifying downstream targets and the functional role of the fusions in prostate cancer development. In summary, we have shown that COPA may be a useful technique for identification of chromosomal rearrangements in tumors that are difficult to study by classical cytogenetic techniques. Importantly, the existence of *TMPRSS2* fusions with ETS family members in prostate cancer suggests that causal gene rearrangements may exist in common epithelial cancers but may be masked by the multiple, non-specific chromosomal rearrangements that occur during tumor progression.

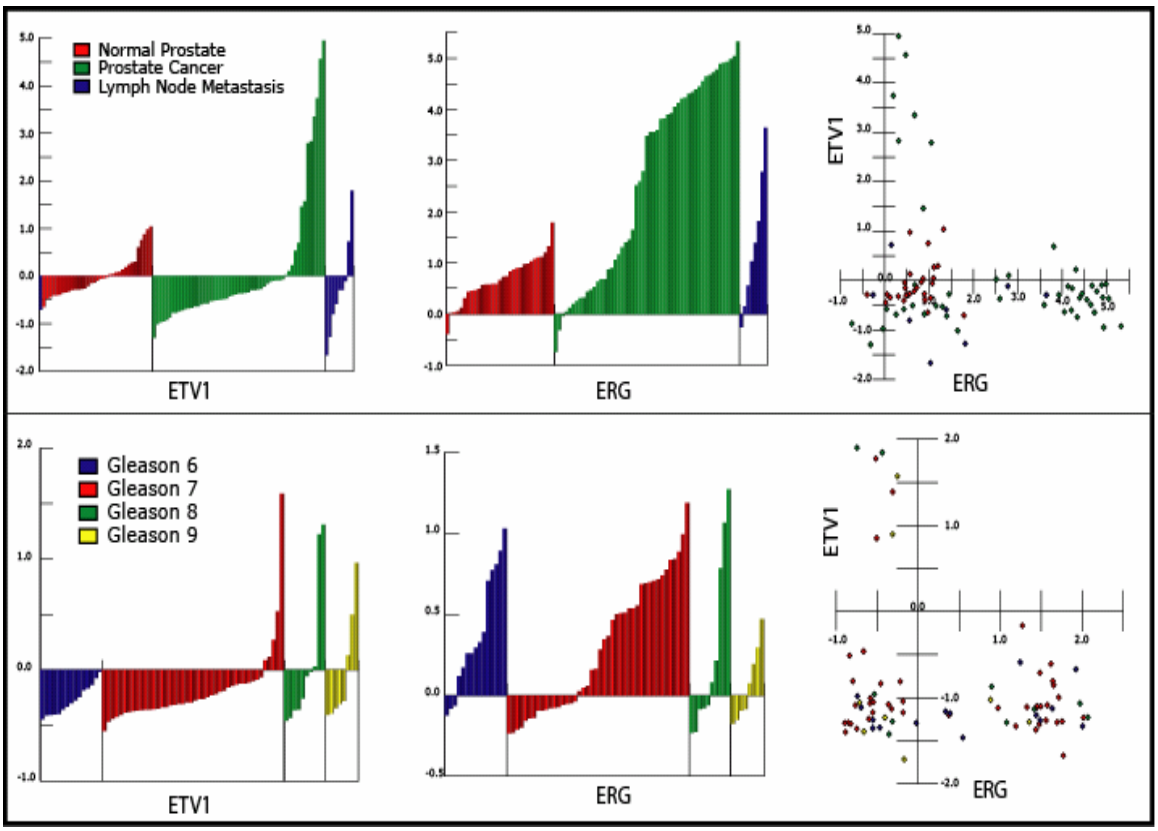

Figure 2.1. Cancer Outlier Profile Analysis. Cancer Outlier Profile Analysis (COPA) of microarray data revealed *ETV1* and *ERG* as outlier genes across multiple prostate cancer gene expression data sets. *ETV1* and *ERG* expression (normalized expression units) are shown from all profiled samples in two-large scale gene expression studies (upper panels, (14); lower panels, (15)). Visualization tools incorporated in Oncomine (www.oncomine.org) were used to generate graphical displays. Sample classes are indicated according to the color scale. In the data set from (15), prostate cancer samples were classified based on Gleason grade. Scatter plots of *ERG* and *ETV1* expression across all of the profiled samples are shown (right panels).

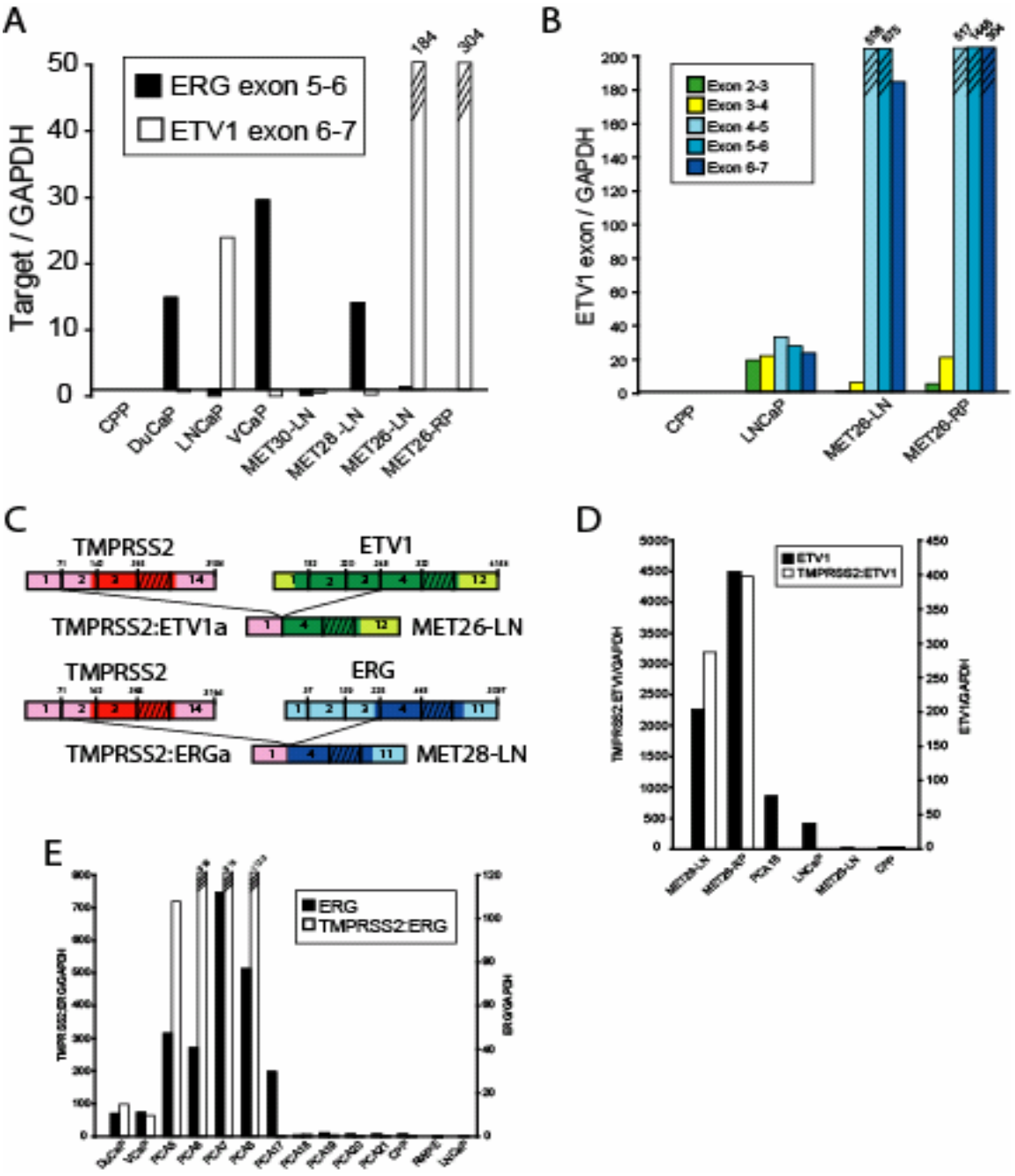

Figure 2.2. Identification and characterization of *TMPRSS2:ETV1* and *TMPRSS2:ERG* gene fusions in prostate cancer (PCA). (**A**) Prostate cancer cell lines, hormone refractory metastatic (MET) prostate cancer tissues and pooled benign prostate tissue (CPP) were analyzed for *ERG* (■) and *ETV1* (☐) mRNA expression by QPCR. (**B**) Reduced over-expression of *ETV1* exons 2 and 3 in MET26 compared with exons 4-7 in MET26 samples as assessed by QPCR. (**C**) Schematic of 5' RLM-RACE revealing fusion of *TMPRSS2* with *ETV1* in MET26-LN and *ERG* in MET28-LN. The numbers above the exons (indicated by boxes) indicate the last base of each exon. Untranslated regions are shown in corresponding lighter shades. (**D**) Validation of *TMPRSS2:ETV1* expression using fusion-specific QPCR in MET26-LN and MET26-RP. (**E**) Validation of *TMPRSS2:ERG* expression using fusion-specific QPCR in cell lines and PCA specimens.

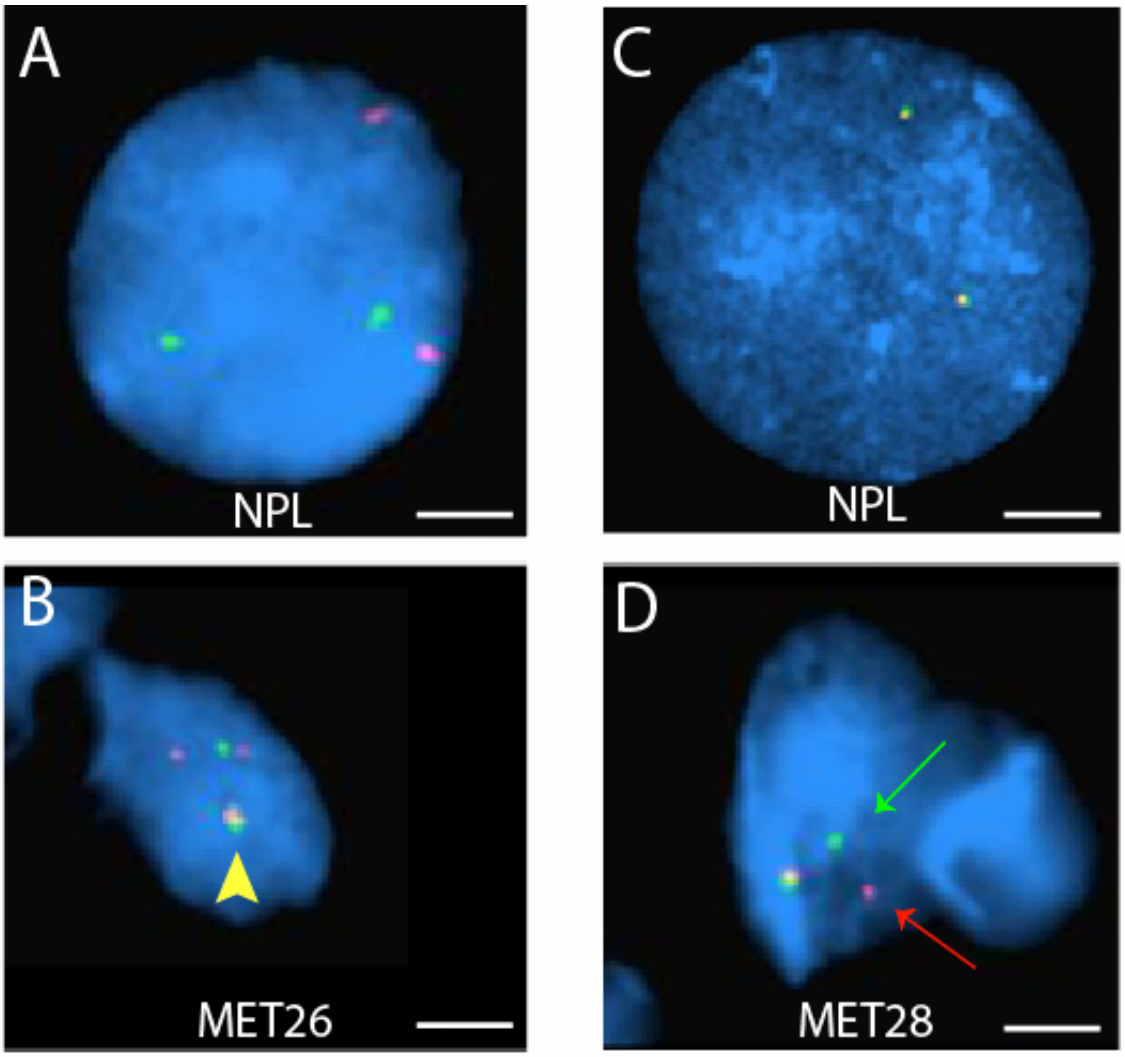

Figure 2.3. Interphase FISH on FFPE tissue sections confirms *TMPRSS2:ETV1* gene fusion and *ERG* gene rearrangement. (**A and B**) Normal peripheral lymphocytes (NPL) showed two *ETV1* (red) and two *TMPRSS2* (green) signals, while MET26 showed fusion of the signals as indicated by the yellow signal (yellow arrowhead). (**C and D**) For detection of *ERG* gene rearrangements, we utilized a split-signal approach, with two probes spanning the *ERG* locus. NPLs showed two yellow signals, indicating overlap of the 5' (green signal) and 3' (red signal) regions of *ERG*, while MET28 shows a rearrangement of *ERG* as indicated by the split signal of the 5' and 3' probes (red and green arrows). Scale bars for all images are 2.5 μm. (**E**) Matrix representation of FISH results using the same probes as (A-D) on an independent tissue microarray containing cores from clinically localized (PCA) and metastatic (MET) prostate cancer. Cores positive for *TMPRSS2:ETV1* probe fusion or split-signal *ERG* probes are indicated by colored cells. All negative findings are indicated by grey cells. The number of positive cases for each feature is indicated to the right of the matrix.

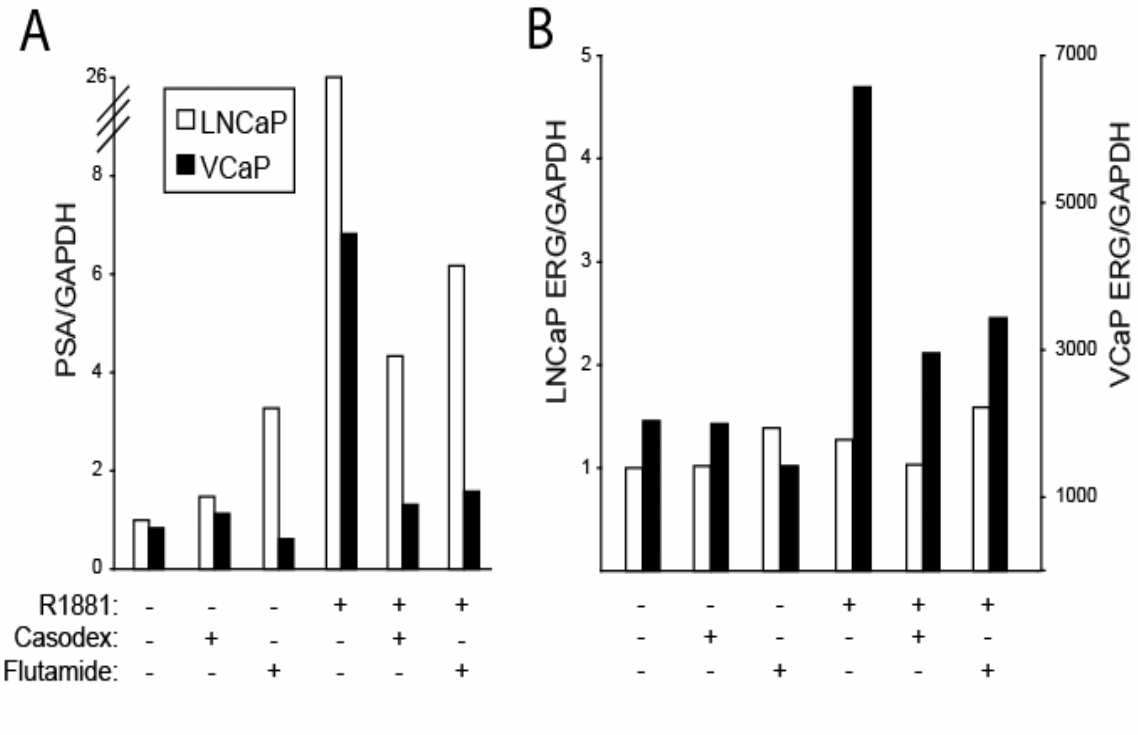

Figure 2.4 Androgen regulation of *ERG* in VCaP prostate cancer cells carrying the *TMPRSS2:ERG* fusion. (A) *PSA* expression relative to *GAPDH* in androgen sensitive LNCaP (□) and VCaP (■) cells was assessed by QPCR. **(B)** *ERG* (exon 5-6) expression relative to *GAPDH* in LNCaP (left axis, □) and VCaP (right axis, ■) cells. Cell lines were incubated with vehicle or 10 μM of the androgen receptor antagonists bicalutamide or flutamide for 2 hours before treatment for 24 hours with 0.5 nM of the synthetic androgen R1881 or vehicle as indicated. Relative *PSA* or *ERG* for each sample was normalized to the amount in the LNCaP control.

REFERENCES

1. Rowley JD. Chromosome translocations: dangerous liaisons revisited. Nat Rev Cancer 2001;*1:*245-50.
2. Rabbitts TH. Chromosomal translocations in human cancer. Nature 1994;*372:*143-9.
3. Rowley JD. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. Nature 1973;*243:*290-3.
4. de Klein A, van Kessel AG, Grosveld G, et al. A cellular oncogene is translocated to the Philadelphia chromosome in chronic myelocytic leukaemia. Nature 1982;*300:*765-7.
5. Deininger M, Buchdunger E, Druker BJ. The development of imatinib as a therapeutic agent for chronic myeloid leukemia. Blood 2005;*105:*2640-53.
6. Mitelman F. Recurrent chromosome aberrations in cancer. Mutat Res 2000;*462:*247-53.
7. Online. Materials and methods are available as supporting material on Science Online.
8. Rhodes DR, Yu J, Shanker K, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia 2004;*6:*1-6.
9. Futreal PA, Coin L, Marshall M, et al. A census of human cancer genes. Nat Rev Cancer 2004;*4:*177-83.
10. Oikawa T, Yamada T. Molecular biology of the Ets family of transcription factors. Gene 2003;*303:*11-34.
11. Hsu T, Trojanowska M, Watson DK. Ets proteins in biological control and cancer. J Cell Biochem 2004;*91:*896-903.
12. Jeon IS, Davis JN, Braun BS, et al. A variant Ewing's sarcoma translocation (7;22) fuses the EWS gene to the ETS gene ETV1. Oncogene 1995;*10:*1229-34.
13. Sorensen PH, Lessnick SL, Lopez-Terrada D, Liu XF, Triche TJ, Denny CT. A second Ewing's sarcoma translocation, t(21;22), fuses the EWS gene to another ETS-family transcription factor, ERG. Nat Genet 1994;*6:*146-51.
14. Lapointe J, Li C, Higgins JP, et al. Gene expression profiling identifies clinically relevant subtypes of prostate cancer. Proc Natl Acad Sci U S A 2004;*101:*811-6.
15. Glinsky GV, Glinskii AB, Stephenson AJ, Hoffman RM, Gerald WL. Gene expression profiling predicts clinical outcome of prostate cancer. J Clin Invest 2004;*113:*913-23.
16. Fonseca R, Barlogie B, Bataille R, et al. Genetics and cytogenetics of multiple myeloma: a workshop report. Cancer Res 2004;*64:*1546-58.
17. Lin B, Ferguson C, White JT, et al. Prostate-localized and androgen-regulated expression of the membrane-bound serine protease TMPRSS2. Cancer Res 1999;*59:*4180-4.
18. Afar DE, Vivanco I, Hubert RS, et al. Catalytic cleavage of the androgen-regulated TMPRSS2 protease results in its secretion by prostate and prostate cancer epithelia. Cancer Res 2001;*61:*1686-92.

19. Jacquinet E, Rao NV, Rao GV, Zhengming W, Albertine KH, Hoidal JR. Cloning and characterization of the cDNA and gene for human epitheliasin. Eur J Biochem 2001;*268:*2687-99.
20. Petrovics G, Liu A, Shaheduzzaman S, et al. Frequent overexpression of ETS-related gene-1 (ERG1) in prostate cancer transcriptome. Oncogene 2005;*24:*3847-52.
21. Jemal A, Murray T, Ward E, et al. Cancer statistics, 2005. CA Cancer J Clin 2005;*55:*10-30.

CHAPTER 3

TMPRSS2:ETV4 GENE FUSIONS DEFINE A THIRD MOLECULAR SUBTYPE OF PROSTATE CANCER

While common in hematological and mesenchymal malignancies, recurrent gene fusions have not been well characterized in epithelial carcinomas. Recently, using a novel bioinformatic approach, we identified recurrent gene fusions between *TMPRSS2* and the *ETS* family members *ERG* or *ETV1* in the majority of prostate cancers. Here, we interrogated the expression of all *ETS* family members in prostate cancer profiling studies and identified marked over-expression of *ETV4* in two of 98 of cases. In one such case, we confirmed the over-expression of *ETV4* using quantitative PCR, and by rapid amplification of cDNA ends (RACE), quantitative PCR and fluorescence in situ hybridization (FISH), we demonstrate that the *TMPRSS2* (21q22) and *ETV4* (17q21) loci are fused in this case. This result defines a third molecular subtype of prostate cancer and supports the hypothesis that dysregulation of *ETS* family members through fusions with *TMRPSS2* may be an initiating event in prostate cancer development.

Despite their well defined role in hematological and mesenchymal malignancies, recurrent gene fusions have not been well characterized in epithelial carcinomas (1-3). Recently, in an effort to nominate candidate oncogenes from DNA microarray data, we developed a novel bioinformatics approach termed Cancer Outlier Profile Analysis (COPA) to identify genes markedly over-expressed in a subset of cancers. Applying the COPA approach to a compendium of tumor gene expression data, the *ETS* family transcription factors *ERG* and *ETV1* were identified as outliers across prostate cancer profiling studies. Using a variety of molecular techniques, we characterized fusions of the 5'-untranslated region of *TMPRSS2* (21q22) with *ERG* (21q22) or *ETV1* (7p21) in cases that over-expressed the respective *ETS* family member (4). *TMPRSS2* has been characterized previously as being both androgen responsive and highly expressed in the

prostate, presumably through androgen response elements (AREs) in the promoter (5-7). As a possible mechanism driving the overexpression of *ETS* family members in cases with the gene fusions, we demonstrated that androgen can induce the over-expression of *ERG* (presumably through AREs) in a *TMPRSS2:ERG* positive cell line (4). Together, these results suggested that dysregulation of *ETS* family activity through AREs upstream of *TMPRSS2* may drive prostate cancer development. Here, we describe a rare third molecular sub-type of prostate cancer, characterized by fusion of *TMPRSS2* to another *ETS* family member, *ETV4*.

While our initial COPA screen led to the characterization of *TMPRSS2* fusions with *ERG* or *ETV1,* we hypothesized that prostate cancers negative for these gene fusions may harbor rearrangements involving other *ETS* family members. By interrogating the expression of all *ETS* family members monitored in prostate cancer profiling studies from the Oncomine database (www.oncomine.org) (8), we identified marked over-expression of the *ETS* family member *ETV4* in a single prostate cancer case from each of two studies—one profiling grossly dissected tissues (9) (**Figure 3.1**) and the other profiling laser capture microdissected (LCM) tissues (**Figure 3.1**). As these cases did not over-express *ERG* or *ETV1*, and no benign prostate tissues showed over-expression, we hypothesized that fusion with *TMPRSS2* may drive the over-expression of *ETV4* in these cases. Although *ELF3* was also over-expressed in a fraction of prostate cancer cases, in both studies normal prostate tissue samples also showed marked *ELF3* over-expression, suggesting that a gene fusion driving expression in both benign and cancerous tissue is unlikely. Thus, we focused on characterizing the *ETV4* over-expressing case (designated here as PCA5) in our LCM cohort.

We isolated total RNA from PCA5 and used an exon-walking quantitative PCR (QPCR) strategy to confirm the over-expression of *ETV4*. QPCR demonstrated that exons 3' to exon 2 of *ETV4* were markedly over-expressed in this case compared to pooled benign prostate tissue (CPP) (~900 fold) and prostate cancers that did not over-express *ETV4* and were either *TMPRSS2:ERG* positive (PCA1-2) or negative (PCA3-4) (**Figure 3.2**). However, we observed a dramatic decrease (>99%) in the expression of exon 2 of *ETV4* relative to distal regions in PCA5, suggesting a possible fusion with

TMPRSS2, as observed previously in *TMPRSS2:ERG* and *TMPRSS2:ETV1* positive cases (4).

To identify the 5' end of the ETV4 transcript in PCA5, we performed RNA-ligase mediated rapid amplification of cDNA ends (RLM-RACE) using a reverse primer in exon 7. RLM-RACE revealed two transcripts, each containing 5' ends consisting of sequence located approximately 8 kb upstream of *TMPRSS2* fused to sequence from *ETV4* (**Figure 3.2**). Specifically, the 5' end of *TMPRSS2:ETV4a* consists of 47 base pairs from this region upstream of *TMPRSS2*, while the 5' end of *TMPRSS2:ETV4b* consists of the same terminal 13 base pairs. These 5' ends of both transcripts were fused to the same contiguous stretch consisting of the 9 base pairs of the intron immediately 5' to exon 3 of ETV4 and the reported reference sequence of exons 3 through the reverse primer in exon 7 of ETV4.

We confirmed the existence of both transcripts in PCA5 and their absence in CPP and PCA1-4 using QPCR, however the results could not be quantified due to no detectable amplification after 40 cycles in CPP and PCA1-4 (**Figure 3.2**). To further exclude the presence of fusion transcripts involving known exons from *TMPRSS2*, we attempted QPCR using a forward primer in exon 1 of *TMPRSS2* and the *ETV4* exon 4 reverse primer, and as expected, no product was detected in CPP or PCA1-5 (data not shown).

Whether other prostate cancers with *ETV4* dysregulation might contain *TMPRSS2:ETV4* fusion transcripts structurally more similar to *TMPRSS2:ERG* and *TMPRSS2:ETV1* transcripts (which involve known exons from *TMPRSS2*) is unknown. It is important to note that the *TMPRSS2:ETV4* fusions reported here would not contain the well characterized AREs immediately upstream of TMPRSS2. However, evidence exists for androgen responsive enhancers located upstream of the TMPRSS2 sequences present in the *TMPRSS2:ETV4* transcripts described here. Nevertheless, the marked over-expression of only *ETV4* exons involved in the fusion transcript strongly suggests that the gene fusion is responsible for the dysregulation of *ETV4*. Together, the structure of the *TMPRSS2:ETV4* fusion transcripts supports the conclusion that the regulatory elements upstream of *TMPRSS2*, rather than transcribed *TMPRSS2* sequences, drive the dysregulation of ETS family members.

To confirm the fusion of the genomic loci surrounding *TMPRSS2* (21q22) and *ETV4* (17q21) as demonstrated by RLM-RACE and QPCR, we used interphase fluorescence in situ hybridization (FISH). Using probes 5' to *TMPRSS2* and 3' to *ETV4*, we identified fusion of *TMPRSS2* and *ETV4* loci in 65% of cancerous cells from PCA5 (**Figure 3.2**). As further confirmation of the rearrangement of *ETV4*, using probes 5' and 3' to *ETV4*, 64% of cancerous cells from PCA5 showed split signals. We also performed FISH on PCA5 using two probes 3' to *ETV4*, *ERG* split signal probes and *TMPRSS2*:*ETV1* fusion probes to exclude additional rearrangements, with negative results obtained for each hybridization.

Taken together, the results from this study highlight the importance of carefully examining outlier profiles in tumor gene expression data, as most analytical methods discount profiles that do not show consistent deregulation (10-12) and would thus fail to identify *ETV4* in prostate cancer, which appears rare (2 of 98 cases). Combined with the identification of *TMPRSS2:ERG* and *TMPRSS2:ETV1* fusions, the results presented here support the hypothesis that dysregulation of *ETS* family members mediated by subversion of AREs or enhancers upstream of *TMPRSS2* is a hallmark of prostate tumorigenesis. While the majority of *ETS* family members were represented in the profiling studies examined, other *ETS* family members that were not monitored may also be rearranged in prostate cancers for which gene fusions have not been ascribed. The reason for the observed frequencies of fusion partners with *TMPRSS2* (*ERG>ETV1>ETV4*), which are consistent across independent sample sets, is unclear, although a similar situation is present in Ewing's sarcoma, where *EWS* partners with *ETS* family members in unequal frequencies (*FLI1>ERG>ETV1*) (13). Lastly, these results establish a third molecular sub-type of prostate cancer which may have prognostic and/or therapeutic relevance in the future.

Methods

To investigate the expression of *ETS* family members in prostate cancer, we selected two prostate cancer profiling studies (Lapointe *et al*. (9) and Tomlins *et al*.) present in the Oncomine database (8). Genes with an ETS domain were identified by the Interpro filter 'Ets' (Interpro ID: IPR000418). Heatmap representations were generated in

Oncomine using the 'median-center per gene' option, and the color contrast was set to accentuate *ERG* and *ETV1* differential expression.

Prostate cancer tissues (PCA1-5) were from the radical prostatectomy series at the University of Michigan, which is part of the University of Michigan Prostate Cancer Specialized Program of Research Excellence (S.P.O.R.E.) Tissue Core. All samples were collected with informed consent of the patients and prior institutional review board approval. Total RNA was isolated with Trizol (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. A commercially available pool of benign prostate tissue total RNA (CPP, Clontech, Mountain View, CA) was also used.

QPCR was performed using SYBR Green dye on an Applied Biosystems 7300 Real Time PCR system (Applied Biosystems, Foster City, CA) as described (4). The amount of each target gene relative to the housekeeping gene glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) for each sample was reported. The relative amount of the target gene was calibrated to the relative amount from the pool of benign prostate tissue (CPP). All oligonucleotide primers were synthesized by Integrated DNA Technologies (Coralville, IA). *GAPDH* primers were as described (14). Primers for exons of *ETV4* were as follows (listed 5' to 3'): ETV4_exon2-f: CCGGATGGAGCGGAGGATGA, ETV4_exon2-r: CGGGCGATTTGCTGCTGAAG, ETV4_exon3-f: GCCGCCCCTCGACTCTGAA,ETV4_exon4-r: GAGCCACGTCTCCTGGAAGTGACT, ETV4_exon11-f: CTGGCCGGTTCTTCTGGATGC, ETV4_exon12-r: CGGGCCGGGGAATGGAGT, ETV4_3'UTR-f: CCTGGAGGGTACCGGTTTGTCA, ETV4_3'UTR-r: CCGCCTGCCTCTGGGAACAC. Exons were numbered by alignment of the RefSeq for *ETV4* (NM_001986.1) with the May 2004 freeze of the human genome using the UCSC Genome Browser. For QPCR confirmation of *TMPRSS2*:*ETV4* fusion transcripts, TMPRSS2:*ETV4*a-f (AAATAAGTTTGTAAGAGGAGCCTCAGCATC) and TMPRSS2:*ETV4*b-f (ATCGTAAAGAGCTTTTCTCCCCGC), which detects both *TMPRSS2*:*ETV4*a and *TMPRSS2*;*ETV4*b transcripts, were used with ETV4_exon4-r. RLM-RACE was performed using the GeneRacer RLM-RACE kit (Invitrogen), according to the manufacturer's instructions as described (4). To obtain the 5' end of ETV4, first-strand cDNA from PCA5 was amplified using the GeneRacer 5' Primer and

ETV4_exon4-r or ETV4_exon7-r (GAAAGGGCTGTAGGGGCGACTGT). Products were cloned and sequenced as described (4). Equivalent 5' ends of the *TMPRSS2*:*ETV4* transcripts were obtained from both primer pairs.

Formalin-fixed paraffin-embedded (FFPE) tissue sections were used for interphase FISH. Deparaffinized tissue was treated with 0.2 M HCl for 10 min, 2x SSC for 10 min at 80^{o} C and digested with Proteinase K (Invitrogen) for 10 min. The tissues and BAC probes were co-denatured for 5 min at 94^{o} C and hybridized overnight at 37^{o} C. Post-hybridization washing was with 2x SSC with 0.1% Tween-20 for 5 min and fluorescent detection was performed using anti-digoxigenin conjugated to fluorescein (Roche Applied Science, Indianapolis, IN) and streptavidin conjugated to Alexa Fluor 594 (Invitrogen). Slides were counterstained and mounted in ProLong Gold Antifade Reagent with DAPI (Invitrogen). Slides were examined using a Leica DMRA fluorescence microscope (Leica, Deerfield, IL) and imaged with a CCD camera using the CytoVision software system (Applied Imaging, Santa Clara, CA).

All BACs were obtained from the BACPAC Resource Center (Oakland, CA) and probe locations were verified by hybridization to metaphase spreads of normal peripheral lymphocytes. For detection of *TMPRSS2:ETV4* fusion, RP11-35C4 (5' to *TMPRSS2*) was used with multiple BACs located 3' to *ETV4* (distal to *ETV4* to proximal: RP11-266I24, RP11-242D8, and RP11-100E5). For detection of *ETV4* rearrangements, RP11-436J4 (5' to *ETV4*) was used with the multiple BACs 3' to *ETV4*. For each hybridization, areas of cancerous cells were identified by a pathologist and 100 cells were counted per sample. The reported cell count for *TMPRSS2:ETV4* fusions used RP11-242D8 and similar results were obtained with all 3' *ETV4* BACs. To exclude additional rearrangements in PCA5, we performed FISH with two probes 3' to *ETV4* (RP11-266I24 and RP11-242D8), *ERG* split signal probes (RP11-95I21 and RP11-476D17) and *TMPRSS2*:*ETV1* fusion probes (RP11-35C4 and RP11-124L22). BAC DNA was isolated using a QIAFilter Maxi Prep kit (Qiagen, Valencia, CA) and probes were synthesized using digoxigenin- or biotin-nick translation mixes (Roche Applied Science).

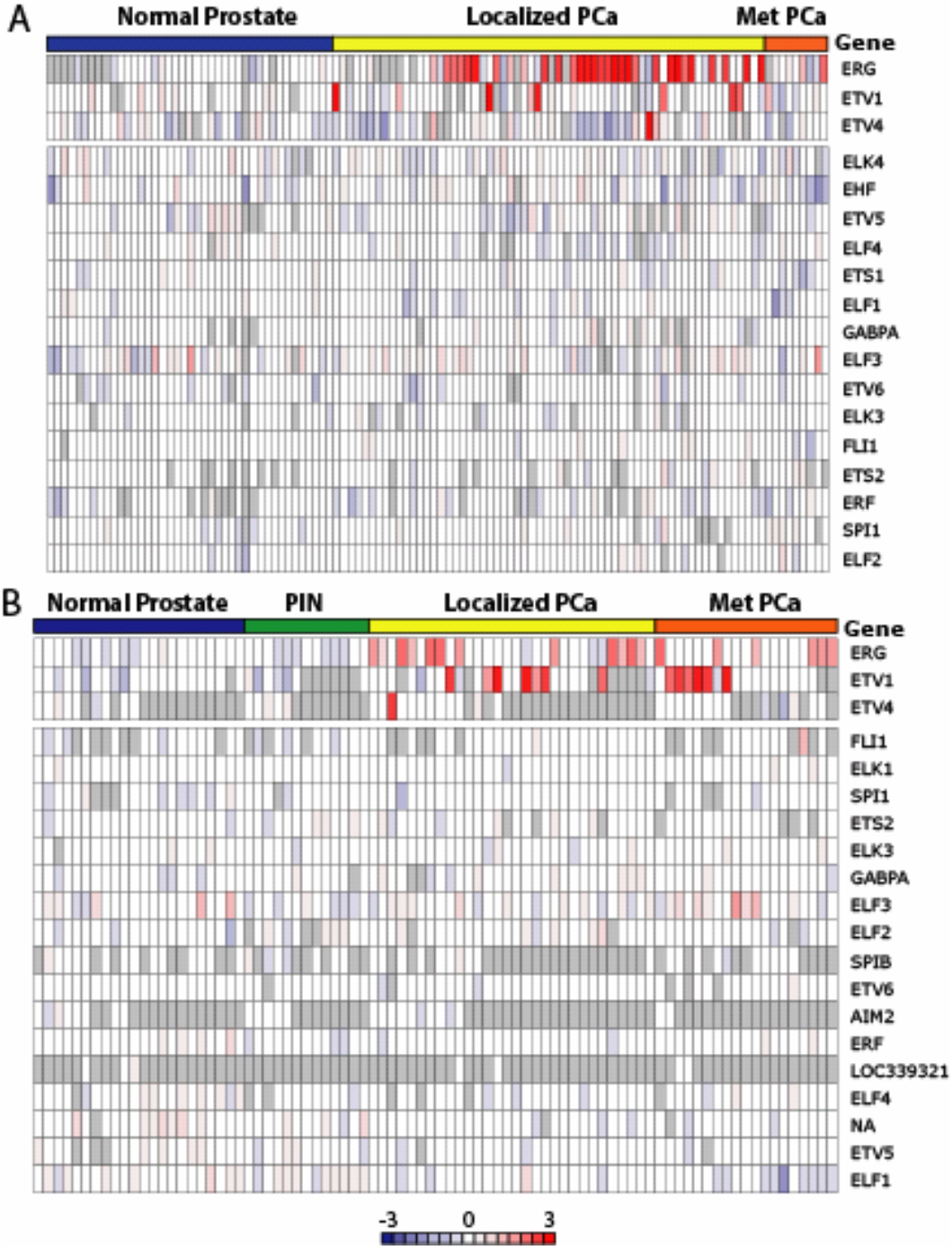

Figure 3.1. Over-expression of *ETS* family members in prostate cancer. Expression of all monitored *ETS* family members in profiled benign prostate (blue), prostatic intraepithelial neoplasia (PIN) (green), clinically localized prostate cancer (PCa, yellow) and metastatic prostate cancer (Met PCa, orange) from grossly dissected tissue (**A**) (Lapointe *et al.* (9)) or tissue isolated by laser capture microdissection (**B**) (Tomlins *et al.*) was visualized using Oncomine (www.oncomine.org). Columns represent samples of the indicated class and rows represent the indicated gene. Blue and red cells indicate relative under- or over-expression (in z-score units, median-centered per gene), respectively, as indicated by the color scale. Grey cells indicate features that did not pass filtering in the original study.

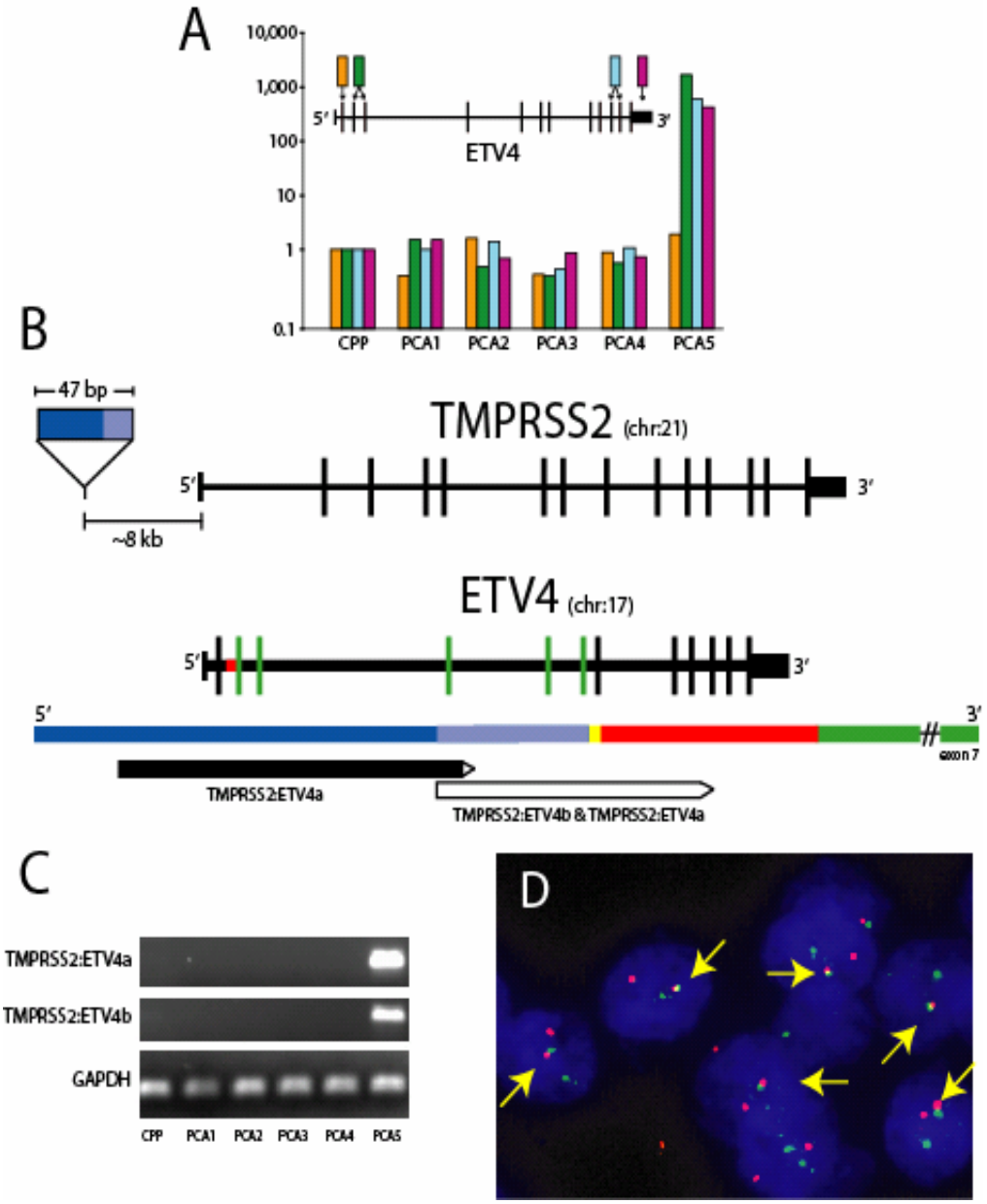

Figure 3.2. Fusion of *TMPRSS2* and *ETV4* loci in a prostate cancer case that over-expresses *ETV4*. A. QPCR was used to determine the relative expression of the indicated exons or region of *ETV4* in pooled benign prostate tissue (CPP) or prostate cancers including the case from our cohort with *ETV4* over-overexpression (PCA5). **B**. RLM-RACE reveals fusion of sequences upstream of *TMPRSS2* with *ETV4* in PCA5. Sequencing of the cloned product revealed two transcripts beginning with 48 base pairs (*TMPRSS2*:*ETV4*a, dark and light blue) or 13 base pairs (*TMPRSS2*:*ETV4*b, light blue) located approximately 8 kb upstream of *TMPRSS2*, a single base pair that did not map to this region or *ETV4* (yellow), and a contiguous region composed of the 19 terminal base pairs in the intron before exon 3 of *ETV4* (red) and the reference sequence of *ETV4* from exons 3 to 7 (green). **C.** Expression of *TMPRSS2*:*ETV4*a and *TMPRSS2*:*ETV4*b in PCA5 by QPCR. **D.** Interphase fluorescence in situ hybridization on formalin-fixed paraffin-embedded tissue confirms fusion of *TMPRSS2* and *ETV4* loci in PCA5. Probes for *TMPRSS2* (red) and *ETV4* (green) demonstrate fusion of the genomic loci (yellow arrows) in cancerous cells from PCA5.

REFERENCES

1. Mitelman F. Recurrent chromosome aberrations in cancer. Mutat Res 2000;*462:*247-53.
2. Rabbitts TH. Chromosomal translocations in human cancer. Nature 1994;*372:*143-9.
3. Rowley JD. Chromosome translocations: dangerous liaisons revisited. Nat Rev Cancer 2001;*1:*245-50.
4. Tomlins SA, Rhodes DR, Perner S, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 2005;*310:*644-8.
5. Afar DE, Vivanco I, Hubert RS, et al. Catalytic cleavage of the androgen-regulated TMPRSS2 protease results in its secretion by prostate and prostate cancer epithelia. Cancer Res 2001;*61:*1686-92.
6. Jacquinet E, Rao NV, Rao GV, Zhengming W, Albertine KH, Hoidal JR. Cloning and characterization of the cDNA and gene for human epitheliasin. Eur J Biochem 2001;*268:*2687-99.
7. Lin B, Ferguson C, White JT, et al. Prostate-localized and androgen-regulated expression of the membrane-bound serine protease TMPRSS2. Cancer Res 1999;*59:*4180-4.
8. Rhodes DR, Yu J, Shanker K, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia 2004;*6:*1-6.
9. Lapointe J, Li C, Higgins JP, et al. Gene expression profiling identifies clinically relevant subtypes of prostate cancer. Proc Natl Acad Sci U S A 2004;*101:*811-6.
10. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 1998;*95:*14863-8.
11. Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999;*286:*531-7.
12. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 2001;*98:*5116-21.
13. Arvand A, Denny CT. Biology of EWS/ETS fusions in Ewing's family tumors. Oncogene 2001;*20:*5747-54.
14. Vandesompele J, De Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 2002;*3:*RESEARCH0034.

CHAPTER 4

INTEGRATIVE MOLECULAR CONCEPTS MODELING OF PROSTATE CANCER PROGRESSION

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 report the profiling of 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, 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) exhibits an attenuated androgen signaling signature, similar to metastatic prostate cancer, which may reflect de-differentiation and explain the clinical association of grade with prognosis. Taken together, we demonstrate that analyzing gene expression signatures in the context of a compendium of molecular concepts has utility in understanding cancer biology.

Prostate cancer is the most common non-cutaneous malignancy in American men (1). Although numerous groups have profiled prostate cancer using DNA microarrays (reviewed in (2)), genetic changes and biological processes mediating important transitions in progression remain undefined. For example, due to 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 (1 being the most differentiated and 5 being the least differentiated). As prostate cancer is often

mutlifocal, the overall Gleason score is the sum of the two most prevalent patterns, and patients 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 (GSEA) 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 Concepts Map (MCM), an analytical framework for exploring the network of inter-relationships among a growing collection of molecular concepts, or biologically related gene sets. In addition to being the largest collection of gene sets for association analysis, the MCM is unique in that it computes pair-wise 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 with the MCM allowed us to systematically link our signatures to over 14,000 molecular concepts.

We used laser capture microdissection (LCM) to isolate 101 specific cell populations from 44 cases representing prostate cancer progression. Isolated total RNA was amplified using OmniPlex Whole Transcriptome Amplification (WTA), before expression profiling on 20,000 element cDNA microarrays. We have recently validated WTA for expression profiling (20), and several lines of evidence support the validity of our signatures as described below.

To identify expression signatures, we loaded our dataset into Oncomine (www.oncomine.org), 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 datasets were collected and analyzed, yielding over 14,000 molecular concepts**.** Our experimental approach is shown in **Figure 4.1**.

We employed 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. To define a comprehensive progression signature, we also sought to incorporate putative precursor lesions. As described below, through several analyses, our results demonstrate that high grade PIN and prostate cancer share remarkably similar expression signatures. However, these same analyses indicate that atrophic lesions, including PIA, share few genetic changes with prostate cancer, suggesting that PIA may only be a very early precursor 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 prostate cancer. Further profiling studies will be needed to define the role of PIA in progression. Thus, we defined our progression signature as genes whose expression correlates with increased or decreased expression from benign epithelium to PIN to prostate cancer to metastatic prostate cancer.

We identified robust over- and under-expressed progression signatures (661 and 862 features at Q <0.05, respectively) (**Figure 4.2**). Several lines of evidence support the accuracy of our progression signature. For example, the most enriched Oncomine signatures in our over- and under-expressed progression signatures are Varmbally *et al.*'s over- (P = 1.7E-68) and under-expressed (P = 2.7E-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 over-expressed in progression signature is 8q (P = 4.5E-4), which is one of the most frequent gains during progression (23). The *MYC* oncogene, located at 8q24, which ranks 24th in our over-expressed progression signature, is thought to be one of the targets of this amplification and has been shown to be amplified during progression (23, 24).

Importantly, stroma confounds the identification of truly under-expressed genes (such as tumor suppressors) in studies using grossly dissected tumors, due 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 under-expressed 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 under-expressed in previous progression signatures from studies using grossly dissected tissue (Lapointe *et al.* (11), P = 2.5E-10, Dhanasekaran *et al.* (29), P = 1.5E-05, Vanaja *et al.* (14), P = 0.005) it ranks no higher than 639th in these studies due to stromal masking. Together, these results indicate that while 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 under-expressed in epithelial cells during progression.

We also validated the differential expression of genes not previously implicated in progression using QPCR on an independent sample set. We validated the over-expression of two transcripts, *ZIC2* (which is involved in the hedgehog signaling pathway (30)) and a noncoding transcript at 11q13.1 (a region where amplification predicts progression to metastatic disease (31))**.** We also validated the under-expression of *NPAL3*, which ranks 10th in our signature and maps to 1p36, a region lost in metastatic prostate cancer (32).

To move beyond the single gene or concept approach, we analyzed our progression signatures and enriched concepts using the MCM, which allows for the identification of enrichment networks. Analyzing our over-expressed progression signature revealed an enrichment network containing proliferation related concepts (including the most enriched gene ontology (GO) Biological Process "cell cycle" (P = 1.6 E-6) and the most enriched literature concept "Differentially expressed genes in Hela cells during the cell cycle" (P = 1.7E-10)) (**Figure 4.2**).

MCM analysis of our under-expressed progression signature revealed an interconnected androgen signaling concept set (**Figure 4.2**). These concepts represent both in vitro (including the most enriched literature concept, “Upregulated genes in prostate cancer cells in response to synthetic androgen R1881” (P = 2E-14)) and in vivo measures of androgen signaling (“Downregulated genes in prostate cancer after androgen ablation therapy” (P = 1.7E-4)). The enrichment of genes normally up-regulated by androgen in the under-expressed 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.

We also attempted to identify molecular correlates of individual histological transitions in prostate cancer progression. For individual transitions, we observed far fewer differentially expressed genes ($Q < 0.05$) between PIN and prostate cancer (1.2% of measured features) than between benign and PIN (13.4%) or 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 and MCM analysis as described below, support this observation, suggesting that key processes driving progression occur from benign prostate epithelium to PIN.

We identified robust expression signatures for genes over- and under-expressed (1,213 and 1,335 features at $Q < 0.05$) in PIN vs. benign epithelium (**Figure 4.3**). MCM analysis of our over-expressed in PIN vs. benign epithelium signature identified an enrichment network of protein biosynthesis concepts (**Figure 4.3**), including the most enriched GO Biological Process ("protein biosynthesis", P = 1.3 E-7) and KEGG Pathway ("ribosome", P = 4.9 E-5). This network also contained the most enriched Transfac promoter binding site concept, the *ETS* transcription factor "Elk-1" (P = 2.1 E-9), representing genes with defined *Elk-1* binding sites in their proximal promoters. Transcriptional targets of other *ETS* family members, including *c-Ets-1* (P = 1.3E-8) and *NRF-2* (also known as *GABPA*) (P = 7.6 E -8), also show strong enrichment due to overlapping transcription factor matrices. These results suggest that a major process occurring from benign epithelium to PIN is increased protein biosynthesis, likely through *ETS* target genes. MCM analysis also supports the genetic similarity of PIN and prostate cancer, as in addition to marked overlap (P < 1E-100), the signatures shared enrichment of the protein biosynthesis and *ETS* target genes concepts. Furthermore, genes over-expressed in PIN vs. benign epithelium also displayed enrichment of concepts indicating increased androgen signaling (**Figure 4.3**), suggesting a link between androgen signaling, *ETS* transcription factors and protein biosynthesis as described below.

We identified a limited number of dysregulated transcripts in localized prostate cancer vs. high grade PIN (199 over- and 23 under-expressed 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 over-expressed in prostate cancer through fusions with the androgen-regulated gene *TMPRSS2* (33, 34) and our observation that *ETS* target genes involved in protein biosynthesis are over-expressed in PIN (and prostate cancer) compared to benign prostatic epithelium. We observed mutually exclusive over-expression 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 in contrast, 0 of the 22 benign or 13 PIN samples. Importantly, our data set contains 3 cases where we captured both PIN and prostate cancer lesions, and an ETS family member (*ERG*) was markedly over-expressed in the prostate cancer samples. The magnitude of *ERG* up-regulation in prostate cancer relative to PIN was dramatic in each case, with *ERG* being the most up-regulated feature in cases 5 and 7 and the second most in case 17. In addition, in cases where multiple prostate cancer foci were profiled, either all or no foci over-expressed *ERG*, *ETV1* or *ETV4*, suggesting that this is a clonally selected event occurring early in prostate cancer development.

It is unclear if tumors with or without *TMPRSS2:ETS* fusions have distinct expression signatures. Thus, in our study and two others (11, 35), we divided localized prostate cancers into *ETS* over-expressing (*ERG*, *ETV1* or *ETV4*) and non-*ETS* over-expressing and attempted to identify expression signatures. In each study, we identified molecular signatures differentiating *ETS* and non-*ETS* tumors (this study = 157 features, Lapointe *et al.* (11) = 524 features, Glinsky *et al.* (35) = 3,328 at $Q < 0.05$). Importantly, the three signatures were highly overlapping (**Figure 4.4).** For example, the over-expressed in *ETS* vs. non-*ETS* signatures from Glinsky *et al.* (35) and this study were the two most enriched Oncomine signatures in Lapointe *et al.*'s (11) over-expressed in *ETS* vs. non-*ETS* signature (P = 3E-74 and 1.6E-54, respectively). Although few concepts were significantly enriched across all three studies, "6q21" was the most enriched chromosome sub-arm in all three over-expressed in ETS vs. non-ETS signatures (this study, 1.20E-04; Lapointe *et al.* (11) = 1.80E-08; Glinsky *et al.* (35) = 8.30E-06) (**Figure 4.4**). This suggests a cooperating amplification at 6q21 in *TMPRSS2:ETS* tumors or loss of 6q21 in non-*TMPRSS2:ETS* tumors; intriguingly, 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 *TMPRSS2:ETS* negative

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* (*$p27^{kip1}$*) and *BCL2L11* (*BIM)* (37, 38).

Metastatic prostate cancer is generally considered incurable and treatment becomes palliative in nature. Treatment with anti-androgens usually results in regression; unfortunately, the cancer almost invariably progresses with a hormone-refractory (HR) phenotype. Thus, identifying signatures and concepts differentiating clinically localized from untreated, or hormone naïve (HN), and HR metastatic prostate cancer is essential to understanding progression. As our study set did not contain enough HN metastatic samples (n =3) for comprehensive analysis, we used two data sets in Oncomine containing benign prostate, localized and HN metastatic prostate cancer (Lapointe *et al.* (11) and Vanaja *et al.* (14)). MCM analysis of these studies revealed two distinct interaction networks—centered on protein biosynthesis and proliferation concepts—enriched in over-expressed in progression (benign to localized to HN metastatic prostate cancer) and in HN metastatic vs. localized prostate cancer signatures (**Figure 4.5**). For example, "protein biosynthesis" was the most enriched GO Biological Process in both Vanaja *et al.* (14) and Lapointe *et al.*'s (11) over-expressed in progression signatures (P = 2.29E-29 and 7.0E-13, respectively). Similarly, "Differentially expressed genes in Hela cells during the cell cycle" was the most enriched literature concept in the over-expressed in HN metastatic vs. localized prostate cancer signatures from both studies (Vanaja *et al.* (14) P = 1.3E-19 and Lapointe *et al.* (11) P = 3.0E-15) (**Figure 4.5**). Importantly, in both studies, progression signatures are more strongly linked to protein biosynthesis, while HN metastatic vs. localized prostate cancer signatures are more strongly linked to proliferation concepts. Increased protein biosynthesis defined our PIN vs. benign and localized prostate cancer vs. benign signatures, suggesting that protein biosynthesis concepts are similarly over-expressed in HN metastatic prostate cancer, while increased proliferation distinguishes HN metastatic from localized prostate cancer.

To understand the transition from HN to HR metastatic prostate cancer, we compared the HN metastatic enrichment networks just described to corresponding HR metastatic signatures from our study and Varmabally *et al.*'s (22) study. Progression to

HR metastatic prostate cancer and HR metastatic vs. localized prostate cancer signatures from both studies were highly enriched with the proliferation network (**Figure 4.5**). For example, "Differentially expressed genes in Hela cells during the cell cycle" was the most enriched literature concept in both over-expressed in progression signatures (this study, P = 1.7E-10; Varambally *et al.* (22) P = 2.4E-36) and in Varambally *et al.*'s HR metastatic vs. localized prostate cancer signature (P = 3.2E-28). However, these signatures did not show significant enrichment of protein biosynthesis concepts. Rather, our under-expressed in HR metastatic vs. localized prostate cancer signature showed strong enrichment with protein biosynthesis concepts (**Figure 4.5**). Additionally, the strongest enrichment in this signature was for decreased androgen signaling concepts, consistent with the castrated HR state. For example, "Upregulated genes in prostate cancer cells in response to synthetic androgen R1881" was the most enriched literature concept (P = 2.1E-31). These results suggest that while HR and HN metastatic prostate cancer share increased proliferation, only HR metastatic cancers show marked decrease of androgen signaling and protein biosynthesis concepts. As protein biosynthesis decreased with decreased androgen signaling, while increased protein biosynthesis and androgen signaling activity occur from benign to PIN, this supports a link between androgen signaling, ETS transcriptional targets and protein biosynthesis. This is in agreement with studies showing that androgen ablation causes reduction in nucleloar size and regression of PIN (39).

Utilizing LCM also allowed us to profile distinct Gleason patterns of cancerous epithelium. We divided our cancer samples into two classes, low (only Gleason pattern 3) and high (samples with Gleason patterns > 3). We did not identify robust signatures distinguishing low and high grade samples (2 features at Q < 0.05) (**Figure 4.6**). However, Lapointe *et. al*'s (11) over- (P = 6.4E-49) and under-expressed (P = 1.5E-36) in high Gleason grade signatures are the most enriched Oncomine signatures in our over- and under-expressed in high Gleason grade signatures, respectively, supporting the existence of more subtle multi-gene signatures. We also validated several of the most differentially expressed features using QPCR on an independent set of grossly dissected tumors (**Figure 4.6**). We further validated the under-expression of SLC22A3 in high Gleason grade tumors by immunohistochemistry (IHC) on tissue microarrays. *SLC22A3*

ranked 15th in our under-expressed in high Gleason grade signature and 47th in our under-expressed in progression signature. IHC confirmed the decrease in SLC22A3 expression during progression (benign to prostate cancer, P = 0.003; localized to metastatic prostate cancer, P = 0.009) as well as decreased expression in high compared to low Gleason grade cancer (P = 1.1E-5) (**Figure 4.6)**. A recent report defining LCM-isolated signatures from low and high Gleason grade prostate cancer also found marked under-expression of *SLC22A3* in high Gleason grade cancer (ranking 3rd by fold change) (9), further validating SLC22A3 as a marker for Gleason grade.

Although few transcripts were differentially expressed between low and high Gleason grade cancer, MCM analysis identified strong enrichment of decreased androgen signaling in high Gleason grade cancers (**Figure 4.6**). For example, "Upregulated genes in prostate cancer cells in response to synthetic androgen R1881" (which includes *SLC22A3*), was the most enriched literature concept in our and Lapointe *et al.*'s (11) under-expressed in high vs. low Gleason grade signature (P = 1E-11 and 1.8E-16).

By combining LCM-based profiling with an integrative molecular concepts analysis, we identified genes, concepts and enrichment networks correlating with the histologic progression of prostate cancer, leading to a unified molecular model (**Figures 7 & 8**). We confirmed the dysregulation 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), the organelle responsible for controlling protein biosynthesis, consistent with our concept analyses.

Our concepts-based analysis allowed us to make several insights into prostate cancer progression. MCM analysis identified 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 up-regulated during the transition from benign epithelium to PIN and down-regulated from localized to HR 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 dysregulated in PIN, possibly through subtle direct dsyregulation 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 dysregulation of this pathway and allow for the development of overt carcinoma. Additionally, as *ETS* targets are already over-expressed in PIN, *TMPRSS2:ETS* fusions may bypass feedback mechanisms, dysregulating a limited number of targets, or result in modest expression changes.

Our study confirms the central role of androgen signaling in prostate cancer. We identified increased androgen signaling from benign to PIN, and decreased androgen signaling from PIN to localized cancer, low to high Gleason grade prostate cancer and localized to HR metastatic prostate cancer. Furthermore, enrichment analysis suggests that decreased androgen signaling also occurs from clinically localized to HN metastatic prostate cancer, as "Upregulated genes (time dependent) in prostate cancer cells in response to androgen" is the most enriched literature concept in Lapointe *et al.*'s under-expressed in HN metastatic vs. localized prostate cancer signature (P = 2.9E-14). A recent report by Hendriksen *et al.* (44) utilized an experimentally derived set of androgen regulated genes to analyze the Lapointe *et al.* study (11), and the authors also observed decreased androgen signaling in high vs. low Gleason grade cancers and HN metastatic vs. localized prostate cancer. The authors proposed a model where localized prostate cancer cells become more aggressive by selectively down-regulating androgen responsive genes resulting in increased proliferation, de-differentiation or reduced apoptosis. An alternative explanation is that the relative amount of androgen signaling during progression reflects the differentiation status of prostatic epithelium, while separate lesions drive the increased proliferation seen late in progression. For example, 8q was the most enriched chromosome arm in both our and Lapointe *et al.*'s (11) over-expressed in high Gleason grade signatures, and 8q is significantly enriched across progression signatures along with proliferation related concepts (see **Figure 4.5**). 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 from localized to HR metastatic prostate cancer is consistent with a recent profiling study (45), which the authors attributed to the castrated state. Importantly, they noted that although dramatically down-regulated in HR metastatic cancer, androgen-regulated genes were still major transcripts in the cancerous cell. HR 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, due to the strong androgen promoter/enhancer elements regulating *TMPRSS2* expression (48). Our results support a model where genetic changes resulting in increased proliferation drive the transition to HN metastatic prostate cancer, while 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 identified concepts correlating with observed histological transitions in prostate cancer progression. By utilizing the MCM, we identified enrichment networks of linked concepts dysregulated during progression, such as *ETS* target genes and protein biosynthesis. As all concepts in the MCM are automatically tested for association, we were also 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 HN 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

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 (S.P.O.R.E.) Tissue Core. Tissues were also obtained from a radical prostatectomy series at the University Hospital Ulm (Ulm, Germany). 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 include: stroma from patients 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 patients with no history of prostatic disease (EPI_NOR), epithelium from nodules of BPH (EPI_BPH), epithelium from patients with prostate cancer (EPI_ADJ_PCA), atrophic epithelium (EPI_ATR) including PIA (EPI_ATR_PIA), PIN, localized prostate cancer, and hormone naïve or refractory metastatic prostate cancer (MET_HN or MET_HR, respectively). For this study, we defined benign epithelium to include EPI_NOR, EPI_BPH and EPI_ADJ_PCA. As described in the text, due to the low number of PIA and HN metastatic samples, we excluded them from all signatures except for the stromal vs. epithelial signature.

Laser Capture Microdissection (LCM) was performed from frozen tissue sections with the SL Microtest device using microCUT 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)..

Exponential RNA amplification was performed using a TransPlex Whole Transcriptome Amplification (WTA) kit (Rubicon Genomics, Ann Arbor, MI) as described (20) and complete details are provided in the MIAME checklist. Complete details of printing the 20K element spotted cDNA microarrays used, 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 auto-gridded 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 WTA amplified samples performed for both print runs utilized 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 utilization of the data.

A complete description of the methods used to identify gene signatures in the Oncomine database and gene set enrichment in the context of the Molecular Concepts Map (MCM) is available ((19, 21, 49) and www.oncomine.org). All gene expression studies catalogued in Oncomine, including this study, are normalized in the same manner. All data were log-transformed, median centered per array, and standard deviation normalized to one per array. For differential expression, a Student's t-test was used for two class differential expression analyses (e.g. high grade vs. low grade) and Pearson's correlation was used for multi-class ordinal analyses (e.g. benign, 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-value < 0.01. P-values < 1E-100 were set to 1E-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 (QPCR) 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 the average 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.

IHC was performed using a goat polyclonal antibody against SLC22A3 (sc-18516, Santa Cruz Biotechnologies, Santa Cruz, CA) 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.

The complete microarray data set is available from GEO (GSE6099) and Oncomine.

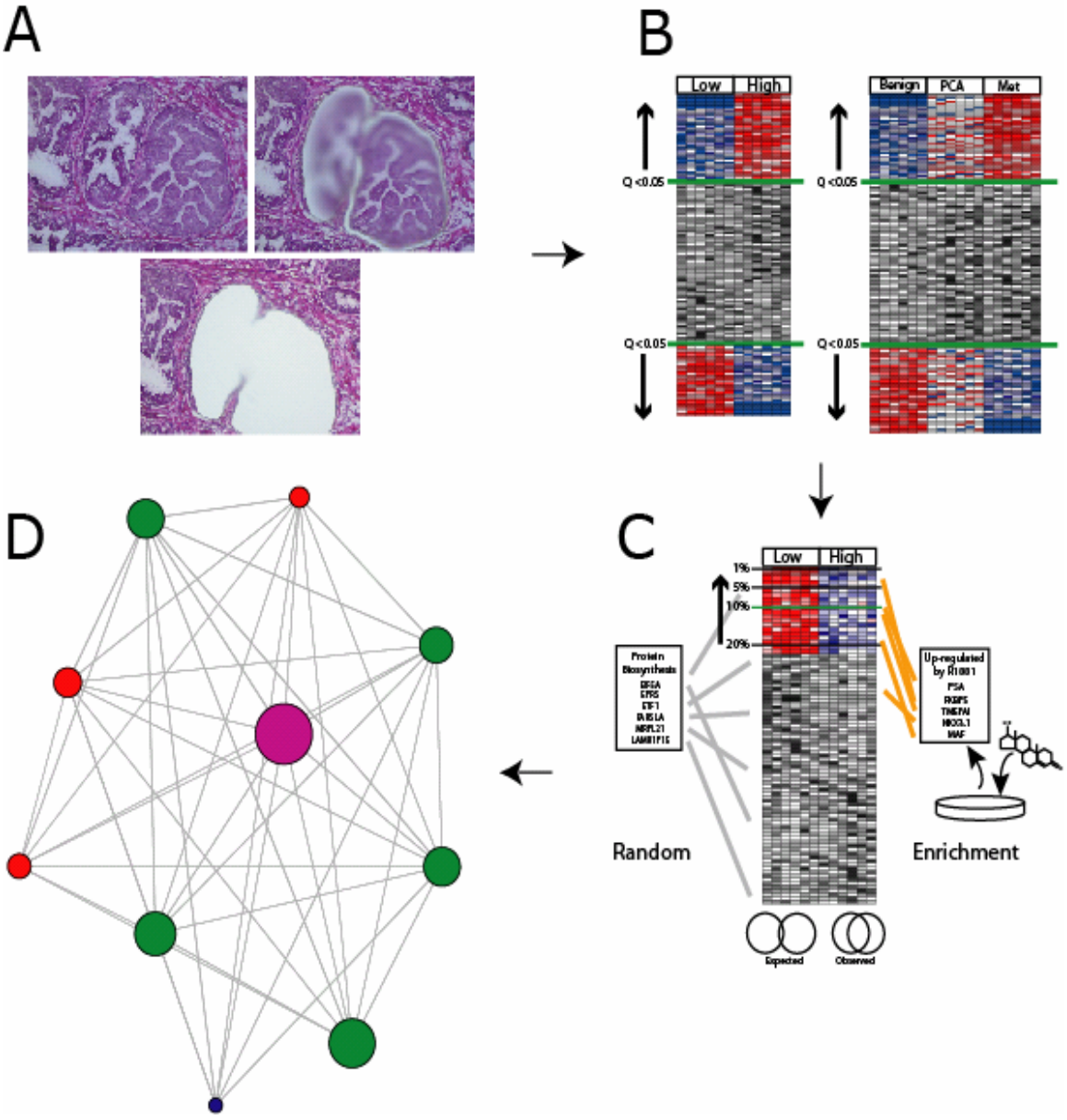

Figure 4.1. Integrative analysis of molecular concepts in prostate cancer progression. A. Samples representing different aspects of progression, such as prostatic intraepithelial neoplasia as shown, were obtained by laser capture microdissection and hybridized to cDNA microarrays. Normalized data was loaded into the Oncomine database for analysis. **B.** Using the tools available in Oncomine, gene signatures were identified for pairwise comparisons, such as low vs. 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 were automatically compared to all concepts in the Molecular Concept Map (MCM), a resource containing approximately 15,000 molecular concepts (biologically related genes), for enrichment by disproportionate overlap. **D.** Enrichment networks were then visualized using the MCM for significant links between concepts.

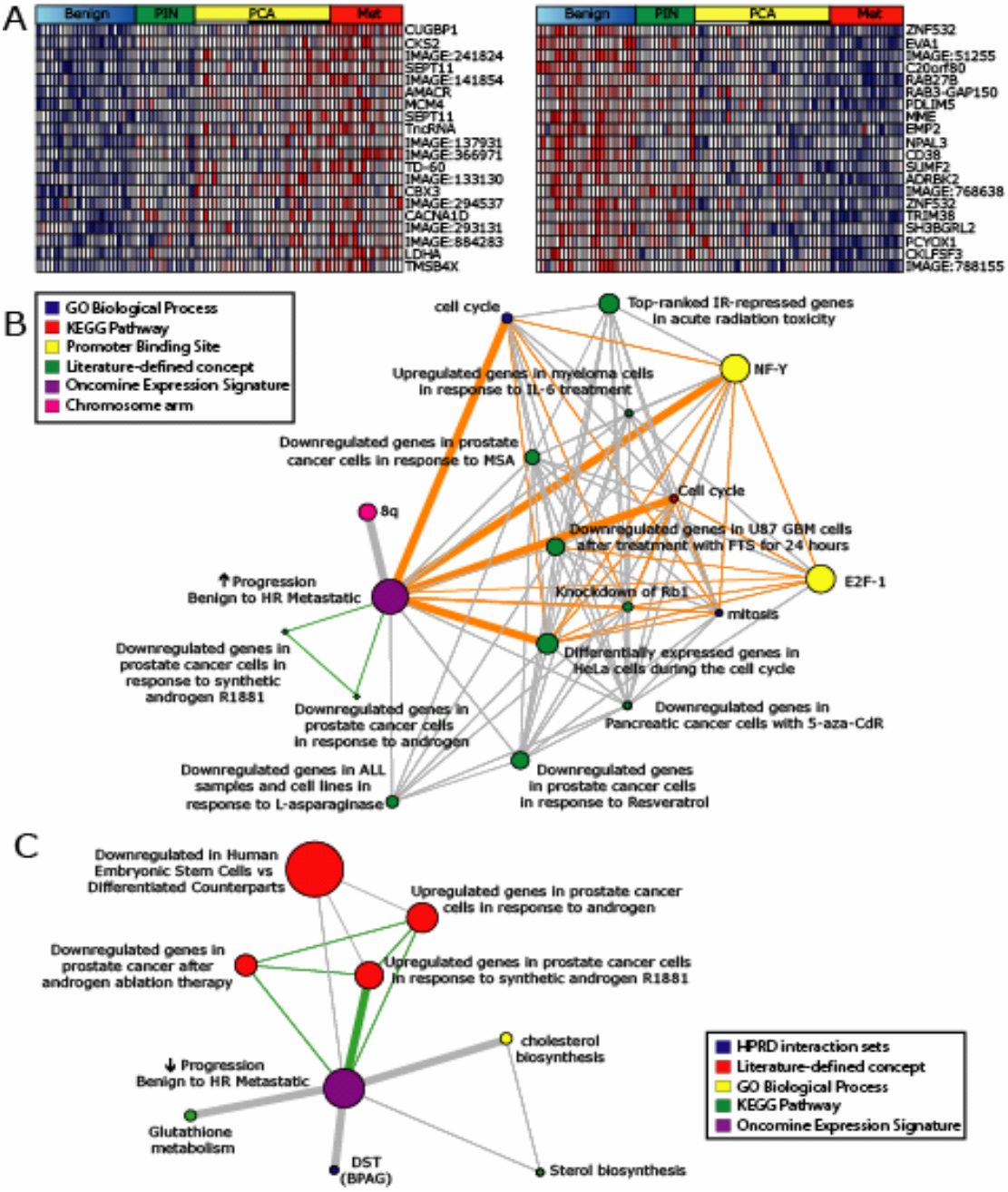

Figure 4.2. Expression signatures and molecular concept analysis of cancer progression in microdissected prostatic epithelia. A. Genes correlating with progression from benign to PIN to prostate cancer (PCA) to hormone refractory metastatic prostate cancer (Met) were identified. **B-C**. Network view of the molecular concept analysis of our over-expressed (**B**) or under-expressed (**C**) during progression signatures 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. Each edge represents a significant enrichment ($P < 5E-4$). 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.

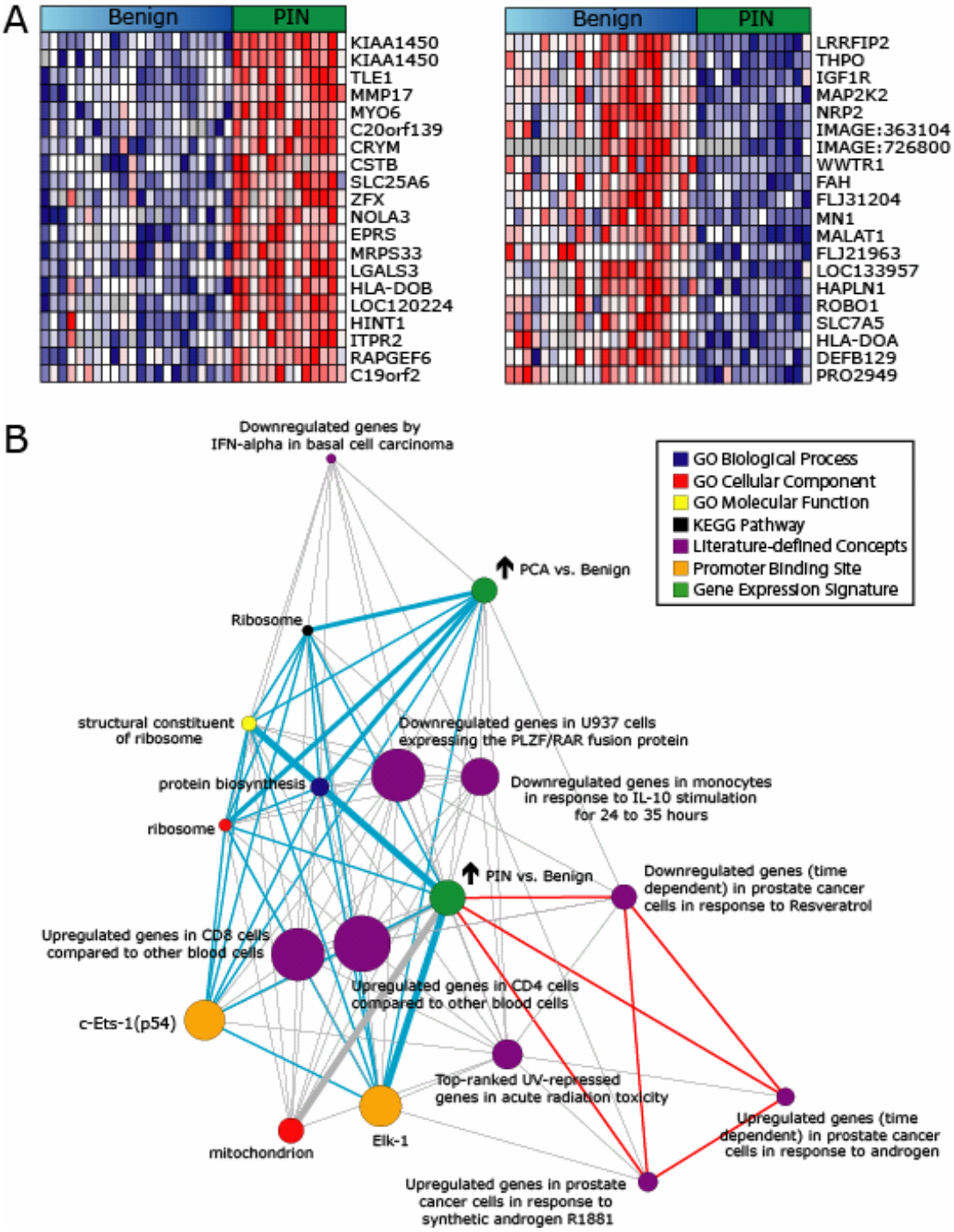

Figure 4.3. Molecular concepts analysis comparing benign epithelium to prostatic intraepithelial neoplasia (PIN). **A**. The top 20 over- (left panel) and under-expressed features (right panel) in PIN compared to benign are indicated. **B**. Network view of the molecular concepts enriched in our over-expressed in PIN vs. benign signature. Enrichments with interconnected "androgen concepts", indicating increased androgen signaling in PIN vs. benign, are indicated by red edges. Enrichments with interconnected "protein biosynthesis concepts", indicating increased protein biosynthesis in PIN vs. benign, are indicated by blue edges.

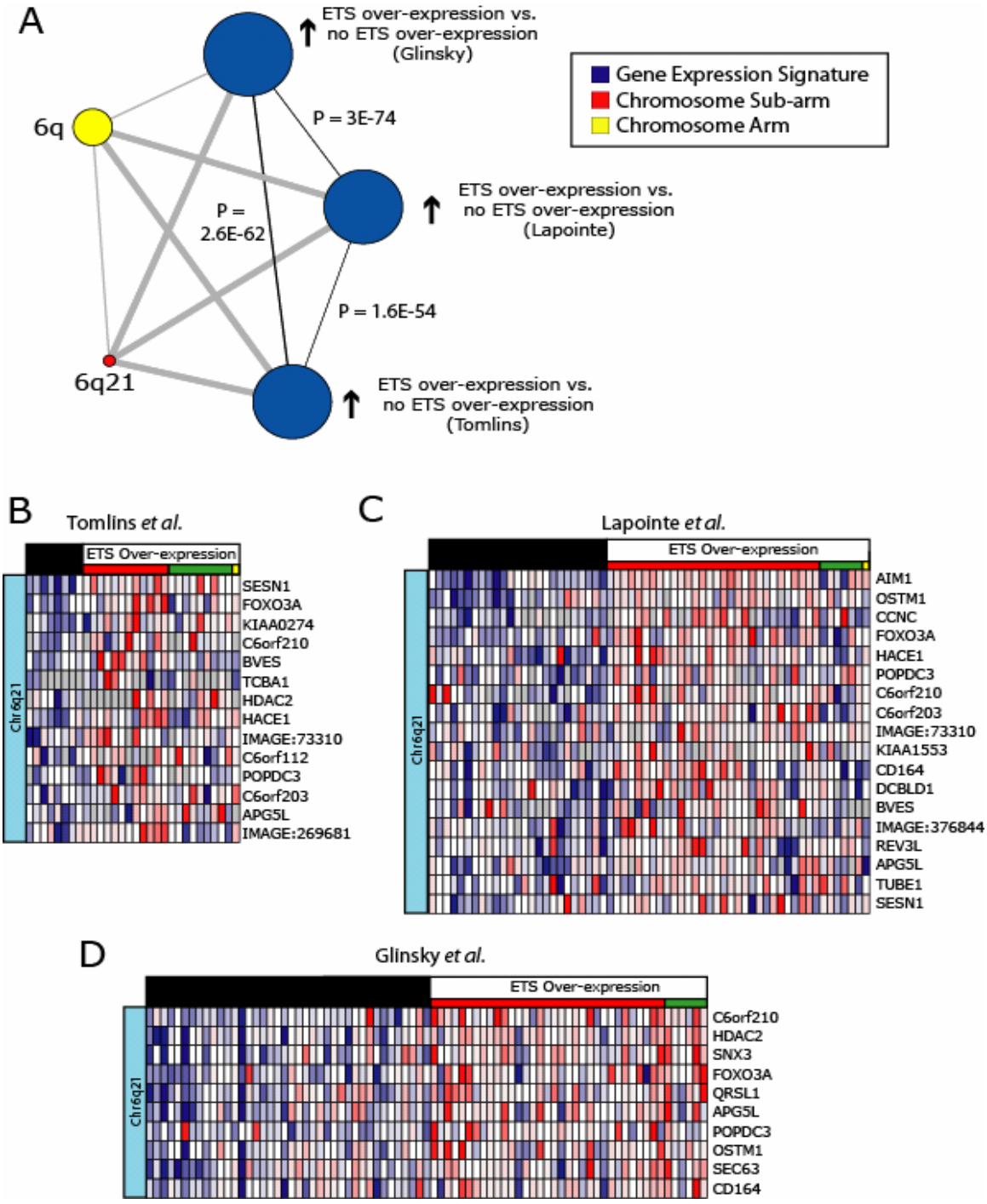

Figure 4.4. Prostate cancers with and without *ETS* family over-expression have distinct expression signatures involving chromosome 6q21. Oncomine was used to determine if prostate cancers with or without over-expression of *ETS* transcription factors (*ERG*, *ETV1* or *ETV4*) have distinct expression signatures. **A**. Network view of molecular concepts enriched in ETS vs. non ETS cancers. **B-D**. Heat map of genes from the 6q21 concept enriched in ETS-over-expressing tumors in (**B**) this study, (**C**) Lapointe *et al.*(11) and (**D**) Glinsky *et al.*(35). In each study, cancers over-expressing *ERG* (red), *ETV1* (green) or *ETV4* (yellow) were compared to cancers without *ETS* family over-expresison (black).

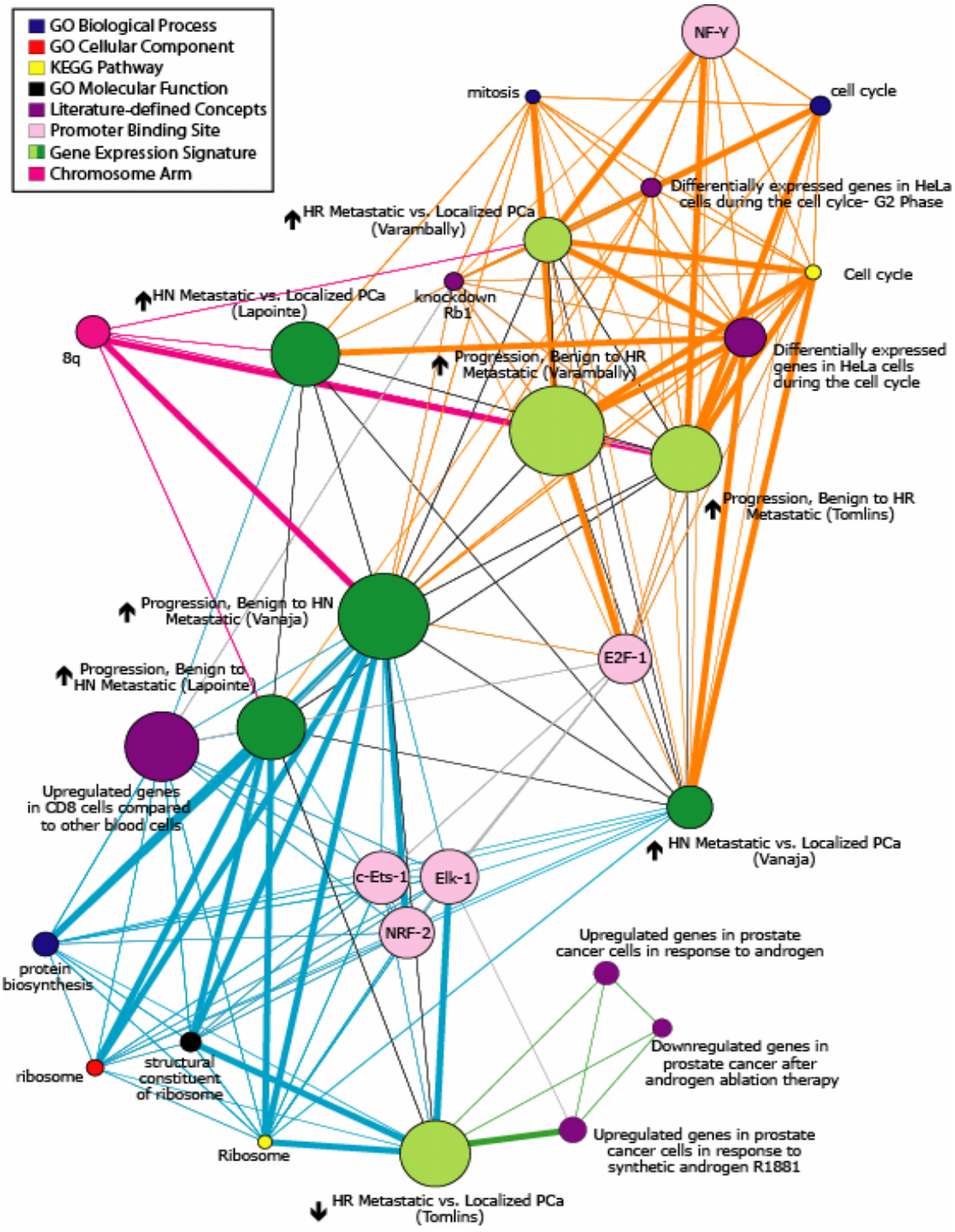

Figure 4.5. Differential expression of proliferation, protein biosynthesis and androgen signaling concepts in clinically localized, hormone naïve metastatic and hormone refractory metastatic prostate cancer. To identify concepts dysregulated between localized, hormone naïve (HN) and hormone refractory (HR) metastatic prostate cancer, we identified enrichment networks for progression signatures (as defined in the text) and metastatic vs. localized prostate cancer signatures from our study and three others indicated by the last name of the first author.

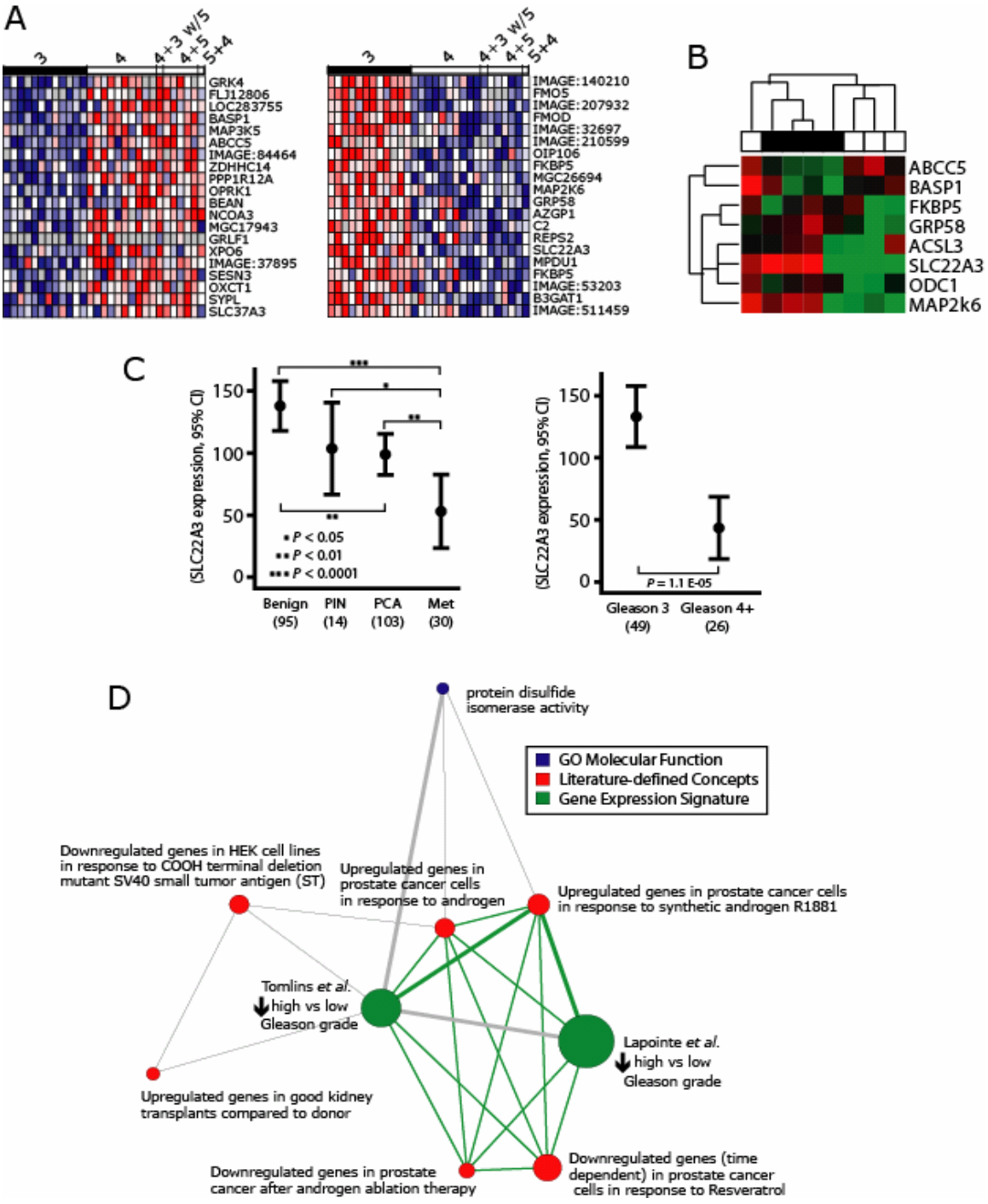

Figure 4.6. **Molecular concepts analysis of low and high Gleason grade prostate cancer. A**. Molecular Gleason signatures. **B.** Heat map of QPCR validation experiments in an independent panel of grossly dissected prostate cancer samples containing > 90% of Gleason pattern 3 (black) or 4 (white) cells. **C.** Validation of decreased SLC22A3 expression in progression and high grade prostate cancer. Expression of SLC22A3, which ranks highly in our under-expressed in high Gleason grade signature (15^{th}) and progression signature (49^{th}), was measured using immunohistochemistry on a tissue microarray. **d.** Androgen signaling is decreased in high Gleason grade prostate cancer.

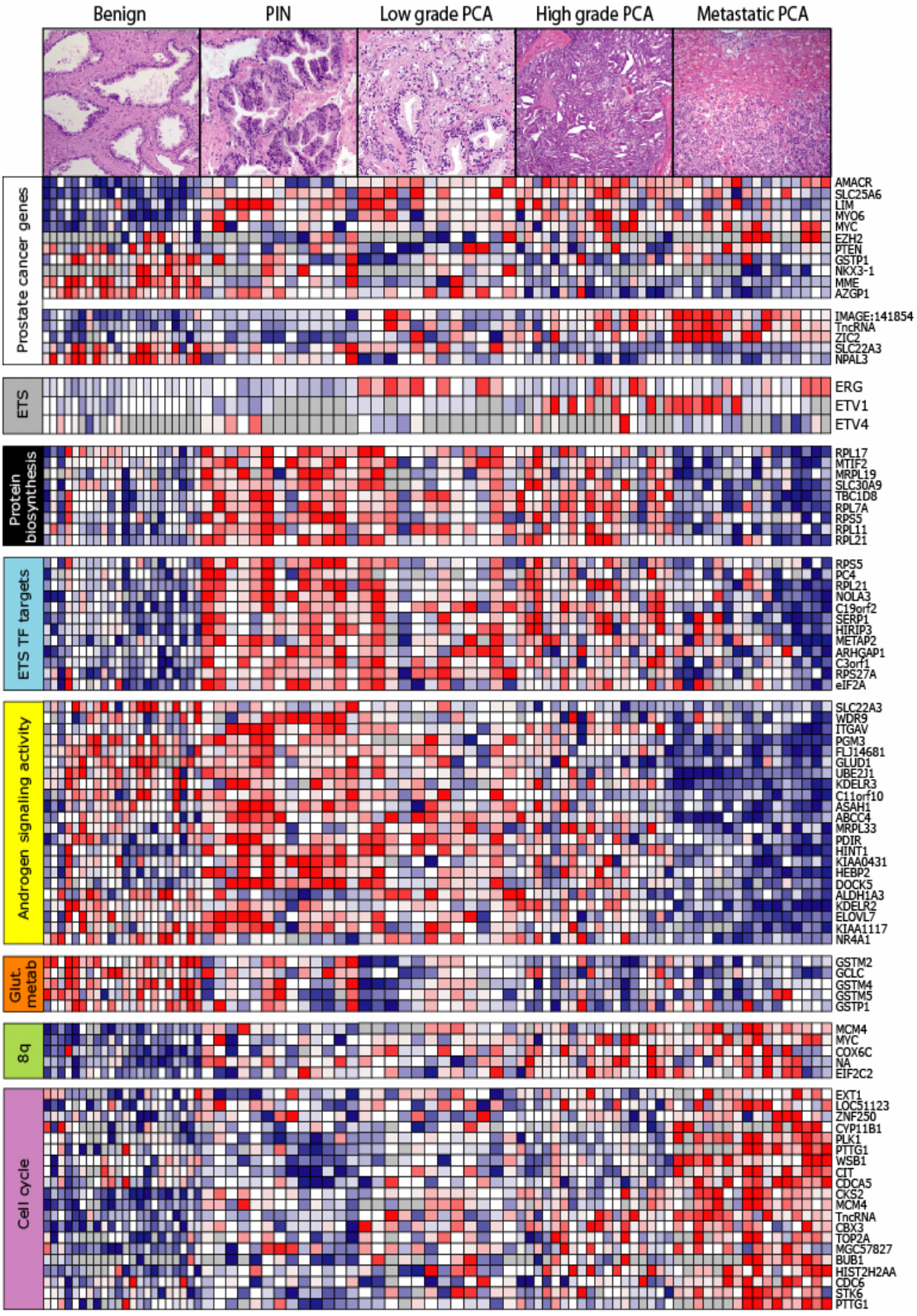

Figure 4.7. Molecular concepts heatmap of prostate cancer progression.
Concept view of expression profiles for epithelial cells from the histological transitions in prostate cancer progression.

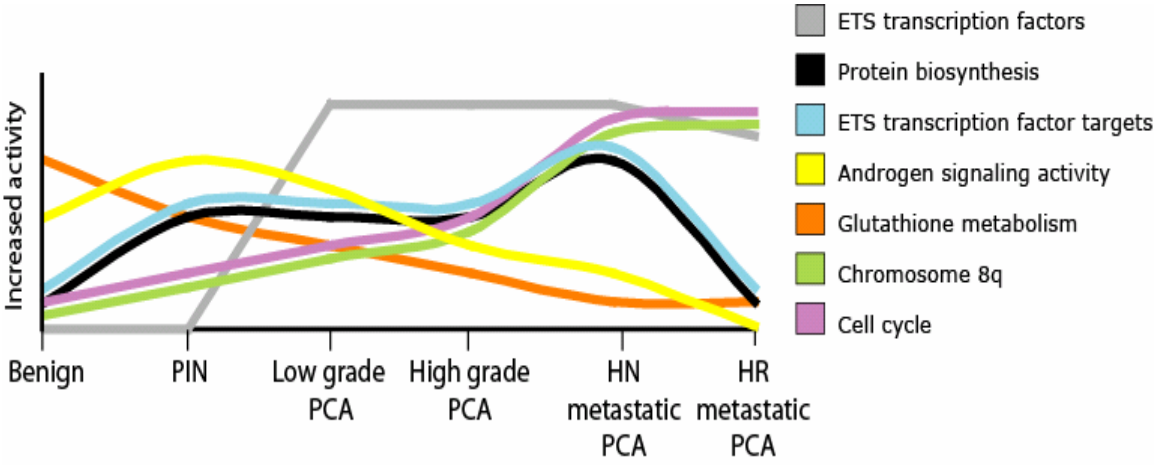

Figure 4.8. Molecular concepts model of prostate cancer progression.
The relative expression of enriched concepts identified by expression profiling of specific cell populations was used to develop a molecular concepts model of prostate cancer (PCA) progression to hormone naïve (HN) and hormone refractory (HR) metastatic PCA. Molecular concepts are indicated according to the legend.

REFERENCES

1. Jemal A, Murray T, Ward E, et al. Cancer statistics, 2005. CA Cancer J Clin 2005;*55:*10-30.
2. Nelson PS. Predicting prostate cancer behavior using transcript profiles. J Urol 2004;*172:*S28-32; discussion S3.
3. Ashida S, Nakagawa H, Katagiri T, 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 2004;*64:*5963-72.
4. Bostwick DG, Qian J. High-grade prostatic intraepithelial neoplasia. Mod Pathol 2004;*17:*360-79.
5. De Marzo AM, Marchi VL, Epstein JI, Nelson WG. Proliferative inflammatory atrophy of the prostate: implications for prostatic carcinogenesis. Am J Pathol 1999;*155:*1985-92.
6. DeMarzo AM, Nelson WG, Isaacs WB, Epstein JI. Pathological and molecular aspects of prostate cancer. Lancet 2003;*361:*955-64.
7. Gleason DF, Mellinger GT. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. J Urol 1974;*111:*58-64.
8. Humphrey PA. Gleason grading and prognostic factors in carcinoma of the prostate. Mod Pathol 2004;*17:*292-306.
9. True L, Coleman I, Hawley S, et al. A molecular correlate to the Gleason grading system for prostate adenocarcinoma. Proc Natl Acad Sci U S A 2006;*103:*10991-6.
10. Best CJ, Leiva IM, Chuaqui RF, et al. Molecular differentiation of high- and moderate-grade human prostate cancer by cDNA microarray analysis. Diagn Mol Pathol 2003;*12:*63-70.
11. Lapointe J, Li C, Higgins JP, et al. Gene expression profiling identifies clinically relevant subtypes of prostate cancer. Proc Natl Acad Sci U S A 2004;*101:*811-6.
12. Singh D, Febbo PG, Ross K, et al. Gene expression correlates of clinical prostate cancer behavior. Cancer Cell 2002;*1:*203-9.
13. Welsh JB, Sapinoso LM, Su AI, et al. Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. Cancer Res 2001;*61:*5974-8.
14. Vanaja DK, Cheville JC, Iturria SJ, Young CY. Transcriptional silencing of zinc finger protein 185 identified by expression profiling is associated with prostate cancer progression. Cancer Res 2003;*63:*3877-82.
15. Mootha VK, Lindgren CM, Eriksson KF, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet 2003;*34:*267-73.
16. Segal E, Friedman N, Koller D, Regev A. A module map showing conditional activity of expression modules in cancer. Nat Genet 2004;*36:*1090-8.
17. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;*102:*15545-50.
18. Rhodes DR, Chinnaiyan AM. Integrative analysis of the cancer transcriptome. Nat Genet 2005;*37 Suppl:*S31-7.

19. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Barrette TR, Ghosh D, Chinnaiyan AM. Mining for regulatory programs in the cancer transcriptome. Nat Genet 2005;*37:*579-83.
20. Tomlins SA, Mehra R, Rhodes DR, et al. Whole Transcriptome Amplification for Gene Expression Profiling and Development of Molecular Archives. Neoplasia 2006;*8:*153-62.
21. Rhodes DR, Yu J, Shanker K, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia 2004;*6:*1-6.
22. Varambally S, Yu J, Laxman B, et al. Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. Cancer Cell 2005;*8:*393-406.
23. Quinn DI, Henshall SM, Sutherland RL. Molecular markers of prostate cancer outcome. Eur J Cancer 2005;*41:*858-87.
24. Sato K, Qian J, Slezak JM, et al. Clinical significance of alterations of chromosome 8 in high-grade, advanced, nonmetastatic prostate carcinoma. J Natl Cancer Inst 1999;*91:*1574-80.
25. Kristiansen G, Pilarsky C, Wissmann C, et al. Expression profiling of microdissected matched prostate cancer samples reveals CD166/MEMD and CD24 as new prognostic markers for patient survival. J Pathol 2005;*205:*359-76.
26. Moore S, Knudsen B, True LD, et al. Loss of stearoyl-CoA desaturase expression is a frequent event in prostate carcinoma. Int J Cancer 2005;*114:*563-71.
27. Stuart RO, Wachsman W, Berry CC, et al. In silico dissection of cell-type-associated patterns of gene expression in prostate cancer. Proc Natl Acad Sci U S A 2004;*101:*615-20.
28. Sumitomo M, Iwase A, Zheng R, et al. Synergy in tumor suppression by direct interaction of neutral endopeptidase with PTEN. Cancer Cell 2004;*5:*67-78.
29. Dhanasekaran SM, Barrette TR, Ghosh D, et al. Delineation of prognostic biomarkers in prostate cancer. Nature 2001;*412:*822-6.
30. Koyabu Y, Nakata K, Mizugishi K, Aruga J, Mikoshiba K. Physical and functional interactions between Zic and Gli proteins. J Biol Chem 2001;*276:*6889-92.
31. Paris PL, Andaya A, Fridlyand J, et al. Whole genome scanning identifies genotypes associated with recurrence and metastasis in prostate tumors. Hum Mol Genet 2004;*13:*1303-13.
32. Nupponen NN, Kakkola L, Koivisto P, Visakorpi T. Genetic alterations in hormone-refractory recurrent prostate carcinomas. Am J Pathol 1998;*153:*141-8.
33. Tomlins SA, Rhodes DR, Perner S, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 2005;*310:*644-8.
34. Tomlins SA, Mehra R, Rhodes DR, et al. TMPRSS2:ETV4 gene fusions define a third molecular subtype of prostate cancer. Cancer Res 2006;*66:*3396-400.
35. Glinsky GV, Glinskii AB, Stephenson AJ, Hoffman RM, Gerald WL. Gene expression profiling predicts clinical outcome of prostate cancer. J Clin Invest 2004;*113:*913-23.
36. Konishi N, Shimada K, Ishida E, Nakamura M. Molecular pathology of prostate cancer. Pathol Int 2005;*55:*531-9.

37. Trotman LC, Alimonti A, Scaglioni PP, Koutcher JA, Cordon-Cardo C, Pandolfi PP. Identification of a tumour suppressor network opposing nuclear Akt function. Nature 2006;*441:*523-7.
38. Tran H, Brunet A, Griffith EC, Greenberg ME. The many forks in FOXO's road. Sci STKE 2003;*2003:*RE5.
39. Vailancourt L, Ttu B, Fradet Y, et al. Effect of neoadjuvant endocrine therapy (combined androgen blockade) on normal prostate and prostatic carcinoma. A randomized study. Am J Surg Pathol 1996;*20:*86-93.
40. Tomlins SA, Rubin MA, Chinnaiyan AM. Integrative Biology of Prostate Cancer Progression. Annual Review of Pathology: Mechanisms of Disease 2006;*1:*243-71.
41. De Marzo AM, DeWeese TL, Platz EA, et al. Pathological and molecular mechanisms of prostate carcinogenesis: implications for diagnosis, detection, prevention, and treatment. J Cell Biochem 2004;*91:*459-77.
42. Fischer AH, Bardarov S, Jr., Jiang Z. Molecular aspects of diagnostic nucleolar and nuclear envelope changes in prostate cancer. J Cell Biochem 2004;*91:*170-84.
43. Ruggero D, Pandolfi PP. Does the ribosome translate cancer? Nat Rev Cancer 2003;*3:*179-92.
44. Hendriksen PJ, Dits NF, Kokame K, et al. Evolution of the androgen receptor pathway during progression of prostate cancer. Cancer Res 2006;*66:*5012-20.
45. Stanbrough M, Bubley GJ, Ross K, et al. Increased expression of genes converting adrenal androgens to testosterone in androgen-independent prostate cancer. Cancer Res 2006;*66:*2815-25.
46. Chen CD, Welsbie DS, Tran C, et al. Molecular determinants of resistance to antiandrogen therapy. Nat Med 2004;*10:*33-9.
47. Feldman BJ, Feldman D. The development of androgen-independent prostate cancer. Nat Rev Cancer 2001;*1:*34-45.
48. Lin B, Ferguson C, White JT, et al. Prostate-localized and androgen-regulated expression of the membrane-bound serine protease TMPRSS2. Cancer Res 1999;*59:*4180-4.
49. Rhodes DR, Yu J, Shanker K, et al. Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. Proc Natl Acad Sci U S A 2004;*101:*9309-14.
50. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 2003;*100:*9440-5.

CHAPTER 5

DISTINCT CLASSES OF CHROMOSOMAL REARRANGEMENTS CREATE ONCOGENIC ETS GENE FUSIONS IN PROSTATE CANCER

Recently, we identified recurrent gene fusions involving the 5' untranslated region of the androgen-regulated gene *TMPRSS2* and the *ETS* family genes *ERG*, *ETV1* or *ETV4*, in the majority of prostate cancers (1, 2). While *TMPRSS2*:*ERG* fusions are predominant, fewer *TMPRSS2:ETV1* cases have been identified than expected based on the frequency of *ETV1* outlier-expression (3-13). Here we explored the mechanism of *ETV1* outlier-expression in prostate tumors and prostate cancer cell lines. We identified novel 5' fusion partners in prostate tumors with *ETV1* outlier-expression, including untranslated regions from a prostate-specific androgen-induced gene (*SLC45A3*) and endogenous retroviral element (*HERV-K_22q11.23*), a prostate-specific androgen-repressed gene (*C15ORF21*), and a strongly expressed housekeeping gene (*HNRPA2B1*). To study aberrant activation of *ETV1*, we identified two prostate cancer cell lines, LNCaP and MDA-PCa 2B, with *ETV1* outlier-expression. Through distinct mechanisms, the entire *ETV1* locus (7p21) is rearranged to a 1.5MB prostate-specific region at 14q13.3-14q21.1 in both LNCaP (cryptic insertion) and MDA-PCa 2B (balanced translocation). As the commonality of these rearrangements is aberrant *ETV1* over-expression, we recapitulated this event *in vitro* and *in vivo*, demonstrating that *ETV1* over-expression in benign prostate cells and the mouse prostate confers neoplastic phenotypes. Identification of distinct classes of ETS gene rearrangements demonstrates that dormant oncogenes can be activated in prostate cancer by juxtaposition to tissue-specific or ubiquitously active genomic loci. Subversion of active genomic regulatory elements may serve as a more generalized mechanism for carcinoma development. Furthermore, the identification of

androgen-repressed and insensitive 5' fusion partners may have implications for the anti-androgen treatment of advanced prostate cancer.

Recurrent chromosomal rearrangements have been causally implicated in hematologic and mesenchymal malignancies; while predicted to occur in common epithelial carcinomas, they have not been well characterized (14, 15). Using a bioinformatics strategy to nominate genes showing high ("outlier") expression in a subset of cancers, we identified fusions of the 5'-untranslated region of *TMPRSS2* (21q22) to *ERG* (21q22), *ETV1* (7p21), or *ETV4* (17q21) in cases that over-expressed the respective *ETS* family member (1, 2). *TMPRSS2* had previously been characterized as androgen-regulated (16), and its androgen-responsive regulatory elements drive *ETS* family member outlier-expression (1, 17). Thus, *TMRPSS2:ETS* fusions are functionally similar to hematological malignancy rearrangements where tissue-specific promoter or enhancer elements of one gene are juxtaposed to proto-oncogenes (15, 18).

Multiple studies have confirmed the presence of *TMPRSS2:ERG* fusions in 36-78% of prostate cancers from PSA-screened surgical cohorts. Approximately 90% of samples with *ERG* outlier-expression harbor *TMPRSS2:ERG* fusions (1, 19), confirming this as the predominant mechanism driving *ERG* over-expression. In contrast, while microarray studies show *ETV1* outlier-expression in 6-16% of prostate cancer samples, only 2 of 205 (1.0%) analyzed samples harbored *TMPRSS2:ETV1* fusions.

Here we addressed this discrepancy between *ETV1*-outlier and *TMPRSS2:ETV1* frequencies. By quantitative PCR (qPCR) across two cohorts, 26 and 3 of 54 localized prostate cancer samples showed *ERG* (48%) and *ETV1* (5.5%) outlier-expression, respectively. Additionally, two hormone-refractory metastatic prostate cancer samples, MET26 (our *TMPRSS2:ETV1* index case (1)) and MET23, showed *ETV1* outlier-expression. However, besides MET26, no samples expressed *TMPRSS2:ETV1* fusion transcripts.

To characterize the *ETV1* transcript in outlier cases, we performed 5' RNA ligase mediated rapid amplification of cDNA ends (RLM-RACE). Rather than 5' exons from *TMPRSS2*, the other four samples contained unique 5' sequences (**Figure 5.1**). In PCa_ETV1_1, exons 1-4 of *ETV1* were replaced with two exons from 22q11.23 with homology to human endogenous retrovirus family K (referred to as *HERV-K_22q11.23*).

In PCa_ETV1_2, exon 1 of *ETV1* was replaced with exon 1 from *HNRPA2B1* (7p15), while in PCa_ETV1_3, exons 1-4 of *ETV1* were replaced with a 5'-extended exon 1 of *SLC45A3* (1q32). In MET23, exons 1-5 of *ETV1* were replaced with exons 1-2 from *C15ORF21* (15q21) (**Figure 5.1**). We confirmed these fusion transcripts by qPCR and genomic fusions by fluorescent in situ hybridization (FISH).

HERV-K_22q11.23:ETV1, *SLC45A3:ETV1* and *C15ORF21:ETV1* fusions contain no predicted translated sequences from the 5' partner, and *HNRPA2B1* would only contribute two residues to a *HNRPA2B1:ETV1* fusion protein. As their regulatory elements likely drive aberrant *ETV1* expression, we characterized the tissue specificity and androgen-regulation of these 5' partners by microarray or massively parallel signature sequencing (MPSS). Similar to *TMPRSS2*, *SLC45A3* showed marked over-expression in prostate cancer (median = 2.45 standard deviations above the median per array) compared to other tumor types (median = 0.33, P = 2.4E-7) in a large DNA microarray study. *C15ORF21* showed similar over-expression in prostate cancer (P = 3.4E-6). By contrast, *HNRPA2B1* showed high expression in prostate and other tumor types (median = 2.36 vs. 2.41, P > 0.05) (**Figure 5.1**). By MPSS, *HERV-K_22q11.23* was highly expressed in normal prostate (94 transcripts per million) compared to the 31 other normal tissues (median = 9 transcripts per million) (**Figure 5.1**). By qPCR, endogenous expression of *SLC45A3* (21.6 fold, P = 6.5E-4) and *HERV-K_22q11.23* (7.8 fold, P = 2.4E-4) in the LNCaP prostate cancer cell line are strongly increased by the synthetic androgen R1881, similar to *TMPRSS2* (14.8 fold, P = 9.95E-7). Conversely, the expression of *C15ORF21* is significantly decreased (1.9 fold, P = 0.0012) and the expression of *HNRPA2B1* is not effected by R1881 stimulation (1.17 fold, P = 0.29)) (**Figure 5.1**).

We next sought to identify cell line models of *ETV1* outlier-expression. Previously, we reported that LNCaP markedly over-expressed *ETV1*, however RLM-RACE revealed expression of only the wild type transcript (1). We hypothesized that LNCaP may harbor a novel rearrangement affecting the expression of *ETV1*, and utilized a split-probe FISH strategy to look for gross rearrangements (**Figure 5.2**). On LNCaP metaphase spreads, this assay revealed two pairs of co-localizing signals at the *ETV1* locus (7p), and two split signals where the 5' signals remained on 7p, while the 3' probes

(overlapping the *ETV1* locus) were inserted into another chromosome (**Figure 5.2**).We identified this rearrangement as a cryptic insertion of a minimal region around *ETV1* into intronic sequence from the *MIPOL1* locus at 14q13.3-14q21.1 in LNCaP (**Figure 5.2**).

Screening additional prostate cancer cell lines for *ETV1* over-expression, we identified *ETV1*-outlier expression in MDA-PCa 2B. A previous analysis of MDA-PCa 2B demonstrated the presence of a balanced t(7;14)(p21;q21) (20), the locations of the *ETV1* and *MIPOL1* loci. We demonstrate that MDA-PCa 2B also harbor a rearrangement involving *ETV1*, as the *ETV1* locus translocates to the d14 (**Figure 5.2**). The 1.5 MB 14q13.3-14q21.1 region is the partner of this balanced translocation, as the telomeric 14q13.3-14q21.1 probe localizes to the d7.

The existence of mechanistically distinct rearrangements resulting in the localization of *ETV1* to 14q13.3-14q21.1 (**Figure 5.2**) in prostate cancer cell lines with *ETV1* outlier-expression suggests that elements at this region mediate the aberrant expression. By characterizing the tissue specificity and androgen regulation of the four contiguous genes at the 14q13.3-14q21.1 breakpoint (*SLC25A21*, *MIPOL1*, *FOXA1* and *TTC6*) (**Figure 5.2**) and *ETV1* in LNCaP and its androgen insensitive derivative C4-2B, we demonstrate that this region is both prostate specific and coordinately regulated by androgen.

As the 5' partners do not contribute coding sequence to the *ETV1* transcript, the common result of the different *ETV1* rearrangements in clinical samples and prostate cancer cell lines is aberrant over-expression of truncated *ETV1*. We recapitulated this event *in vitro* and *in vivo* to determine the role of aberrant *ETS* family member expression in prostate cancer. We designed adenoviral and lentiviral constructs to over-express *ETV1* as expressed in our index *TMPRSS2:ETV1* fusion positive case, MET26. In RWPE and PrEC cells, *ETV1* over-expression had no detectable effect on proliferation. *ETV1* over-expression had no effect on the percentage of RWPE cells in S phase of the cell cycle and was not sufficient for transformation.

However, *ETV1* over-expression markedly increased invasion in a basement membrane invasion assay in RWPE (3.4 fold, P = 0.0005) (**Figure 5.3**) and PrEC (6.3 fold, P = 0.0006). Additionally, *ETV1* knockdown in LNCaP using either siRNA or shRNA inhibited invasion (**Figure 5.3**), consistent with previous work (21). To

investigate the transcriptional program regulated by *ETV1*, we profiled stable RWPE-*ETV1* cells and analyzed the expression signatures using the Oncomine Concepts Map (OCM, www.oncomine.org). The OCM is a resource to look for associations between more than 20,000 biologically related gene sets by disproportionate overlap (10, 22). As shown in **Figure 5.3**, OCM analysis identified a network of molecular concepts related to cell invasion that were enriched in our *ETV1* over-expressed signature, consistent with the phenotypic effects described above.

We next determined the effects of *ETV1* over-expression in vivo. We generated transgenic mice expressing FLAG-tagged, truncated version of *ETV1* under the control of the modified probasin promoter (ARR2Pb-*ETV1*), which drives strong transgene expression exclusively in the prostate under androgen regulation (23). This transgene is functionally analogous to the androgen-induced gene fusions of *ETV1* we identified in human prostate cancer. By 12-14 weeks of age, 6 of 8 (75%) ARR2Pb-*ETV1* mice developed mouse prostatic intraepithelial neoplasia (mPIN) (**Figure 5.4**). Consistent with the definition of mPIN (24), we observed focal proliferative lesions contained within normal glands in the prostates of ARR2Pb-*ETV1* mice (**Figure 5.4**), exhibiting nuclear atypia, including stratification, hyperchromasia and macronucleoli. mPIN was observed in all three prostatic lobes (anterior, ventral and dorsolateral) of ARR2Pb-*ETV1* mice, and most commonly in the ventral lobe (7/11, 63.6%). By immunohistochemistry in ARR2Pb-*ETV1* mice, we observed strong ETV1-FLAG expression exclusively in mPIN foci and not benign glands (data not shown). While we have not observed progression to invasive carcinoma in ARR2Pb-*ETV1* mice, we have only characterized 4 mice at greater than 19 weeks of age, 3 of which (75%) also had mPIN, suggesting that additional genetic lesions are required for the development of carcinoma. Combined with our *in vitro* observations, these results demonstrate that *ETV1* induces a neoplastic phenotype in the mouse prostate and supports an oncogenic role for *ETS* gene fusions in human prostate cancer.

Including *TMPRSS2:ETS* fusions, we have now identified five classes of *ETS* rearrangements in prostate cancer. Identification of untranslated regions from the prostate-specific androgen-induced gene *TMPRSS2* provided a mechanism for aberrant *ETS* family member expression. Thus, *TMPRSS2:ETS* gene fusions (Class I) represent

the predominant class of *ETS* rearrangements in prostate cancer. Rearrangements involving fusions with untranslated regions from other prostate-specific androgen-induced 5' partner genes (Class IIa) or endogenous retroviral elements (Class IIb), are likely functionally similar to *TMPRSS2:ETS* rearrangments. Similar to 5' partners in class I and II *ETS* rearrangements, *C15ORF21* is markedly over-expressed in prostate cancer, however as *C15ORF21* is repressed by androgen, this represents a novel class of rearrangements (Class III) involving prostate-specific androgen-repressed 5' partners.

By contrast, *HNRPA2B1*, which encodes a member of the ubiquitously expressed heteronuclear ribonuclear proteins, did not show prostate-specific expression or androgen-responsiveness. Thus, *HNRPA2B1:ETV1* represents a novel class of *ETS* rearrangements (Class IV), where non-tissue specific promoter elements drive *ETS* expression. While class I-III *ETS* rearrangements are functionally analogous to *IGH-MYC* rearrangements in B cell malignancies, *HNRPA2B1:ETV1* is more analogous to inv(3)(q21q26) and t(3;3)(q21;q26) in acute myeloid leukemia, which are thought to place *EVI* under the control of enhancer elements of the constitutively expressed *RPN1* gene (25, 26).

Screening prostate cancer cell lines with *ETV1* outlier-expression, we identified rearrangements in LNCaP and MDA-PCa 2B resulting in the localization of *ETV1* to 14q13.3-14q21.1. As this aberration is recurrent in prostate cancer cell lines, we hypothesize that characterizing additional prostate cancer cases with *ETV1*-oultier expression will identify clinical specimens with similar rearrangements (class V), where the entire ETS gene is rearranged to prostate-specific regions. The identification of distinct classes of 5' fusion partners has implications for the detection of gene fusions in prostate cancer and may be important for management, particularly with regard to the effects of androgen ablation on the expression of the different ETS rearrangement classes.

Multiple classes of gene rearrangements in prostate cancer suggest a generalized role for chromosomal rearrangements in common epithelial cancers. For example, tissue specific promoter elements may be fused to oncogenes in other hormone driven cancers, such as breast cancer. Additionally, while prostate specific fusions would not provide a growth advantage and be selected for in other epithelial cancers, fusions involving strong

promoters of ubiquitously expressed genes, such as *HNRPA2B1*, could result in the aberrant expression of oncogenes across tumor types. In summary, this study supports a role for chromosomal rearrangements in common epithelial tumor development through a variety of mechanisms, similar to hematological malignancies.

Methods

Quantitative PCR, RLM-RACE for ETV1 fusions, androgen stimulation of LNCaP cells and interphase FISH were performed essentially as described (1, 2). Tissue specific expression of 5' fusion partners was determined using the International Genomics Consortium's expO dataset accessed in the Oncomine database (27) and the Lynx Therapeutics MPSS dataset (GSE1747). Expression profiling was performed using Agilent Whole Human Genome Oligo Microarrays. Adenoviruses and lentiviruses expressing *ETV1* were generated by the University of Michigan Vector Core. Transgenic ARR2Pb-*ETV1* mice were generated by the University of Michigan Transgenic Animal Model Core.

The primary microarray data have been deposited in NCBIs Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) and are accessible through GEO Series accession numbers GSE7701 and GSE7702. Sequences of the ETV1 fusion transcript junctions identified by RACE have been deposited in GenBank with accessions EF632109-EF632112.

Prostate tissues were from the radical prostatectomy series at the University of Michigan and from the Rapid Autopsy Program (28), which are both part of University of Michigan Prostate Cancer Specialized Program of Research Excellence (S.P.O.R.E.) Tissue Core. All samples were collected with informed consent of the patients and prior institutional review board approval.

The benign immortalized prostate cell line RWPE and the prostate cancer cell lines LNCaP, Du145 NCI-H660 and PC3 were obtained from the ATCC. Primary benign prostatic epithelial cells (PrEC) were obtained from Cambrex Bio Science (Walkersville, MD). The prostate cancer cell lines C4-2B, LAPC4 and MDA-PCa 2B were provided by Evan Keller (University of Michigan). The prostate cancer cell line 22-RV1 was provided

by Jill Macoska (University of Michigan). VCaP was derived from a vertebral metastasis from a patient with hormone-refractory metastatic prostate cancer (29).

For androgen stimulation experiments, LNCaP cells were grown in charcoal-stripped serum containing media for 24 hours before treatment for 24 hours with 1% ethanol or 1 nM of methyltrienolone (R1881, NEN Life Science Products, Boston, MA) dissolved in ethanol. For all samples, total RNA was isolated with Trizol (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions.

Quantitative PCR (QPCR) was performed using Power SYBR Green Mastermix (Applied Biosystems, Foster City, CA) on an Applied Biosystems 7300 Real Time PCR system as described (1, 2). All oligonucleotide primers were synthesized by Integrated DNA Technologies (Coralville, IA). *HMBS* and *GAPDH* (30), and *PSA* (31) primers were as described. Androgen stimulation reactions were performed in quadruplicate, siRNA knockdown reactions were performed in triplicate and all other reactions were performed in duplicate.

RLM-RACE was performed using the GeneRacer RLM-RACE kit (Invitrogen), according to the manufacturer's instructions as described (1, 2). To obtain the 5' end of *ETV1*, first-strand cDNA was amplified with Platinum Taq High Fidelity (Invitrogen) using the GeneRacer 5' primer and ETV1_exon4-5-r. For amplification from MET23, ETV1_exon7-r was used with the GeneRacer 5' primer. Products were cloned and sequenced bidirectionally as described (1, 2). RLM-RACEd cDNA was not used for other assays.

Interphase FISH on formalin-fixed paraffin-embedded (FFPE) tissue sections was performed as described (2). A minimum of 50 nuclei per assay were evaluated. For metaphase FISH, spreads of LNCaP and MDA-PCa 2B were prepared using standard cytogenetic techniques. Slides were pre-treated in 2x SSC for 2 min, 70% ethanol for 2 min and 100% ethanol for 2 min, and air dried. Slides and probes were co-denatured at 75 degrees for 2 min, and hybridized overnight at 37 C. Post-hybridization was in 0.5x SSC at 42 C for 5 min, followed by 3 washes in PBST. Fluorescent detection was performed using anti-digoxigenin conjugated to fluorescein (Roche Applied Science, Indianapolis, IN) and streptavidin conjugated to Alexa Fluor 594 (Invitrogen). Slides were counterstained and mounted in ProLong Gold Antifade Reagent with DAPI (Invitrogen).

Slides were examined using a Zeiss Axio Imager Z1 fluorescence microscope (Zeiss, Thornwood, NY) and imaged with a CCD camera using ISIS software (Metasystems, Altlussheim, Germany). At least 5 metaphases were assessed, and reported aberrations were observed in all interpretable spreads. BACs were obtained from the BACPAC Resource Center (Oakland, CA), and probes were prepared as described (2). Pre-labeled chromosome 7 centromere and 7p telomeric probes were obtained from Vysis (Des Plaines, IL). The integrity and correct localization of all probes were verified by hybridization to metaphase spreads of normal peripheral lymphocytes.

To determine the tissue specific expression of 5' fusion partners and genes at 14q13-q21, we interrogated the International Genomics Consortium's expO dataset (https://expo.intgen.org/expo/public/downloaddata.jsp), consisting of expression profiles from 630 tumors of 29 distinct types, using the Oncomine database (www.oncomine.org) (27). To interrogate the expression of *HERV-K_22q11.23*, which is not monitored by commercial array platforms, we queried the Lynx Therapeutics normal tissue massively parallel signature sequencing (MPSS) dataset (GSE1747) with the MPSS tag "GATCTTTGTGACCTACT", which unambiguously identifies HERV-K_22q11.23, as described (32).

Expression profiling of LNCaP, C4-2B, RWPE-*ETV1* and RWPE-*GUS* cells were performed using the Agilent Whole Human Genome Oligo Microarray (Santa Clara, CA). Total RNA isolated using Trizol was purified using the Qiagen RNAeasy Micro kit (Valencia, CA). One ug of total RNA was converted to cRNA and labeled according to the manufacturer's protocol (Agilent). Hybridizations were performed for 16 hrs at 65°C, and arrays were scanned on an Agilent DNA microarray scanner. Images were analyzed and data extracted using Agilent Feature Extraction Software 9.1.3.1, with linear and lowess normalization performed for each array. For the LNCaP and C4-2B hybridizations, the reference for each cell line was pooled benign prostate total RNA (Clontech, Mountain View, CA). A dye flip for each cell line was also performed. Features were ranked by average expression (Log Ratio) in the two LNCaP arrays divided by the average expression in the two C4-2B arrays after correction for the dye flip. For RWPE cells, four hybridizations were performed (duplicate RWPE-*ETV1*/RWPE-*GUS* and RWPE-*GUS*/RWPE-*ETV1* hybridizations). Over and under-

expressed signatures were generated by filtering to include only features with significant differential expression (PValueLogRatio < 0.01) in all four hybridizations and 2 fold average over- or under-expression (Log Ratio) after correction for the dye flip. Over-and under-expressed in RWPE-*ETV1*/RWPE-*GUS* signatures were loaded into the Molecular Concepts Map (MCM) (22), resulting in concepts containing 527 and 558 unique genes, respectively. Each signature was tested against all contained concepts in the MCM for association using Fisher's exact test as described (10, 22).

Genomic DNA (10 μg) from LNCaP, VCaP, pooled normal human male DNA (Promega, Madison, WI) and normal placental DNA (Promega) was digested with EcoRI or PstI (New England Biologicals, Ipswich, MA) overnight. Fragments were resolved on a 0.8% agarose gel at 40 V overnight, transferred to Hybond NX nylon membrane, prehybridized, hybridized with probe and washed according to standard protocols. A series of 22 probes spanning the region of chr 7 implicated by FISH (between RP11-313C20 and RP11-703A4) were generated by PCR amplification with Platinum Taq High Fidelity on pooled normal human male genomic DNA. Twenty-five ng of each probe was labeled with dCTP-P^{32} and used for hybridization.

To identify the *ETV1* breakpoint in LNCaP cells, we utilized an inverse PCR strategy based on the rearrangement identified by Southern blotting as described previously (33). Primers A1, A2, A3, which are reverse complemented from the wildtype sequence and are divergent to primers B1, B2, B3, were used for inverse PCR on PstI digested and religated (in order to promoter intramolecular ligation) LNCaP genomic DNA template. Nested PCRs were performed in the following order of primer combinations: A1-B1, A2-B2 and A3-B3. Expand 20 kbplus PCR system (Roche) was used for amplifying the fusion product according to the manufactures suggestions. The enriched 3Kb band observed in nested PCRs was cloned into pCR8/GW/TOPO (Invitrogen), miniprep DNA was screened for inserts and positive clones were sequenced (University of Michigan DNA Sequencing Core, Ann Arbor, MI). The *ETV1* insertion was confirmed by PCR with Platinum Taq High Fidelity using primers from chromosomes 7 and 14.

cDNA of *ETV1*, as present in the *TMPRSS2:ETV1* fusion to the reported stop codon of *ETV1* (269-1521, NM_004956.3), was amplified by RT-PCR from MET26 (1)

and TOPO cloned into the Gateway entry vector pCR8/GW/TOPO (Invitrogen), yielding pCR8-*ETV1*. To generate adenoviral and lentiviral constructs, pCR8-*ETV1* and a control entry clone (pENTR-*GUS*) were recombined with pAD/CMV/V5 (Invitrogen) and pLenti6/CMV/V5 (Invitrogen), respectively, using LR Clonase II (Invitrogen). Control pAD/CMV/*LACZ* clones were obtained from Invitrogen. Adenoviruses and Lentiviruses were generated by the University of Michigan Vector Core. The benign immortalized prostate cell line RWPE was infected with lentiviruses expressing *ETV1* or *GUS*, and stable clones were generated by selection with blasticidin (Inivtrogen). Benign PrEC were infected with adenoviruses expressing *ETV1* or *LACZ*, as stable lines could not be generated in primary PrEC cells. Cell counts were estimated by trypsinizing cells and analysis by Coulter counter at the indicated time points in triplicate. For invasion assays, PREC-*ETV1* and -*LACZ* (48 hours after infection) or stable RPWE-*ETV1* and -*GUS* cells were used. Representative results from three separate experiments are shown.

For siRNA knockdown of *ETV1* in LNCaP cells, the individual siRNAs composing the Dharmacon SMARTpool against *ETV1* (MU-003801-01, Chicago, IL) were tested for *ETV1* knockdown by qPCR, and the most effective single siRNA (D-003801-05) was used for further experiments. siCONTROL Non-Targeting siRNA #1 (D-001210-01) or siRNA against *ETV1* was transfected into LNCaP cells using Oligofectamine (Invitrogen). After 24 hours we carried out a second identical transfection and cells were harvested 24 hours later for RNA isolation and invasion assays as described below. For shRNA knockdown of *ETV1* in LNCaP cells, the shRNAmir construct against *ETV1* from the pMS2 retroviral vector (V2HS_61929, Open Biosystems, Huntsville, AL) was cloned into an empty pGIPZ lentivral vector (RHS4349, Open Biosystems) according to the manufacturer's protocol. pGIPZ lentiviruses with shRNAmirs against *ETV1* or a non silencing control (RHS4346) were generated by the University of Michigan Vector Core. LNCaP cells were infected with lentiviruses, and 48 hours later cells were used for invasion assays as described below. Representative results from 6 independent experiments are reported.

Equal numbers of the indicated cells were seeded onto the basement membrane matrix (EC matrix, Chemicon, Temecula, CA) present in the insert of a 24 well culture plate, with fetal bovine serum added to the lower chamber as a chemoattractant. After 48

hours, non-invading cells and EC matrix were removed by a cotton swab. Invaded cells were stained with crystal violet and photographed. The inserts were treated with 10% acetic acid and absorbance was measured at 560nm.

RWPE-*ETV1* and RWPE-*GUS* cells were assessed by FACS for cell cycle characterization. Cells were washed with 2x PBS and approximately 2 x 10^6 cells were resuspended in PBS before fixation in 70% ethanol. Pelleted cells were washed and treated with RNase (100ug/ml final concentration) and propidium iodide (10ug/ml final concentration) at 37 C for 30 min. Stained cells were analyzed on a LSR II flow cytometer (BD Biosciences, San Jose, CA) running FACSDivia, and cell cycle phases were calculated using ModFit LT (Verity Software House, Topsham, ME).

A 0.6% (wt/vol) bottom layer of low melting point agarose in normal medium was prepared in six-well culture plates. On top, a layer of 0.3% agarose containing 1 x 10^4 RWPE-*GUS*, RWPE-*ETV1* or *DU145* (positive control) cells was placed. After 12 days, foci were stained with crystal violet and counted.

For in vivo over-expression of *ETV1*, a C terminal 3XFLAG-epitope tagged construct was generated by PCR using pCR8-*ETV1* as the template with the reverse primer encoding a triple FLAG tag before the stop codon. The product was TOPO cloned into pCR8. To generate a prostate specific *ETV1* transgenic construct, 3xFLAG-*ETV1* was inserted into pBSII (Stratagene, La Jolla, CA) downstream of a modified small composite probasin promoter (ARR2PB) and upstream of a bovine growth hormone polyA site (PA-BGH). The ARR2PB sequence contains the original probasin sequence PB (-426/+28) plus two additional androgen response elements (23). The construct was sequenced and tested for promoter inducibility by androgen in LNCaP cells upon transient transfection before microinjection into FVB mouse eggs. The ARR2PB-*ETV1* plasmid was linearized with PvuI/KpnI//SacII and microinjected into fertilized FVB mouse eggs and surgically transplanted into a pseudo-pregnant female by the University of Michigan Transgenic Animal Model Core. Transgenic founders were screened by PCR using genomic DNA isolated from tail snips. Multiple ARR2Pb-*ETV1* transgenic founders were obtained and crossed with FVB mice, and transgene-positive male mice offspring were sacrificed at various time points.

Prostates from transgenic mice were dissected using a Nikon dissection scope, fixed in 10% buffered formalin and embedded in paraffin. Five um sections were stained with hematoxylin and eosin, and evaluated by three pathologists (R.M., M.A.R. and R.B.S.) according to the criteria provided in The Consensus Report from the Bar Harbor Meeting of the Mouse Models of Human Cancer Consortium Prostate Pathology Committee (24).

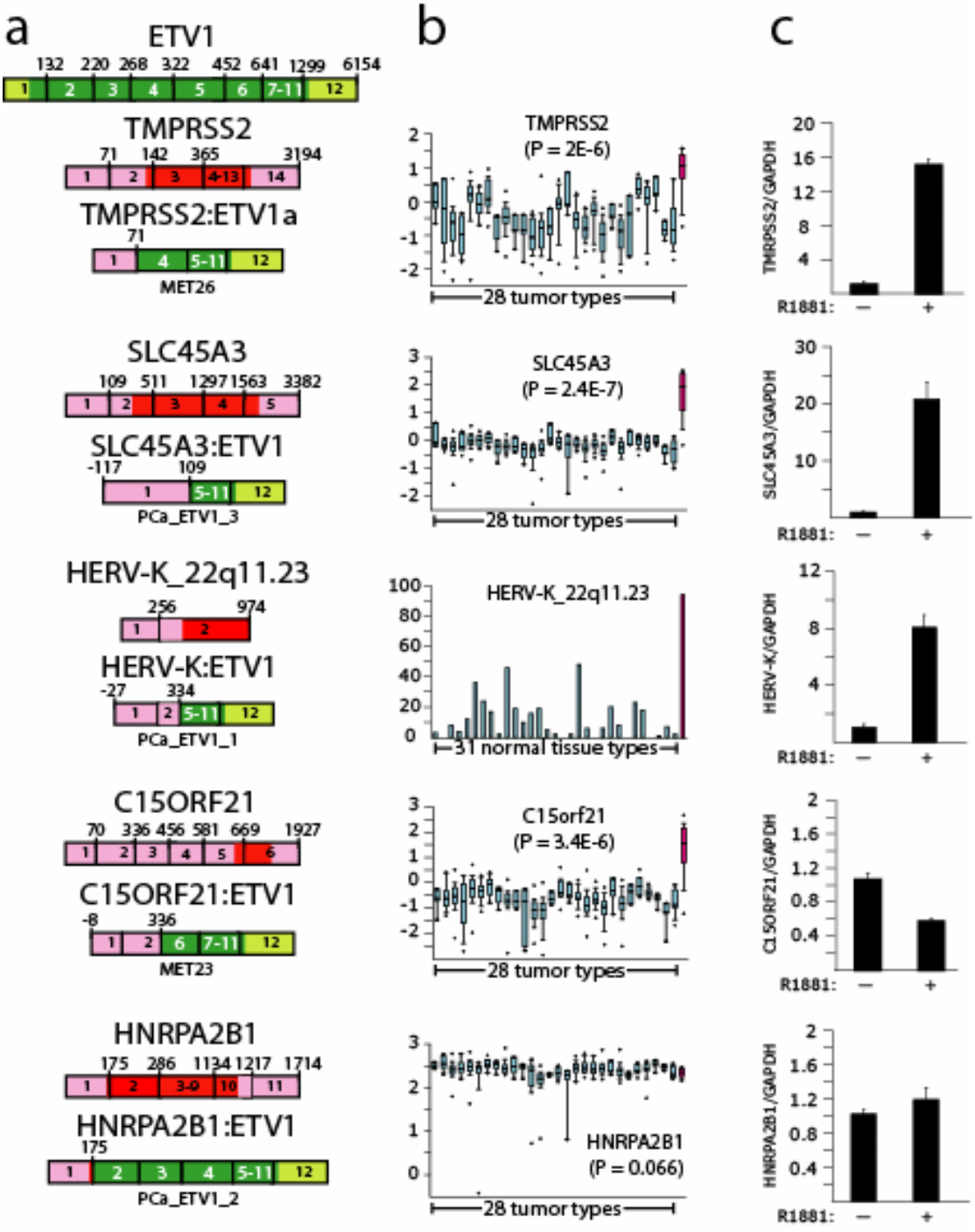

Figure 5.1. Identification of prostate-specific or ubiquitously active regulatory elements fused to *ETV1*. **a**. Structure of novel 5' partners fused to *ETV1* in outlier cases. Numbers above the exons (boxes) indicate the last base of each exon. Untranslated regions are in lighter shades. **b**. Tissue specificity of 5' fusion partners was determined in normal tissues or cancers (blue) and normal prostate or prostate cancer (magenta) (Methods). **c**. Assessment of androgen regulation of 5' fusion partners. Endogenous expression of 5' fusion partners was assessed by qPCR in LNCaP cells with (+) or without (-) stimulation by the synthetic androgen R1881, (mean (n = 4) + S.E.).

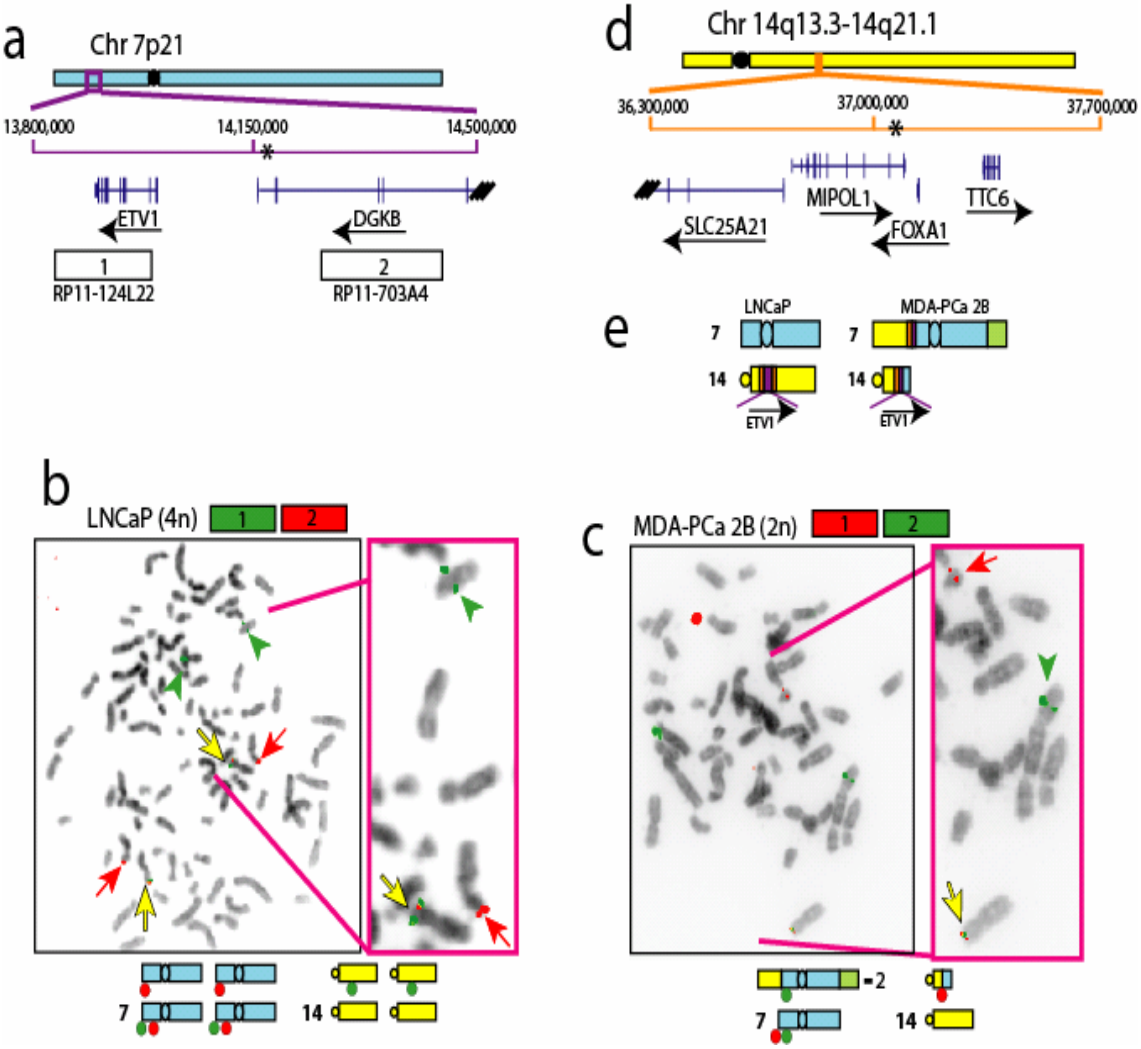

Figure 5.2. *ETV1* is rearranged to 14q13.3-14q21.1 in LNCaP and MDA-PCa 2B. a. Schematic of the *ETV1* locus (purple) on chromosome 7 (blue) and BACs (rectangles) used for FISH (adapted from the UCSC Genome Browser. Genes are shown with the direction of transcription indicated by the arrowhead. **b-c.** FISH using the indicated BACs on **b**) LNCaP (tetraploid) and **c**) MDA-PCa 2B (diploid) metaphase spreads. Co-localized signals are indicated by yellow arrows and red and green split signals are indicated by red arrows and green arrowheads. Schematics of probe localization and chromosome structures as determined by spectral karyotyping(20, 34) are indicated. **d**. Schematic of 14q13.3-14q21.1 (orange) on chromosome 14 (yellow). The LNCaP breakpoints are indicated by asterisks. **e**. Structure of *ETV1* and 14q13.3-14q21.1 in LNCaP and MDA-PCa 2B.

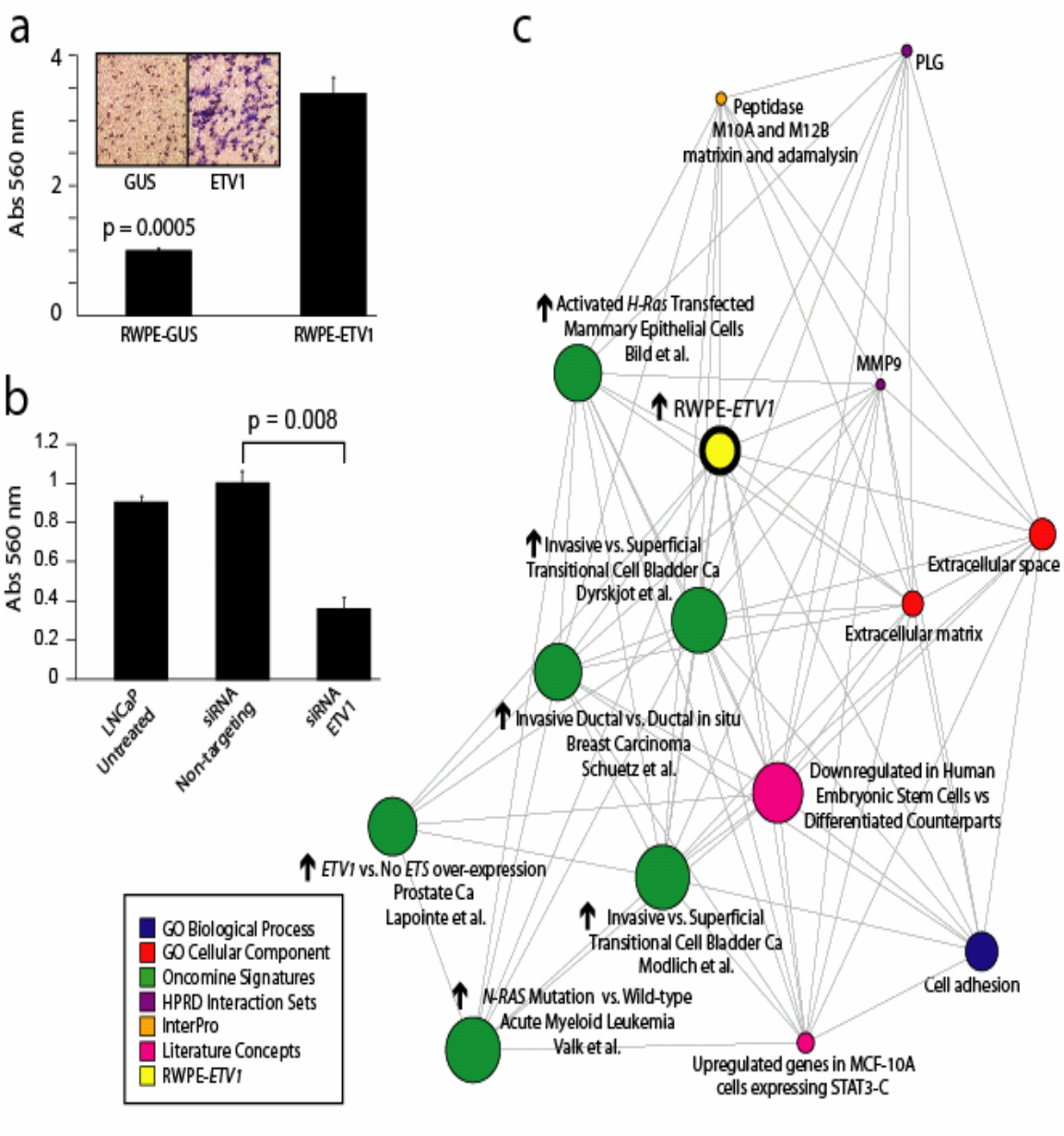

Figure 5.3. ***ETV1*** **over-expression in prostate cells confers invasiveness. a**. We infected the benign prostate cell line RWPE with lentiviruses expressing *ETV1* or control (*GUS*). Stable clones were assayed for invasion through a modified basement membrane, mean (n = 3) + S.E.. **b**. LNCaP cells were treated with transfection reagent alone (Untreated), or transfected with non-targeting or *ETV1* siRNA and assessed for invasion as in **a**; mean (n = 3) + S.E.. **c**. Oncomine concepts map of genes over-expressed in RWPE-*ETV1* compared to RWPE-*GUS* cells (yellow node).

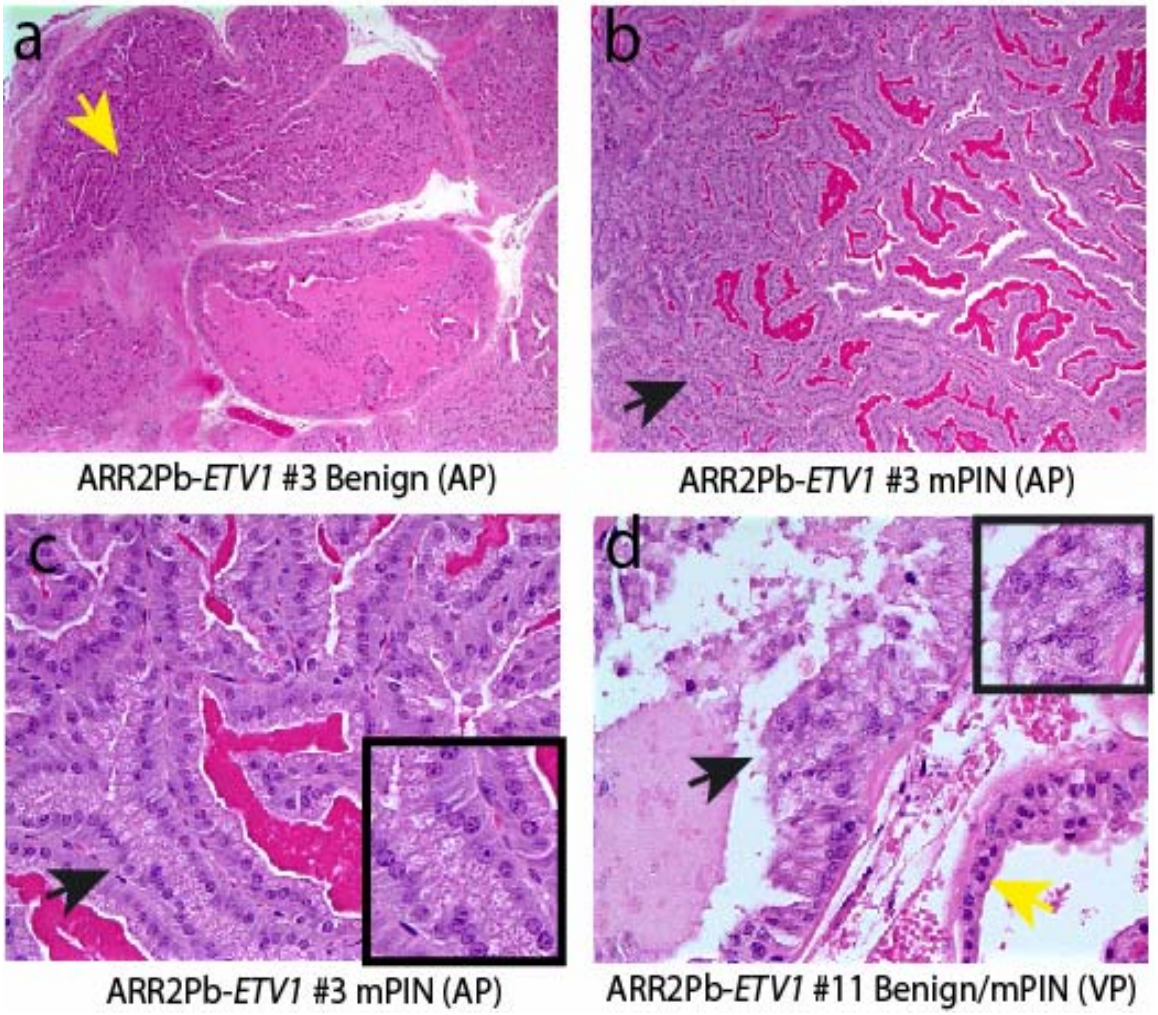

Figure 5.4. Transgenic mice expressing *ETV1* develop mouse prostatic intraepithelial neoplasia (mPIN). We generated transgenic mice expressing *ETV1* under the control of the modified probasin promoter (ARR2Pb-*ETV1*). Mice were sacrificed at 12-43 weeks and mPIN was observed in 75% of ARR2Pb-*ETV1* mice. **a-d**. Hematoxylin and eosin stained ARR2Pb-*ETV1* prostates. Benign epithelia and areas of mPIN are indicated by yellow and black arrows, respectively. Consistent with the definition of mPIN, **a**) normal areas and **b**) mPIN were observed in the anterior prostate (AP) of an ARR2Pb-*ETV1* mouse. **c**. High power view of **b**. **d**. Normal glands and mPIN foci in the ventral prostate (VP) of an ARR2Pb-*ETV1* mouse. Original magnification for **a** & **b** is 100x and **c** & **d** is 400x.

REFERENCES

1. Tomlins SA, Rhodes DR, Perner S, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 2005;*310:*644-8.
2. Tomlins SA, Mehra R, Rhodes DR, et al. TMPRSS2:ETV4 gene fusions define a third molecular subtype of prostate cancer. Cancer Res 2006;*66:*3396-400.
3. Cerveira N, Ribeiro FR, Peixoto A, et al. TMPRSS2-ERG gene fusion causing ERG overexpression precedes chromosome copy number changes in prostate carcinomas and paired HGPIN lesions. Neoplasia 2006;*8:*826-32.
4. Glinsky GV, Glinskii AB, Stephenson AJ, Hoffman RM, Gerald WL. Gene expression profiling predicts clinical outcome of prostate cancer. J Clin Invest 2004;*113:*913-23.
5. Hermans KG, van Marion R, van Dekken H, Jenster G, van Weerden WM, Trapman J. TMPRSS2:ERG fusion by translocation or interstitial deletion is highly relevant in androgen-dependent prostate cancer, but is bypassed in late-stage androgen receptor-negative prostate cancer. Cancer Res 2006;*66:*10658-63.
6. Lapointe J, Li C, Higgins JP, et al. Gene expression profiling identifies clinically relevant subtypes of prostate cancer. Proc Natl Acad Sci U S A 2004;*101:*811-6.
7. Mehra R, Tomlins SA, Shen R, et al. Comprehensive assessment of TMPRSS2 and ETS family gene aberrations in clinically localized prostate cancer. Mod Pathol 2007.
8. Perner S, Demichelis F, Beroukhim R, et al. TMPRSS2:ERG Fusion-Associated Deletions Provide Insight into the Heterogeneity of Prostate Cancer. Cancer Res 2006;*66:*8337-41.
9. Soller MJ, Isaksson M, Elfving P, Soller W, Lundgren R, Panagopoulos I. Confirmation of the high frequency of the TMPRSS2/ERG fusion gene in prostate cancer. Genes Chromosomes Cancer 2006;*45:*717-9.
10. Tomlins SA, Mehra R, Rhodes DR, et al. Integrative molecular concept modeling of prostate cancer progression. Nat Genet 2007;*39:*41-51.
11. Winnes M, Lissbrant E, Damber JE, Stenman G. Molecular genetic analyses of the TMPRSS2-ERG and TMPRSS2-ETV1 gene fusions in 50 cases of prostate cancer. Oncol Rep 2007;*17:*1033-6.
12. Yoshimoto M, Joshua AM, Chilton-Macneill S, et al. Three-Color FISH Analysis of TMPRSS2/ERG Fusions in Prostate Cancer Indicates That Genomic Microdeletion of Chromosome 21 Is Associated with Rearrangement. Neoplasia 2006;*8:*465-9.
13. Yu YP, Landsittel D, Jing L, et al. Gene expression alterations in prostate cancer predicting tumor aggression and preceding development of malignancy. J Clin Oncol 2004;*22:*2790-9.
14. Mitelman F, Johansson B, Mertens F. Fusion genes and rearranged genes as a linear function of chromosome aberrations in cancer. Nat Genet 2004;*36:*331-4.
15. Rowley JD. Chromosome translocations: dangerous liaisons revisited. Nat Rev Cancer 2001;*1:*245-50.
16. Lin B, Ferguson C, White JT, et al. Prostate-localized and androgen-regulated expression of the membrane-bound serine protease TMPRSS2. Cancer Res 1999;*59:*4180-4.

17. Mertz KD, Setlur SR, Dhanasekaran SM, et al. Molecular characterization of the TMPRSS2-ERG gene fusion in the NCI-H660 prostate cancer cell line – a new perspective for an old model. Neoplasia;*In press*.
18. Rabbitts TH. Chromosomal translocations in human cancer. Nature 1994;*372:*143-9.
19. Demichelis F, Fall K, Perner S, et al. TMPRSS2:ERG gene fusion associated with lethal prostate cancer in a watchful waiting cohort. Oncogene 2007.
20. van Bokhoven A, Caires A, Maria MD, et al. Spectral karyotype (SKY) analysis of human prostate carcinoma cell lines. Prostate 2003;*57:*226-44.
21. Cai C, Hsieh CL, Omwancha J, et al. ETV1 Is a Novel Androgen Receptor-Regulated Gene That Mediates Prostate Cancer Cell Invasion. Mol Endocrinol 2007.
22. Rhodes DR, Kalyana-Sundaram S, Tomlins SA, et al. Molecular concepts analysis links tumors, pathways, mechanisms, and drugs. Neoplasia 2007;*9:*443-54.
23. Ellwood-Yen K, Graeber TG, Wongvipat J, et al. Myc-driven murine prostate cancer shares molecular features with human prostate tumors. Cancer Cell 2003;*4:*223-38.
24. Shappell SB, Thomas GV, Roberts RL, et al. Prostate pathology of genetically engineered mice: definitions and classification. The consensus report from the Bar Harbor meeting of the Mouse Models of Human Cancer Consortium Prostate Pathology Committee. Cancer Res 2004;*64:*2270-305.
25. Suzukawa K, Parganas E, Gajjar A, et al. Identification of a breakpoint cluster region 3' of the ribophorin I gene at 3q21 associated with the transcriptional activation of the EVI1 gene in acute myelogenous leukemias with inv(3)(q21q26). Blood 1994;*84:*2681-8.
26. Wieser R. Rearrangements of chromosome band 3q21 in myeloid leukemia. Leuk Lymphoma 2002;*43:*59-65.
27. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, et al. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. Neoplasia 2007;*9:*166-80.
28. Rubin MA, Putzi M, Mucci N, et al. Rapid ("warm") autopsy study for procurement of metastatic prostate cancer. Clin Cancer Res 2000;*6:*1038-45.
29. Korenchuk S, Lehr JE, L MC, et al. VCaP, a cell-based model system of human prostate cancer. In Vivo 2001;*15:*163-8.
30. Vandesompele J, De Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 2002;*3:*RESEARCH0034.
31. Specht K, Richter T, Muller U, Walch A, Werner M, Hofler H. Quantitative gene expression analysis in microdissected archival formalin-fixed and paraffin-embedded tumor tissue. Am J Pathol 2001;*158:*419-29.
32. Stauffer Y, Theiler G, Sperisen P, Lebedev Y, Jongeneel CV. Digital expression profiles of human endogenous retroviral families in normal and cancerous tissues. Cancer Immun 2004;*4:*2.
33. Wiemels JL, Greaves M. Structure and possible mechanisms of TEL-AML1 gene fusions in childhood acute lymphoblastic leukemia. Cancer Res 1999;*59:*4075-82.

34. Beheshti B, Karaskova J, Park PC, Squire JA, Beatty BG. Identification of a high frequency of chromosomal rearrangements in the centromeric regions of prostate cancer cell lines by sequential giemsa banding and spectral karyotyping. Mol Diagn 2000;5:23-32.

CHAPTER 6

THE ROLE OF THE *TMPRSS2-ERG* GENE FUSION IN PROSTATE CANCER PROGRESSION

Multiple groups have confirmed the high prevalence of *TMPRSS2-ERG* gene fusions in prostate cancer (1-15), suggesting that this lesion is a predominant molecular subtype. Here we explored the role of the *TMPRSS2-ERG* gene fusion product using *in vitro* and *in vivo* model systems. Transgenic mice expressing the *ERG* gene fusion product under androgen-regulation develop mouse prostatic intraepithelial neoplasia (mPIN), a precursor lesion of prostate cancer. These results suggest that *TMPRSS2-ERG* may not be sufficient for transformation in the absence of secondary molecular lesions. In human tumors, *TMPRSS2-ERG* gene fusions occur early during prostate cancer progression in PIN lesions associated with or in close proximity to invasive cancer (4, 16). Introduction of the *ERG* gene fusion product into primary or immortalized benign prostate epithelial cells induced an invasion- associated transcriptional program but did not increase cellular proliferation or anchorage-independent growth. To investigate the role of *TMPRSS2-ERG* in the context of pre-existing genetic lesions, we used RNA interference to knockdown *ERG* in the *TMPRSS2-ERG* positive cell line VCaP (1). In addition to inhibiting invasion, transcriptional profiling revealed decreased expression of genes over-expressed in ETS positive prostate cancers. VCaP cells and benign prostate cells over-expressing *ERG* directly engage the plasminogen activation pathway to mediate cellular invasion, potentially representing a downstream ETS target susceptible to therapeutic intervention. *ERG* knockdown in VCaP cells also induced a transcriptional program consistent with prostate differentiation, suggesting that *TMPRSS2-ERG* maintains prostate cancer cells in a progenitor-like state. Together, these results provide roles and functional models for *TMPRSS2-ERG* gene fusions in prostate cancer.

Based on a bioinformatics strategy that nominated genes showing high expression in a subset of cancer cases, we identified fusions of the 5'-untranslated region of *TMPRSS2* (21q22) to *ERG* (21q22), *ETV1* (7p21), or *ETV4* (17q21) in prostate cancer cases that over-expressed the respective *ETS* family member (1, 17). *TMPRSS2-ERG* fusions are the most predominant molecular subtype, with multiple studies showing that approximately 50% of prostate cancers from prostate specific antigen (PSA) screened surgical cohorts are *TMPRSS2-ERG* fusion positive, and greater than 90% of prostate cancers over-expressing *ERG* harbor *TMPRSS2-ERG* fusions (1-15, 18).

As *TMPRSS2* had previously been characterized as an androgen-regulated gene (19), and *TMPRSS2* only contributes untranslated sequence to many *TMPRSS2-ERG* transcripts, we hypothesized that the androgen responsive regulatory elements of *TMRPSS2* drive *ERG* over-expression in fusion positive cases. In support of this hypothesis, we observed that treatment of the *TMPRSS2-ERG* positive prostate cancer cell line VCaP with the synthetic androgen R1881 resulted in increased expression of the *TMPRSS2-ERG* fusion product (1, 20). Additionally, castration of mice with androgen dependent *TMPRSS2-ERG* positive xenografts resulted in decreased expression of *ERG* in the xenograft (21).

Genomic fusion of the *TMPRSS2* and *ERG* loci resulting in over-expression of *ERG* and *TMPRSS2-ERG* transcripts is highly specific for prostate cancer (1, 4, 18). However, functional roles for *TMPRSS2-ERG* fusions in prostate cancer are lacking. Here we recapitulated *TMPRSS2-ERG* fusions in vivo and in vitro and used an integrative expression profiling strategy to determine functional roles for *TMRPSS2-ERG* in prostate cancer.

Fusion transcripts juxtaposing exon 1 of *TMPRSS2* (NM_005656.2) to exon 2 of *ERG* isoform 1 (NM_182918.2; identical to exon 4 of *ERG* isoform 2, NM_004449.3) are the most commonly detected transcripts in *TMPRSS2-ERG* positive cases (*TMPRSS2-ERGa*) (1, 2, 6). As exon 1 of *TMPRSS2* is entirely non-coding, this fusion transcript likely results in a truncated *ERG* protein product. Thus, we generated transgenic mice expressing the truncated *ERG* product from *TMPRSS2-ERGa* (beginning at exon 2 through the reported stop codon (base 1533) of NM_182918.2, C terminal FLAG-tagged) under the control of the modified probasin promoter (ARR2Pb-*ERG*) (**Figure 6.1**), which

drives androgen-regulated transgene expression exclusively in the prostate (22, 23). This transgene is functionally analogous to the *TMPRSS2-ERGa* fusion product. We obtained multiple ARR2Pb-*ERG* founders and lines were expanded for phenotypic analysis. By 12-14 weeks of age, 3 of 8 (37.5%) ARR2Pb-*ERG* mice developed mouse prostatic intraepithelial neoplasia (mPIN) (**Figure 6.1**), the candidate precursor lesion of prostate cancer (24).

We observed normal glands in the prostates of ARR2Pb-*ERG* mice containing focal proliferative lesions displaying nuclear atypia, including stratification, hyperchromasia and macronucleoli (**Figure 6.1**), consistent with the definition of mPIN (24). In 12-14 week old ARR2Pb-*ERG* mice, foci of mPIN were observed exclusively in the ventral lobe. Immunohistochemistry in ARR2Pb-*ERG* mice demonstrated strong ERG-FLAG expression primarily in mPIN foci and not benign glands (**Figure 6.1**), and qPCR confirmed that transgene expression was limited to the prostate (data not shown).

All lesions were confirmed to be *in situ* by the presence of an intact fibromuscular layer, as demonstrated by contiguous smooth muscle actin staining (**Figure 6.1**). However, immunohistochemistry with the basal cell markers cytokeratin 5 and p63 demonstrated loss of the circumferential basal epithelial layer in ARR2Pb-*ERG* mPIN compared to benign glands (**Figure 6.1**), indicating disruption of the basal cell layer. As loss of the basal layer is a hallmark of prostate carcinoma development in both mice and humans (25), ARR2Pb-*ERG* mice will be closely monitored for the development of invasive carcinoma at later time points. While we have not observed progression to invasive carcinoma in ARR2Pb-*ERG* mice, we have only characterized 3 mice at greater than 20 weeks of age, 1 of which (33.3%) also had mPIN in both the ventral and dorsolateral lobes. These results demonstrate that although *ERG* induces a neoplastic phenotype in the mouse prostate, providing support for an oncogenic role in human prostate cancer, it is not sufficient for the development of cancer at early time points.

Next, we determined the effects of *ERG* over-expression in vitro, by generating adenoviruses and lentiviruses that express the same truncated *ERG* product from *TMPRSS2-ERGa* as in the ARR2Pb-*ERG* mice (**Figure 6.2**). We infected the benign immortalized prostate epithelial cell line RWPE with lentivirus expressing *ERG* and selected for stable RWPE-*ERG* cells, and transiently over-expressed *ERG* in primary

benign prostate epithelial cells (PrEC) by infection with adenovirus expressing *ERG*. By immunoblotting, we confirmed the expression of a protein product recognized by a commercial anti-ERG antibody in both RWPE and PrEC.

In both RWPE and PrEC cells, over-expression of *ERG* did not increase proliferation, and *ERG* did not affect the percentage of RWPE cells in S phase by cell cycle analysis. Additionally, soft agar transformation assays showed that *ERG* over-expression was not sufficient to transform RWPE cells. Finally, orthotopic xenograft assays using RWPE-*ERG* cells did not result in tumor formation (data not shown). However, *ERG* over-expression markedly increased invasion in a modified basement membrane invasion assay in both RWPE (5 fold, P = 0.001) (**Figure 6.2**) and PrEC cells (6.9 fold, P = 0.0016) (**Figure 6.2**). Transient over-expression of *ERG* in RWPE using *ERG* adenovirus similarly increased invasion. These results are similar to over-expression of *ETV1*, which we have previously shown to increase invasion in PrEC and RWPE cells (26).

To investigate the transcriptional program regulated by *ERG*, we profiled stable RWPE-*ERG* and transiently expressing RWPE-*ERG* and PrEC-*ERG* cells using Agilent Whole Genome Oligo Expression Arrays, and identified 865, 854 and 221 features that were over-expressed in the respective cell lines (as described in the Methods). We have recently developed a resource termed the Oncomine Concepts Map (OCM, www.oncomine.org) to look for associations between more than 20,000 biologically related gene sets by disproportionate overlap (16, 27). Thus, we uploaded these expression signatures into the OCM to identify transcriptional programs induced by *ERG*. We began by seeding the OCM analysis with the 'over-expressed in stable RWPE-*ERG*' signature. OCM analysis identified the most significantly enriched concept as our previous 'over-expressed in stable RWPE-*ETV1* signature' (26) (odds ratio (OR) = 59.43 , *P* = 1E-100) (**Figure 6.2**), consistent with their similar phenotypes and supporting the functional redundancy of these ETS family members in gene fusions.

The stable RWPE-*ERG* signature also shared significant enrichment with the 'over-expressed in transient RWPE-*ERG*' (OR = 19.43, *P* = 1.1E-100) and 'transient PrEC-*ERG*' (OR = 5.77, *P* = 3.1E-10) signatures, demonstrating similarities in these transcriptional programs, as well as several molecular concepts related to invasion. These

concepts include the Interpro concept of gene products containing 'Peptidase M10A and M12B, matrixin or adamalysin domains' (OR = 5.27, *P* = 0.002), which includes matrix metalloproteinases (MMPs) and a disintegrin and metalloproteinasedomains (ADAMs), and a signature of genes 'over-expressed in benign breast epithelial cells (HMLHT) over-expressing the *STAT3-C* oncogene' (OR = 4.04, *P* = 6.3E-5). In this system, *STAT3-C* over-expression did not increase proliferation, but increased invasion in an *MMP9* dependent manner (28).

We identified several genes over-expressed in RWPE-*ERG* that were present in multiple concepts in this enrichment network and have been directly implicated in invasion in multiple cancers and models, including the metalloproteinases *MMP3*, *MMP9* and *ADAM19,* and urokinase plasminogen activator (*PLAU*) and the plasminogen activator inhibitor type 1 (*SERPINE1*, also known as *PAI-1*) (29, 30). Both MMPs and the urokinase plasminogen pathway have been reported to be direct target of ETS transcription factors (29-31). By qPCR, we confirmed the over-expression of these genes, as well as the MMP cleavage target *IGFBP3* in RWPE-*ERG* cells (**Figure 6.2**).

By immunoblotting, we confirmed the over-expression of PLAU and MMP3 in RWPE-*ERG* cells (**Figure 6.2**). To determine if these genes are direct targets of ERG, we performed chromatin immunoprecipitation (ChIP), which demonstrated that ERG binds to the proximal promoter of both *PLAU* and *MMP3* (**Figure 6.2**), but not *PLAT* (tissue plasminogen activator, which did not show altered expression in RWPE-ERG by microarray). No enrichment of ERG binding was observed in RWPE-*GUS* cells or LNCaP (*ETV1* rearrangement positive (26)) for *MMP3* or *PLAU*.

We next assessed the role of both MMPs and the plasminogen activator pathways in the invasive phenotype of RWPE-*ERG* cells using small molecule MMP inhibitors, amiloride (a specific PLAU inhibitor (32)), ectopic PAI-1 (which inhibits plasminogen activators (33)) and siRNA knockdown of *PLAU*. As shown in **Figure 6.2**, while MMP inhibitors did not significantly inhibit invasion, both amiloride and PAI-1 significantly inhibited the invasiveness of RWPE-*ERG* cells. Similarly, siRNA knockdown of *PLAU* significantly inhibited the invasion of RWPE-*ERG* cells, while siRNA knockdown of the tissue plasminogen activator (*PLAT*) had no effect on RWPE-*ERG* invasion (**Figure 6.2**). Similar effects on invasion were seen with independent siRNA duplexes directed against

PLAU or *PLAT*. Cytosine arabinoside (ARA-C), which has recently been identified as an inhibitor of the EWS-FLI fusion found in Ewing's sarcoma (34), also showed no effect on RWPE-*ERG* invasion (**Figure 6.2**). Together, this work demonstrates that *ERG* directly induces *PLAU* expression in RWPE cells, and inhibition of *PLAU* blocks *ERG* mediated invasion.

Together, our in vivo and in vitro studies show that the most common *TMPRSS2-ERG* fusion product is unable to transform benign prostatic epithelial cell lines or induce the development of frank adenocarcinoma in the mouse prostate. These results support our previous work, including expression profiling on laser captured microdissected cell populations and a FISH-based study on prostate cancer progression, which suggest that *TMPRSS2-ERG* gene fusions occur during the PIN to carcinoma transition or very early in carcinoma development (4, 16). Thus, in human prostate cancer development, *TMPRSS2-ERG* gene fusions likely occur in the context of earlier lesions that induce the development of PIN.

To investigate the role of *TMPRSS2-ERG* in this context of pre-existing genetic lesions, we utilized short interfering (si) RNA to knock down *ERG* in VCaP (VCaP-si*ERG*) cells that harbor the TMPRSS2:ERG gene fusion (1). Immunoblotting confirmed that siRNA directed against ERG reduced expression compared to non-targeting control siRNA (**Figure 6.3**). qPCR also demonstrated a 63% decrease in ERG transcript expression in VCaP-si*ERG* ($P = 0.009$). *ERG* knockdown also significantly inhibited the invasion of VCaP cells (**Figure 6.3**) without affecting proliferation, similar to *ERG* over-expression in RWPE cells. Similar results were seen using a second siRNA targeting an independent sequence in *ERG*.

To determine the transcriptional profile mediated by *TMPRSS2-ERG* in VCaP, we profiled VCaP-si*ERG* cells. We identified 265 and 291 features over- and under-expressed (as described in the Methods), respectively, in VCaP-si*ERG* compared to VCaP treated with non-targeting siRNA, and uploaded these signatures into the OCM. The two most significantly enriched concepts in our 'under-expressed in VCaP-si*ERG*' signature were two signatures of genes 'over-expressed in *ETS* positive vs. negative prostate cancers' (GSE8218, OR = 5.73, $P = 2.5E-19$ and Vanaja et al (35), OR = 3.49, $P = 3.9E-11$) (**Figure 6.3**). All other 'over-expressed in *ETS* positive vs. negative prostate

cancer' signatures (16, 36, 37) in the Oncomine database were also enriched in our 'under-expressed in VCaP-si*ERG*" signature, supporting VCaP as a highly relevant model of *TMPRSS2-ERG* positive prostate cancers. Our under-expressed in VCaP-si*ERG* signature also shared enrichment with our previous signature of genes 'over-expressed in laser captured prostate cancer vs. PIN' (OR = 3.79, *P* = 4.5E-6). In that study, we observed that PIN and prostate cancer had very similar expression signatures and hypothesized that *TMPRSS2-ERG* fusions occurred during the PIN to prostate cancer transition and dysregulated a limited number of transcripts, likely involved in invasion (16).

Our under-expressed in VCaP-si*ERG* signature also shared significant enrichment with our 'over-expressed in transient PrEC-*ERG*' and 'transient RWPE-*ERG*' signatures (**Figure 6.3**, OR = 6.89 and 3.21, P = 1.4E-5 and 7E-5, respectively), suggesting common transcriptional programs controlled by *ERG* across cell types and genetic context. However, the PrEC and RWPE signatures were not strongly linked to *ETS* positive vs negative prostate cancer signatures, supporting the use of VCaP as a more relevant system to identify downstream targets of *TMPRSS2-ERG* in human prostate cancers.

Interestingly, the most strongly down regulated feature in VCaP-si*ERG* cells was the tissue plasminogen activator *PLAT*. Similar to *PLAU*, which we showed to be strongly over-expressed and a direct target of *ERG* in RWPE cells, we confirmed that *PLAT* was strongly down-regulated and a direct target of *ERG* in VCaP-si*ERG* cells (**Figure 6.3**). Intriguingly, although microarray analysis showed that VCaP-si*ERG* cells did not have reduced expression of *PLAU* (and VCaP cells express very low levels of *PLAU* at baseline) ChIP identified *PLAU* as a direct target of ERG in VCaP-si*ERG* cells (**Figure 6.3**). Additionally, while ectopic PAI-1, amiloride (which inhibits *PLAU* but not *PLAT* (32)) (**Figure 6.3**) and siRNA knockdown of *PLAU* inhibited the invasion of VCaP cells, siRNA knockdown of *PLAT* had no effect on VCaP invasion (**Figure 6.3**). Additionally, inhibitors of MMPs and ARA-C had no significant effect on VCaP invasion (**Figure 6.3**), similar to RWPE-*ERG*. Together, these results support plasminogen activators as direct targets of *ERG* across multiple *TMPRSS2-ERG* model systems, and demonstrate that inhibition of *PLAU* blocks *ERG* induced invasion across *TMPRSS2-ERG* cell line models.

Our 'under-expressed in VCaP-si*ERG*' signature also shared significant enrichment with a cluster of 18 genes co-expressed across 72 prostate cancer tissue samples (36), with 8 genes shared (OR = 56.65, *P* = 7.2E-10). As this cluster contains *ERG*, this result supports *ERG* knockdown in VCaP modulating genes regulated by *ERG* in *TMPRSS2-ERG* positive tumors. To identify such genes, we examined genes co-expressed with *ERG* across multiple prostate cancer profiling studies in the Oncomine database. We identified 4 genes, *CACNA1D*, *KCNS3*, *LAMC2* and *PLA1A,* that showed greater than 0.5 correlation with *ERG* across multiple studies that were also down-regulated in *VCaP*-si*ERG* cells. *CACNA1D* was significantly down regulated in 3 of 4 arrays, with the fourth array showing 0.54 fold expression in VCaP-si*ERG* (P = 0.06). In addition, we also identified decreased expression of *ARGHDIB* in VCaP-si*ERG* cells and over-expression in all ETS positive vs. negative expression signatures. By qPCR, we confirmed the decreased expression of these genes in VCaP-si*ERG* cells and ChIP identified *LAMC2*, *KCNS3*, and *PLA1A* as direct targets of *ERG*. By qPCR, we also confirmed the co-expression of *ERG* and *PLA1A* (R = 0.72, P = 6.1E-8) in an independent set of prostate tissues. Thus, our work provides direct *ERG* target genes over-expressed in *TMPRSS2-ERG* positive prostate cancers for further functional study.

We next examined our 'over-expressed in VCaP-si*ERG*' signature using the OCM. Consistent with the results described above, all 'under-expressed in *ETS* positive vs. negative prostate cancer' signatures in the Oncomine database (GSE8218 and (16, 35-37)) were enriched in our 'over-expressed in VCaP-si*ERG*' signature (OR = 6.41-2.71, *P* = 5.2E-15 - 7.0E-5). Intriguingly, OCM analysis revealed that the most significantly enriched concept in our 'over-expressed in VCaP-si*ERG*' signature was a signature of genes 'over-expressed in prostate cancers compared to 28 other cancer types' (GSE2109) (OR = 4.46, *P* = 5.8E-18) (**Figure 6.4**). Several other concepts representing genes over-expressed in prostate cancer compared to other cancers, normal prostate tissue compared to other normal tissues and normal prostate compared to prostate cancer were also strongly enriched in our signature. Examining the genes common to these concepts and VCaP-si*ERG*, we identified numerous archetypal prostate epithelial cell transcripts, including *KLK3* (*PSA*), *MSMB*, *NKX3-1*, *TMPRSS2*, *TRGV9* (also known as *TARP*) (38), *SLC30A4* (also known as ZnT4) (39) and *SLC45A3* (26) (**Figure 6.4**). We confirmed the

over-expression of this transcriptional program by qPCR (**Figure 6.4**), and confirmed that these genes are normally expressed specifically in luminal epithelial prostate cells using a dataset containing expression profiling data from magnetically sorted prostate luminal epithelial, basal epithelial, stromal fibromuscular and endothelial cells (**Figure 6.4**).

As *ERG* knockdown in VCaP results in the increased expression of genes associated with differentiated luminal prostate epithelial cells, we hypothesized that *TMPRSS2-ERG* fusions may function to keep prostate cancer cells in a dedifferentiated state. Supporting this hypothesis, expression profiling revealed decreased expression of *CD44* in VCaP-si*ERG* cells (n = 40 features, median = 0.51 VCaP-siERG/VCaP-siNT, maximum *P* value = 0.0006). *CD44* has been proposed as a prostate cancer stem cell marker, and prostate cancer cell populations with enriched for CD44 expression show cancer stem cell like properties (40) . We confirmed the decreased expression of *CD44* in VCaP-siERG cells by qPCR (**Figure 6.4)**. Future experiments will be needed to further address this hypothesis.

In this study, we demonstrated that over-expression of *ERG* in RWPE cells induces invasion through a *PLAU* dependent pathway, however, *ERG* over-expression was not sufficient for transformation. Over-expression of *ERG* in the mouse prostate resulted in mPIN, without the development of adenocarcinoma. These results are consistent with our previous work suggesting that in human prostate cancer development, *TMPRSS2-ERG* fusions likely occur in the context of potentially earlier lesions, such as loss of single *NKX3-1* and/or *PTEN* alleles (41). The development of mPIN in *ERG* transgenic mice without early progression to carcinoma is similar to mouse models of these other early events in human prostate cancer, such as *NKX3-1*$^{+/-}$ and *PTEN*$^{+/-}$ mice (42-44). Crosses between ARR2Pb-*ERG* mice and these mice should produce highly relevant oncogene/tumor suppressor models mimicking early events in human prostate cancer development. In an effort to study *TMPRSS2-ERG* function in a more realistic cellular context, we confirmed VCaP as a highly relevant cell line model, as siRNA knockdown of *ERG* modulates transcriptional programs apparent in *TMRPSS2-ERG* positive tumors. Gene expression profiling and ChIP demonstrate that *ERG* regulates distinct genes in different cellular context, likely related to the expression of other ETS

transcription factors, which bind to similar consensus sequences, and the expression of other transcription factors required for DNA binding (45). For example, *PLA1A* and *PLAT* were both direct targets of ERG and down-regulated exclusively in VCaP but not RWPE-*ERG* cells.

Our work here also supports the functional similarity between *ERG* and *ETV1* gene fusions, consistent with our initial observation of mutually exclusive *ERG* or *ETV1* over-expression in prostate cancers (1). This includes the similar phenotypic and transcriptional programs induced by over-expression in benign prostate cells, the similar phenotype of transgenic mice expressing *ERG* or *ETV1* in the prostate (26), and the enrichment of genes over-expressed in *ERG* or *ETV1* positive vs. ETS negative prostate cancers in our VCaP-siERG signature (see **Figure 6.3c**). In summary, our work has identified in vivo and in vitro models of *TMPRSS2-ERG* positive prostate cancer, highlighting the plasminogen activator pathway as crucial to *ERG* mediated invasion in multiple in vitro systems. This pathway warrants further investigation as a therapeutic target for *TMPRSS2-ERG* positive prostate cancer.

Methods

cDNA of *ERG*, as present in the *TMPRSS2-ERGa* fusion transcript to the reported stop codon (exon 2 to base 1533 of NM_182918.2), was amplified by RT-PCR from the VCaP cell line and TOPO cloned into the Gateway entry vector pCR8/GW/TOPO (Invitrogen), yielding pCR8-*ERG*. For in vivo over-expression of *ERG*, a C terminal 3XFLAG-epitope tagged construct was generated by PCR using pCR8-*ERG* as the template with the reverse primer encoding a triple FLAG tag before the stop codon. The product was TOPO cloned into pCR8, generating pCR8-3xFLAG-ERG. To generate a prostate specific *ERG* transgenic construct, 3xFLAG-*ERG* was inserted into pBSII (Stratagene, La Jolla, CA) downstream of a modified small composite probasin promoter (ARR2PB) and upstream of a bovine growth hormone polyA site (PA-BGH). The ARR2PB sequence contains the original probasin sequence PB (-426/+28) plus two additional androgen response elements (22, 23). The construct was sequenced and tested for promoter inducibility by androgen in LNCaP cells upon transient transfection before microinjection into FVB mouse eggs. The ARR2PB-*ERG* plasmid was linearized with

PvuI/KpnI//SacII and microinjected into fertilized FVB mouse eggs and surgically transplanted into a pseudo-pregnant female by the University of Michigan Transgenic Animal Model Core. Transgenic founders were screened by PCR using genomic DNA isolated from tail snips. Multiple ARR2Pb-*ERG* transgenic founders were obtained and crossed with FVB mice, and transgene-positive male mice offspring were sacrificed at various time points.

Prostates from transgenic mice were dissected using a Nikon dissection scope, fixed in 10% buffered formalin and embedded in paraffin. Five um sections were stained with hematoxylin and eosin, and evaluated by three pathologists (R.M., M.A.R. and R.B.S.) according to the criteria provided in The Consensus Report from the Bar Harbor Meeting of the Mouse Models of Human Cancer Consortium Prostate Pathology Committee (24).

For immunohistochemical detection of Erg-FLAG, the basal cell markers p63 and cytokeratin 5 (CK5), and smooth muscle actin, deparaffinized slides were subjected to microwave-citrate antigen retrieval and incubated with rabbit anti-FLAG polyclonal antibody (1:50 dilution, overnight incubation, Cell Signaling Technology, #2368), mouse monoclonal anti-p63 antibody (1:100 dilution, 45' incubation, LabVision, MS1081P1), rabbit polyclonal anti-CK5 antibody (1:500 dilution, 30' incubation, AbCam, ab24647) and mouse monoclonal anti-smooth muscle actin antibody (1:50 dilution, 30' incubation, DakoAb M0851), respectively. Visualization of p63 and SMA was performed using a standard biotin-avidin complex technique using M.O.M Immunodetection kit (PK2200, Vector Laboratories). FLAG and CK5 were detected using Envision+System-HRP (DAB) kit (K4011, DakoCytomation).

The benign immortalized prostate cell line RWPE was obtained from the ATCC. Primary benign prostatic epithelial cells (PrEC) were obtained from Cambrex Bio Science (Walkersville, MD). VCaP was derived from a vertebral metastasis from a patient with hormone-refractory metastatic prostate cancer (46), and provided by Kenneth Pienta (University of Michigan).

Prostate 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 (S.P.O.R.E.)

Tissue Core. All samples were collected with informed consent of the patients and prior institutional review board approval. For all samples and cell lines, total RNA was isolated with Trizol (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions.

To generate adenoviral and lentiviral constructs, pCR8-*ERG* and a control entry clone (pENTR-*GUS*) were recombined with pAD/CMV/V5 (Invitrogen) and pLenti6/CMV/V5 (Invitrogen), respectively, using LR Clonase II (Invitrogen). Control pAD/CMV/*LACZ* clones were obtained from Invitrogen. Adenoviruses and Lentiviruses were generated by the University of Michigan Vector Core. The benign immortalized prostate cell line RWPE was infected with lentiviruses expressing *ERG* or *GUS*, and stable clones were generated by selection with blasticidin (Inivtrogen). Benign PrEC cells were infected with adenoviruses expressing *ERG* or *LACZ*, as stable lines could not be generated in primary PrEC cells. RWPE cells were also infected with *ERG* or *LACZ* adenoviruses for transient over-expression.

Cells were homogenized in NP40 lysis buffer containing 50 mM Tris-HCl (pH 7.4), 1% NP40 (Sigma, St. Louis, MO), and complete proteinase inhibitor mixture (Roche). Fifteen ug of protein extracts were mixed with SDS sample buffer and electrophoresed onto a 10% SDS-polyacrylamide gel under reducing conditions. The separated proteins were transferred onto nitrocellulose membranes (Amersham Pharmacia Biotech, Piscataway, NJ). The membrane was incubated for 1 h in blocking buffer [Tris-buffered saline with 0.1% Tween (TBS-T) and 5% nonfat dry milk]. Primary antibody was applied at the indicated dilution in blocking buffer overnight at 4°C. After washing three times with TBS-T buffer, the membrane was incubated with horseradish peroxidase-linked donkey anti-mouse IgG antibody or donkey anti-rabbit IgG antibody (Amersham Pharmacia Biotech) at a 1:5,000 dilution for 1 h at room temperature. The signals were visualized with the enhanced chemiluminescence detection system (Amersham Pharmacia Biotech) and autoradiography.

Rabbit polyclonal anti-ERG (sc-354, Santa Cruz Biotechnology, Santa Cruz) was applied at 1:500 dilution, mouse monoclonal anti-MMP-3 (IM36L, Calbiochem, San Diego) was applied at 1:500 dilution, mouse monoclonal anti-uPA (IM13L, Calbiochem) was applied at 1:500 dilution, and mouse anti-GAPDH antibody (Abcam, Cambridge, MA) was applied at 1:30,000 dilution for loading control.

Cell counts were estimated by trypsinizing cells and analysis by Coulter counter (Beckman Coulter, Fullerton, CA) at the indicated time points in triplicate.

RWPE-*ERG* and RWPE-*GUS* cells were assessed by FACS for cell cycle characterization. Cells were washed with 2x PBS and approximately 2 x 10^6 cells were resuspended in PBS before fixation in 70% ethanol. Pelleted cells were washed and treated with RNase (100ug/ml final concentration) and propidium iodide (10ug/ml final concentration) at 37 C for 30 min. Stained cells were analyzed on a LSR II flow cytometer (BD Biosciences, San Jose, CA) running FACSDivia, and cell cycle phases were calculated using ModFit LT (Verity Software House, Topsham, ME).

A 0.6% (wt/vol) bottom layer of low melting point agarose in normal medium was prepared in six-well culture plates. On top, a layer of 0.3% agarose containing 1 x 10^4 RWPE-*GUS*, RWPE-*ERG* or *DU145* (positive control) cells was placed. After 12 days, foci were stained with crystal violet and counted.

For invasion assays, PrEC and RWPE-*ERG* and –*LACZ* cells (48 hours after infection with adenoviruses), stable RWPE-*ERG* and –*GUS* cells, or VCaP cells were used. Equal numbers of the indicated cells were seeded onto the basement membrane matrix (EC matrix, Chemicon, Temecula, CA) present in the insert of a 24 well culture plate, with fetal bovine serum added to the lower chamber as a chemoattractant. After 48 hours, non-invading cells and EC matrix were removed by a cotton swab. Invaded cells were stained with crystal violet and photographed. The inserts were treated with 10% acetic acid and absorbance was measured at 560nm.

For inhibitor studies, amiloride (20 uM, EMD Biosciences, San Diego), MMP3 inhibitor (10 uM, EMD Biosciences), MMP2/9 inhibitor (10 uM, EMD Biosciences), MMP8 inhibitor (10 uM EMD Biosciences), the pan MMP inhibitor GM 6001 (10 uM EMD Biosciences), the EWS:FLI inhibitor cytosine arabinoside (250 nM) (34) or vehicle control was added to VCaP and stable RWPE-*ERG* or -*GUS* cells for 24 hours, prior to trypisinization and seeding for invasion assays as described above. For PAI-1, VCaP and stable RWPE-*ERG* or -*GUS* cells were trypsinized and treated with the indicated amount of recombinant PAI-1 (EMD Biosciences, San Diego) for 15 minutes at indicated concentrations, before seeding as described above.

For siRNA knockdown of *ERG*, *PLAT*, or *PLAU*, the individual siRNAs composing the Dharmacon SMARTpool against *ERG* (MQ-003886-01, Lafayette, CO), *PLAT* (LQ-005999-00), or *PLAU* (LQ-006000-00), were tested for knockdown by qPCR, and the most effective single siRNA (*ERG*, D-003886-01; *PLAT*, J-005999-05; *PLAU*, J-006000-07;) was used for further experiments. siCONTROL Non-Targeting siRNA #1 (D-001210-01) or siRNA against *ERG*, *PLAT*, or *PLAU* was transfected into VCaP or RWPE-ERG cells as indicated using Oligofectamine (Invitrogen). After 24 hours we carried out a second identical transfection and cells were harvested 24 hours later for RNA isolation, invasion assays or proliferation assays as described above.

Expression profiling was performed using the Agilent Whole Human Genome Oligo Microarray (Santa Clara, CA). Total RNA isolated using Trizol was purified using the Qiagen RNAeasy Micro kit (Valencia, CA). One μg of total RNA was converted to cRNA and labeled according to the manufacturer's protocol (Agilent). Hybridizations were performed for 16 hrs at 65°C, and arrays were scanned on an Agilent DNA microarray scanner. Images were analyzed and data extracted using Agilent Feature Extraction Software 9.1.3.1, with linear and lowess normalization performed for each array. For all hybridizations involving *ERG* over-expression by adenovirus or lentivirus, the reference was the same cell line expressing *LACZ* or *GUS*, respectively. For profiling of *ERG* knockdown in VCaP, the reference was VCaP treated with non-targeting siRNA. All hybridizations were performed in duplicate with duplicate dye flips, for a total of four arrays, except for transiently expressing RWPE-*ERG*, which consisted of duplicate hybridizations and a single dye flip. Over and under-expressed signatures were generated by filtering to include only features with significant differential expression (PValueLogRatio < 0.01) in all hybridizations and 2 fold average over- or under-expression (Log Ratio) after correction for the dye flip. For VCaP profiling, all features with significant differential expression (PValueLogRatio < 0.01) and Cy5/Cy3 ratios > or < 1 in all hybridizations were included in the over- and under-expressed signatures, respectively.

All expression signatures were uploaded into the Oncomine Concepts Map (OCM, www.oncomine.org) (27) as molecular concepts, using all features on the Agilent Whole Human Genome Oligo Microarray as the null set. Each signature was tested

against all contained concepts in the OCM for association using Fisher's exact test as described (16, 27). Molecular concept maps and overlay maps of genes included in concepts were generated by the OCM. Datasets included in Oncomine without associated publications include the International Genomics Consortium's expO dataset https://expo.intgen.org/expo/public/downloaddata.jsp (GSE2109) and the Yang et al. "Gene expression data from prostate cancer samples" dataset (GSE8218).

For the assessment of prostate specific gene expression, the expO and Shyamsundar normal tissue (47) datasets were accessed using the Oncomine database. For the assessment of prostate cell type expression, the Oudes et al. (48) "Prostate cell specific expression" Affymetrix dataset was downloaded from GEO (GSE3998). Data is reported as RMA normalized fluorescent intensity.

Quantitative PCR (QPCR) was performed using Power SYBR Green Mastermix (Applied Biosystems, Foster City, CA) on an Applied Biosystems 7300 Real Time PCR system as described (1, 17). All oligonucleotide primers were synthesized by Integrated DNA Technologies (Coralville, IA). All reactions were performed in duplicate unless otherwise indicated.

ChIP was performed according to published protocols with slight modifications (49). Briefly, formaldehyde was added directly to the cultured cells to a final concentration of 1%. Crosslinking was stopped by adding 1/20V of 2.5M Glycine to culture medium and cells were washed with 1xPBS and harvested in 1xPBS with proteinase inhibitors. Cells were then pelleted, washed once with 1xPBS plus proteinase inhibitors, and resuspended in cell lysis buffer containing protease inhibitors. After incubation in cell lysis buffer for 10 min, the samples were pelleted, resuspended in nuclei lysis buffer and sonicated to chromatins with an average size of 500bp. Chromatin was then precleared using Salmon sperm DNA/Protein A Agarose-50% flurry (Upstate) and incubated with anti-ERG (Santa Cruz, sc-354x) or rabbit anti-IgG (Santa Cruz, sc-2027) antibodies overnight. The next day, the antibody-bound chromatin was pulled down using protein A/agarose, washed extensively and reverse-crosslinked. Immunoprecipitated DNA and whole cell extract input DNA were purified by treatment with RNase A, proteinase K and purified using a Qiaquick PCR purification kit, and and eluted in 25ul EB buffer (Qiagen, Valencia, CA). For PCR analysis of enrichment of

target gene promoters, 2ul each of input DNA, ERG-enriched, or IgG-enriched DNA were subjected to PCR using Platinum PCR Supermix (Invitrogen) and primers specific for target gene promoters. PCR conditions were 95^{o}C for 3 min, followed by 25-35 cycles of 30 sec denaturation at 95^{o}C, 30 sec annealing at 55-57^{o}C, and 1 min extension at 75^{o}C. PCR products were then analyzed by gel electrophoresis.

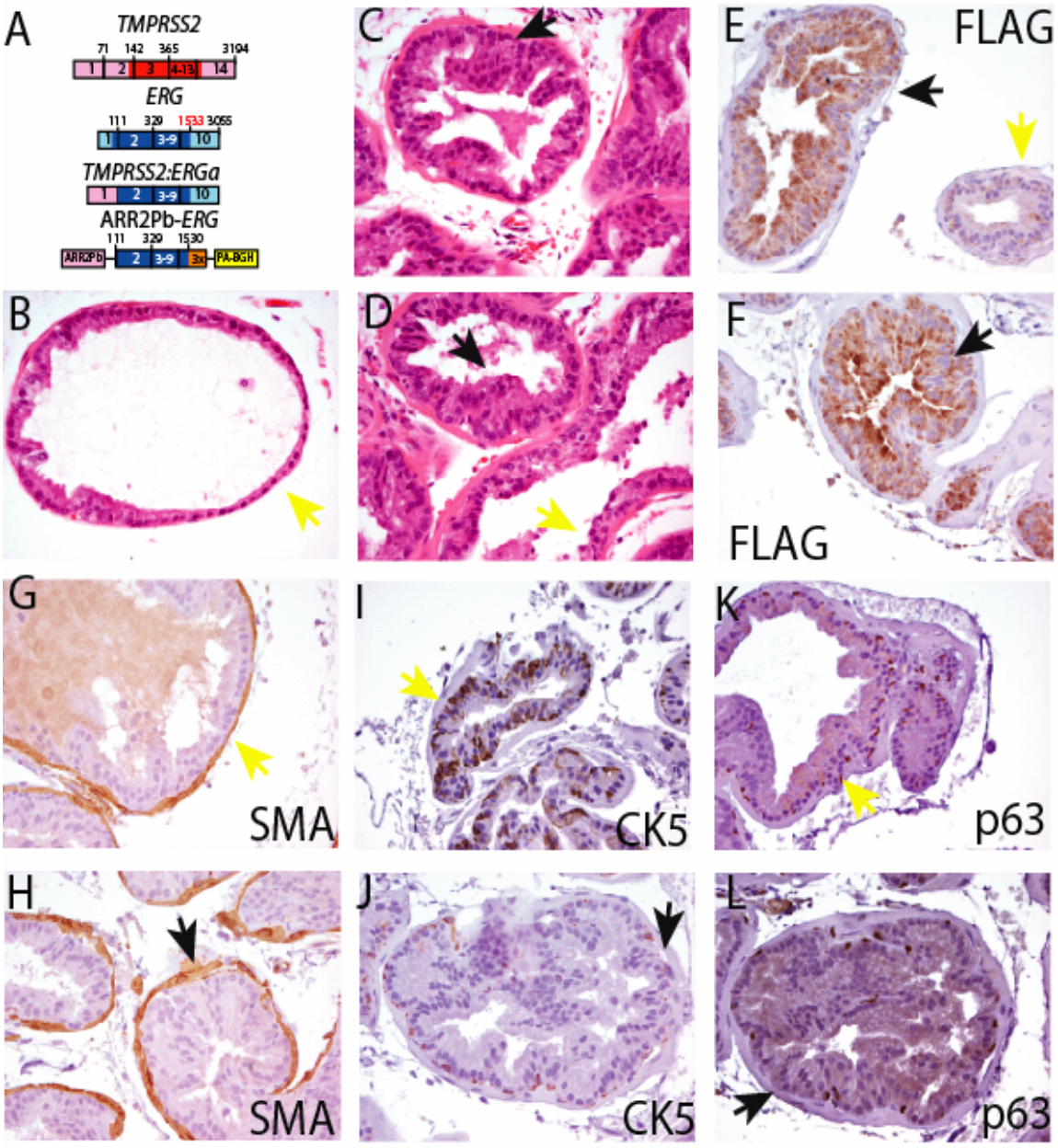

Figure 6.1. Transgenic mice recapitulating *TMPRSS2-ERG* in the prostate develop mPIN. A. We generated transgenic mice over-expressing the *ERG* gene fusion product (C-terminal 3X-FLAG epitope tag) under the control of the enhanced probasin promoter (ARR2Pb). Mouse prostatic intraepithelial neoplasia (mPIN) was observed in 4 of 11 ARR2Pb-*ERG* mice. Benign epithelia and areas of mPIN are indicated by yellow and black arrows, respectively. **B-D**. Hematoxylin and eosin staining of ARR2Pb-*ERG* prostates for morphological assessment. Consistent with the focality of mPIN, **B**) benign glands and **C-D**) mPIN were observed in the ventral prostate (VP) of ARR2Pb-*ERG* mice. **E-F**. Immunohistochemistry (IHC) confirmed ERG-FLAG expression exclusively in areas of mPIN and not benign glands in ARR2Pb-*ERG* mice. IHC with smooth muscle actin (SMA) demonstrates a continuous fibromuscular layer around **G**) benign glands and **H**) all mPIN lesions, while the basal cell markers **I-J**) cytokeratin 5 (CK5) and **K-L**) p63 demonstrate loss of circumferential basal cells in mPIN foci (**J,L**) compared to normal glands (**I,K**).

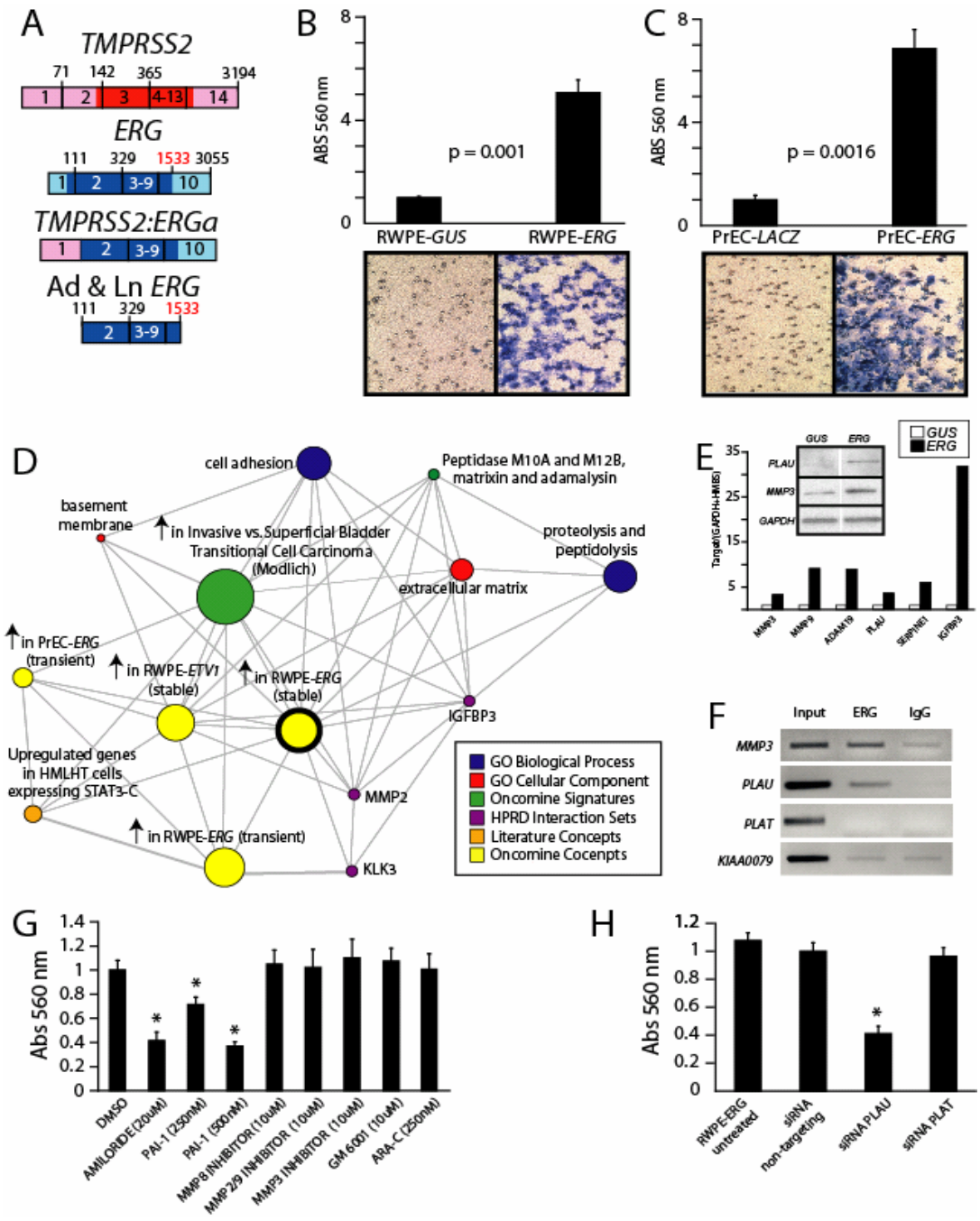

Figure 6.2. Over-expression of *ERG* in RWPE cells increases invasion through the plasminogen activator pathway. A. We generated adenoviruses and lentiviruses expressing the ERG gene fusion product. **B-C**. Infected **B**) RWPE and **C**) PrEC were assayed for invasion through a modified basement membrane. **D**. MCM analysis of the 'over-expressed in RWPE-*ERG* compared to RWPE-*GUS*' signature (ringed yellow node). **E**. qPCR confirmation of increased expression of genes involved in invasion. Inset shows immunoblot confirmation. **F.** Chromatin immunoprecipitation shows enrichment of ERG binding to the proximal promoters of *PLAU* and *MMP3*, but not *PLAT*, compared to IgG control. **G**. RWPE-*ERG* cells were treated with *PLAU* inhibitors, MMP inhibitors, or the EWS:FLI inhibitor ARA-C as indicated and assayed for invasion as in **C**. **H**. RWPE-ERG cells were treated with trasnfection reagent alone (untreated), or transfected with non-targeting, *PLAU* or *PLAT* siRNA as indicated and assayed for invasion. For all invasion assays mean (n = 3) + S.E. are shown, asteriks indicate $P < 0.05$.

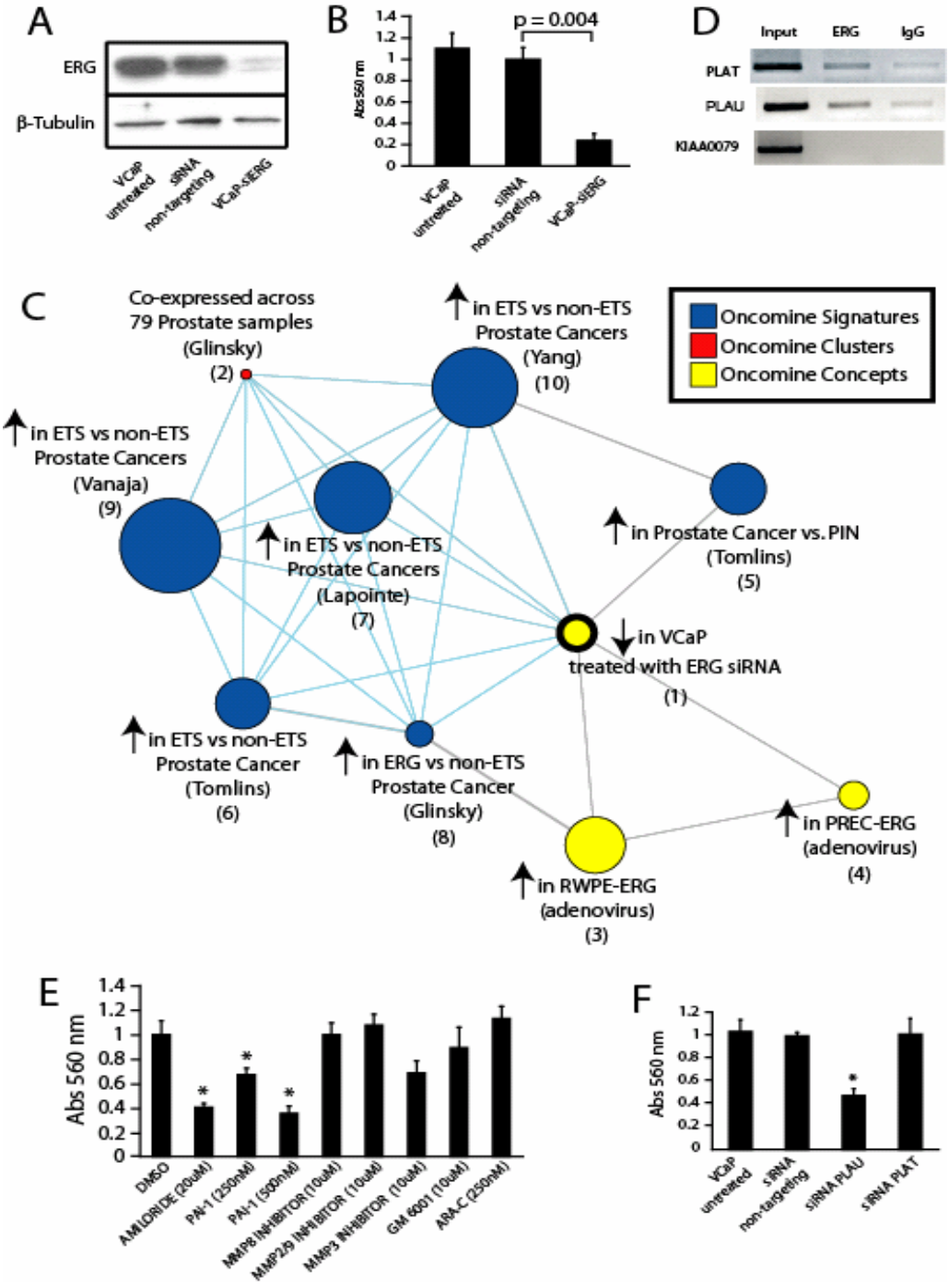

Figure 6.3. Knockdown of ERG in VCaP cells. A. The *TMPRSS2-ERG* positive prostate cancer cell line VCaP was treated with transfection reagent alone (untreated), or transfected with non-targeting or *ERG* siRNA (VCaP-si*ERG*) as indicated. *ERG* knockdown was confirmed by immunoblotting. **B**. VCaP cells as indicated were assayed for invasion through a modified basement membrane. **C**. VCaP-si*ERG* and VCaP cells treated with non-targeting siRNA were profiled and a molecular concept map of the 'under-expressed in VCaP-si*ERG*' signature (ringed yellow node) was generated. Each edge represents a significant enrichment ($P < 0.001$). Blue edges indicate enrichments with in vivo *ETS* positive vs. negative prostate cancer signatures. **D**. ChIP identifies *PLAT* and *PLAU* as direct targets of ERG in VCaP cells. **E**. VCaP cells were treated with **E**) inhibitors or **F**) siRNAs as in **Figure 6.2** and assessed for invasion.

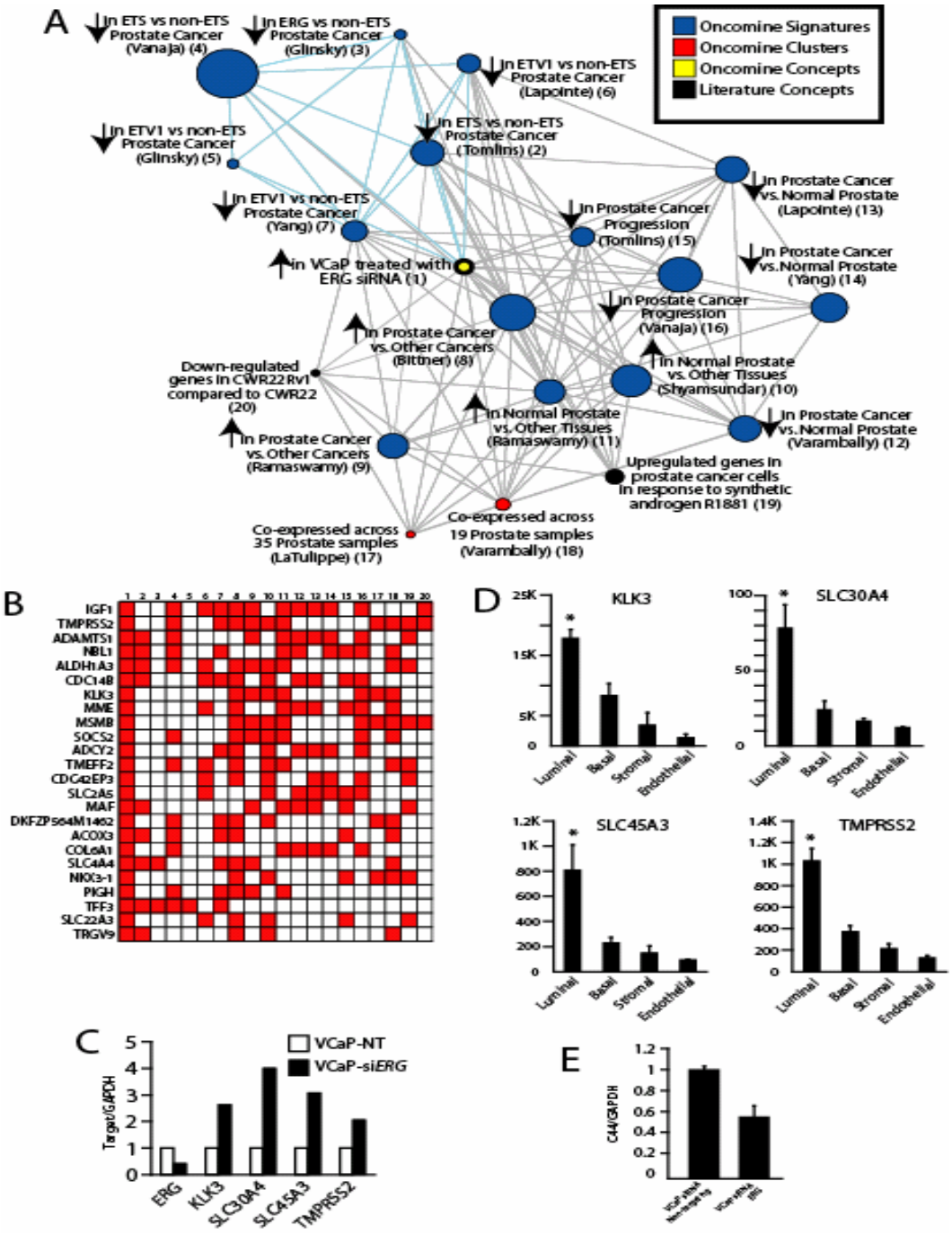

Figure 6.4. *ERG* knockdown in VCaP cells de-represses a transcriptional program associated with normal prostatic epithelial cell differentiation. A. Molecular concept map of the 'over-expressed in VCaP-si*ERG*' signature (ringed yellow node) was generated. Blue edges indicate enrichments with in vivo '*ETS* positive vs. negative prostate cancer' signatures. **B**. Overlay map identifying genes present (red cells) across multiple concepts in the over-expressed in VCaP-si*ERG* enrichment network **C**. qPCR confirmation of increased expression in VCaP-siERG cells compared to VCaP-NT cells of transcripts strongly expressed in prostatic epithelial cells. **D**. Analysis of prostate cell type specificity using a microarray dataset profiling magnetically sorted prostate cell populations. Asterisks indicate $P < 0.05$ for all pairwise t-tests involving luminal cells. **E**. Decreased expression of the proposed prostate cancer stem cell marker *CD44* in VCaP-si*ERG* cells was identified by microarray analysis and confirmed by qPCR.

REFERENCES

1. Tomlins SA, Rhodes DR, Perner S, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 2005;*310:*644-8.
2. Lapointe J, Kim YH, Miller MA, et al. A variant TMPRSS2 isoform and ERG fusion product in prostate cancer with implications for molecular diagnosis. Mod Pathol 2007.
3. Perner S, Demichelis F, Beroukhim R, et al. TMPRSS2:ERG Fusion-Associated Deletions Provide Insight into the Heterogeneity of Prostate Cancer. Cancer Res 2006;*66:*8337-41.
4. Perner S, Mosquera JM, Demichelis F, et al. TMPRSS2-ERG fusion prostate cancer: an early molecular event associated with invasion. Am J Surg Pathol 2007;*31:*882-8.
5. Soller MJ, Isaksson M, Elfving P, Soller W, Lundgren R, Panagopoulos I. Confirmation of the high frequency of the TMPRSS2/ERG fusion gene in prostate cancer. Genes Chromosomes Cancer 2006;*45:*717-9.
6. Wang J, Cai Y, Ren C, Ittmann M. Expression of Variant TMPRSS2/ERG Fusion Messenger RNAs Is Associated with Aggressive Prostate Cancer. Cancer Res 2006;*66:*8347-51.
7. Cerveira N, Ribeiro FR, Peixoto A, et al. TMPRSS2-ERG gene fusion causing ERG overexpression precedes chromosome copy number changes in prostate carcinomas and paired HGPIN lesions. Neoplasia 2006;*8:*826-32.
8. Clark J, Merson S, Jhavar S, et al. Diversity of TMPRSS2-ERG fusion transcripts in the human prostate. Oncogene 2006.
9. Hermans KG, van Marion R, van Dekken H, Jenster G, van Weerden WM, Trapman J. TMPRSS2:ERG fusion by translocation or interstitial deletion is highly relevant in androgen-dependent prostate cancer, but is bypassed in late-stage androgen receptor-negative prostate cancer. Cancer Res 2006;*66:*10658-63.
10. Iljin K, Wolf M, Edgren H, et al. TMPRSS2 Fusions with Oncogenic ETS Factors in Prostate Cancer Involve Unbalanced Genomic Rearrangements and Are Associated with HDAC1 and Epigenetic Reprogramming. Cancer Res 2006;*66:*10242-6.
11. Yoshimoto M, Joshua AM, Chilton-Macneill S, et al. Three-Color FISH Analysis of TMPRSS2/ERG Fusions in Prostate Cancer Indicates That Genomic Microdeletion of Chromosome 21 Is Associated with Rearrangement. Neoplasia 2006;*8:*465-9.
12. Mehra R, Tomlins SA, Shen R, et al. Comprehensive assessment of TMPRSS2 and ETS family gene aberrations in clinically localized prostate cancer. Mod Pathol 2007.
13. Nam RK, Sugar L, Wang Z, et al. Expression of TMPRSS2 ERG Gene Fusion in Prostate Cancer Cells is an Important Prognostic Factor for Cancer Progression. Cancer Biol Ther 2007;*6*.
14. Rajput AB, Miller MA, De Luca A, et al. Frequency of the TMPRSS2:ERG gene fusion is increased in moderate to poorly differentiated prostate cancers. J Clin Pathol 2007.

15. Winnes M, Lissbrant E, Damber JE, Stenman G. Molecular genetic analyses of the TMPRSS2-ERG and TMPRSS2-ETV1 gene fusions in 50 cases of prostate cancer. Oncol Rep 2007;*17:*1033-6.
16. Tomlins SA, Mehra R, Rhodes DR, et al. Integrative molecular concept modeling of prostate cancer progression. Nat Genet 2007;*39:*41-51.
17. Tomlins SA, Mehra R, Rhodes DR, et al. TMPRSS2:ETV4 gene fusions define a third molecular subtype of prostate cancer. Cancer Res 2006;*66:*3396-400.
18. Demichelis F, Fall K, Perner S, et al. TMPRSS2:ERG gene fusion associated with lethal prostate cancer in a watchful waiting cohort. Oncogene 2007.
19. Lin B, Ferguson C, White JT, et al. Prostate-localized and androgen-regulated expression of the membrane-bound serine protease TMPRSS2. Cancer Res 1999;*59:*4180-4.
20. Mertz KD, Setlur SR, Dhanasekaran SM, et al. Molecular characterization of TMPRSS2-ERG gene fusion in the NCI-H660 prostate cancer cell line: a new perspective for an old model. Neoplasia 2007;*9:*200-6.
21. Hendriksen PJ, Dits NF, Kokame K, et al. Evolution of the androgen receptor pathway during progression of prostate cancer. Cancer Res 2006;*66:*5012-20.
22. Zhang J, Thomas TZ, Kasper S, Matusik RJ. A small composite probasin promoter confers high levels of prostate-specific gene expression through regulation by androgens and glucocorticoids in vitro and in vivo. Endocrinology 2000;*141:*4698-710.
23. Ellwood-Yen K, Graeber TG, Wongvipat J, et al. Myc-driven murine prostate cancer shares molecular features with human prostate tumors. Cancer Cell 2003;*4:*223-38.
24. Shappell SB, Thomas GV, Roberts RL, et al. Prostate pathology of genetically engineered mice: definitions and classification. The consensus report from the Bar Harbor meeting of the Mouse Models of Human Cancer Consortium Prostate Pathology Committee. Cancer Res 2004;*64:*2270-305.
25. DeMarzo AM, Nelson WG, Isaacs WB, Epstein JI. Pathological and molecular aspects of prostate cancer. Lancet 2003;*361:*955-64.
26. Tomlins SA, Laxman B, Dhanasekaran SM, et al. Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. Nature 2007;*448:*595-9.
27. Rhodes DR, Kalyana-Sundaram S, Tomlins SA, et al. Molecular concepts analysis links tumors, pathways, mechanisms, and drugs. Neoplasia 2007;*9:*443-54.
28. Dechow TN, Pedranzini L, Leitch A, et al. Requirement of matrix metalloproteinase-9 for the transformation of human mammary epithelial cells by Stat3-C. Proc Natl Acad Sci U S A 2004;*101:*10602-7.
29. Laufs S, Schumacher J, Allgayer H. Urokinase-receptor (u-PAR): an essential player in multiple games of cancer: a review on its role in tumor progression, invasion, metastasis, proliferation/dormancy, clinical outcome and minimal residual disease. Cell Cycle 2006;*5:*1760-71.
30. Fingleton B. Matrix metalloproteinases: roles in cancer and metastasis. Front Biosci 2006;*11:*479-91.

31. de Launoit Y, Baert JL, Chotteau-Lelievre A, et al. The Ets transcription factors of the PEA3 group: transcriptional regulators in metastasis. Biochim Biophys Acta 2006;*1766:*79-87.
32. Vassalli JD, Belin D. Amiloride selectively inhibits the urokinase-type plasminogen activator. FEBS Lett 1987;*214:*187-91.
33. Gils A, Declerck PJ. Plasminogen activator inhibitor-1. Curr Med Chem 2004;*11:*2323-34.
34. Stegmaier K, Wong JS, Ross KN, et al. Signature-based small molecule screening identifies cytosine arabinoside as an EWS/FLI modulator in Ewing sarcoma. PLoS Med 2007;*4:*e122.
35. Vanaja DK, Cheville JC, Iturria SJ, Young CY. Transcriptional silencing of zinc finger protein 185 identified by expression profiling is associated with prostate cancer progression. Cancer Res 2003;*63:*3877-82.
36. Glinsky GV, Glinskii AB, Stephenson AJ, Hoffman RM, Gerald WL. Gene expression profiling predicts clinical outcome of prostate cancer. J Clin Invest 2004;*113:*913-23.
37. Lapointe J, Li C, Higgins JP, et al. Gene expression profiling identifies clinically relevant subtypes of prostate cancer. Proc Natl Acad Sci U S A 2004;*101:*811-6.
38. Wolfgang CD, Essand M, Vincent JJ, Lee B, Pastan I. TARP: a nuclear protein expressed in prostate and breast cancer cells derived from an alternate reading frame of the T cell receptor gamma chain locus. Proc Natl Acad Sci U S A 2000;*97:*9437-42.
39. Henshall SM, Afar DE, Rasiah KK, et al. Expression of the zinc transporter ZnT4 is decreased in the progression from early prostate disease to invasive prostate cancer. Oncogene 2003;*22:*6005-12.
40. Tang DG, Patrawala L, Calhoun T, et al. Prostate cancer stem/progenitor cells: identification, characterization, and implications. Mol Carcinog 2007;*46:*1-14.
41. Tomlins SA, Rubin MA, Chinnaiyan AM. Integrative Biology of Prostate Cancer Progression. Annual Review of Pathology: Mechanisms of Disease 2006;*1:*243-71.
42. Kim MJ, Cardiff RD, Desai N, et al. Cooperativity of Nkx3.1 and Pten loss of function in a mouse model of prostate carcinogenesis. Proc Natl Acad Sci U S A 2002;*99:*2884-9.
43. Abdulkadir SA, Magee JA, Peters TJ, et al. Conditional loss of Nkx3.1 in adult mice induces prostatic intraepithelial neoplasia. Mol Cell Biol 2002;*22:*1495-503.
44. Di Cristofano A, De Acetis M, Koff A, Cordon-Cardo C, Pandolfi PP. Pten and p27KIP1 cooperate in prostate cancer tumor suppression in the mouse. Nat Genet 2001;*27:*222-4.
45. Li R, Pei H, Watson DK. Regulation of Ets function by protein - protein interactions. Oncogene 2000;*19:*6514-23.
46. Korenchuk S, Lehr JE, L MC, et al. VCaP, a cell-based model system of human prostate cancer. In Vivo 2001;*15:*163-8.
47. Shyamsundar R, Kim YH, Higgins JP, et al. A DNA microarray survey of gene expression in normal human tissues. Genome Biol 2005;*6:*R22.
48. Oudes AJ, Campbell DS, Sorensen CM, Walashek LS, True LD, Liu AY. Transcriptomes of human prostate cells. BMC Genomics 2006;*7:*92.

49. Boyd KE, Wells J, Gutman J, Bartley SM, Farnham PJ. c-Myc target gene specificity is determined by a post-DNAbinding mechanism. Proc Natl Acad Sci U S A 1998;*95:*13887-92.

CHAPTER 7

CONCLUSION

Chromosomal aberrations accompanying carcinogenesis have been documented for almost half a century, with gene fusions being the most prevalent type of aberration. Gene fusions leading to generation of aberrant fusion proteins or aberrant expression of normal proteins provide a potent route to carcinogenesis, and have recently emerged as attractive therapeutic targets. Intriguingly, while gene fusions have been widely observed in hematological malignancies, they have been far less frequently described in the more common epithelial carcinomas. Recently, scientists have proposed that technical issues, rather than any fundamental dichotomy between hematological and solid cancers, account for under-reporting of gene fusions in epithelial cancers. Recent reports from our group support this contention and provide evidence of widespread recurrent gene fusions in prostate cancer using a novel analysis of gene expression profiles. This review provides an appraisal of the state of our knowledge of gene fusions in epithelial cancers. Future implications of gene fusions in common epithelial cancers are also discussed.

A wide variety of recurrent chromosomal aberrations have been associated with cancers including polymorphisms, changes in gene copy number (amplifications/ deletions), point mutations, epigenetic modifications, and most commonly, gene fusions due to structural chromosomal aberrations. Of the 291 genes named as ‘cancer genes’ based on documented causal association with oncogenesis, as many as 90% have recognized somatic mutations (1), and the most common class of somatic mutations affecting these cancer genes involve structural aberrations that often result in a gene fusion. This makes recurrent structural aberrations the most prevalent class of genetic aberrations causally associated with cancers. Two types of structural aberrations have been described. One involves the regulatory elements of a gene (promoter and/or enhancer) becoming aberrantly apposed to a proto-oncogene, driving deregulated

expression of the oncogene; for example, immunoglobulin and T cell receptor regulatory regions driving aberrant expression of the *MYC* oncogene in B and T cell malignancies, respectively (2). The other type of structural aberration results when the coding regions of two genes are juxtaposed, resulting in a chimeric transcript that produces a fusion protein with a new or altered activity; for example the t(9;22) resulting in the *BCR-ABL1* gene fusion in chronic myelogenous leukemia (CML) (2,3). Recurrent chromosomal aberrations have causal roles in oncogenesis, therefore, it is not surprising that they provide specific therapeutic targets, including Trastuzumab for breast carcinoma with *ERBB2* (*HER2*) amplification (4) and Imatinib for CML (5, 6).

A puzzling observation for cancer biologists has been that the recurrent structural aberrations and gene fusions typifying the majority of the 'cancer genes' have been predominantly observed in hematological malignancies (leukemias and lymphomas) and soft tissue tumors (sarcomas), which together represent only 10% of all human cancers. However, recurrent structural aberrations are relatively rare in the most common epithelial carcinomas, which account for 80% of cancer related deaths. Does this apparent discrepancy indicate that recurrent structural aberrations are truly rare in the epithelial carcinomas, perhaps due to some fundamental difference from hematological malignancies? Or could this discrepancy simply indicate that epithelial carcinomas are not amenable to the same analytical techniques that are successfully used for detecting chromosomal aberrations in liquid or soft tissue tumors? This review summarizes the current status of this argument, favoring the latter view, as well as findings from our group that provide evidence of prevalent gene fusions in a common epithelial tumor.

Recently, our group has reported the discovery of recurrent structural aberrations resulting in gene fusions in prostate cancer, a common epithelial carcinoma (7-9). This discovery was made using gene expression profiling data, which provides an alternative approach to identifying chromosomal aberrations including gene fusions and amplifications. Furthermore, this work provides compelling support to the idea that recurrent structural aberrations and gene fusions are as much of an etiological factor in epithelial carcinomas as in hematological malignancies.

Tracing the history of recurrent gene fusions in cancer, a clear skew towards hematological malignancies emerges. We examine possible reasons for such a skew,

provide a brief description of the well documented gene fusions in epithelial carcinoma, and discuss why such gene fusions are not more frequently described in epithelial cancers. The review then describes in detail the recent discovery of gene fusions in one of the most common epithelial carcinomas, prostate cancer, using a novel analytical approach involving gene expression profiling data. We conclude with a consideration of the open questions, and the outlook for future studies.

In 1960, Peter Nowell, a pathologist at the University of Pennsylvania and David Hungerford, a cytogeneticist at the Institute for Cancer Research in Philadelphia, reported an abnormally small chromosome in all seven patients with chronic granulocytic leukemia, but not in normal cells or in the cells of other types of leukemia (10, 11). This was the first time a recurrent chromosomal aberration was described in a human neoplasm. Remarkably, Nowell and Hungerford also noted the presence of this aberration, later termed the "Philadelphia chromosome", in the leukemia cells at the onset of the disease, pre-treatment, as well as in the leukemia cells of patients with persistent disease.

Nowell and Hungerford used painstaking observations of metaphase chromosome preparations for their watershed work. About 10 years later, in 1970, quinacrine fluorescence/Giemsa based chromosome banding techniques were described. These techniques facilitated precise identification of specific chromosomal landmarks (12). Using chromosomal banding techniques, the Philadelphia chromosome was found to be generated as a result of a translocation of a large portion of the long arm of chromosome 22 to chromosome 9. This was shown to be accompanied with a reciprocal translocation between a truncated portion of the proto-oncogene protein-tyrosine kinase *c-abl* from chromosome 9 to the *bcr* gene locus on chromosome 22 (13). The resulting fusion gene *bcr-ab*l was found to encode a chimeric protein with tyrosine kinase activity (14) which could produce leukemia when expressed in mouse bone marrow cells (15). This molecular insight formed the basis for development of specific inhibitors of the *bcr-abl* kinase, including the rationally designed drug Imatinib for the treatment of leukemia patients harboring the *bcr-abl* fusion (5, 6). The discovery of the *bcr-abl* fusion and it's association with a specific malignancy, followed by causal linkage with a specific cancer

and molecular characterization has resulted in an enormous interest in identifying similar chromosomal aberrations and targeted therapeutics.

Karyotype analysis of leukemias and lymphomas using chromosome banding techniques led to widespread identification of numerous prevalent, recurrent chromosomal aberrations associated with specific clinical entities. In 1973, the *AML1-ETO*/t(8;21) translocation was described in acute myelogenous leukemia (16), *AML1-ETO* (*RUNX1/RUNX1T1*)positive cases accounting for over 10% of AML patients. AML1 is a sequence-specific DNA binding protein that activates transcription of genes involved in myeloid maturation (17). AML1-ETO, on the other hand, inhibits transcription of AML1-responsive genes (18) presumably through the action of ETO which recruits transcriptional repressing machinery to the regulatory regions of AML1 responsive genes (19).

The fusion of the retinoic acid receptor alpha (*RARA*) gene to the *PML* gene as a result of translocation t(15;17) was associated with acute promyelocytic leukemia in 1991 (20). Interestingly, the promyelocytic leukemia (*PML*) gene actually encodes a tumor suppressor that controls apoptosis, cell proliferation, and senescence (21), but the PML-RARA fusion acts as a dominant-negative retinoic acid receptor, inhibiting the response to retinoic acid and blocking retinoic-induced differentiation of promyelocytes (22) through aberrant chromatin acetylation (23, 24).

Numerous distinct gene fusions involving the *MLL* gene at 11q23 and approximately 20 partner genes have been identified in patients with AML and ALL (25, 26). MLL is believed to regulate homeotic genes that are transcriptional activators of developmentally regulated gene families, and the different fusion partners, bring about modulation of MLL target genes through their effects on chromatin remodelling activity of MLL involving the SWI/SNF or histone acetylase complexes (27).

A subset of AML is characterized by pericentric inversion of chromosome 16 [inv(16)(p13q22)], resulting in fusion of upstream *CBFB (PEBP2B)* gene to downstream *MYH11(SMHC)*, a smooth muscle myosin heavy chain gene, leading to synthesis of chimeric CBFB-MYH11 protein (28).

More recently, identification and fine mapping of additional recurrent chromosomal aberrations has been improved with refinements of molecular cytogenetic

tools, including fluorescence in situ hybridization (FISH) (29, 30) especially multicolor-FISH techniques (31, 32), resulting in increasingly higher resolution and sensitivity.

The development of specific diagnoses and classification of hematological malignancies using gene-fusion specific probes as well as development of targeted therapy specifically directed against fusion protein product (for example Imatinib against *BCR-ABL1* protein) represent rare success stories in cancer biology where basic information has heralded clinically relevant developments.

Since the advent of chromosome banding techniques over 30 years ago, more than 600 acquired, recurrent, balanced chromosomal aberrations have been documented from more than 47,000 neoplasms representing all tumor tissue types (33, 34). The existing data strongly highlights that cancers are characterized by specific, recurrent chromosomal rearrangements with diagnostic and therapeutic implications. In the Mitelman Database of Chromosome Aberrations in Cancer, the veritable 'Yellow Pages' of recurrent gene fusions, one remarkable discrepancy stands out: hematological malignancies, which constitute less than 10% of all human cancers, account for 74% of all known recurrent balanced chromosomal aberrations. On the other hand, malignant epithelial tumors which cause 80% of human cancer deaths, constitute only 10% of all known recurrent balanced chromosomal aberrations (35). Recurrent rearrangements and gene fusions have not been described in common epithelial cancers such as colon, breast, lung and prostate adenocarcinoma--until very recently. This disproportionately meager representation of structural aberrations in solid tumors, especially epithelial tumors, has lead to speculation that structural aberrations are not causally important in solid tumors. Alternatively, this discrepancy may simply reflect a technological limitation that has prevented the identification of recurrent structural aberrations in epithelial tumors.

Before a discussion on the likely reasons for the lack of widespread identification of gene fusions in epithelial tumors, we will briefly survey the few but well documented structural aberrations resulting in gene fusions that characterize specific epithelial tumors. A comprehensive, tumor-wise account of cytogenetic aberrations have been compiled (36).

ETV6-NTRK3: A gene fusion between the dimerization domain of the transcription factor *ETV6* (*TEL*) with the tyrosine kinase domain of Neurotrophic tyrosine

kinase, receptor, type 3, *NTRK3* (*TRKC*), originally associated with congenital fibrosarcoma (37), was identified in secretory breast carcinoma, a rare subset of breast cancer (38). Also expressed in hematopoietic adult AML (39) and mesenchymal cellular mesoblastic nephroma (40), the ETV6-NTRK3 protein has been shown to display potent transforming activity through activation of the Ras-MAP kinase and PI3 kinase-AKT pathways (41, 42).

RET and NTRK1: As many as 50% of thyroid papillary carcinomas are believed to harbor structural aberrations between receptor tyrosine kinases *RET* (Rearranged during Transfection proto-oncogene) or *NTRK1* (Neurotrophic tyrosine kinase, receptor, type 1) and various different partners (43-45). These fusion proteins are comprised of the C-terminal halves of *RET* or *NTRK1* containing their tyrosine kinase domains linked to the N-terminal portions of various partner proteins. RET fusion protein partners include *H4* (unknown function) (46, 47), *R1alpha* (PKA regulatory subunit) (48), *ELE1* (Androgen receptor-associated cotranscription factor) (49, 50) and RFG5 (golgi autoantigen) (51). Similarly, NTRK1 fusions have been reported with *TPR* (Translocated promoter region; also known to be involved in tumorigenic rearrangements with the *met* and *raf* genes) (52, 53) and *TFG* (TRK-fused gene) (54) etc. The native RET and NTRK1 proteins are membrane bound, and RET tyrosine kinase activity is activated as a result of multimerization upon ligand binding. The gene fusions remove the N-terminal signal peptide of RET and NTRK1, rendering them cytoplasmic (55), and the various fusion partners with their dimerization or multimerization domains presumably account for their constitutive tyrosine kinase activity (56).

PAX8:PPARgamma: Follicular thyroid carcinoma (FTC) accounts for approximately 20% of all thyroid cancers, and up to 40% of the deaths associated with this disease (57). Up to 50% of FTCs undergo a translocation event between chromosome regions 3p25 and 2q13 that fuses the DNA binding domain of the thyroid-specific transcription factor, paired box gene 8 (*PAX8*) with the peroxisome proliferator-activated receptor gamma, (*PPARG*), a ubiquitously expressed transcription factor. The *Pax8/PPARG* fusion gene (designated *PPFP*) acts as an oncogene in cell line experiments, where it accelerates cell growth, reduces rates of apoptosis and permits anchorage independent and contact uninhibited growth. PPFP is believed to act as a

dominant-negative inhibitor of the wild-type PPARG transcription factor, likely through competition for the genomic *PPARG* response elements, the endogenous ligand (57). Indeed, in a recent microarray analysis, PPAR target genes have been shown to be upregulated in *PPFP*-positive follicular carcinomas (58).

PRCC:TFE3: A number of sporadic cases of papillary renal cell carcinoma affecting children and young adults have been found to harbor chromosomal translocations involving the *TFE3* gene at chromosome Xp11.2.49-51 with the N-terminal region of a novel gene designated *PRCC* (for papillary renal cell carcinoma) at 1q21.2 (59-61). The *TFE3* gene encodes a helix–loop–helix transcription factor related to the proto-oncogene product *c-myc*, and is suspected to be the oncogenic element fused to the novel gene *PRCC* containing proline rich regions.

PLAG1 and HMGA2: Recurrent gene fusions resulting in pathogenetically relevant fusion oncogenes have been frequently found in both benign and malignant salivary gland tumors (62). The majority of translocations involve DNA-binding transcription factors PLAG1 (pleomorphic adenoma gene 1) and *HMGA2* (High mobility group family protein), notably involved in growth factor signaling and cell cycle regulation. Other important carcinoma associated fusion oncogenes described in salivary tumors include the *CRTC1* (*TORC1)-MAML2* fusion involving *CREB* regulated transcription coactivator 1, *CRTC1* (*TORC1*) fused to mastermind like gene, *MAML2*, a coactivator of Notch. These fusions have also been observed in other soft tissue tumors as well as benign tumors (62).

BRD4 and NUT: Some of the poorly differentiated carcinoma of midline, head and neck or thoracic structures affecting young individuals, characterized by a highly aggressive and fatal clinical course have been reported to carry a translocation between chromosomes 15 and 19, i.e. t(15;19), resulting in the fusion oncogene *BRD4* (bromodomain-containing gene) and *NUT* (nuclear protein in testis) (63-65). French et al, have also described rarer, and somewhat less aggressive *NUT*-rearranged carcinomas (NRCs) among young patients with midline cancers that appear to have less aggressive clinical course and had longer average survival (96 weeks, n = 3) as compared to *BRD4-NUT* carcinomas (28 weeks, n = 8) (66). Haruki et al, have cloned the fusion transcript of *BRD4-NUT* with the 3' end of *BRD4* (chromosome 19p) fused to the 5' end of *NUT*

(chromosome 15q) from a lung cancer cell line, although none of the 128 lung cancer tissues screened with that probe using FISH demonstrated t(15;19) (67).

The above examples of gene fusions in epithelial carcinomas, although rare in absolute terms, provide tantalizing indication of a more widespread occurrence of similar gene fusions in other more common epithelial cancers. The big puzzle has been, why don't we observe them as often?

Overall, other than the above mentioned, relatively rare examples, the representation of gene fusions in more common epithelial carcinoma has been so scarce, it merits asking why? Previously identified aberrations in hematological malignancies have been identified by karyotyping, and candidate translocations followed up later by further molecular techniques. Karyotyping carcinomas has not been as feasible. The chromosome quality in epithelial neoplasms is particularly poor, often yielding only partial and poor quality karyotypes, leading to inaccurate analysis. Further, unlike hematological malignancies, which routinely display limited cytogenetic changes, carcinomas often present with multiple aberrations apparently acquired in the course of tumorigenesis. Whole genome analyses by recent high throughput techniques like array comparative genomic hybridization (aCGH) also reveal karyotypic complexity in epithelial tumors, complicating the identification of primary, causal aberrations from secondary acquired aberrations. Further, clonal heterogeneity due to the presence of cytogenetically unrelated clones, seen in less than 5% of leukemias, lymphomas, and mesenchymal tumors, but in up to 80% of epithelial carcinomas (68), renders epithelial tumors uniquely problematic in terms of identifying causal aberrations and gene fusions (69). While these difficulties may have impeded the determination of the true contribution of gene fusions in carcinomas, is there evidence to support the contention that the problem is only technical, not conceptual? Is there any evidence, for example, to suggest that the solid epithelial carcinomas are in reality not fundamentally different from the liquid and semi-solid cancers? This question was addressed by Mitelman *et al*. (see below), using all available data, and their conclusions have proved prophetic.

In a study involving all of the more than 450 recurrent balanced aberrations documented from more than 45,000 cytogeneticaly characterized human neoplasms from different tissue types (35), Mitelman *et al.* noted that the number of fusion genes as well

as the broader subset of rearranged genes in each tumor category were "strictly proportional" to the total number of cases with an abnormal karyotype. In other words, since the incidence of recurrent chromosomal abnormalities, including gene fusions, in different tumor types is "simply a function of the number of cases with an abnormal karyotype", fewer recurrent gene fusions observed in epithelial neoplasms is simply because far fewer accurately determined abnormal karyotypes have been documented from epithelial tumors, and not due to any fundamental tissue-specific differences. These somewhat unexpected observations lead Mitelman *et al*. to proclaim, rather boldly, in their own words, "the unorthodox concept", that "cytogenetic aberrations resulting in deregulated or rearranged genes may be of greater importance as an initial step in epithelial tumorigenesis than generally believed". Another profound speculation made by the authors is that "the epithelial tumors are characterized by numerous but individually rare, pathogenetically important gene rearrangements that have not yet been identified". The passages that follow bear out these predictions substantially.

Using microarrays, gene expression signatures corresponding to different subgroups of cancers have been widely reported, including those associated with some known oncogenes (70, 71). Could a converse strategy be applied to discover novel rearranged or amplified oncogenes based on gene expression patterns? Recently, our group developed a novel bioinformatics algorithm in an effort to identify oncogenes. Rather than focusing on the routine broad based signatures of gene expression used to shortlist potential biomarkers, we chose to focus on the so called 'outliers' (7). As cancers are heterogeneous and also often display a heterogeneous pattern of oncogene expression, the conventional search for broad gene expression signatures across a class of cancer samples would fail to identify oncogene profiles that may be highly active in only a subset of cases (similar to the argument of Mitelman *et al* (35), referred in the previous subheading). For example, the breast cancer oncogene *Her2/neu* is amplified in 20-30% of breast cancers. So a '*HER2* status naive' microarray study involving 50 samples would likely show *HER2* overexpression in only 10-15 samples, which would get overlooked using conventional analytical techniques in favor of genes uniformly over-expressed in all breast cancers. Similarly, as *E2A-PBX1* translocations are present in only 5-10% of

leukemia, it would be even more unlikely to be discovered by conventional analysis if cytogenetic studies were not performed.

In our recent study (7), we hypothesized that the expression profiles of cancers likely reflect marked overexpression of oncogenes (resulting from gene fusions or amplifications), albeit only in small subsets of samples. To systematically identify such small subsets with highly elevated levels of expression we developed a novel algorithm termed Cancer Outlier Profile Analysis (COPA) to interrogate gene expression profiles. The COPA transformation involves median centering all the expression values, setting each gene's median expression value to zero. This is followed by calculation of the median absolute deviation (MAD) for each gene and the gene expression value is divided by its MAD to obtain the transformed expression value. This COPA transformation effectively compresses typical biomarker profiles characterized by general overexpression in cancer relative to normal tissue, whereas it accentuates the 'outlier' profiles characterized by general low expression with marked overexpression in a fraction of cancer samples. The 75th, 90th, and 95th percentiles of the transformed expression values are then tabulated and rank-ordered to provide a list of outlier genes.

We applied COPA to 132 gene expression data sets comprising more than 10,000 microarray experiments (available in the Oncomine (72) database, www.oncomine.org). This analysis successfully identified several well known cancer genes in specific cancer types with well documented recurrent chromosomal rearrangements or amplifications. In addition, we observed that two ETS transcription factors known to be involved in gene fusions in Ewing's sarcoma, *ERG* (21q22.3) and *ETV1* (7p21.2), were highly ranked outliers in multiple independent prostate cancer profiling studies.

The discovery of marked overexpression of *ERG* and *ETV1* in prostate cancer was followed up by another intriguing observation; the outlier profiles of *ERG* and *ETV1* were mutually exclusive, similarly to Ewing's sarcoma, where *EWSR1-ERG* and *EWSR1-ETV1* fusion genes are mutually exclusive in different cases (73). We used exon walking QPCR to show that samples with overexpression of the 3' regions of *ERG* or *ETV1* showed markedly reduced expression of the 5' regions, suggesting a possible gene fusion. Finally, using 5' RNA ligase mediated RACE to identify the 5' end of the *ETV* and *ERG* transcripts in cases over-expressing the respective ETS gene, we identified fusions with

the 5' untranslated region (UTR) of a prostate specific, androgen responsive, transmembrane serine protease gene (*TMPRSS2*, 21q22.2) (74-76). The presence of these novel gene fusions in prostate cancer was confirmed at the chromosomal level when interphase fluorescence in situ hybridization (FISH) on tissue microarrays showed 23 out of 29 (79%) randomly selected prostate cancers were positive for one of the two gene fusions. These numbers, while small in absolute terms, are staggering in that prostate cancer is one of the most common epithelial carcinomas (affecting more than 200,000 men per year in the US alone). Thus, *TMPRSS2:ETS* fusions might represent the most common recurrent structural aberration and gene fusion among all human malignancies! The finding of TMPRSS2:ETS fusions in the majority of prostate cancers has subsequently been confirmed by multiple indepdent groups.

More recently, our group has reported another rare chromosomal translocation involving *TMPRSS2* and another ETS family member, *ETV4*, in a prostate cancer sample that did not harbor the previously identified fusions involving *ERG* or *ETV1* (9). Additionally, we have also identified novel 5' partners which result in distinct functional classes (8). While these recurrent aberrations result in the production of fusion genes, their structure indicates that *TMPRSS2* and other 5' partners do not contribute to the coding sequence. Instead, the androgen responsive and other promoter elements upstream of *TMPRSS2* or other 5' partners appear to drive the aberrant expression of the ETS gene in the prostate (**Figure 7.1**). Thus, these novel gene fusions appear to be more similar to the class of structural aberrations characterized by tissue specific regulatory element usurping control of proto-oncogenes, as has been well documented in lymphoid malignancies.

The discovery of the first recurrent structural aberrations in the form of gene fusions in prostate cancer is exciting in and of itself, but the novel use of gene expression data to rationally identify putative gene rearrangements in epithelial tumors is a conceptual advance of profound impact, as it provides a long sought tool for discovery of gene fusions in all common epithelial tumors (9). It may be envisaged that using this approach, more gene fusions might be identified in other common epithelial carcinomas, such as breast, lung, colon, and brain tumors etc, where quality gene expression data is available. Recently, using a retroviral transformation assay from a cDNA library, EML4-

ALK gene fusions have been identified in ~6% of non-small cell lung carcinoma (77), supporting a more generalized role for gene fusions in epithelial cells. Intriguingly, most tumors with fusion genes, such as secretary breast carcinoma, renal carcinoma, thyroid and myeloid malignancies, all tend to present more frequently in adolescence and young adulthood (78), unlike prostate and lung cancer, which are late onset malignancies. This could be yet another reflection of the global prevalence of gene fusions, as well as potentially reflect a more fundamental early role of gene fusions in carcinogenesis (**Figure 7.2**).

Gene fusions accompanying cancers represent very useful diagnostic markers and potent therapeutic targets. The prototype gene fusion *BCR-ABL1* described for over four decades in chronic myelogenous leukemia has contributed enormously to our understanding of such chromosomal aberrations in carcinogenesis as well as has served as a target for effective therapeutic intervention. Similar progress is envisioned with respect to gene fusions in other hematological malignancies and soft-tissue tumors, and as described here, now even in epithelial carcinomas. It is tempting to propose that with increasing observations of gene fusions in solid cancers, some common gene fusion(s) might be found to be represented across different tumor types, potentially providing a broad-based therapeutic target. Indeed, gene fusions may provide a common mechanism for carcinogenesis.

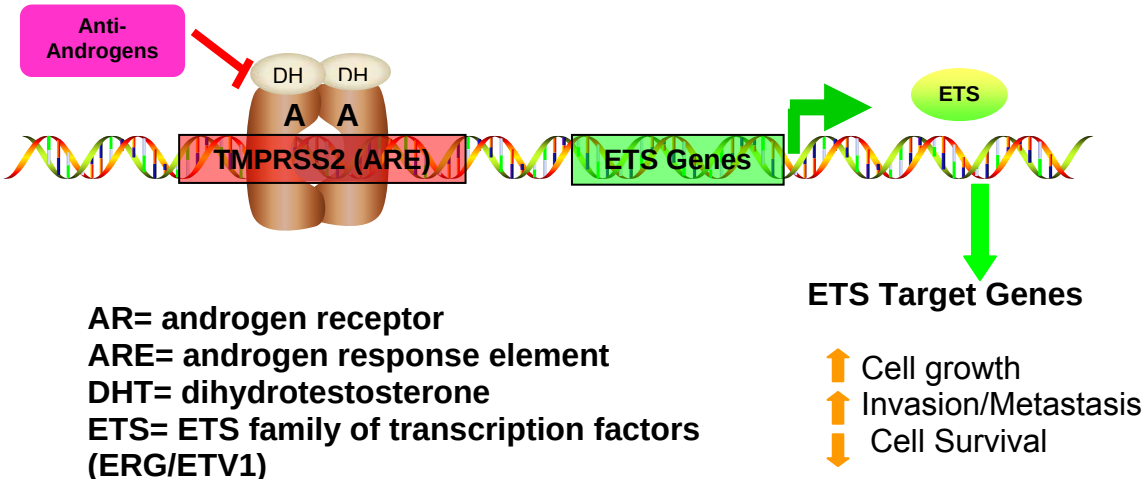

Figure 7.1: A molecular basis for prostate cancer. The androgen responsive upstream regulatory sequences of TMPRSS2 driving ETS family of transcription factors could be representing a common theme for prostate carcinogenesis. A molecular analysis of this nexus might lead to a better understanding of the role of androgens in prostate cancers.

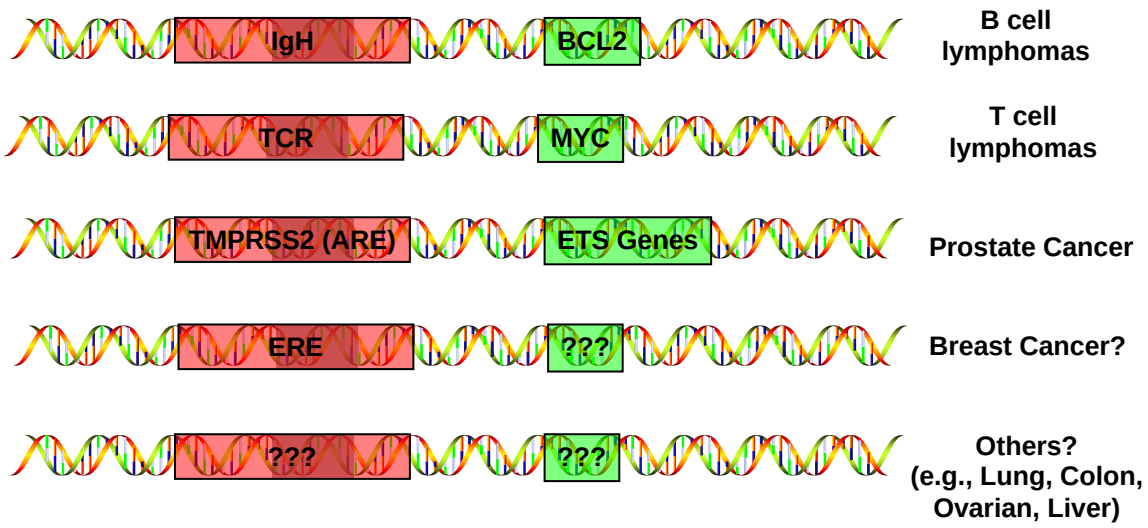

Figure 2: Subversion of tissue-specific promoters/enhancers to cause cancer. It is tempting to hypothesize that a genomic rearrangement causing a tissue specific regulatory element to usurp the control of an oncogene may be a common theme in tumorigenesis. A focused search for combinations of such tissue specific regulators and their oncogene counterparts, involved in common epithelial cancers would provide both, insights into the molecular basis of tumorigenesis as well as novel therapeutic targets.

REFERENCES

1. Futreal PA, Coin L, Marshall M, et al. A census of human cancer genes. Nat Rev Cancer 2004;*4:*177-83.
2. Rabbitts TH. Chromosomal translocations in human cancer. Nature 1994;*372:*143-9.
3. Rowley JD. Chromosome translocations: dangerous liaisons revisited. Nat Rev Cancer 2001;*1:*245-50.
4. Baselga J, Tripathy D, Mendelsohn J, et al. Phase II study of weekly intravenous recombinant humanized anti-p185HER2 monoclonal antibody in patients with HER2/neu-overexpressing metastatic breast cancer. J Clin Oncol 1996;*14:*737-44.
5. Druker BJ, Sawyers CL, Kantarjian H, et al. Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome. N Engl J Med 2001;*344:*1038-42.
6. Druker BJ, Talpaz M, Resta DJ, et al. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. N Engl J Med 2001;*344:*1031-7.
7. Tomlins SA, Rhodes DR, Perner S, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 2005;*310:*644-8.
8. Tomlins SA, Laxman B, Dhanasekaran SM, et al. Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. Nature 2007;*448:*595-9.
9. Tomlins SA, Mehra R, Rhodes DR, et al. TMPRSS2:ETV4 gene fusions define a third molecular subtype of prostate cancer. Cancer Res 2006;*66:*3396-400.
10. Nowell PCHungerford DA. A minute chromosome in human chronic granulocytic leukemia. Science 1960;*132:*1497-501.
11. Nowell PCHungerford DA. Chromosome studies in human leukemia. II. Chronic granulocytic leukemia. J Natl Cancer Inst 1961;*27:*1013-35.
12. Caspersson T, Zech L, Johansson C. Differential binding of alkylating fluorochromes in human chromosomes. Exp Cell Res 1970;*60:*315-9.
13. Rowley JD. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. Nature 1973;*243:*290-3.
14. Lugo TG, Pendergast AM, Muller AJ, Witte ON. Tyrosine kinase activity and transformation potency of bcr-abl oncogene products. Science 1990;*247:*1079-82.
15. Daley GQ, Van Etten RA, Baltimore D. Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome. Science 1990;*247:*824-30.
16. Rowley JD. Identificaton of a translocation with quinacrine fluorescence in a patient with acute leukemia. Ann Genet 1973;*16:*109-12.
17. Hiebert SW, Downing JR, Lenny N, Meyers S. Transcriptional regulation by the t(8;21) fusion protein, AML-1/ETO. Curr Top Microbiol Immunol 1996;*211:*253-8.

18. Frank R, Zhang J, Uchida H, Meyers S, Hiebert SW, Nimer SD. The AML1/ETO fusion protein blocks transactivation of the GM-CSF promoter by AML1B. Oncogene 1995;*11:*2667-74.
19. Wang J, Hoshino T, Redner RL, Kajigaya S, Liu JM. ETO, fusion partner in t(8;21) acute myeloid leukemia, represses transcription by interaction with the human N-CoR/mSin3/HDAC1 complex. Proc Natl Acad Sci U S A 1998;*95:*10860-5.
20. Goddard AD, Borrow J, Freemont PS, Solomon E. Characterization of a zinc finger gene disrupted by the t(15;17) in acute promyelocytic leukemia. Science 1991;*254:*1371-4.
21. Salomoni PPandolfi PP. The role of PML in tumor suppression. Cell 2002;*108:*165-70.
22. Grignani F, Ferrucci PF, Testa U, et al. The acute promyelocytic leukemia-specific PML-RAR alpha fusion protein inhibits differentiation and promotes survival of myeloid precursor cells. Cell 1993;*74:*423-31.
23. Grignani F, De Matteis S, Nervi C, et al. Fusion proteins of the retinoic acid receptor-alpha recruit histone deacetylase in promyelocytic leukaemia. Nature 1998;*391:*815-8.
24. Lin RJ, Nagy L, Inoue S, Shao W, Miller WH, Jr., Evans RM. Role of the histone deacetylase complex in acute promyelocytic leukaemia. Nature 1998;*391:*811-4.
25. Cimino G, Rapanotti MC, Sprovieri T, Elia L. ALL1 gene alterations in acute leukemia: biological and clinical aspects. Haematologica 1998;*83:*350-7.
26. Downing JRLook AT. MLL fusion genes in the 11q23 acute leukemias. Cancer Treat Res 1996;*84:*73-92.
27. Redner RL, Wang J, Liu JM. Chromatin remodeling and leukemia: new therapeutic paradigms. Blood 1999;*94:*417-28.
28. Claxton DF, Liu P, Hsu HB, et al. Detection of fusion transcripts generated by the inversion 16 chromosome in acute myelogenous leukemia. Blood 1994;*83:*1750-6.
29. Cremer T, Lichter P, Borden J, Ward DC, Manuelidis L. Detection of chromosome aberrations in metaphase and interphase tumor cells by in situ hybridization using chromosome-specific library probes. Hum Genet 1988;*80:*235-46.
30. Gray JWPinkel D. Molecular cytogenetics in human cancer diagnosis. Cancer 1992;*69:*1536-42.
31. Schrock E, du Manoir S, Veldman T, et al. Multicolor spectral karyotyping of human chromosomes. Science 1996;*273:*494-7.
32. Speicher MR, Gwyn Ballard S, Ward DC. Karyotyping human chromosomes by combinatorial multi-fluor FISH. Nat Genet 1996;*12:*368-75.
33. Mitelman, Johansson B, F M. Mitelman Database of Chromosome Aberrations in Cancer. 2005.
34. Mitelman F. Recurrent chromosome aberrations in cancer. Mutat Res 2000;*462:*247-53.
35. Mitelman F, Johansson B, Mertens F. Fusion genes and rearranged genes as a linear function of chromosome aberrations in cancer. Nat Genet 2004;*36:*331-4.

36. Mitelman F. Cancer cytogenetics update 2005. Atlas Genet Cytogenet Oncol Haematol. March 2005 2005.
37. Knezevich SR, McFadden DE, Tao W, Lim JF, Sorensen PH. A novel ETV6-NTRK3 gene fusion in congenital fibrosarcoma. Nat Genet 1998;*18:*184-7.
38. Tognon C, Knezevich SR, Huntsman D, et al. Expression of the ETV6-NTRK3 gene fusion as a primary event in human secretory breast carcinoma. Cancer Cell 2002;*2:*367-76.
39. Eguchi M, Eguchi-Ishimae M, Tojo A, et al. Fusion of ETV6 to neurotrophin-3 receptor TRKC in acute myeloid leukemia with t(12;15)(p13;q25). Blood 1999;*93:*1355-63.
40. Rubin BP, Chen CJ, Morgan TW, et al. Congenital mesoblastic nephroma t(12;15) is associated with ETV6-NTRK3 gene fusion: cytogenetic and molecular relationship to congenital (infantile) fibrosarcoma. Am J Pathol 1998;*153:*1451-8.
41. Tognon C, Garnett M, Kenward E, Kay R, Morrison K, Sorensen PH. The chimeric protein tyrosine kinase ETV6-NTRK3 requires both Ras-Erk1/2 and PI3-kinase-Akt signaling for fibroblast transformation. Cancer Res 2001;*61:*8909-16.
42. Wai DH, Knezevich SR, Lucas T, Jansen B, Kay RJ, Sorensen PH. The ETV6-NTRK3 gene fusion encodes a chimeric protein tyrosine kinase that transforms NIH3T3 cells. Oncogene 2000;*19:*906-15.
43. Pierotti MA. Chromosomal rearrangements in thyroid carcinomas: a recombination or death dilemma. Cancer Lett 2001;*166:*1-7.
44. Nikiforov YE. RET/PTC rearrangement in thyroid tumors. Endocr Pathol 2002;*13:*3-16.
45. Bongarzone I, Vigneri P, Mariani L, Collini P, Pilotti S, Pierotti MA. RET/NTRK1 rearrangements in thyroid gland tumors of the papillary carcinoma family: correlation with clinicopathological features. Clin Cancer Res 1998;*4:*223-8.
46. Grieco M, Santoro M, Berlingieri MT, et al. PTC is a novel rearranged form of the ret proto-oncogene and is frequently detected in vivo in human thyroid papillary carcinomas. Cell 1990;*60:*557-63.
47. Pierotti MA, Santoro M, Jenkins RB, et al. Characterization of an inversion on the long arm of chromosome 10 juxtaposing D10S170 and RET and creating the oncogenic sequence RET/PTC. Proc Natl Acad Sci U S A 1992;*89:*1616-20.
48. Sozzi G, Bongarzone I, Miozzo M, et al. A t(10;17) translocation creates the RET/PTC2 chimeric transforming sequence in papillary thyroid carcinoma. Genes Chromosomes Cancer 1994;*9:*244-50.
49. Fugazzola L, Pierotti MA, Vigano E, Pacini F, Vorontsova TV, Bongarzone I. Molecular and biochemical analysis of RET/PTC4, a novel oncogenic rearrangement between RET and ELE1 genes, in a post-Chernobyl papillary thyroid cancer. Oncogene 1996;*13:*1093-7.
50. Santoro M, Dathan NA, Berlingieri MT, et al. Molecular characterization of RET/PTC3; a novel rearranged version of the RETproto-oncogene in a human thyroid papillary carcinoma. Oncogene 1994;*9:*509-16.

51. Klugbauer S, Demidchik EP, Lengfelder E, Rabes HM. Detection of a novel type of RET rearrangement (PTC5) in thyroid carcinomas after Chernobyl and analysis of the involved RET-fused gene RFG5. Cancer Res 1998;*58:*198-203.
52. Greco A, Miranda C, Pagliardini S, Fusetti L, Bongarzone I, Pierotti MA. Chromosome 1 rearrangements involving the genes TPR and NTRK1 produce structurally different thyroid-specific TRK oncogenes. Genes Chromosomes Cancer 1997;*19:*112-23.
53. Greco A, Pierotti MA, Bongarzone I, Pagliardini S, Lanzi C, Della Porta G. TRK-T1 is a novel oncogene formed by the fusion of TPR and TRK genes in human papillary thyroid carcinomas. Oncogene 1992;*7:*237-42.
54. Greco A, Mariani C, Miranda C, et al. The DNA rearrangement that generates the TRK-T3 oncogene involves a novel gene on chromosome 3 whose product has a potential coiled-coil domain. Mol Cell Biol 1995;*15:*6118-27.
55. Greco A, Fusetti L, Miranda C, et al. Role of the TFG N-terminus and coiled-coil domain in the transforming activity of the thyroid TRK-T3 oncogene. Oncogene 1998;*16:*809-16.
56. Borrello MG, Alberti L, Arighi E, et al. The full oncogenic activity of Ret/ptc2 depends on tyrosine 539, a docking site for phospholipase Cgamma. Mol Cell Biol 1996;*16:*2151-63.
57. McIver B, Grebe SK, Eberhardt NL. The PAX8/PPAR gamma fusion oncogene as a potential therapeutic target in follicular thyroid carcinoma. Curr Drug Targets Immune Endocr Metabol Disord 2004;*4:*221-34.
58. Giordano TJ, Au AY, Kuick R, et al. Delineation, Functional Validation, and Bioinformatic Evaluation of Gene Expression in Thyroid Follicular Carcinomas with the PAX8-PPARG Translocation. Clin Cancer Res 2006;*12:*1983-93.
59. Meloni AM, Dobbs RM, Pontes JE, Sandberg AA. Translocation (X;1) in papillary renal cell carcinoma. A new cytogenetic subtype. Cancer Genet Cytogenet 1993;*65:*1-6.
60. Sidhar SK, Clark J, Gill S, et al. The t(X;1)(p11.2;q21.2) translocation in papillary renal cell carcinoma fuses a novel gene PRCC to the TFE3 transcription factor gene. Hum Mol Genet 1996;*5:*1333-8.
61. Weterman MA, Wilbrink M, Geurts van Kessel A. Fusion of the transcription factor TFE3 gene to a novel gene, PRCC, in t(X;1)(p11;q21)-positive papillary renal cell carcinomas. Proc Natl Acad Sci U S A 1996;*93:*15294-8.
62. Stenman G. Fusion oncogenes and tumor type specificity--insights from salivary gland tumors. Semin Cancer Biol 2005;*15:*224-35.
63. Engleson J, Soller M, Panagopoulos I, Dahlen A, Dictor M, Jerkeman M. Midline carcinoma with t(15;19) and BRD4-NUT fusion oncogene in a 30-year-old female with response to docetaxel and radiotherapy. BMC Cancer 2006;*6:*69.
64. French CA, Miyoshi I, Kubonishi I, Grier HE, Perez-Atayde AR, Fletcher JA. BRD4-NUT fusion oncogene: a novel mechanism in aggressive carcinoma. Cancer Res 2003;*63:*304-7.
65. Mertens F, Wiebe T, Adlercreutz C, Mandahl N, French CA. Successful treatment of a child with t(15;19)-positive tumor. Pediatr Blood Cancer 2006.
66. French CA, Kutok JL, Faquin WC, et al. Midline carcinoma of children and young adults with NUT rearrangement. J Clin Oncol 2004;*22:*4135-9.

67. Haruki N, Kawaguchi KS, Eichenberger S, et al. Cloned fusion product from a rare t(15;19)(q13.2;p13.1) inhibit S phase in vitro. J Med Genet 2005;*42:*558-64.
68. Gorunova L, Hoglund M, Andren-Sandberg A, et al. Cytogenetic analysis of pancreatic carcinomas: intratumor heterogeneity and nonrandom pattern of chromosome aberrations. Genes Chromosomes Cancer 1998;*23:*81-99.
69. Johansson B, Mertens F, Mitelman F. Primary vs. secondary neoplasia-associated chromosomal abnormalities--balanced rearrangements vs. genomic imbalances? Genes Chromosomes Cancer 1996;*16:*155-63.
70. Fine BM, Stanulla M, Schrappe M, et al. Gene expression patterns associated with recurrent chromosomal translocations in acute lymphoblastic leukemia. Blood 2004;*103:*1043-9.
71. Kauraniemi P, Barlund M, Monni O, Kallioniemi A. New amplified and highly expressed genes discovered in the ERBB2 amplicon in breast cancer by cDNA microarrays. Cancer Res 2001;*61:*8235-40.
72. Rhodes DR, Yu J, Shanker K, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia 2004;*6:*1-6.
73. Oikawa TYamada T. Molecular biology of the Ets family of transcription factors. Gene 2003;*303:*11-34.
74. Afar DE, Vivanco I, Hubert RS, et al. Catalytic cleavage of the androgen-regulated TMPRSS2 protease results in its secretion by prostate and prostate cancer epithelia. Cancer Res 2001;*61:*1686-92.
75. Jacquinet E, Rao NV, Rao GV, Zhengming W, Albertine KH, Hoidal JR. Cloning and characterization of the cDNA and gene for human epitheliasin. Eur J Biochem 2001;*268:*2687-99.
76. Lin B, Ferguson C, White JT, et al. Prostate-localized and androgen-regulated expression of the membrane-bound serine protease TMPRSS2. Cancer Res 1999;*59:*4180-4.
77. Soda M, Choi YL, Enomoto M, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature 2007;*448:*561-6.
78. Kroll TG. Molecular events in follicular thyroid tumors. Cancer Treat Res 2004;*122:*85-105.

APPENDIX

Supplementary Information for the individual chapters is available online at the following addresses:

CHAPTER 2
http://www.sciencemag.org/cgi/content/full/310/5748/644/DC1

CHAPTER 4
http://www.nature.com/ng/journal/v39/n1/suppinfo/ng1935_S1.html

CHAPTER 5:
http://www.nature.com/nature/journal/v448/n7153/suppinfo/nature06024.html

Multiple individuals contributed to the work presented in these chapters and resulting manuscripts. Contributions of individuals for each chapter are as follows:

CHAPTER 2

Scott Tomlins, Daniel Rhodes and Arul Chinnaiyan conceived the experiments and wrote the manuscript represented in this chapter. Scott Tomlins, with technical assistance from Xuhong Cao, isolated RNA and performed expression profiling from LCM tissue samples, performed all qPCR on prostate cancer tissue samples and cell lines, performed RACE, qPCR and RT-PCR confirmation of gene fusions, and performed qPCR on androgen treated cells.

Daniel Rhodes and Scott Tomlins developed the COPA algorithm. James Montie, Rainer Kuefer, Rajal Shah, and Rohit Mehra obtained and evaluated prostate tissues. Rohit Mehra isolated tissue samples by LCM for expression profiling. Saravana

Dhanasekaran and Sooryanarayana Varambally isolated RNA from prostate cancer tissue samples and cell lines, and performed androgen stimulation on prostate cancer cell lines. Saravana Dhanasekaran performed Southern blotting. Sven Perner, Xiao-Wei Sun, Joelle Tchinda, Charles Lee and Mark Rubin, generated FISH probes and performed hybridizations.

CHAPTER 3

Scott Tomlins, Daniel Rhodes and Arul Chinnaiyan conceived the experiments and Scott Tomlins and Arul Chinnaiyan wrote the manuscript represented in this chapter. Scott Tomlins, with technical assistance from Beth Helgeson and Xuhong Cao, performed qPCR for ETV4 expression, performed RLM-RACE, designed and confirmed the localization of FISH probes, and performed confirmatory qPCR.

Mark Rubin, John Wei and Rajal Shah obtained and/or evaluated prostate tissue samples. Lisa Smith and Diane Roulston helped optimize FISH. Rohit Mehra performed and evaluated interphase FISH. Daniel Rhodes and Scott Tomlins performed the in silico analysis.

CHAPTER 4

Scott Tomlins and Arul Chinnaiyan conceived the experiments and wrote the manuscript represented in this chapter. Scott Tomlins, with technical assistance from Xuhong Cao, isolated RNA from LCM samples, amplified the RNA, performed expression profiling and the Oncomine and Molecular Concepts Map analyses, and performed confirmatory qPCR.

Rajal Shah and Rohit Mehra evaluated prostate tissue samples. Rohit Mehra, with technical assistance from Lei Wang, performed LCM on prostate tissue samples and sectioned tissue samples. Saravana Dhanasekarn isolated RNA from confirmatory prostate tissue samples. Ken Pienta, John Wei and Mark Rubin obtained prostate tissue samples. Rohit Mehra evaluated immunohistochemistry on tissue microarrays. Daniel Rhodes and Shanker Kalyana-Sundaram developed the Molecular Concepts Map and myOncomine.

CHAPTER 5

Scott Tomlins and Arul Chinnaiyan conceived the experiments and wrote the manuscript represented in this chapter. Scott Tomlins, with technical assistance from Beth Helgeson, Xuhong Cao, David Morris, Anjana Menon, isolated RNA and performed qPCR for ETV1 expression from prostate tissues and cell lines, performed RLM-RACE, qPCR validation of ETV1 fusions and in silico characterization of fusion partners, generated and confirmed the localization of FISH probes, performed metaphase FISH on LNCaP and MDA-PCa 2B cells, generated probes for Southern blotting, performed microarray profiling of LNCaP and C4-2B cell lines, generated constructs for adenovirus and lentivirus generation, profiled RWPE-ETV1 cell lines and performed the Oncomine concepts analysis, and generated the transgenic mouse construct. The University of Michigan Vector Core generated the adenoviruses and lentiviruses. The University of Michigan Transgenic Model Core generatred the transgenic mice.

Bharathi Laxman maintained, genotyped and dissected transgenic mice. Bharathi Laxman infected RWPE and PrEC cells with adenovirus and lentiviruses and generated stable cell lines. Saravana Dhanasekaran performed Southern blotting and inverse PCR to identify the LNCaP ETV1 breakpoint. Qi Cao performed invasion assays. Bo Han, Lucy Wang and Rohit Mehra performed interphase FISH hybridizations. James Montie and Ken Pienta obtained prostate cancer tissue samples. Mark Rubin, Rohit Mehra and Rajal Shah evaluated transgenic mice for prostate pathology. Diane Roulston provided interpretation for cytogenetic studies and generated LNCaP metaphase preparations. Jindan Yu performed CHIP on CHIP. Sooryanarayana Varambally performed siRNA, shRNA, inhibitor and androgen stimulation studies in LNCaP and RWPE cells.

CHAPTER 6

Scott Tomlins and Arul Chinnaiyan conceived the experiments and wrote the manuscript represented in this chapter. Scott Tomlins, with technical assistance from Xuhong Cao, Beth Helgeson and John Prensner, generated the transgenic mouse, adenovirus and lentivirus constructs, performed qPCR and expression profiling and performed the Molecular Concepts Map analyses.

Mark Rubin, Rajal Shah and Rohit Mehra evaluated transgenic mice pathology and performed and evaluated immunohistochemistry. Jindan Yu performed CHIP analyses. Bharathi Laxman maintained, genotyped and dissected the transgenic mice, infected RWPE and PrEC cells with adenoviruses and lentiviruses and generated stable clones, and performed proliferation and soft agar assays. Sooryanarayana Varambally performed siRNA and inhibitor studies, and performed Western blotting. Qi Cao performed invasion assays.