Title: Deregulation of a Hox Protein Regulatory Network Spanning Prostate Cancer Initiation and Progression

Authors: James L. Chen1, Jianrong Li2, Kyle J. Kiriluk4, Alex M. Rosen4, Gladell P. Paner3, Tatjana Antic3, Yves A. Lussier2, and Donald J. Vander Griend4

1 Department of Medicine, Section of Hematology/Oncology; The University of Chicago, Chicago, IL.
2 Departments of Medicine and Bioengineering; University of Illinois at Chicago, Chicago, IL.
3 Department of Pathology; The University of Chicago, Chicago, IL.
4 Department of Surgery, Section of Urology; The University of Chicago, Chicago, IL.

Please address all correspondence to: Donald J. Vander Griend, Ph.D., The University of Chicago Department of Surgery, The Section of Urology, 5841 S. Maryland Ave, MC6038, Chicago, IL 60637; prostate@uchicago.edu

Running Title: A novel Hox regulator biomodule for prostate cancer

Key Words: Meis1, Meis2, Pbx1, HoxA9, prostate cancer, seminal vesicle, protein interaction networks, bioinformatics

Abbreviations: ADT (Androgen Deprivation Therapy); AR (Androgen Receptor); CRPC (Castration-Resistant Prostate Cancer); GEO (Gene Expression Omnibus); GO (Gene Ontology); HOX (homeobox); Meis (Myeloid Ecotropic Viral Integration Site); PAM (Prediction Analysis for Microarrays); PBX1 (Pre-B-Cell Leukemia Transcription Factor 1); SPAN (Single Protein Analysis of Network); TALE (Three Amino Acid Loop Extension).

Funding: The University of Chicago Department of Surgery, the Section of Urology; the University of Chicago Comprehensive Cancer Center (UCCCC); a Career Development Award from the NCI P50 CA090386 SPORE in Prostate Cancer at the Robert H. Lurie Comprehensive Cancer Center of Northwestern University and the Cancer Research Center of the University of Chicago; The Brinson Foundation; the Alvin Baum Family Fund; The University of Chicago Cancer Research Foundation Women's Board; The University of Chicago CTSA/ITM Pilot for Collaborative Translational and Clinical Studies Awards UL1 RR024999.
ABSTRACT

Purpose: The aberrant activity of developmental pathways in prostate cancer may provide significant insight into predicting tumor initiation and progression, as well as identifying novel therapeutic targets. To this end, despite shared androgen-dependence and functional similarities to the prostate gland, seminal vesicle cancer is exceptionally rare.

Experimental Design: We conducted genomic pathway analyses comparing patient-matched normal prostate and seminal vesicle epithelial cells to identify novel pathways for tumor initiation and progression. Derived gene expression profiles were grouped into cancer biomodules using a protein-protein network algorithm to analyze their relationship to known oncogenes. Each resultant biomodule was assayed for its prognostic ability against publically available prostate cancer patient gene array datasets.

Results: Analyses demonstrate that the embryonic developmental biomodule containing four homeobox gene family members (Meis1, Meis2, Pbx1, and HoxA9) detects a survival difference in a set of watchful-waiting patients (n = 172, p = 0.05), identify men who are more likely to recur biochemically post-prostatectomy (n = 78, p= 0.02), correlate with Gleason score (r = 0.98, p = 0.02), and distinguish between normal prostate, primary tumor, and metastatic disease. In contrast to other cancer types, Meis1, Meis2, and Pbx1 expression is decreased in poor-prognosis tumors, implying that they function as tumor suppressor genes for prostate cancer. Immunohistochemical staining documents nuclear basal-epithelial and stromal Meis2 staining, with loss of Meis2 expression in prostate tumors.

Conclusion: These data implicate deregulation of the Hox protein cofactors Meis1, Meis2, and Pbx1 as serving a critical function to suppress prostate cancer initiation and progression.

TRANSLATIONAL RELEVANCE

While prostate cancer is the most common malignancy in men, its neighboring sex accessory organ the seminal vesicle is very rarely the site of malignant tumor formation. We exploited this profound difference to identify novel cellular pathways that may be important in prostate tumor initiation and progression, and then tested these pathways against publically available datasets for their expression in human prostate tissues. Our studies identified a set of Hox protein co-factors which are known to regulate and modulate Hox protein transcriptional specificity and are significantly down-regulated at various stages of prostate tumor initiation and progression. Thus, these Hox gene co-regulators appear to function as essential tumor-suppressor proteins in prostate cancer.
INTRODUCTION

The prostate gland is the most predominant site of neoplasia in aging men. In the United States alone in 2011, approximately 240,890 men were estimated to be diagnosed with prostate cancer (1). Less than 14% (33,720) of men, however, actually died of the disease (1). Clearly the advent of more aggressive screening programs has reduced mortality from prostate cancer, but has concomitantly detected tumors that often have no impact on overall patient morbidity or survival. Our understanding of prostate cancer treatment traces back to the seminal discovery by Charles Huggins and Clarence Hodges in 1941 that castration significantly benefits patients with advanced prostate cancer (2). Researchers have since identified the critical role deregulations in the androgen receptor (AR) pathway play in both oncogenesis and eventual resistance to castration. Although there are newer and more effective strategies targeting the androgen receptor, the majority of men continue to have disease progression. Thus, it is critical to first identify novel molecular pathways which are deregulated within prostate tumors that will permit discrimination of low-grade tumors at risk of progression. Secondly, pharmacologic targeting of these pathways in prostate cancer may provide disease control in the advanced, metastatic setting.

As the AR pathway is developmentally critical, we hypothesized that other developmental pathways may also play a key role in oncogenesis and tumor progression. To provide us insight into what pathways may be deregulated, we took advantage of the unique similarities and differences between seminal vesicle and prostate tissues. In contrast to over 2 million U.S. cases of prostate cancer diagnosed in the past decade U.S., there are fewer than 50 cases of primary seminal vesicle cancer in the English literature (3, 4). Similar to the prostate, the development and maintenance of the seminal vesicles is androgen-dependent, and the seminal vesicles synthesize and secrete molecules and proteins that contribute to semen (5). This implies that there are unique differences between prostate and seminal vesicle tissues which strongly pre-dispose prostate tissue to neoplasia and cancer. These differences could be accounted for by their specific developmental differences and their distinct embryologic origins, as well as exposure to infectious and inflammatory agents, distinct secreted factors produced by their gland of origin, or differences in the kinetics of tissue turnover (6).

Here we report the identification of a developmental transcription factor biomodule that is unique to prostate epithelial cells when compared to patient-matched seminal vesicle epithelial cells. A “biomodule” is defined as a subnetwork of genes connected by their protein-protein interactions. The biomodule described here consists of the three-amino acid loop extension (TALE) and homeodomain transcription factors Meis1, Meis2, and Pbx1, and the homeodomain-
containing DNA-binding protein HoxA9. Meis1, Meis1, and Pbx1 are critical Hox gene co-
factors which modify Hox gene specificity (7), and have critical roles during development (7-9).
In addition, Meis proteins and Pbx1 have a well-defined role in promoting certain leukemias (10-
14). The clinical relevance of this biomodule in prostate cancer was tested against multiple, comprehensive, and independent publically available prostate tissue databases. Aberrant expression of these four genes was significant in normal prostate as compared to tumor, predictive for both biochemical relapse (i.e. PSA recurrence) and overall patient survival.

MATERIALS AND METHODS
Tissue preparation and Cell Culture
Fresh prostate and seminal vesicle cells were isolated from patients undergoing radical prostatectomy at our institution according to an Institution Review Board-approved protocol. Dissociation of prostate tissue and growth of epithelial cells has been previously described (15, 16). Briefly, 4mm biopsy punches were taken from prostate and seminal vesicle tissue; half of this tissue was fixed and analyzed by a pathologist to confirm the absence of tumor. The remaining portion was digested overnight at 37°C in collagenase solution [0.28% collagenase 1 (Sigma-Aldridge, St. Louis, MO), 1% DNAse I (Sigma), 10% FCS, 1x antibiotic/antimycotic (Life Technologies-Invitrogen), in RPMI-1640]. The following day, the cell suspension was washed in PBS, and epithelial cell clusters were isolated by density sedimentation, whereby cells in 10mL PBS were allowed to settle for 10 minutes at room temperature and the top 9ml of medium (containing fibroblasts) were removed; this was repeated two more times. Epithelial cell clusters were further dissociated into single cells via treatment with DTT (1mmol/L for 30 minutes at 37°C), a PBS wash, and trypsin/EDTA (0.25% for 30 minutes at 37°C). The trypsin was neutralized with RMPI 1640 plus 10% FCS, and the cells were washed twice in PBS. Prostate epithelial cell cultures (PrECs) and seminal vesicle epithelial cell cultures (SVECs) were established and grown using Keratinocyte Serum Free defined media supplemented with growth factors (GFS) (standard K-SFM, Invitrogen Life Technologies). All cultures were screened for the absence of mycoplasma contamination using the ATCC Universal Mycoplasma Detection Kit (Manassas, VA).

RNA Isolation and Gene Array Analyses
Cells were plated at a density of 50,000 cells per cm², and after 48 hours mRNA was isolated using the Qiagen RNeasy Mini kit (Valencia, CA) per the manufacturers’ instructions. All RNA used for analyses had an RNA integrity number (RIN) of ten when measured using an
Agilent 2100 Bioanalyzer (Santa Clara, CA). Double-stranded cDNA and cRNA synthesis was conducted according to the GeneChip Expression Analyses Technical Manual (Rev5), and samples hybridized to Affymetrix HG 133+2.0 Arrays (Santa Clara, CA; Lot #4087104 for all chips).

**Analysis of epithelial tissue gene expression data**

Affymetrix expression data was analyzed using BRB-ArrayTools developed by Dr. Richard Simon and the BRB-ArrayTools Development Team (http://linus.nci.nih.gov/BRB-ArrayTools.html) (17). We performed standard RMA normalization. Probes were filtered based on the requirement of a two-fold change in the dataset and no less than 50% missing data. Paired analysis using Significance Analysis of Microarrays (SAM) (18) was performed with a false discovery rate (FDR) of 0.05 over the dataset to generate an epithelial-specific gene signature. For purposes of this article, we define gene signature as any collection of genes, be they a collection of differentially expressed genes or a set of refined genes. We derived the final gene expression signature by converting the probes to their unique corresponding gene names.

**Historical datasets used in the analyses**

Six historical prostate cancer datasets were used in the analyses: Thompson (GSE11376 (19)), Cai (GSE32269 (20)); Yu (GDS2545 (21)), Sboner (GSE16560) (22)) Taylor (GSE21032 (23)) and Glinsky (24). The first five were downloaded from the Gene Expression Omnibus (GEO) database on the NCBI and the Glinsky set was obtained from the author. The Thompson, Cai, and Yu datasets were used as points of comparison for their sets of differentially expressed genes. The Sboner and Glinsky datasets were used for Kaplan-Meier analysis regarding prognostic survival and recurrence capabilities. The Taylor dataset was used for clustering analysis and biomodules validation. **Supplemental Table 1** provides a detailed description of these datasets and their functions.

**Differential gene signature generation from historical datasets**

GSE32269, GDS2545 and GSE11376 were loaded into BRB-ArrayTools. A two-fold change filter was implemented. Paired SAM analysis was performed on the GSE11376, and unpaired analysis on the GSE33269 and GDS2545 dataset samples. A false discovery rate (FDR) of 0.05 was used to control for multiplicity. Resulting genes in each signature were analyzed for their fold-change in the comparison conditions.
Protein-protein interaction network analysis and pathway enrichment

To evaluate whether distinct derived gene signatures connected to cancer-related genes greater than expected by chance, we used a Single Protein Analysis of Network (SPAN) methodology (25-27). As our reference cancer gene gold standard, we used the Wellcome Trust Sanger Institute Cancer Gene Census downloaded on August 5th, 2011 from http://www.sanger.ac.uk/genetics/CGP/Census/ (28). The Cancer Gene Census contains a catalog of genes for which mutations have been causally implicated in cancer, acquired and updated through literature-based methods. SPAN analysis has been extensively previously described (26, 27). In brief, each gene expression signature was compared with the Sanger Cancer Gene set using SPAN. We compared the expected distribution of linkages between the gene expression signatures and the Sanger cancer genes through 10,000 permutation re-samplings. The unadjusted p-value of each signature gene’s connectivity was further adjusted for multiplicity using a Bonferroni correction (29).

The SPAN method prioritizes genes at the protein level. We first evaluated whether there were cancer genes more connected to each gene signature than expected by conservative empirical controls. A converse calculation was then performed where each single Sanger cancer gene was analyzed for its total number of interactions with each independent, unique gene in the gene signatures. Prioritized genes and their interactors that had a FDR<0.05 were retained. The resulting statistically significant genes were then retained and deemed a “biomodule”. As each SPAN protein keeps an equal number of partners in the empirical distribution the network is conservatively prioritized. Visualization of the SPAN prioritized biomodule was then performed using Cytoscape software (30). Biomodules were then enriched for Gene Ontology pathways using the Database for Annotation, Visualization and Integrated Discovery (DAVID) tool v6.7 (31).

Evaluation of clinical datasets

To test the clinical relevance of the cancer gene centered biomodules, we first examined the ability of individual biomodules to find a survival or recurrence difference in three independent retrospective datasets. This survival analysis using a separate dataset serves as a clinical evaluation of the biomodule signature. In the first dataset (Glinsky), we examined the disease-free interval after prostatectomy of 79 samples (24). This dataset contained 37 patients with recurrent and 42 patients with non-recurrent disease. After loading the dataset into R/Bioconductor, we used the Prediction Analysis for Microarrays (32) (PAM) algorithm for each biomodule gene list to provide an unbiased division of the prostate cancer cohort into two
groups. Kaplan-Meier analysis was performed using time from diagnosis until recurrence. We then validated the Glinsky results in a second dataset (Taylor) that contained 131 prostatectomy specimens with longitudinal data as to their PSA recurrence (27 recurrent, 104 non-recurrent). We performed a PAM analysis similar to the Glinsky procedure outlined above. The third dataset (Sboner) contained 281 Swedish men who underwent a course of 'watchful waiting' after being diagnosed with prostate cancer (GSE16560) (22). This set of 281 only included patients who were alive or had died from prostate cancer-specific causes. In addition to the PAM algorithm described previously, we also dichotomized the patient cohort into those who were above or below the average mRNA expression for each gene or gene set of interest. Kaplan-Meier analysis was performed using time from diagnosis until death. In a subsequent analysis, we examined only the Gleason Score 6 and 7 cohorts (n = 83, n = 117, respectively).

In a second analysis, we explored genes from relevant biomodules as possible independent predictors of recurrence or survival using Cox regression analysis. We again examined the datasets detailed above. Loading the datasets into R/Bioconductor, we explored biomodule gene expression levels first in an univariate analysis and then adjusted for Gleason score and age.

**Biomodule classification ability in prostate cancer**

To graphically evaluate the ability of a biomodule to classify prostate cancer, we computed the score of the used a large independent dataset of 185 gene expression profiles including normal tissue, primary prostate tissue, metastatic tumor specimens, and immortalized prostate cancer cell lines (23). A composite biomodule score was generated via the summation of the up-regulated genes and subtracting sum of the scores of the down-regulated genes. The unadjusted score for each sample was compared to the average score across the dataset to produce a percentage change from mean. Samples were ordered by percentage change from mean and graphed using Excel.

**Immunohistochemistry**

Immunostaining for Meis2 (63-T, Santa Cruz Biotechnology; mouse monoclonal) was performed on formalin-fixed, paraffin-embedded (FFPE) sections managed either by the University of Chicago Human Tissue Resource Core facility or the Northwestern University/University of Chicago Prostate SPORE and their Specimen Procurement Program. Tissue samples were de-identified and there was no exclusion criteria applied toward sample selection. After deparaffinization and rehydration, tissues were treated with antigen retrieval
buffer (S1699 from DAKO; Glostrup, Denmark) in a steamer for 20 minutes. Anti-MEIS2 antibody (1:50 dilution) was applied on tissue sections for 1 hour at room temperature in a humidity chamber. Following TBS wash, the antigen-antibody binding was detected with Envision + system (DAKO, K4001 for mouse primary antibodies) and DAB + Chromogen (DAKO, K3468). Tissue sections were briefly immersed in hematoxylin for counterstaining and were cover-slipped. Tissues were analyzed by a trained Genitourinary Pathologist and scored on percentage of cells with positive nuclear staining (0 = no staining; 1 = 1-10% positive cells; 2 = 11-50% positive cells; and 3 = >50% positive cells); as well as the intensity of staining (0 = no staining; 1 = weak staining; 2 = moderate staining; 3 = strong staining). For images, slides were digitized using a Pannoramic Scan Whole Slide Scanner (Cambridge Research and Instrumentation; Hopkinton, MA) and images captured using the Pannoramic Viewer software version 1.14.50 (3DHistech; Budapest, Hungary).

**Statistical Analyses**

Gene expression profiling of the PrEC and SVEC samples were performed using standard RMA normalization and SAM analysis to control for multiplicity as described above. SPAN analysis, described above, uses a permutation sampling methodology of the network controlling for connectivity. This permits us to derive empiric p-values for all SPAN results. Biomodule testing and statistical analysis was performed using the PAM algorithm in the R/Bioconductor environment using the survival package. Meis protein module signature scores were generated simply by taking algorithmic sum of genes that were decreased in expression and subtracting from it the expression level of genes that were up-regulated. No training algorithms were used. Descriptive statistics were used to summarize the TMA data. Yate’s chi-squared test was then used to compare tumor versus normal tissue staining. Cox proportional hazards regression models were used to examine the effect of gene expression parameters as continuous variables, and to estimate hazard ratios. Kaplan-Meier was used to estimate overall survival and disease free recurrence in dichotomized patient cohorts. Kaplan-Meier analysis, Cox regression analysis, and proportional hazards assumption testing were performed using the “survival” package in R.
RESULTS

Comparison of prostate to seminal vesicle epithelium reveals widespread embryonic differences.

Gene array analyses were conducted on a set of three matched normal human prostate epithelium (PrEC) and seminal epithelium (SVEC) cultures. For clarity, we define “gene signature” as any purposeful collection of genes. For example, the set of differentially expressed genes derived from the SVEC v PrEC epithelial comparison will be termed a signature in this article. An overview of our workflow is displayed in Figure 1. The epithelium-only signature, after correcting for multiplicity, resulted in 447 differentially-expressed genes (Supplemental Table 2). Strikingly, there were 15 homeobox (Hox) genes that were differentially expressed (HoxA13, B2-3, B5-9, B13, C6, D1, D3-4, and D10-11), as well as Meis1 and Meis2 which were 25.43 and 51.4-fold over-expressed in PrECs, respectively (Supplemental Table 2). Differential expression of both Meis1 and Meis2 was validated using quantitative real-time PCR (Supplemental Figure 1). This observation was further evaluated using the functional enrichment formed using the Biological Process branch of Gene Ontology (GO) (33). We used a false discovery rate (FDR) of 5 percent and the GO enrichment yielded annotations primarily involved in development (Supplemental Table 3). In fact, the top five branches related to embryonic development. In addition, we observed GO branches relating to epithelium development and differentiation, as well as urogenital system development.

Presence of homeobox and cell cycle prioritized genes distinguish epithelial signature from mixed tissue gene signature

Previous studies have evaluated the difference between seminal vesicle and normal prostate using whole and unfractionated prostate and seminal vesicle tissue but did not report on a preponderance of homeobox proteins (19). We hypothesized that careful gene profiling of matched primary epithelial cultures from normal prostate and seminal vesicles would reduce confounding noise during analysis and provide a more robust pathway prioritization. In order to evaluate this difference more thoroughly, we re-analyzed the gene expression data from Thompson et al. and derived a gene signature (mixed tissue signature) in exactly the same manner (19). The resulting mixed tissue signature contained 502 probes that were significant by SAM analysis with a FDR of 5 percent. Of note, Gene Ontology enrichment did not significantly enrich for embryonic development within whole tissue analyses.

To see where our epithelium-only signature overlapped and differed with the whole-tissue signature, we used a validated methodology to compare them using Single Protein
Analysis of Network (SPAN). We have previously successfully analyzed both breast and prostate cancer signatures in this manner (27). The SPAN methodology relies on comparing gene signatures to a common background based on statistical prioritization of protein-protein interactions. For this analysis, we compared both the mixed tissue and epithelial signatures to the Wellcome Trust Cancer Gene Census (28). We asked whether there were members of each signature that were connected to known cancer genes more than expected by chance (FDR ≤ 0.05). The results of the comparison are represented graphically in Figure 2A.

Based on each signature’s connectivity to the Wellcome Trust Cancer Gene Census, we noted excellent overlap between the two signatures in well-established cancer genes such as the Janus kinase (JAK) family, phosphatidylinositol 3-kinase (PI3K), and H-Ras. Of significance to the epithelium-only signature was the additional prioritization of two genes, Cdkn2c and Meis1. Cdkn2c is a cell cycle gene whereas Meis1 is a homeobox gene co-factor and corresponded with our GO enrichment findings documenting significant differential Meis1 expression between prostate and seminal vesicle epithelial cells.

Protein-protein network evaluation of epithelial only signature confirms Meis biomodule

Given our original hypothesis that differences in prostate versus seminal vesicle gene expression are important for oncogenesis, we performed a second SPAN analysis. In contrast to our first analysis, we now asked whether there were individual Wellcome Trust Cancer Gene Census members that were connected to our epithelial only signature more than expected by chance. In other words, we prioritized known cancer genes that may not have been differentially expressed but are implicated by significant pathway connectivity to our gene signature. For simplicity, we defined a “biomodule” as a prioritized cancer gene and the members of the gene signature that were connected to it. Biomodules that interconnected were considered as one biomodule. To clarify the function of these biomodules, we performed a GO enrichment as described prior. Three biomodules emerged from our analysis as displayed in Figure 2B. First, Biomodule A centered on Myd88 (myeloid differentiation primary response gene) and enriched for Myd88-dependent Toll-Like receptor signaling pathway. Second, Biomodule B centered on Eln (Elastin) and enriched for proteinaceous extracellular matrix. Third, Biomodule C centered on homeobox genes Pbx1 and Hoxa9 and enriched for embryonic skeletal system development. Consistent with our previous observations, Biomodule C also contained the homeobox genes Meis1 and Meis2. Thus, we observed involvement of Meis proteins using two different and independent prioritizations. We termed this biomodule the Meis biomodule.
The Meis biomodule genes are prognostic for both prostate cancer recurrence and survival

To explore and refine the revealed biomodules for their clinical relevance and potential clinical translation, we adopted a two-step approach. First, we evaluated each biomodule’s ability to divide patients into statistically different cohorts based on Kaplan-Meier analysis using the Prediction Analysis for Microarrays (PAM) algorithm (32). PAM provides an unbiased, unsupervised approach to determining whether there are different cohorts based on gene expression patterns. If the PAM methodology showed indications of curve separation with Chi-squared p-values ≤ 0.1, we would then proceed with a second step – that of biomodule refinement. This second step is critical. As we have previously noted (27), the most informative (pathway-enriched) biomodules are not usually the most discriminatory (i.e. able to find survival/recurrence differences). Therefore, incorporation of mRNA data is needed to modify our informative biomodules into discriminatory biomodules. To perform the refinement, we systematically evaluated at the mRNA expression scale of biology gene combinations within the selected biomodules using independent datasets. We specifically trained in one dataset and validated our findings using independent datasets.

In particular, we initially examined two independent datasets: the first being a well-studied dataset of 78 patients status post-prostatectomy (24) (Glinsky data set) where we evaluated the disease-free interval (i.e. PSA recurrence after treatment of the primary tumor). The second dataset was a cohort of 281 men who were diagnosed with prostate cancer who were treated with watchful-waiting in which we evaluated their overall survival (Sboner Swedish Watchful Waiting cohort) (22). Of the biomodules evaluated, only the Meis biomodule demonstrated an association in the disease-free interval dataset. Finally, we re-verified our Meis biomodule in another recurrence dataset (Taylor) that contains 131 patients with biochemical recurrence data. Only PAM analysis with the Meis biomodule demonstrated statistical significance (Chi-square, p = 0.039) (Supplemental Figure 2).

Because the Meis biomodule now had demonstrated clinical import in the disease-free interval dataset, we proceeded to refine the biomodule using the same approach. Whereas the biomodule consisting of five genes (Meis1, Meis2, Pbx1, HoxA9 and HoxB7) approached statistical significance, we noted that a smaller combination of four genes (HoxA9, Meis1, Meis2, and Pbx1) provided the most significant p-value (Chi-square) of 0.02 (Figure 3A). We then validated this refined biomodule in the independent Swedish Watchful Waiting dataset in an unbiased manner using PAM and found this biomodules to have a p-value of 0.05.

This same curve separation persisted when we examined low-grade prostate cancer tumors. In fact, when we looked at patients with Gleason 7 disease in the Swedish Watchful
Waiting cohort, we saw marked differences in the two groups (n = 117, Chi-square p = 0.03) (22). Interestingly, the four-gene signature was not able to split Gleason 6 tumors. However, when we examined the individual signature genes, we noted that decreased expression (less than the mean of all samples) of only Meis1 and Meis2 combined were able to define cohorts with differences in survival (n = 83, p = 0.008, Figure 3B). Differential expression of Hoxa9 and Pbx1 individually did not demonstrate any measureable survival difference among low-grade tumors and were only significant in disease recurrence and patient survival from advanced tumors. Impressively, patients with differential expression in both Meis1 and Meis2 Gleason 6 tumors demonstrated an 18-month decrease in overall survival (157 months vs. 175 months, Figure 3C), supporting a role for these proteins in the formation of poor-prognosis tumors. We saw a similar pattern among the Gleason 7 cohort with a 40-month survival difference (n = 117, p = 0.01, Figure 3D).

We next asked whether Meis1, Meis2, or a combination could serve as independent predictors of recurrence and survival using Cox regression analysis. Recapitulating our Kaplan-Meier analysis experience, univariate analysis of Meis1 and Meis2 expression scores (or their combination) demonstrated significant hazard ratios in all categories (Supplemental Table 4). Even after adjusting for Gleason score and age, elevated levels of Meis2 remained an independent predictor of decreased rate of recurrence (HR 0.69, p = 0.052). Similarly, both elevated levels of Meis1 and the combination of Meis1 and Meis2 were highly associated with improved survival (HR 0.14, p = 0.017; and HR 0.33, p = 0.025, respectively).

**Decreased Expression of Meis1, Meis2, Pbx1 and increased Expression of HoxA9 characterize Meis biomodule**

Having established clinical relevance in this Meis-driven biomodule, we then set out to determine the changes in the biomodule within different clinical stages. To do this, we examined the fold-change of our genes of interest from two other independent datasets: the first compared primary tumor versus normal prostate (GDS2547: 64 versus 18 specimens, respectively) (34) and the second castrate resistant prostate cancer as compared to primary tumor specimens (GSE32269: 22 specimens versus 29 specimens, respectively) (20). Table 1 provides a synopsis of our findings. These data document a consistent pattern of decreased expression of Meis1, Meis2, and Pbx1 during tumor initiation and progression. Indeed, from normal prostate to primary prostate tumor, the greatest decrease is in Meis1, suggesting that decreased expression of Meis1 plays a role in tumor initiation. HoxA9 demonstrated a 1.1-fold increase (not significant). Although it exhibited only a subtle change, our initial Kaplan-Meier
analysis of the recurrence dataset demonstrated that HoxA9 was essential for proper separation of the population into two cohorts. When HoxA9 was left out of the PAM algorithm, no significant curve separation was seen. Therefore, we reasoned that subtle changes in HoxA9 appear to have profound changes in phenotype and have maintained it as part of the Hox biomodule. Given that these were prostatectomy specimens, we further reasoned that HoxA9 potentially is involved in tumor initiation. The fold-change comparing primary prostate tumor to castrate-resistant tumor documents the greatest decline in Meis2 and Pbx1, suggesting that these genes play a role in tumor progression and castration resistance.

Meis biomodule signature effectively classifies prostate oncogenic states and is associated with Gleason grade

Given that the four-gene signature could, in an unsupervised fashion, identify cohorts of patients with decreased survival and decreased time to disease relapse, we tested the hypothesis that this four-gene signature could indeed separate groups of patients based on their expression profile. To do this, we used another independent dataset of 185 gene expression profiles comprising normal prostate, primary prostate cancer, prostate metastasis, and immortalized prostate cancer cell lines (23). We computed a composite score by simply adding the normalized expression levels of Meis1, Meis2, Pbx1 and subtracting that of HoxA9. We subtracted HoxA9 since, when we examined the two cohorts identified by PAM in the relapse post-prostatectomy dataset, we saw a non-significant trend toward an increase in HoxA9 as opposed to significant decreases in Meis1, Meis2 and Pbx1. We then ordered the specimens based on this composite signature score. Importantly, we did not employ any training algorithm that may influence or bias our analyses. Data shown in Figure 4A demonstrates that normal prostate was separated furthest from immortalized cell lines and metastatic disease. Although there was some overlap between the normal prostate and primary prostate tumor, they are graphically separated. Thus, without the use of a training algorithm, we are able to cluster prostate tissue in different stages of oncogenesis using our Meis1/Meis2/PBX1/HoxA9 biomodule.

We repeated this simple graphical analysis using only primary prostate cancer tumor specimens and grouped them based on Gleason score. The pattern that emerges is that Gleason 6 and Gleason 7 prostate cancers are quite heterogeneous with wide-ranging Meis module scores (Figure 4B). As noted in the figure, the mean Meis module score decreases with increasing Gleason with a Pearson correlation coefficient of 0.98 (p = 0.02).
**Immunohistochemical detection of Meis2 in prostate tissues.**

To validate the bioinformatic analyses of our Meis module using clinical specimens, we evaluated the histologic pattern and prevalence of Meis2 expression in 110 prostate tissue specimens. We focused on Meis2 since mRNA expression appeared to decrease in two stages, first during oncogenesis and second during metastasis and castration-resistance, as shown in Table 1; this indicated that Meis2 has a role during both oncogenesis and tumor progression. Immunohistochemical analyses of normal human prostate tissue documents strong nuclear Meis2 expression in basal-epithelial cells and stromal fibroblasts, and weak nuclear staining in luminal-epithelial cells (Figure 5A). Analysis of a series of prostate tumors documented a profound loss of Meis2 expression in prostate cancer cells, with retention of stromal Meis2 expression (Figure 5A and B). This loss of Meis2 expression from normal prostate tissue to tumor was highly statistically significant (Chi-square, Yates p-value < 0.001). This difference in staining is displayed graphically in Figure 5B. Analyses of metastatic lesions also demonstrated a lack of Meis2 expression within the cancer cells.

The robust expression of Meis2 in the basal cell layer of normal prostate epithelia suggested that the observed decrease in Meis2 mRNA expression during cancer initiation and progression, as observed in our gene array analyses, may simply be due to the characteristic loss of basal cells and enrichment of malignant luminal epithelial cells in prostate tumors (35). To address this, we conducted further analyses of the expression of basal-specific Cytokeratins 5 and 14, and luminal-specific Cytokeratin 8 in patients split by Meis1 and/or Meis2 differences. These data document that, in Gleason 6 watchful waiting patients, there is no statistical difference in the expression of basal Cytokeratins 5 and 14, or luminal Cytokeratin 8 between tumors that were grouped based on Meis expression (Figure 5C). Thus, decreases in Meis2 expression during cancer initiation and progression is not simply an artifact of changes in basal to luminal epithelial cell content. These data validate our bioinformatic analyses of mRNA data demonstrating decreased expression of Meis2 in poor prognosis prostate tumor specimens, and confirm that deregulation of the homeobox gene co-factors Meis1, Meis2, and Pbx1 may be significant molecular events during prostate cancer initiation and progression.

**DISCUSSION**

Here we document the identification of a single developmental signaling pathway that spans both prostate oncogenesis and progression. Our study documents molecular differences between seminal vesicle and prostate epithelial cells that highlight critical determinants in the vast disparity between cancer incidences for these sex accessory glands. The role of Meis1 and
2 in low-grade prostate tumors suggests that these genes play a critical function in the formation of poor prognosis tumors and may be a useful biomarker or therapeutic target.

Indeed, a recent study of familial prostate cancer identified a germline mutation of HoxB13 (G84E) in men with early-onset, familial prostate cancer (36). The G84E mutation and other additional mutations all occurred within the Meis interacting domain of HoxB13. Thus, interference with appropriate Meis/Hox protein interactions, either by germline mutation or decreased expression, is clearly a critical molecular event in prostate cancer etiology. As shown in the protein-protein network of Supplemental Figure 3, our data demonstrates a clear connectivity among the Meis and Pbx proteins and HoxB13.

Concordant to our observations, Meis2 was identified in an array screen as one of the top 50 down-regulated genes in prostate cancer, and changes in HoxA9 mRNA expression was reported in a prostate cancer cell line in response to selenium treatment (37). The relative weak DNA-binding specificity of Hox proteins themselves, along with their dependence upon co-factors to regulate gene targets, implies two critical points (7). First, Hox proteins may not be as important as the co-factors which are expressed and which regulate Hox protein DNA-binding specificity and transcriptional activity. Second, changes in co-factor expression may profoundly alter the prostate gene-specificity of Hox protein transcriptional regulation. Hox proteins may function to promote differentiation and suppress proliferation of normal prostate epithelial cells or promote a less-aggressive cancer phenotype. Thus, compounds that restore Meis1, Meis2, and/or Pbx1 may be effective toward the prevention or treatment of aggressive prostate tumors.

In addition, our work provides a cautionary tale to the analysis of genomic profiling of unselected tumor tissue. The persistence of stromal Meis2 expression provides an explanation as to why previous gene array analyses comparing whole tissue prostate and seminal vesicle did not identify Meis2 as being differentially expressed (19). It may explain why Meis2 mRNA expression appeared to decrease further in metastatic tissues, as there was no longer any Meis2-positive stroma present. Thus, the overall decrease in Meis2 expression from gene array analyses of primary prostate tumors may be even more profound than observed as there is often contaminating stroma present during the tissue preparation.

From a translational standpoint, the Meis module is exciting on two counts. First, loss of the Meis1 and/or Meis2 expression may be a useful biomarker to discern patients with low-grade tumors that would benefit from treatment. PCR analysis of patient prostate needle biopsies may provide critical prognostic information as to the necessity of prostatectomy in low-risk disease. Indeed, patients who underwent radical prostatectomy for Gleason 6 prostate cancer trended toward an improvement in overall survival with an absolute risk reduction of 13%
and absolute decrease in distant metastasis by 11.4% (38). Thus it is clear that some patients are benefiting from treatment of their tumors, and proper selection of these low-risk patients may spare others undue morbidity. Secondly, targeting the Meis module could function as an efficacious clinical target for the development of new therapies to prevent prostate tumor initiation and treat aggressive prostate tumors.

Conclusion

Loss of Hox protein co-factors Meis1, Meis2, and Pbx1 are associated with both prostate cancer oncogenesis and tumor progression. Not only may these genes potentially serve as useful therapeutic targets, but they may also serve as powerful biomarker in identifying men for treatment of low-grade disease.

Acknowledgements

We wish to acknowledge the support of the University Of Chicago Section Of Urology led by Dr. Arieh Shalhav, and the Director of Urologic Research Dr. Carrie Rinker-Schaeffer. We would also like to acknowledge the support of the University of Chicago Comprehensive Cancer Center (UCCCC) led by Dr. Michelle Le Beau. We also wish to thank expert technical assistance of Kristen Otto in the Urology Research Laboratory, and the Human Tissue Resource Center core facility led by Dr. Mark Lingen, and the assistance of Leslie Martin and Mary Jo Fekete. We also thank the Immunohistochemistry Core Facility run by Terri Li, and the Functional Genomics Core facility run by Dr. Pieter W. Faber. We wish to thank Dr. Karen Kaul and Dr. Robin Leikin of the Prostate Cancer SPORE of Northshore University and Northwestern University, respectively. Finally, for providing comments on our manuscript we wish to thank Dr. Walter Stadler of the University of Chicago Department of Hematology and Oncology, and Dr. William Isaacs at Johns Hopkins University and the James Buchanan Brady Urological Institute.
REFERENCES

FIGURE LEGENDS

Figure 1: Biomodule Discovery and Validation Workflow.

Step 1: We first examined the differentially expressed genes between normal prostate epithelium and seminal vesicle epithelium. Matched tissue samples were profiled using standard Affymetrix gene arrays. Genes with a false discovery rate $\leq 5\%$ were retained. Step 2: Using the protein-network prioritization algorithm SPAN (Single Protein Analysis of Networks), we compared our set of differentially expressed genes from Step 1 to a reference set of known cancer-related genes from the Wellcome Trust Sanger Cancer Gene Census. SPAN identifies Cancer Genes that are connected to the differentially expressed genes more than expected by chance (False Discovery Rate $< 5\%$). Each set (Cancer Gene and its connected differentially expressed genes) are termed biomodules. Step 3: To determine if these biomodules from Step 2 are clinically relevant and of prognostic import, we evaluated their ability to identify patients with poor prognosis using the PAM (Prediction Analysis for Microarrays) algorithm in historical prostate cancer datasets. Biomodules that were prognostic were then retained. Step 4: Genes from the prognostic biomodule are then validated and refined in silico (using other independent historical datasets), in vitro (via protein and qPCR analysis of genes), and on tissue microarrays (TMA).

Figure 2: Prioritization of a Meis Regulation Module in Prostate Epithelial Cells.

A. Differences between the mixed tissue and the epithelial-only signatures. Analyses of the genes signatures from the mixed tissue (blue circles) and epithelial-only (red circles) signatures based on their protein-protein interactions (black lines). We used SPAN, a protein network prioritization algorithm, to identify significant connections between gene signatures to known cancer genes (green circles) as defined by the Wellcome Trust Cancer Gene Census. Large circles indicate a false discovery rate of less than 5%. In other words, differentially expressed genes from the mixed tissue and epithelial only sets that are highly connected to multiple cancer genes are displayed. The network demonstrated significant network connectivity between the signatures with important exceptions: Meis1 and CDKN2C. These genes were seen only detected in the epithelium-only signature. Dashed lines indicate membership to the gene signature (gray squares).

B. Identifying prioritized cancer-specific biomodules. To identify cancer-related genes (green circles) that were connected to the epithelial-only signature members (red circles) more than expected by chance, we again used the SPAN algorithm. In all, four cancer-related genes
Myd88, Eln, Pbx1, HoxA9 were prioritized resulting in 3 biomodules (HoxA9 and Pbx1 are in the same biomodule). Circle size indicates that the prioritized gene was statistically significant (FDR < 5%). Each biomodule was annotated using Gene Ontology (GO) with statistically significant annotations listed in the figure along with their GO ID number. Dashed lines indicate membership to the gene signature (gray squares).

**Figure 3:** Changes in Meis Regulatory Biomodule is Prognostic for PSA Recurrence and Patient Survival.

A. Gene expression profiles from prostatectomy specimens of 79 patients (Glinsky dataset) using the PAM (Prediction Analysis for Microarrays) algorithm to stratify patients into two groups based on changes in expression of Meis1, Meis2, Pbx1, and HoxA9. PAM computationally separates patients into two groups based on available gene expression data. The result is two computation clusters (PAM cluster 1 and PAM cluster 2). Time to PSA recurrence was analyzed using Kaplan-Meier analysis and detected an 8-month decrease in median disease free interval (logrank p-value=0.02).

B. Gene expression profiles of biopsy specimens from 281 patients (GSE16560) who underwent watchful-waiting until death were analyzed using Kaplan-Meier techniques. Decreased expression of either Meis1 or Meis2 (below the mean Meis levels, green lines and bars) significantly decreased patient survival compared to controls (at or above mean Meis levels, blue lines and bars) and demonstrated a 33-month difference in median survival (p=0.01).

C. Gene expression profile of 83 Gleason 6 tumors biopsy specimens (GSE16560) for their changes in Meis1 and Meis2 using Kaplan-Meier techniques. Decreased expression of either Meis1 or Meis2 (below the mean Meis levels, green lines and bars) significantly decreased patient survival compared to controls (at or above mean Meis levels, blue lines and bars) and demonstrated an 18-month survival difference (p=0.009).

D. Gene expression profile of 117 men with Gleason 7 tumors biopsy specimens (GSE16560) analyzed using Kaplan-Meier techniques. Decreased expression of either Meis1 or Meis2 (below the mean Meis levels, green lines and bars) significantly decreased patient survival compared to controls (at or above mean Meis levels, blue lines and bars) and demonstrated a 40-month difference in median survival (p=0.01).

**Figure 4:** Classification of Prostate Oncogenic States Using Meis1, Meis2, Pbx1, and HoxA9 Expression.
A. The Meis biomodule is capable of classifying an independent set (GSE21032) of 185 prostate tissues (from normal prostate, cancer, metastasis, and immortalized cell lines) based on a simple score. Each biomodule score is the sum of Meis1, Meis2, and Pbx1 subtracted by the expression value of HoxA9. The scores are normalized to the average score and then ordered by percentage change from mean. Different tissue types are colored graphically as indicated in the legend. Similar tissue types clustered together without any need for a training algorithm.

B. The Meis biomodule score correlates with Gleason score. We analyzed 130 prostate cancer gene expression profiles (GSE21032) and calculated their scores. These scores were then ordered by their % change from mean and grouped by Gleason score. The yellow lines depict the mean score. Mean scores were correlated with Gleason Score (p = 0.02).

Figure 5: Loss of Meis2 Protein Expression in Prostate Tumors.
A. Immunohistochemical detection of Meis2 in normal prostate tissue documented strong nuclear staining of basal-epithelial cells with weak luminal-epithelial cell staining (left panel). Furthermore, Meis2 was expressed in stromal fibroblasts. In all prostate tumors, Meis2 expression was undetectable (middle and right panel). Stromal and adjacent normal tissue retained Meis2 expression.

B. Summary of epithelial Meis2 staining of prostate tissues. Bars represent the percentage of normal glands (Normal), tumors (GG3, GG4, GG5), or lymph-node metastases (LN-Met) which had more than 1% (grey bars) or 10% (black bars) of cells staining positive for Meis2. This loss of epithelial Meis2 expression from normal prostate tissue to tumor was highly statistically significant (p-value < 0.001).

C. Changes in Meis1 and Meis2 are not associated with changes in basal and luminal epithelial cell-specific cytokeratins. To control for changes in basal and luminal epithelial cell content during tumor initiation and progression, we analyzed the associated expression of Cytokeratins 5 and 14 (basal, CK5 and CK14) and 8 (luminal, CK8) between patients separated via Meis expression within the Swedish Watchful Waiting Gleason 6 cohort. Decreased expression of either Meis1 or Meis2 (below the mean Meis levels, green lines and bars) significantly decreased patient survival compared to controls (at or above mean Meis levels, blue lines and bars). No significant change in cytokeratin expression, however, was observed between tumors derived from these patient populations (NS=not significant).
**DISCOVERY**

1. Find differentially expressed genes

   - Normal Prostate
   - Normal Seminal Vesicle Epithelium

   **Differentially Expressed Genes**

2. Identify cancer gene centered biomodules via protein networks

   **Unselected Cancer Gene Centered Biomodules**

3. Identify prognostic biomodules

   **Prognostic Cancer Gene Centered Biomodule**

**VALIDATION**

4. Validate on multiple scales of biology

   - Tissue Microarray
   - In vitro
   - In silico

   **INDEPENDENT Prostate Cancer Dataset**
A. Recurrence (All Gleason)  
- p = 0.02, n = 79  
- median DFI 23m vs. 15m

B. Survival (All Gleason)  
- p = 0.01, n = 281  
- median survival 129m vs. 96m

C. Survival (Gleason 6 pts)  
- p = 0.009, n = 83  
- median survival 175m vs. 157m

D. Survival (Gleason 7 pts)  
- p = 0.01, n = 117  
- median survival 136m vs. 96m
### Table 1: Fold-Change Univariate Analysis

<table>
<thead>
<tr>
<th></th>
<th>Epithelial Only Signature (Prostate vs. Sem. Ves.)</th>
<th>Primary Prostate vs. Normal Tissue</th>
<th>CRPC vs. Primary Tumor</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEIS1</td>
<td>25.43</td>
<td>0.55</td>
<td>N.S.</td>
</tr>
<tr>
<td>MEIS2</td>
<td>52.63</td>
<td>0.64</td>
<td>0.09</td>
</tr>
<tr>
<td>PBX1</td>
<td>2.78</td>
<td>0.69</td>
<td>0.28</td>
</tr>
<tr>
<td>HOXA9</td>
<td>N.S.</td>
<td>N.S.</td>
<td>N.S.</td>
</tr>
</tbody>
</table>

N.S. = not significant (FDR >0.05)
A. Meis Module Signature by Tissue Type

- Decreasing Meis Biomodule Score
- Increasing Meis Biomodule Score

- Metastatic Tumor
- Normal Tissue
- Immortalized Cell Line
- Primary Prostate Tumor

B. Meis Module Signature by Gleason Score

- G6
- G7
- G8
- G9

Correlation with Gleason:
- r = 0.98, p = 0.02

Mean Meis Module Score
Deregulation of a Hox Protein Regulatory Network Spanning Prostate Cancer Initiation and Progression

James L Chen, Jianrong Li, Kyle Kiriluk, et al.

Clin Cancer Res  Published OnlineFirst June 21, 2012.

Updated version  Access the most recent version of this article at: doi:10.1158/1078-0432.CCR-12-0373

Supplementary Material  Access the most recent supplemental material at: http://clincancerres.aacrjournals.org/content/suppl/2012/06/21/1078-0432.CCR-12-0373.DC1

Author Manuscript  Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.