Award Number: W81XWH-11-1-0744

TITLE: Development of Assays for Detecting Significant Prostate Cancer Based on Molecular Alterations Associated with Cancer in Non-Neoplastic Prostate Tissue

PRINCIPAL INVESTIGATOR: Farhad Kosari, Ph.D.

CONTRACTING ORGANIZATION: Mayo Clinic
Rochester, MN 55905

REPORT DATE: December 2016

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
December 2016	**Final**	09/30/2011 - 10/29/2016

4. TITLE AND SUBTITLE

Development of Assays for Detecting Significant Prostate Cancer Based on Molecular Alterations Associated with Cancer in Non-Neoplastic Prostate Tissue

5a. CONTRACT NUMBER
W81XWH-11-1-0744

5b. GRANT NUMBER
10623678

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Farhad Kosari F, Cheville JC, Vasmatzis G, Karnes RJ, Manemann M, Murphy SJ

E-Mail: kosari.farhad@mayo.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Mayo Clinic
Rochester, MN 55905

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT The goal of this project was to develop molecular models to distinguish significant (Gleason score 7 and higher) prostate cancer (PCa) from indolent (Gleason score 6) PCa using biomarkers in the benign (non-neoplastic) prostate tissue and in high grade PIN (HGPIN) samples. Aim1 focused in gene expression and methylation analyses by sequencing in a large set of samples to identify gene expression and epigenetic candidate biomarkers. Most promising gene expression biomarkers which were validated in multiple independent sets were selected to develop two logistic regression models. The model for testing in the bulk tissue incorporated 6 genes, including NAV1, LYST, ADD3, SMC5, CEP350, and KIAA2026. This model was validated in 4 of the 5 datasets we tested. Another model to distinguish HGPIN samples in indolent and significant cases was based on three genes including LRRC4C, SUGT1, and KLHL28 and was validated in an independent dataset. This career development grant was an excellent opportunity for us to study prostate cancer field and to realize the clinical potential of cancer field effect. We were able to obtain a sizable set of data which will be used in future research in projects aimed at prevention of PCa and also early detection of significant PCa.

15. SUBJECT TERMS
LCM AACR

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					USAMRMC
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
U	U	U	UU	12	

Table of Contents

Introduction:

Each year in the U.S. more than a million men with an elevated serum PSA or abnormal digital rectal exam undergo a prostate biopsy, and nearly 200,000 are found to have prostate cancer (PCa). Decisions to treat PCa are heavily influenced by the Gleason score (GS) of the tumor in the needle biopsy specimen. Gleason score is a measure of tumor differentiation based on the two most prevalent patterns of tumor growth. Patients whose entire tumor is composed of GS6 rarely progress, and recently, more men diagnosed with GS6 tumors on needle biopsy are selecting active surveillance rather than surgery or radiation therapy. In contrast, men with more poorly differentiated tumors (GS7 and higher) have a significantly increased risk of progression, and require treatment. Choosing the best treatment options for patients with biopsy GS6 is complicated by the fact that a biopsy procedure only samples a very small part of the prostate, and in about 30% of men, it underestimates the GS. In those cases, men with GS7 and higher ($GS7^+$) prostate cancer are assumed to have GS6 tumors potentially leading to inappropriate treatment. In addition, because of the limited sampling and 30% false negative rate for detecting cancer, many men with a negative biopsy result may have clinically significant prostate cancer. Because of that, many of the 800,000 patients with a negative biopsy undergo repeat biopsies which can be frustrating for both patients and urologists. When a pathologist examines a prostate needle biopsy specimen, the focus is on the identification of prostate cancer and appropriate Gleason scoring. Very little attention is paid to the "normal" areas which often comprise the majority of biopsy samples. This is despite a considerable body of evidence suggesting that molecular alterations associated with tumor in adjacent non-neoplastic cells, the so called "tumor field effect", can provide valuable clues regarding the status of the tumor. Remarkably, the field effect alterations have also been associated with aggressive prostate cancer.

Body / Results:

The objective was to develop clinically relevant molecular models to predict significant prostate cancer with $GS7^+$ based on the prostate cancer field effect markers. This proposal focused only on identification of significant tumors with $GS7^+$ because Gleason score is the single strongest predictor of outcome in men with prostate cancer, and has the greatest influence on the clinical management of men with prostate cancer. This proposal concentrated on the "omics" areas where prostate cancer field effect has been best demonstrated, namely transcriptomic and epigenomics. There were two Aims. The goal of Aim I was to identify and validate prostate cancer field effect markers associated with $GS7^+$ tumors and the goal of Aim II was to develop molecular models for stratification of indolent and significant tumors in biopsies.

Table 1: Samples for methylation sequencing by RRBS

		Laser Capture Microdissection (LCM)		
	Bulk	**HGPIN**	***GP3 tumor**	***GP4+ tumor**
BP	16			
GS 6	19	19	20	
GS 7	6	5	19	16
GS 8^+	18	17		4
Total	59	41	39	20

* Funding through Mayo Clinomics Program

Research Accomplishments:

Task 1: NextGen Sequencing (NGS)

Table 2: Samples for gene expression analysis by RNA-seq

		Laser Capture Microdissection		
	Bulk	**HGPIN**	****GP3 tumor**	****GP4+ tumor**
BP	19			
GS 6	*237	18	15	
GS 7	*56	12	18	18
GS 8$^+$	*176	22		6
Total	*488	52	33	24

* Collaboration with Drs. Thibodeau and Wang
** Funding through Mayo Clinomics Program

1.1 – 1.5 NextGen Sequencing sample selection and processing: Tables 1 and 2 describe the samples that were selected and processed by sequencing for methylation (by RRBS) and gene expression (by RNA-seq) analyses, respectively. To expand the discovery phase of our project in this task, we developed collaborations with Drs. Thibodeau and Wang. The processed set is considerably larger than the original set proposed in the application and its large size contributes to more confidence in the robustness of the biomarkers selected.

1.6 Analysis - Mapping of methylated DNA was performed by the bioinformatics core facility (BIC) at the Mayo Clinic. All samples (Table 1) were mapped to the latest reference genome (HG38).

1.7 Analysis - Mapping of transcriptome data was performed within our group. All samples were mapped to HG38 using the "tophat" aligner.

1.8 Analysis - Selection of transcriptomic markers

We analyzed 50 N_{8+} and 50 N_6 by the Cufflink program to select promising genes based on False Discover rate (FDR) and the quality of sequence reads. This strategy identified hundreds of genes from which the most promising markers were selected based on other criteria, such as signal to noise ratio (SNR). At the end, we finalized a small set of genes as shown in Table 3 to move forward for further validation. We used a similar strategy to identify differentially expressed genes in HGPIN samples. Selected HGPIN markers are shown in Table 4. Interestingly, a known prostate cancer over-expressed gene, ERG, was among the significantly over-expressed genes in HGPINs in high grade (GS8$^+$) cases.

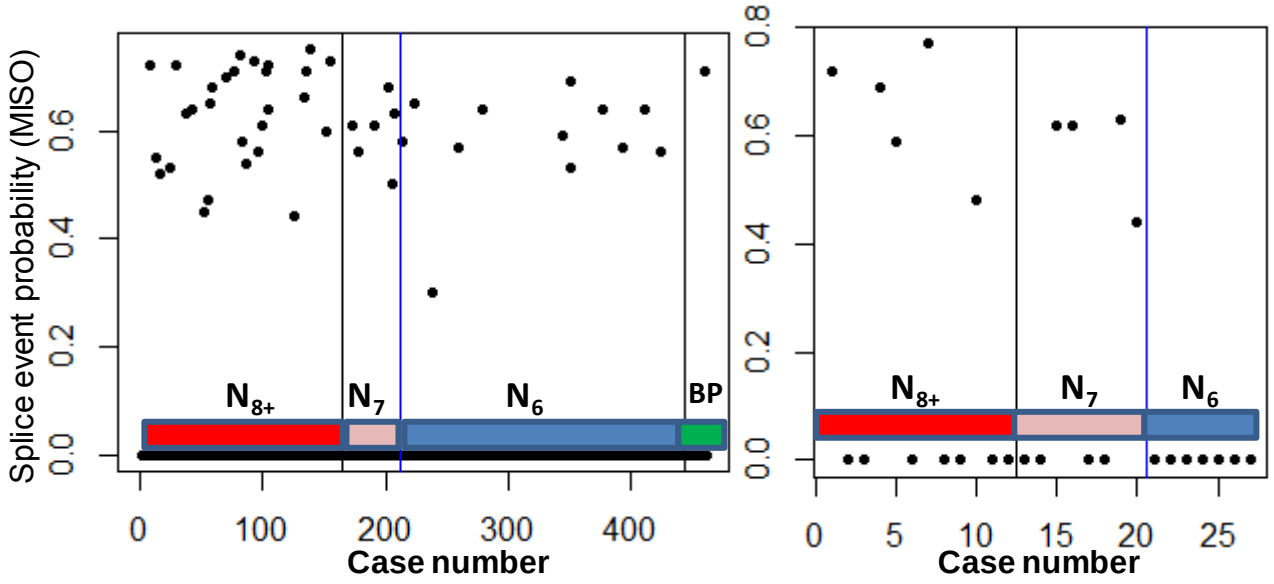
Table 3: Differentially expressed genes in N_{8+} compared with N_6

Down regulated in N_{8+}		Up regulated in N_{8+}	
CNNM2	LIN54	ADAMTSL3	KIAA2026
NUDT19	FAHD1	ADD3	LYST
SPAST	NQO1	ADI1	MPC1
TMEM127	SEC14L3	AP2M1	NAV1
RUNDC1	SMCR8	ARHGAP22	NEDD4
TBC1D20	PTK6	CEP350	RPAP2
CCDC36	ASB6	CYSLTR1	SFT2D2
FBXL20	SMC5	ERVW-1	TMEM182
		FRK	VPS13B
		KIAA1328	

Table 4: Differentially expressed genes in HGPIN from GS8$^+$ compared with GS6

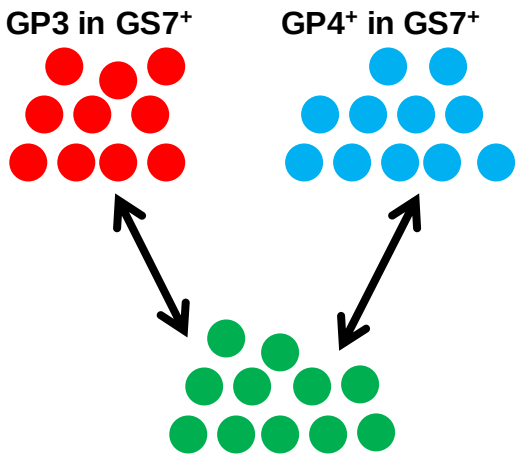
Down regulated in GS8$^+$ HGPIN	Up regulated in GS8$^+$ HGPIN	
PDCD11	AAR2	IFI6
MIR133A1	APOC1,APOE	KLHL28
TRPM2	CNTN4	LPL
CCDC186,MIR2110	CTC-575D19.1	LRRC4C
GTSF1	ELAVL2	USP15
FAM73A,RNA5SP21	ELN	NEB
MUTYH	**ERG**	NRXN3
GP1BB,SEPT5	FXYD1,FXYD7	PRH1-PRR4
	HLA-DQB1	SLC8A1-AS1

Figure 1: Representative splicing event with a more than 3 times frequency in N_{8+} and N_7 than in N_6. RNA-seq dataset in the left has lower sequencing depth but higher number of samples than the dataset on the right. The relative frequency between N_{8+} and N_6 did not change appreciably between the two dataset according to our expectations.



We also aimed to identify biomarkers based on skipped exons in bulk samples which can distinguish indolent from significant PCa. These biomarkers can be incorporated in sensitive PCR assays that rely on the presence or absence of the skipped exon, instead of quantitative measurement of expression levels. Because of that, these biomarkers are highly desirable in our project. We developed a novel bioinformatics approach and also used an existing program (MISO) to identify potential candidates. Both approaches identified about 15 potential candidates which we are examining. We anticipate the prevalence of these skipped exon events to be higher than the observed rates in RNA-seq data mostly because of the insufficient depth of coverage in standard RNA-seq experiments. In line with this expectation, we observed that the prevalence of these events in RNA-seq of higher depth and longer sequence fragment lengths were considerably higher while the ratio of these events comparing significant and indolent cases did not change significantly. Figure 1 illustrates one such event which was identified by the MISO program.

Figure 2: RRBS dataset in the Mayo Clinomics Project that were used in this proposal to identify field effect methylation biomarkers in the bulk and HGPIN samples



1.9 Analysis - selection of epigenetic markers

Selection of methylation markers highly benefited from the RRBS methylation data we obtained through a separate project that was funded through the Mayo Clinomics Project. In line with our field effect project, the goal of the Clinomics project was to identify methylation changes that would distinguish Gleason pattern 3 (GP3) in Gleason score

7 and higher PCa from GP3 in indolent GS6 tumors (Figure 2). We expect that such methylation changes occur in large areas in the prostate tissue and overlap with methylation changes in non-neoplastic areas. Selection of the field effect methylation biomarkers were based on the following criteria:

1- At least 5 CpG having concordance changes.

2- False discovery rate ≤ 0.1.

3- Concordant changes in comparisons of $GP4^+$ in significant cancers (i.e. $GS7^+$) against GP3 in insignificant (GS6) PCa.

4- Concordant changes in comparisons of GP3 in significant cancers against GP3 in insignificant (GS6) PCa.

Table 5: Differentially methylated gene loci in N7+ compared with N6. Entries are FDR q-values. Hyper and hypo methylated loci are shown in light green and blue shadings, respectively.

Gene Locus	N_{7+} vsN_6	$GP4^+$ vs GP3 in GS6	GP3 in GS7+ vs GP3 in GS6
C2orf81	4.03E-06	0.000034	0.002
COL18A1	0.0139	0.0008	0.0001
FOXN4	0.0753	0.00005	0.0058
GSC	0.0641	0.0000007	0.0003
LTBP4	0.0128	0.01	0.007
OSMR	0.0005	1.6E-09	0.0007
PRKAR1B	0.0347	0.004	0.008
RASAL3	0.0023	0.001	0.002
SDHAF1	0.0295	0.00007	2.00E-06
SLC22A3	0.0467	0.0000007	0.0005
TLE3	0.0947	8.8E-09	0.0005
TPPP	0.0356	0.0003	0.0002
RPTOR	0.0230	0.002	0.0006

In the bulk RRBS data in N_{8+} comparisons with N_6, this analysis identified 12 hyper-methylated and 1 hypo-methylated CpG islands (Table 5). Also, in HGPIN methylation data we found 11 hyper-methylated and over 12 hypo-methylated CpGs (Table 6). Interestingly, FOXN4, a member of the FOX family with important role in many cancers were common in the short list of hyper-methylated genes in both tables.

Task 2: Validation on surgery samples

2.1-2.3 Case selection and sample processing of validation samples.

Validation experiments used 20, 18, and 33 samples in each of the N_{8+}, N_6, and BP categories, respectively. Also, we processed 16 and 8 HGPIN samples in prostate tissues from patients with GS6 and GS8+

Figure 3: Multiplexing targeted DNA methylation assay that was tested for validating the identified epigenetic markers.

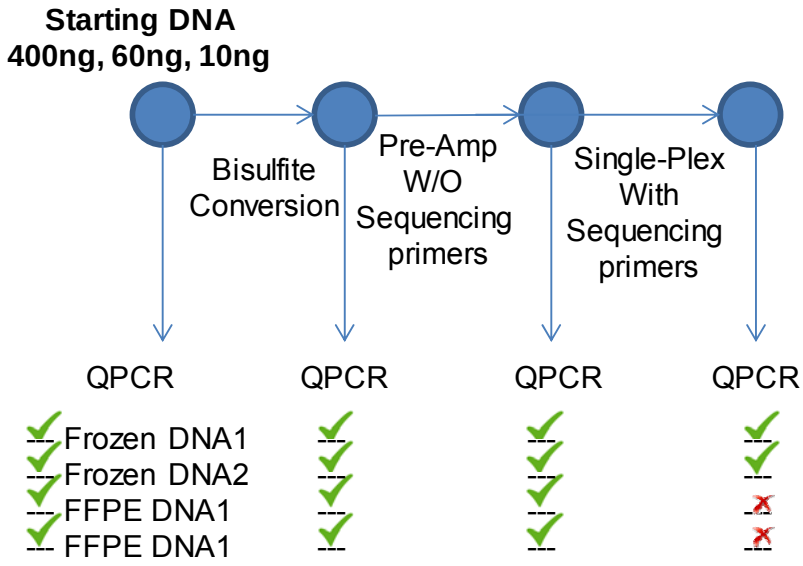


Table 6: Differentially methylated gene loci in HGPIN in GS8$^+$ compared with HGPIN in GS6. Entries are FDR q-values. Hyper and hypo methylated loci are shown in light green and blue shadings, respectively.

Gene Locus	HP in GS8+ vs HP in GS6	GP4$^+$ vs GP3 in GS6	GP3 in GS7+ vs GP3 in GS6
C12orf49	0.00012	0.00012	0.00003
CBX4	0.00200	0.00062	0.00009
CRACR2B	0.00427	0.00018	0.00066
FOXN4	**0.00051**	0.00005	0.00508
GLTSCR1	0.00098	0.04011	0.00656
GSE1	0.00029	0.00335	0.00253
NKX2-6	0.00218	0.01722	0.00579
PRKCZ	0.00105	0.00062	0.00628
SORBS2	0.00955	0.00021	0.00000
WIPI2	0.00748	0.00000	0.00114
WNT7B	0.00128	0.00136	0.00966
AQP12A	0.00006	0.01726	0.00807
ATP6AP1L	0.00463	0.00139	0.00014
CTDP1	0.00025	0.00626	0.00807
GALNS	0.00038	0.00651	0.00642
IL15RA	0.00045	0.00006	0.00092
IRF4	0.00680	0.00011	0.00681
NARS2	0.00001	0.000003	0.00517
NCS1	0.00516	0.00100	0.00040
PCBD1	0.00961	0.00017	0.00858
PFKP	0.00122	0.00000	0.00314
SOX7	0.00006	0.00250	0.00227
UBE2E2	0.00012	0.00007	0.00030

prostate cancers, respectively.

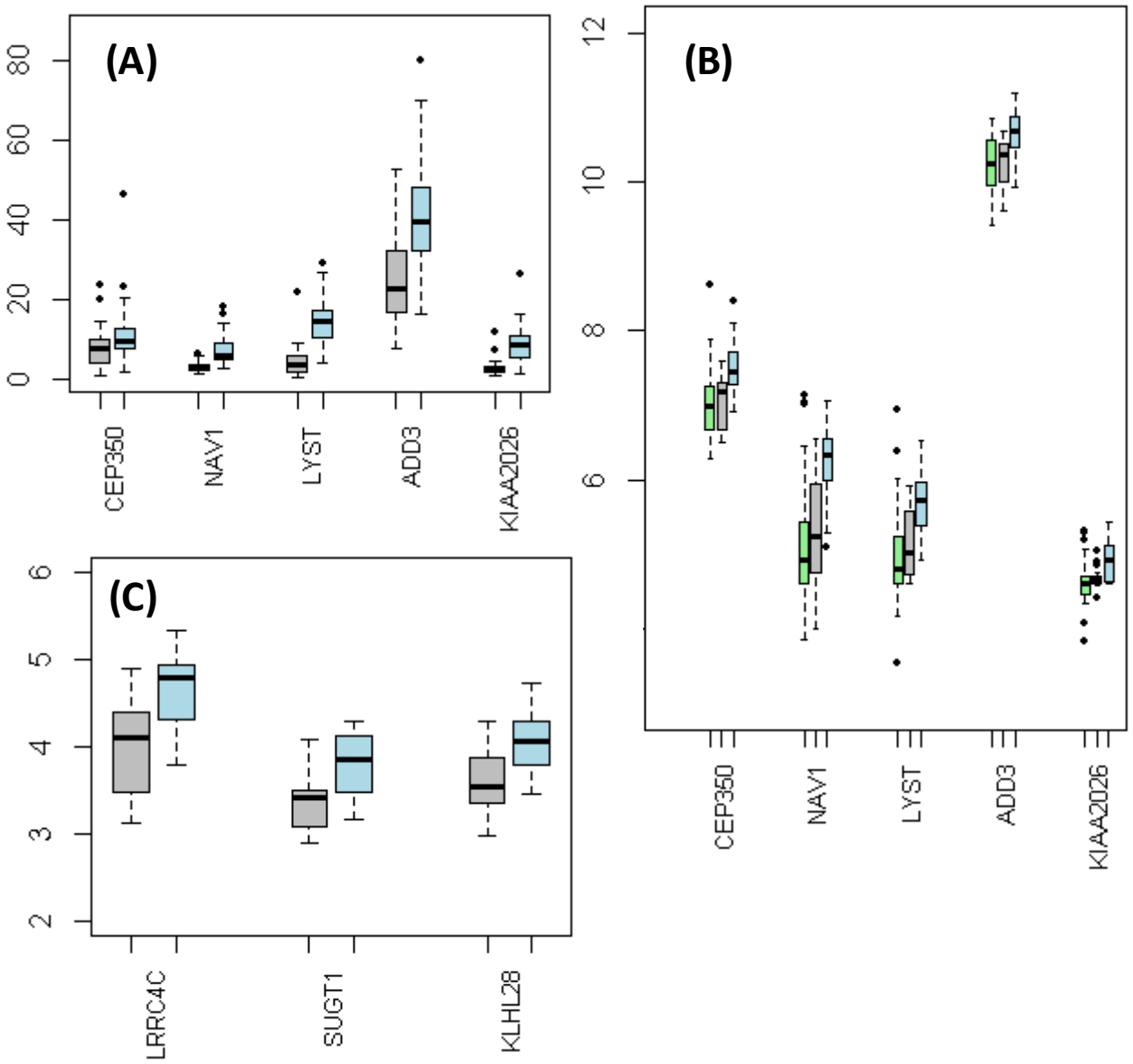
2.4-2.5 Development of validation Assays and selection of biomarkers:

Methylation assays: In this step we encountered challenges. We considered several assays for testing the methylation markers and the one which was most promising was a targeted sequencing approach for simultaneous analyses of many CpG loci in tens of samples. Figure 3 describes the basic scheme which included a pre-amplification of target loci in bisulfite converted DNA followed by single-plex PCR with primers that were compatible with Illumina sequencing. This approach was successful in testing FFPE DNA of higher quantity (400 ng) and DNA from fresh frozen samples of all quantities that we tested. However, it failed in the last step in generating sequencing libraries from biopsy compatible (6 and 60 ng) FFPE DNA. Despite this setback, we still believe this approach to be the most appropriate for this project and are continuing to trouble shoot this design so we can test the identified methylation markers in Tables 5 and 6.

Gene expression assays: Candidate gene expression markers (Tables 3 and 4) were analyzed by additional RNA sequencing and by Affymetrix arrays (ClariomD and U133PLUS2).

Selection of biomarkers: Selected gene expression markers for distinguishing indolent from significant PCa in non-neoplastic bulk prostate and HGPIN samples (Tables 3&4) were analyzed and the most promising markers were identified. Figure 4 illustrates independent validation of 5 biomarkers in bulk data in two independent sets of validation samples. These markers, plus one additional marker (SMC5) were used in task 4 to develop a prediction model. Also, Figure 4C displays three biomarkers in HGPIN samples that were used to develop a prediction model in task 4.

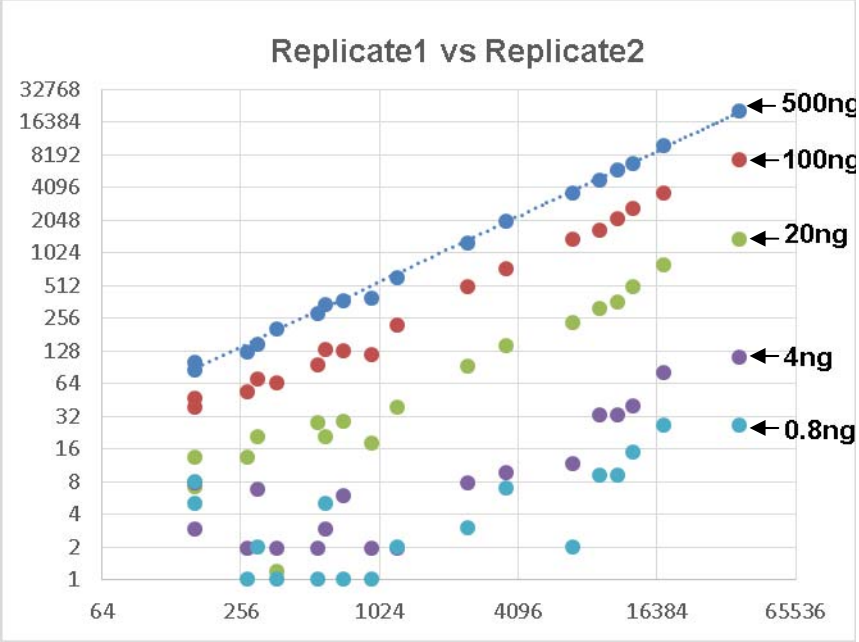
Figure 4 Validation of biomarkers in independent samples. A & B display validation of 5 over- expressed genes in N_{8+} compared with N_6. Box plot in (A) is a non-overlapping N_{8+} (n=50) and N_6 (n=50) data set of from the discovery set and (B) is another data set containing BP (n= 33), N_6 (n= 18), and N_{8+} (n=20). Box plot in (C) displays validation of HGPIN biomarkers. Light green, blue and grey bars are BP, N_6, and N_{8+} samples, respectively.



Task 3: Validation on biopsy samples

Despite our efforts, validation of biomarkers in biopsies was not completed. That was because method development for expression and methylation assays came close but did not reach RNA/DNA quantities that are compatible with biopsies. Our efforts for developing multiplexed methylation assays were described above. For gene expression analysis, we tested a nanostring assay and found it robust in FFPE RNA down to about 100 ng (Figure 5). However, most of the biopsy samples we processed in this task did not produce more than 50-60ng

Figure 5 Nanostring assay showing reproducible data in 100 ng FFPE RNA for genes at all expression levels tested. Lower input RNA only produced reliable expression data for highly expressed genes.



RNA. Furthermore, three of the selected genes (NAV1, LYST, and KIAA2026) had relatively low expression levels which required even a higher starting RNA. We have tested several pre-amp approaches and hope to be able to test these biomarkers in biopsies in the near future. Especially, we anticipate that the pre-amp based approaches will be successful in testing the skipped exon biomarkers described in task-1.

Task 4: Processing of case control and cohort biopsy samples

4.1-4.6 Processing of biopsy samples and testing the biomarkers: We developed working models in surgical samples as described below and are planning to test them in biopsy samples once we have working assays for analyzing selected biomarkers and reference housekeeping genes in biopsies.

4.6-4.7 Developing and testing models applicable in non-neoplastic bulk prostate tissue and in HGPIN:

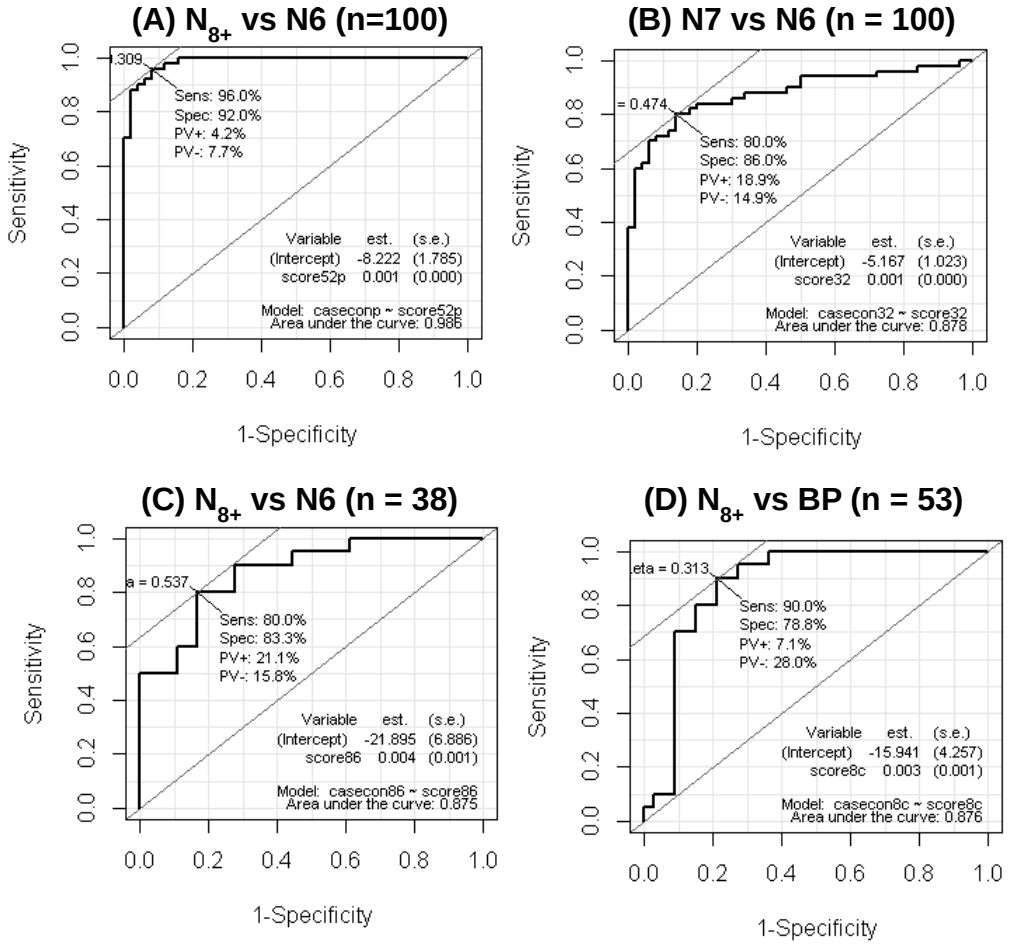
A molecular model for bulk samples: A logistic regression model based on 6 biomarkers were developed in our discovery set and tested in multiple validation datasets. Model coefficients were calculated in the original discovery set containing 50 N_{8+} and 50 N_6 as follows:

Score = (449.2* NAV1) + (141.9*LYST) + (56.9*ADD3) - (7.8*SMC5) - (67.3* CEP350) + (307.7* KIAA2026)

This model was tested in three validation sets, including one with 50 N_{8+} and 50 N_6 which were non-overlaping with the discovery set (Figure 6A), and another set of 50 N_7 and original 50 N_6 (Figure 6B), and a third and a fourth independent set containing N_{8+} (n=20), N_6 (n=18), and BP (n=33) (Figure 6C and 6D). In these tests, the model performance was acceptable. However, this model was not predictive in a another set containing 50 N_7 and N_6. We are examining this data to identify whether samples were correctly categorized or the biomarkers did not perform as expected.

A molecular model applicable in HGPIN samples: Based on three markers described in Figure 4C, we developed a logistic regression in 20 and 22 HGPIN samples in prostates containing GS6 and $GS8^+$ tumors, respectively. The score was calculated as:

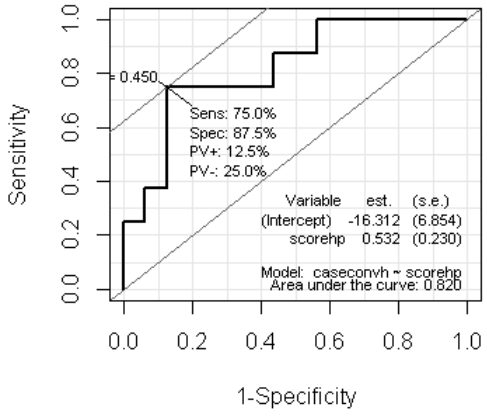
Figure 6: Testing the logistic regression model for bulk tissue in validation samples. In the 4 datasets tested, the performance of the model was acceptable.



HGPIN score = (7.0* LRRC4C) + (0.31* SUGT1) + (0.34* KLHL28)

This model was applied to an independent sample containing 16 and 8 HGPIN in prostates containing GS6 and GS8^{+} tumors, respectively, and had an acceptable performance (Figure 8)

Figure 8: AUC plot of the HGPIN model in independent samples (n = 24)



Additional related research activities

We received additional funding through Mayo Clinomics program (2015) as described in the report. Also, we were engaged in several studies that examined genomic DNA abnormalities associated with significant PCa. Through these activities, we developed a clinical model based on copy number variations of PTEN and another set of 4 loci to identify patients with GS6 biopsies who are at increased risk or PCa recurrence. This model is expected to be available to Mayo Clinic patients in 2017.

Reportable Outcomes:

- Set of gene expression markers distinguishing significant from indolent PCa which were validated in independent samples
- Novel methodology to identify skipped exons and a set of potential candidates
- Set of methylation markers that overlap with tumor methylation markers in distinguishing indolent from significant PCa.
- Two logistic regression models which for stratification of non-neoplatic prostate tissue and HGPIN samples in significant and indolent PCa
- Two manuscripts in preparation describing our work in the bulk and HGPIN samples.

Conclusion: This project provided an excellent opportunity to study prostate cancer field effect. We were able to analyze a sizeable set of gene expression and methylation data which was considerably larger than in the proposed application. These analyses have convinced us about the clinical utility of prostate cancer field effect and believe that we will have a testable molecular model based on expression biomarkers in the near future. We are especially interested in the skipped exon biomarkers which can be incorporated in sensitive PCR based assays. We are currently in the process of submitting two manuscripts describing our results in bulk and HGPIN samples. This project also provided the opportunity to develop collaborations with other prominent investigators in the field. With the opportunities that were provided in this career development grant and the sizable data that was obtained, we are in good standing to develop research projects aimed at prevention of PCa or early diagnosis of significant prostate cancer in over a million men who are at risk of prostate cancer in the US. each year