AWARD NUMBER: W81XWH-16-1-0737

TITLE: Developing a PTEN-ERG Signature to Improve Molecular Risk Stratification in Prostate Cancer

PRINCIPAL INVESTIGATOR: Tamara Lotan

CONTRACTING ORGANIZATION: Johns Hopkins University

REPORT DATE: Oct 2019

TYPE OF REPORT: Annual

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

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE
Oct 2019

2. REPORT TYPE Annual

3. DATES COVERED
09/30/2018 - 09/29/2019

4. TITLE AND SUBTITLE
Developing a PTEN-ERG Signature to Improve Molecular Risk Stratification in Prostate Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-16-1-0737

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Luigi Marchionni and Tamara L. Lotan (partnering PI)

E-Mail: marchion@jhu.edu and tlotan1@jhmi.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

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

Johns Hopkins University 1550 Orleans
Street CRB2, Rm 1M52 Baltimore, MD
21231

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Development 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

Prostate cancer (PCA) is a clinically and genetically heterogeneous and the development of a molecular classification is critical to distinguish lethal from indolent tumors and minimize overtreatment. Genomic alterations of the PTEN and ERG genes are among the most common in PCA and there is an interest in exploiting these alterations for routine risk assessment. We found that PTEN loss is most strongly associated with PCA death in patients whose tumors do not carry an ERG gene rearrangement, suggesting that ERG absence strengthens PTEN loss association with lethal progresion. Despite the widely accessible PTEN/ERG molecular classification, our understanding of their biological interaction along PCA progression remains very limited. Hence, in our study we will perform a comprehensive molecular profiling of well-annotated PCA samples in relation to PTEN and ERG status. Our goals are threefold: 1) to confirm that PTEN/ERG double negative tumors are the most aggressive; 2) to characterize the expression profiles associated with PTEN and ERG alterations; and 3) to determine whether such expression profiles can be used to improve PCA patient stratification into different risk groups.

15. SUBJECT TERMS

Prostate cancer, PTEN, ERG, ETS, MYC, cell cycle, gene expression, Cap Analysis of Gene Expression (CAGE)

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

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

TABLE OF CONTENTS

1. INTRODUCTION

Prostate cancer (PCA) is a clinically and genetically heterogeneous and the development of a molecular classification is critical to distinguish lethal from indolent tumors and minimize overtreatment. Recent technological advances have enabled extraordinary insights into molecular changes occurring in PCA and the *PTEN* and *ERG* genomic alterations have emerged as the most common in PCA. Furthermore, we have found that PTEN loss is associated with PCA death most strongly in patients carrying *ERG* rearrangements, hence there is an interest in exploiting such alterations for routine risk assessment. Furthermore, despite the fact that PTEN and ERG molecular classification is widely accessible, our understanding of their interaction during disease progression is very limited, and a molecular signature of PTEN/ERG loss in PCA is still lacking.

To address these issues, we have formed a collaborative, multi-disciplinary team – led by a urologic pathologist and computational biologist with expertise in PCA molecular pathology and cancer genomics – to perform a comprehensive molecular assessment of well-annotated prostate cancers in relation to PTEN and ERG status using existing and novel data. Our objectives are threefold: 1) to confirm that the tumors with loss of PTEN and lacking *ERG* rearrangement are among the most aggressive; 2) to characterize the *expression profiles* associated with PTEN and ERG alterations; and 3) to determine whether these *expression profiles* can improve the way we stratify prostate cancer patients into different risk groups.

Findings from our proposed research have the potential for both immediate and long-term clinical and translational research applicability. First, by analyzing several large clinical cohorts from multiple institutions, we will be able to confirm the performance of these biomarkers in patient risk stratification. Second, we will also be able to assess if and how PTEN/ERG *molecular signatures* correlate with lethal disease risk in comparison to currently available prognostic assays. Third, we expect to identify novel molecular alterations responsible for the distinct clinical and biological behavior of tumors based on PTEN and ERG status. Lastly, we will also generate a wealth of information about the biologic drivers of prostate cancer behavior, which shall then be utilized by the entire PCA research community.

2. KEYWORDS

Prostate cancer, PTEN, ERG, ETS, MYC, cell cycle, gene expression, RNA sequencing, Cap Analysis of Gene Expression (CAGE)

3. ACCOMPLISHMENTS

Below are listed tasks, subtasks, and accomplishments for research sites 1 (coordinated by the initiating PI, Dr. Marchionni), and site 2 (coordinated by the partnering PI Dr. Lotan).

SPECIFIC AIM 1 (Dr. Lotan)

Expected tasks and milestones are summarized below.

Specific Aim 1: Validate association of PTEN and ETS status with risk of lethal prostate cancer	**Timeline (Months)**
Major Task 1: Assessing prostatectomy cohorts on multiple tissue microarrays (TMA) for PTEN, ETS, and cell proliferation rate	1-36
Subtask 1: Perform immunostaining for PTEN, ERG and Ki-67 and in situ hybridization on tissue microarrays (TMAs) from JHU and MSKCC cohorts; immunostaining for Ki-67 on HPFS/PHS cohort	1-12
Subtask 2: Score immunostaining and in situ hybridization from Subtask 1 • Digitally scan all slides using Aperio CS slide scanning system in Johns Hopkins OTS Core facility • Segment TMAs and upload to web-based browser, TMAJ (http://tmaj.pathology.jhmi.edu/) • Dr. Lotan, Dr. Gopalan (MSKCC), and pathology fellow supervised by Dr. Lotan perform scoring. Image analysis software (FRiDA on TMAJ) to be used for Ki-67 scoring	13-24

Subtask 3: Analysis of immunostaining and in situ hybridization data from Subtask 2 • Multivariate models to assess association of PTEN/ETS status with metastasis and survival in JHU and MSKCC cohorts • Correlation of PTEN/ETS status with proliferation in JHU, MSKCC and HPFS/PHS cohorts	18-30
Milestone #1: Co-author manuscript on association of PTEN/ETS status with cell cycle gene expression, proliferation rate and risk of metastasis and death in multiple validation cohorts	31-36

Progress on Major Task 1 – Subtask 1: All immunostaining and *in situ* hybridization has been completed on the JHU and MSKCC cohorts.

Progress on Major Task 1 – Subtask 2: The scoring for ETS gene rearrangements (ETV1/4/5) (as well as PTEN and ERG) on the JHU tissue microarrays has been completed and analyzed. Ki-67 immunostaining on those arrays is completed. PTEN and ERG staining as well as Ki-67 staining is completed for these arrays as well. All arrays have been digitally scanned and are viewed on our TMAJ viewer (**Figure A**). However digital automated scoring of Ki-67 has been challenging using TMAJ since it is difficult to normalize the number of positive detected nuclei (brown) to the negative tumor nuclei (blue). This is because it is difficult to detect all negative tumor nuclei. After examining the possibility of simply normalizing the number of positive tumor cells to the total area of the spot (ie, density of positive cells per mm^2 of tumor), which is the easiest method, we have decided that this is suboptimal since it can be confounded by the amount of tumor nuclei sampled in the TMA spot (**Figure A**). In this analysis, we find that lower grade tumors, with fewer tumor nuclei will have inappropriately low Ki-67 density scores. Thus, we must manually annotate the tumor-containing region on all TMA spots and we are in the process of completing this (there are at least 400 spots on each TMA and 12 TMAs in the MSKCC cohort and 9 TMAs in the JHU cohort.

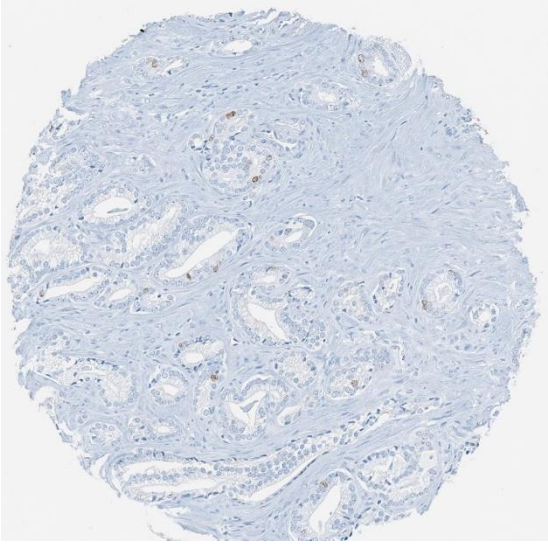

Figure A: Ki-67 labelling in JHU cohort (200x magnification)

Progress on Major Task 1 – Subtask 3: For the HPFS/PHS tissue microarrays, we have completed and analyzed Ki-67 immunostaining. Data comparing the percent of tumor cells labeling for Ki-67, stratified by PTEN and ERG status were presented in the prior progress report. While there was a statistically significant increase in Ki-67 labelling when comparing tumors that have PTEN loss and those that are PTEN intact, there is not a significant difference in Ki-67 labeling between tumors with PTEN loss and ERG expression and those with PTEN loss that do not express ERG (p=0.68).

We have finalized the analysis of PTEN and ERG on the MSKCC cohort and examined the correlation of these molecular alterations with clinical outcomes (lethal prostate cancer) in multivariable models. These results are presented in **Table A** and **Figure B** below, and were recently published in *Journal of Urology* (Haney NM, Faisal FA, Lu J, Guedes LB, Reuter VE, Scher HI, Eastham JA, Marchionni L, Joshu C, Gopalan A*, Lotan TL*. PTEN loss with ERG-negative status is associated with lethal disease after radical prostatectomy. *J Urol*. 2020, 203(2):344-350. *Equal Contribution. PMID: 31502941).

Table A: Multivariable models of association of PTEN-ERG status with lethal prostate cancer in the MSKCC cohort.

			Univariable		Multivariable*	
	No. Cases	No. Controls	HR (95% CI)	p Value	HR (95% CI)	p Value
PTEN:						
Intact	46	542	Referent	—	Referent	—
Loss	47	156	3.25 (2.16—4.88)	<0.001	1.87 (1.15—3.04)	0.012
ERG:						
Neg	62	384	Referent	—	Referent	—
Pos	30	300	0.64 (0.41—0.99)	0.043	0.64 (0.36—1.11)	0.113
PTEN/*ERG*:						
PTEN intact/*ERG* neg	35	328	Referent	—	Referent	—
PTEN intact/*ERG* pos	10	200	0.47 (0.23—0.96)	0.037	0.48 (0.18—1.26)	0.136
PTEN loss/*ERG* neg	27	56	3.76 (2.27—6.21)	<0.001	2.31 (1.29—4.14)	0.005
PTEN loss/*ERG* pos	20	100	1.84 (1.06—3.18)	0.030	1.09 (0.56—2.12)	0.809

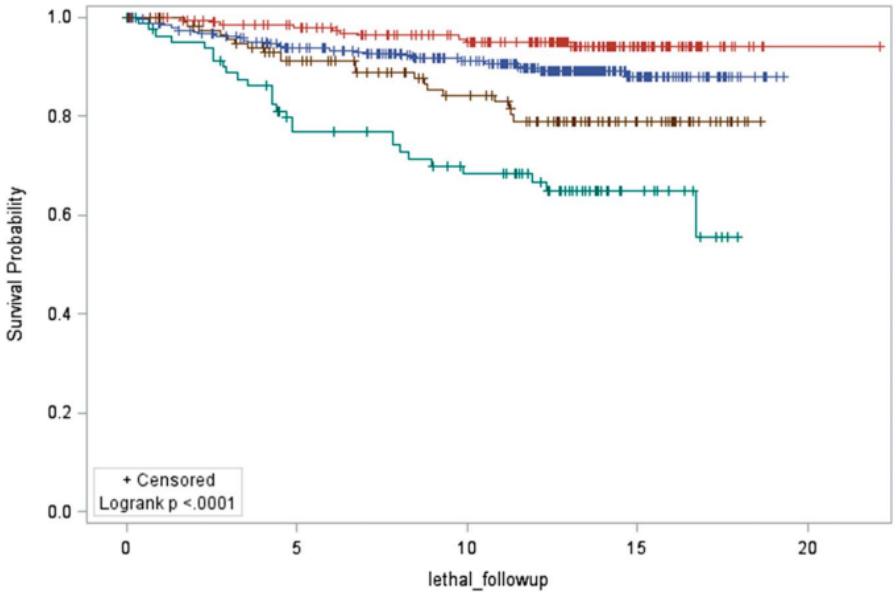

Figure B: Kaplan-Meier survival curves of freedom from lethal prostate cancer by PTEN and ERG status. Blue curve indicates PTEN intact and ERG negative in 363 patients. Red curve indicates PTEN intact and ERG positive in 210 patients. Green curve indicates PTEN loss and ERG negative in 83 patients. Brown curve indicates PTEN loss and ERG positive in 120 patients.

Training and professional development: Nothing to report

Results dissemination to communities of interest: Results from Major Task 1 – Subtask 3 were recently published in *Journal of Urology* (Haney NM, Faisal FA, Lu J, Guedes LB, Reuter VE, Scher HI, Eastham JA, Marchionni L, Joshu C, Gopalan A*, Lotan TL*. PTEN loss with ERG-negative status is associated with lethal disease after radical prostatectomy. *J Urol*. 2020, 203(2):344-350. *Equal Contribution. PMID: 31502941).

SPECIFIC AIM 2 (Dr. Marchionni)

Expected tasks and milestones are summarized below.

Specific Aim 2: Leverage multi-dimensional public domain data to discover genomic features and signaling pathways associated with PTEN loss in ERG-positive and ERG-negative PCa.	**Timeline (Months)**
Major Task 1: Exploratory analysis of genomics datasets	1-6
Subtask 1: Examine gene expression distributions and identify outliers and other potential problems: • Use statistical summaries and visualizations (*e.g.*, principal component analysis, hierarchical clustering) • Apply appropriate transformation to the data if required	1-6
Major Task 2: Classify tumors based on PTEN, ETS, and MKI67 status.	6-24

Subtask 1: Use the EM-algorithm to classify tumors as positive or negative based on the expression levels of PTEN, ETS family members, and MKI67	6-12
Subtask 2: Compare expression-based classification to IHC and in-situ based status from in Specific Aim 1	12-30
Subtask 3: Analysis of PTEN and ETS status in cohorts available from GenomeDX and the public domain • Multivariate models to assess association of PTEN/ETS status based on genes expression dichotomization with metastasis and survival in all cohorts • Correlation of PTEN/ETS status based on genes expression dichotomization with proliferation in all cohorts	12-24
Major Task 3: Comprehensive meta-analysis of differential gene expression programs modulated by PTEN and ETS status in prostate cancer and characterization of their biological and clinical correlates	12-30
Subtask 1: Use generalized linear model to identify genes differentially expressed and differentially modulated by PTEN and ETS in prostate cancer	12-24
Subtask 2: Identification of relevant biological processes and signaling pathways associated with PTEN/ETS molecular signatures in prostate cancer	18-30
Subtask 3: Development and validation of predictive models based on associated with PTEN/ETS molecular signatures in prostate cancer	24-36
Milestone #2: Co-author manuscript on comprehensive meta-analysis of genes and signaling pathways associated with PTEN/ETS status in prostate cancer	24-36
Milestone #3: Co-author manuscript on prognostic values of PTEN/ETS molecular signatures in prostate cancer	24-36

Progress on Major Task 1 – Subtask 1 and 2: These activities have been successfully completed.

Progress on Major Task 2 – Subtasks 2: These activities have been successfully completed.

Progress on Major Task 2 – Subtasks 3: This analysis is currently underway, in conjunction with the development of a prognostic signature (see below, section **Progress on Major Task 3 – Subtasks 3)**.

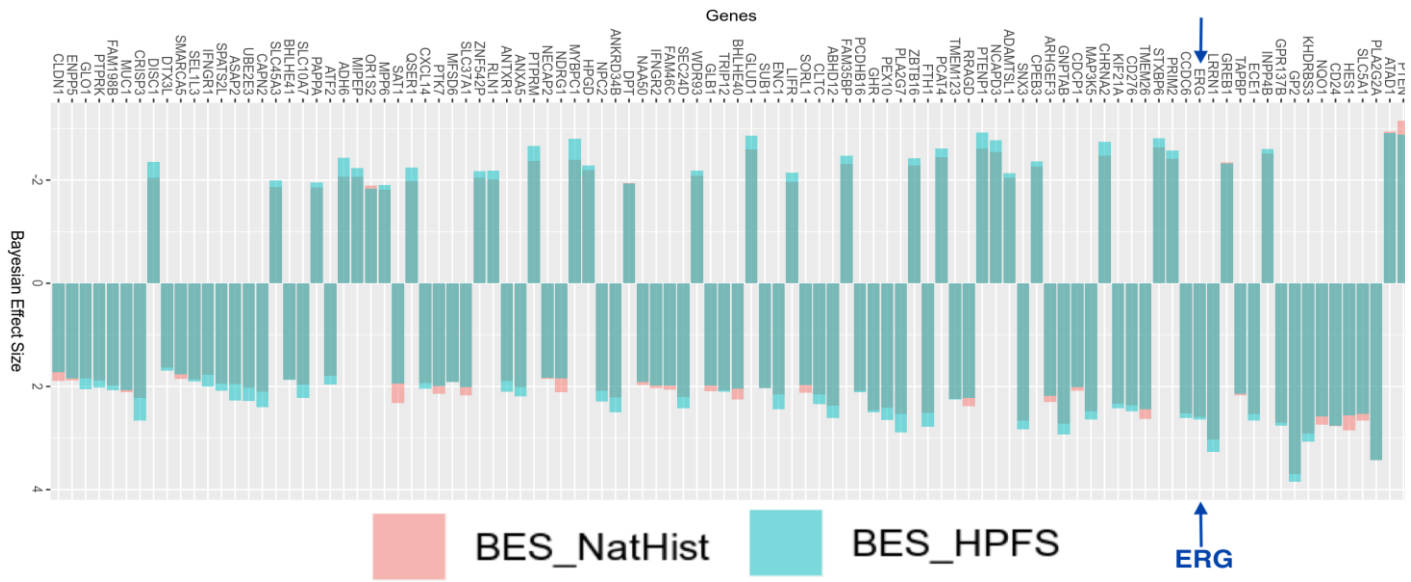

Figure C. PTEN signature from meta-analysis. PTEN signature obtained by multi-level model for cross-study detection of differential gene expression based on IHC calls on Natural History and HPFS cohorts. Figure shows the effect size of each cohort. ERG is one of the most upregulated genes associated with PTEN loss (blue arrows).

Progress on Major Task 3 – Subtask 1 and 2: During the first two years of project, we have developed and characterized in depth a consensus signature for PTEN loss using a meta-analytic approach. In the third year of

the project, we have investigated the association of this signature with ERG status. This analysis has revealed that the ERG gene itself is among the top upregulated genes in our PTEN loss sigature (**Figure C**). Based on this observation, we have therefore hypothesized that our PTEN signature could be heavily influenced by the ERG rearrangement, since this gene encodes a transcription factor. In order to test this hypothesis, we have therefore repeated the meta-analysis by splitting the samples by ERG status and then by fitting two separate Bayesian hierarchical models for differential expression by PTEN status. In the samples with ERG rearrangement, we observed a signature similar to the overall PTEN consensus signature we previously developed in year 2. On the contrary, in the samples without ERG rearrangement, we could not find any significant differences between samples with PTEN loss and PTEN intact.

This finding was surprising, given that PTEN is a powerful tumor suppressor capable of triggering important molecular and functional changes. We speculated that this result could be caused by two reasons: 1) PTEN loss in the absence of ERG rearrangement, does not impact the cell in any significant way; or 2) The absence of ERG rearrangement generates a high level of heterogeneity that makes it hard to estimate difference between PTEN-null and PTEN intact samples. The first hypothesis, however, is highly unlikely, given the fact that it is well-established that PTEN loss triggers deep changes in cellular metabolism and growth. Therefore, we performed experiments to test if the second hypothesis was true. To this end, we computed the hamming distance between samples with different ERG rearrangement status across distinct molecular data types (expression levels, copy number alterations, and mutations) both in the TCGA and HPFS cohorts. The mean hamming distance was computed by bootstrapping a subset of samples 1000 times. For all molecular data types and for both cohorts, we observed that the intra-group distances between the ERG positive samples (*i.e.*, those with ERG rearrangement) were always significantly higher than between ERG negative tumors, thus confirming our hypothesis (**Figure D**).

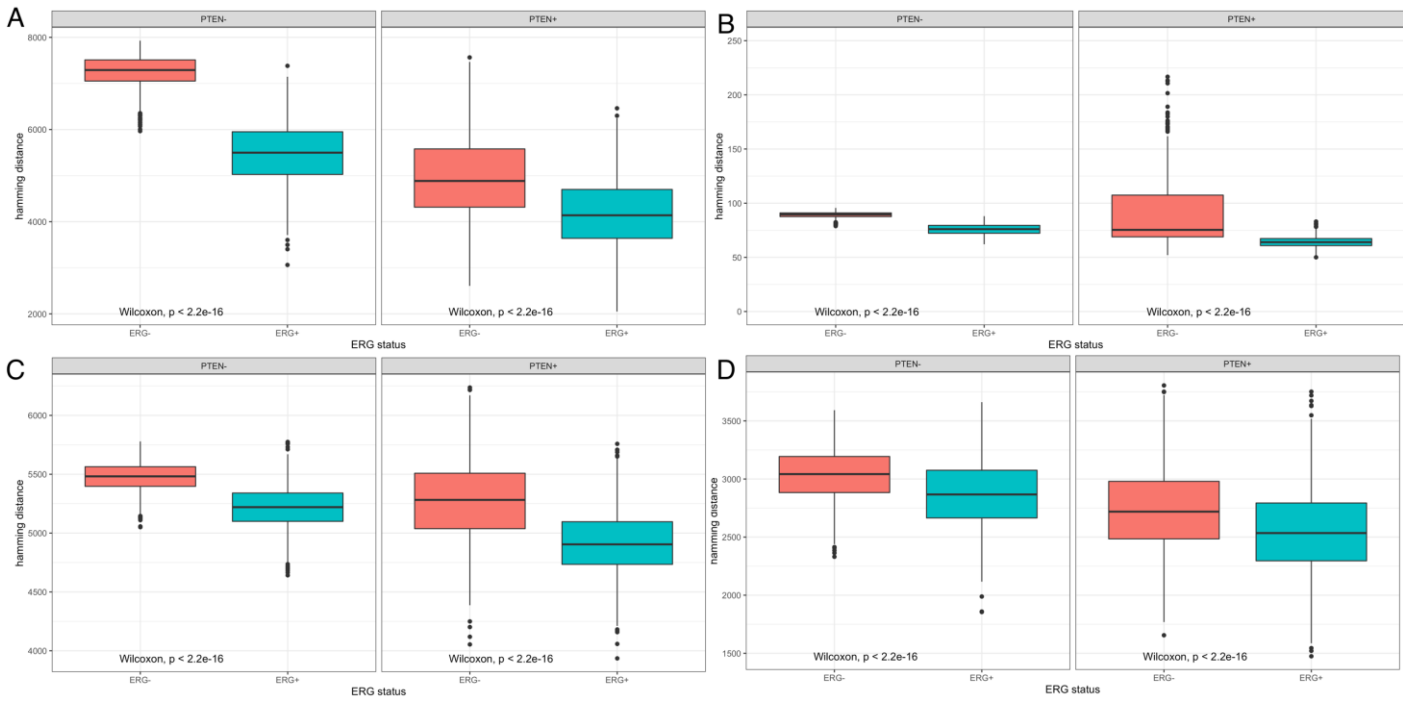

Figure D. Heterogeneity analysis in ERG positive and negative tumors. Average hamming distance based on 1000 bootstraps intra-samples between each group, showing that samples in absence of ERG rearrangement presented higher levels of heterogeneity (higher distances). **Panel A)** Hamming distance based on CNV in TCGA; **Panel B)** Hamming distance based on mutation in TCGA; **Panel C)** Hamming distance based on divergence expression levels in TCGA and **Panel D)** Hamming distance based on divergence expression levels in HPFS.

Progress on Major Task 2 and 3 – Subtask 3: During the third year of research, however, we have generated a prognostic gene expression signature for prostate cancer progression using a combination of gene expression data from the public domain, as detailed below. To this end, a total of 674 primary prostate cancer samples (from 3

distinct studies) were used for discovery of the gene signature, while an independent cohort of 248 samples was used for validation and signature performance assessment (see **Table B**).

GEO accession	GPL	Number of samples (metastasis cases)	Training/Testing
GSE55935	GPL10558	44(8)	Training
GSE51066	GPL5188	85(51)	Training
GSE46691	GPL5188	545(212)	Training
GSE116918	GPL25318	248(22)	**Testing**

Table B. Collected data sets showing the number of samples and the number of metastasis cases. 3 datasets were used for training and one data set (GSE116918) was used as an independent validation cohort.

First, we have performed a large scale differential expression analysis of gene expression data from different microarray platforms. We have identified 49 up-regulated and 26 down-regulated genes in prostate cancer metastasis cases. We have then further optimized this signature using a ”forward search” process reducing the original list to just 14 up-regulated and 17 down-regulated genes. Finally, we have combined the gene expression levels for these genes into a meta-score for use in subsequent analyses, including multivariable Cox proportional hazard model analyses with other clinical and pathological variables (Age, PSA, T-stage ,and Gleason grade).

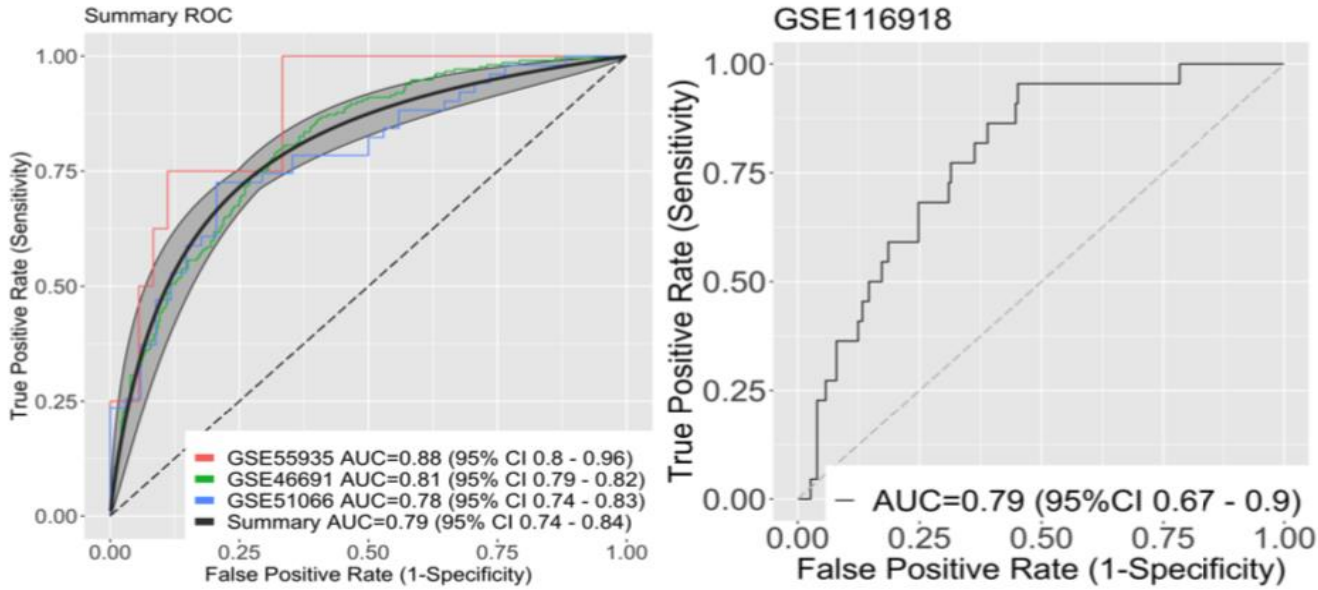

Figure E. Performance of the prognostic signature. ROC curves in the 3 training data sets with a summary ROC curve of all data sets combined (Left) and ROC curve in the independent testing data set (Right).

To assess the performance of our signature, we measured the area under the receiver operating characteristic curve (AUC). In the training the AUC ranged from 0.78 to 0.88 (**Figure E**, right panel), while in the testing cohort the AUC was 0.79 (**Figure E**, left panel), confirming the prognostic value of the signature.

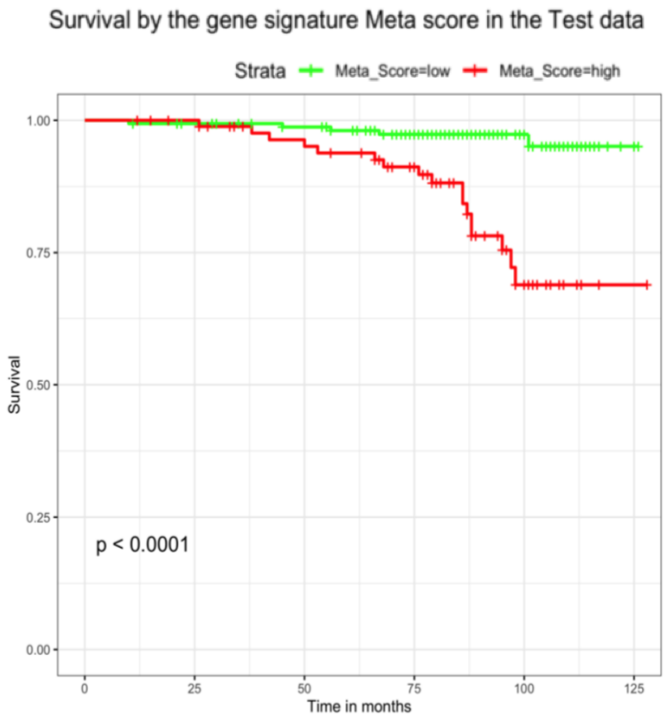

Figure F. Kaplan Meier analysis in the testing cohort. Kaplan Meier survival curves based on signature meta-score. Patients with low score have a better metastasis-free survival than those with a high score (p-value < 0.0001).

We performed Kaplan-Meier analyses in the testing cohort. Patients with higher signature meta-score had worse metastasis-free survival than those with lower score (p-value < 0.0001, see **Figure F**). Additionally, we also performed survival analyses using individual gene expression levels rather than the signature meta-score. In this analysis, 7 out of 14 up-regulated genes (TMSB10, IQGAP3, CST2, STC2, FOXH1, PTDSS1, HES6) were significantly associated with lower survival, while 8 out of the 17 down-regulated genes (AZGP1, NT5DC1, KCTD14, PTPRN2, UFM1, CCK, KIAA1210, POTEG) were significantly associated with better survival when highly-expressed.

Most importantly, the signature meta-score was the only significant variable in the multivariable Cox regression analysis performed in the testing cohort. The model included the meta-score together with age, PSA (prostate specific antigen), Gleason grade and T-stage, with a hazard ratio of 5.67 (95% CI : 2.02 - 15.9, see **Figure G**)

Collectively, these analyses show the importance of integrating gene expression data from multiple studies to identify accurate and consistent prognostic signatures. We are currently integrating this signature with PTEN and ERG classification obtained by the EM-algorithm, as previously described.

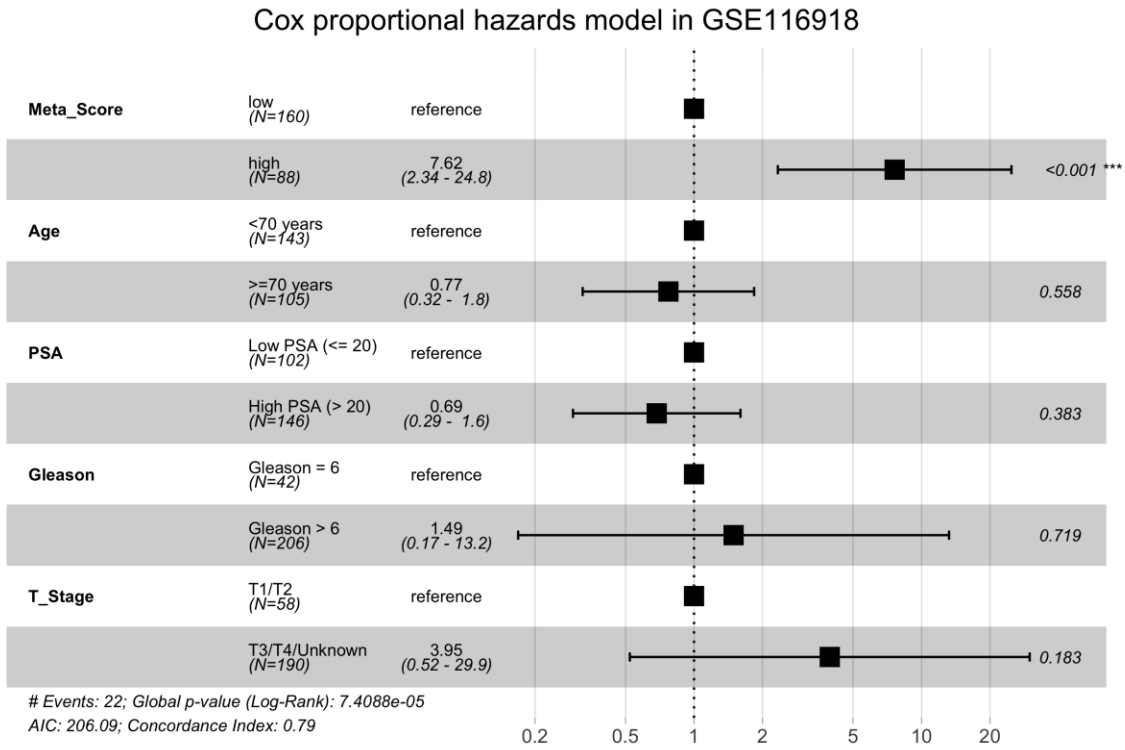

Figure G. Forest plot for Cox proportional hazards model results in the testing cohort. The signature meta-score is the only significant variable, outperforming other clinical and pathological variables.

Training and professional development: Nothing to Report.

Results dissemination to communities of interest: Nothing to Report.

SPECIFIC AIM 3 (Drs. Lotan and Marchionni)

Expected tasks and milestones are summarized below.

Specific Aim 3: Discover and validate gene regulatory and expression signatures associated with PTEN loss on genetically homogeneous ERG-positive and ERG-negative backgrounds.	**Timeline (Months)**
Major Task 1: Select 40 FFPE tumors from Johns Hopkins Surgical Pathology archives (20 ERG-positive and 20 ERG-negative, ETV1-negative). Within each group 10 have heterogeneous PTEN loss, 5 have homogeneous PTEN loss and 5 have intact PTEN by IHC	1-12
Subtask 1: Immunostaining 100 index tumors from Gleason 3+4=7 radical prostatectomies	1-6
Subtask 2: Score staining and select cases	4-8
Subtask 2: Punch blocks and prepare RNA for CAGE	8-12
Major Task 2: Perform CAGE analysis of the tumors resulting from Major Task 1 of Specific Aim3. Technology assessment and troubleshooting in collaboration with Dr. Carninci (RIKEN, Japan)	6-24
Subtask 1: CAGE library preparation, quality assessment, and sequencing • Performed at the Next Generation Sequencing Center (NGSC, Dr. Yegnasubramanian)	6-18
Major Task 3: Bioinformatics analysis of CAGE data generated in Major Task 2 of Specific Aim 3. Technology assessment and troubleshooting in collaboration with Dr. Carninci (RIKEN, Japan)	12-36
Subtask 1: CAGE short reads quality evaluation and alignment to the reference genome • Performed using NGSC computing cluster (Dr. Wheelan)	12-24
Subtask 2: Quantification of expressed genomic regions using CAGE tags • Performed using the School of Public Health (SPH) High Performance Computing Cluster (HPCC)	18-30
Subtask 3: Classification of expressed genomic regions, identification of active enhancers, promoters, and transcript • Performed using SPH HPCC	24-30
Subtask 4: Gene expression regulatory network reconstruction and analysis • Performed using SPH HPCC	24-36
Milestone #4: Co-author manuscript on CAGE analysis of PTEN/ETS status in prostate cancer	30-36

Progress on Major Task 1 – Subtask 1-3 (Dr. Lotan): These activities have been successfully completed.

Progress on Major Task 2 – Subtasks 1 (Dr. Marchionni): During years 1 and 2 of the proposal, we have tested CAGE and nanoCAGE sequencing protocols using high quality RNA obtained from several prostate cancer cell lines. These protocols were optimized for an Illumina mySeq instrument. In year 3 of the proposal, we have focused on optimizing the protocols for RNA samples prepared from tissue specimens. We also worked on developing optimal multiplexing protocols, in order to take advantage of the higher sequencing throughput of the Illumina HiSeq2500 instrument. To this end, we have obtained RNA from 12 tumor samples, prepared the nanoCAGE libraries, and then performed sequencing, as detailed below.

Tumor samples from Major Task 1 were multiplexed and the nanoCAGE protocol was used to the prepare the pooled libraries for sequencing. Before processing the samples on the Illumina HiSeq2500 instrument, we also performed after a successful mini-run on a mySeq instrument. For an unknown reason, however, the sequencer analytical pipeline failed to demultiplex the sequenced samples. We therefore extensively reviewed the experiments and performed an in depth troubleshooting. The quality control analysis in the whole dataset revealed that although the overall sequence quality was good (> 30 Phred Score), there was a very high level of duplicated reads (82.6% and 64.2% for R1 and R2, respectively). We therefore attempted to analyze the sequencing data using an alternative pipeline. Specifically, we tried to process the libraries using the TagDust2 software, which

also failed in demultiplexing the libraries. Next, we also aligned the reads to the human genome (hg38) in order to check if the sequences obtained were originating from the tumor RNA or from the sequencing kit by-products. In this analysis, only about ~6% of the reads aligned uniquely to the human genome, and around ~17% aligned to multiple loci, indicating that most of the sequences obtained were not originating from the human RNA from the tumors. Finally, we tried to align the sequences to the PhiX genome since this DNA was used during the library preparation to increase the library complexity. This analysis revealed that around ~46% of the reads aligned to the PhiX genome, highlighting potential problems during library preparation and/or sequencing (*e.g.*, incorrect primer loading in the Illumina HiSeq2500). For this reason, we are currently repeating the analysis using a mySeq sequencing instrument.

Progress on Major Task 2 – Subtasks 2, 3, and 4 (Dr. Marchionni)

In year 1 and 2 of the project, we have developed a comprehensive atlas of gene expression based on recent annotations from the FANTOM consortium based on CAGE-sequencing data (CAGE Associated Transcriptome, referred as FANTOM-CAT) and the recount2 database. This resource – called FC-R2 – accounts for gene expression summaries for over 109,000 genes across over 70,000 human samples.

In year 3 of the project, we have leveraged the FC-R2 resource and we have performed differential expression analysis between PTEN-null and PTEN-intact samples (see Aim 2 – Major task 3 – subtask 2). In this analysis, we have identified several novel lncRNAs associated with PTEN-loss that exclusively annotated or significantly expanded in the FANTOM-CAT gene models.

Among the most downregulated FANTOM-CAT exclusive genes were CATG00000038715, CATG00000079217 and CATG00000000330. Notably, CATG00000038715 is in close proximity of cytochrome P450 enzymes, more specifically CYP4F2 and CYP4F11. Both enzymes are involved in the process of inactivating and degrading leukotriene B4 (LTB4). LTB4 is a key gene in inflammatory response which are produced in leukocytes in response to inflammatory mediators and is able to induce the adhesion and activation of leukocytes on the endothelium. It has been recently demonstrated that leukotrienes can provide a selective proliferative advantage to cancer cells with intrinsically higher tumorigenicity, therefore downregulation of the intergenic promoter CATG00000038715 together with CYP4F2 and CYP4F11 could lead to increased leukotrienes levels upon PTEN loss resulting in the selection of PCa cells with higher tumorigenicity.

Expression of CATG00000038715 and CYP4F2 are highly correlated ($R=0.91$, $p < 2.2e-16$) and expression of the former was shown to be highly specific for prostate cancer (see **Figure H**). CATG00000079217 is closely located with the coding gene FBXL7 which has been shown to regulate Survivin stability which overexpression is known to lead to poor prognosis in several cancers. FBXL7 is also known to regulate mitotic arrest and mediate Class I MHC antigen processing and presentation. While expression of FBXL7 and CATG00000079217 showed only a weak correlation ($R=0.14$, $p < 7.4e-4$), CATG00000079217 expression was notably higher in prostate and breast cancer than in other cancers, and it showed to be moderately correlated with several prostate cancer biomarkers (e.g. KLK2, KLK3, STEAP2, PCGEM1, SLC45A3) ($R=0.37-0.57$, $p < 2.2e-16$). CATG00000000330 is located in the same loci of PTEN which has already been extensively document as a key oncogene in PCa.

Among the upregulated lncRNA FANTOM-CAT genes, CATG00000117664 was among the most upregulated lncRNA. Located in close proximity with the androgen regulated gene GPR158 which is reported to stimulate cell proliferation in prostate cancer cell lines, and it is linked to neuroendocrine differentiation. Expression between these genes are correlated ($R=0.54$, $p < 2.2e-16$) and CATG00000117664 expression was shown to be highly specific to prostate cancer.

Altogether, we have shown that these novel lncRNAs harmoniously track together with several coding mRNAs and lncRNAs already reported to be involved in PCa development and progression. This analysis reveals a plethora of lncRNAs, known or novel, that have never been associated with PCa and therefore empower further studies on the mechanisms leading to the development of PCa as well its more aggressive subtypes and aids in the future development of potential biomarkers and drug targets.

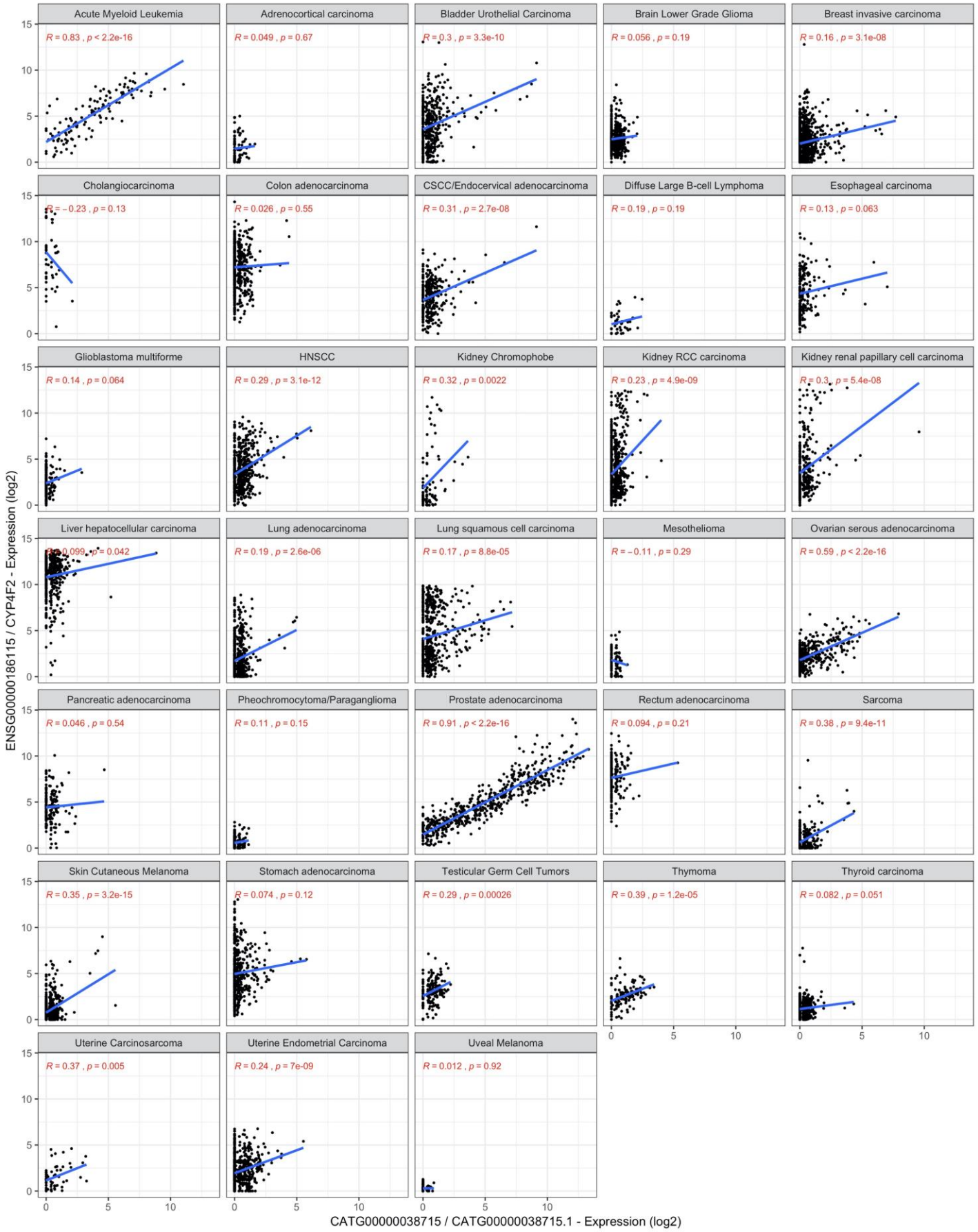

Figure H - Person correlation of the unknown gene CATG00000038715 and CYP4F2 across cancer types. CATG00000038715 and CYP4F2 expression are shown to be highly correlated in prostate cancer. Moreover, CATG00000038715 expression is shown to be highly specific to prostate cancer. With exception of leukemia cells, none of the other tumors expressed high levels of CATG00000038715.

Training and professional development: Nothing to Report.

Results dissemination to communities of interest: : Results from Major Task 2 – Subtasks 2, 3, and 4 were recently published in the following bioRxiv pre-print article: "Recounting the FANTOM Cage Associated Transcriptome", by Eddie-Luidy Imada, Diego Fernando Sanchez, Leonardo Collado-Torres, Christopher Wilks, Tejasvi Matam, Wikum Dinalankara, Aleksey Stupnikov, Francisco Lobo-Pereira, Chi-Wai Yip, Kayoko Yasuzawa, Naoto Kondo, Masayoshi Itoh, Harukazu Suzuki, Takeya Kasukawa, Chung-Chau Hon, Michiel JL de Hoon, Jay W Shin, Piero Carninci, FANTOM consortium, Andrew E Jaffe, Jeffrey T Leek, Alexander Favorov, Gloria R Franco, Ben Langmead, and Luigi Marchionni. doi: https://doi.org/10.1101/659490

This article has been provisionally accepted for publication in Genome Research.

4. IMPACT

Impact on prostate cancer research

We have successfully classified ERG status in all available datasets analyzed. Furthermore, we have successfully reproduced in an independent cohort our previous findings indicating that PTEN loss is associated with a worst prognosis in ERG/ETS-negative patients. *These results were published this year* (see above).

We have successfully applied highly validated IHC and in situ hybridization assays to determine PTEN and ETS status in 2 additional cohorts (MSKCC and JHU) with accompanying gene expression data for future analysis. Association of PTEN with Ki-67 proliferation index has been performed and analyzed for two datasets.

We have developed a consensus molecular signatures of PTEN loss in prostate, identifying the signaling pathways and biological processes associated with PTEN loss. We have also revealed that such molecular response to PTEN loss is mostly present in ERG positive tumors, and revealed a higher heterogeneity among samples in the ERG negative group.

We have generated a comprehensive catalog of expression of coding and non-coding genes using the FANTOM-CAT annotation and the recount2 atlas. Suing this resource will have identified several lncRNAs associated with PTEN and ERG status.

This project will add significantly to prostate cancer research by further refinement and validation of this prognostic biomarker as we develop expression signatures in the next reporting periods.

Impact on other disciplines: Nothing to Report.

Impact on technology transfer: Nothing to Report.

Impact on society beyond science: Nothing to Report.

5. CHANGES/PROBLEMS

Nothing to Report.

6. PRODUCTS

A manuscript on the PTEN consensus signature is under preparation.

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

Name:	Luigi Marchionni
Project role:	Initiating Principle Investigator
Researcher Identifier:	0000-0002-7336-8071 (ORCID)
Institution:	Johns Hopkins University
Nearest person month worked:	4 (rounded to 4)
Contribution to Project:	Dr. Marchionni coordinated the project, provided supervision of research activities provided by the fellows, and directly performed the analyses
Funding Support:	NA

Name:	Tamara Lotan
Project role:	Partnering Principle investigator
Researcher Identifier:	0000-0002-0494-9067 (ORCID)
Institution:	Johns Hopkins University
Nearest person month worked:	1
Contribution to Project:	Dr. Lotan coordinated the project, provided supervision of research activities provided by the fellows, and directly performed the analyses
Funding Support:	NA

Name:	Anne Jedlicka
Project role:	Co-investigator
Researcher Identifier:	NA
Institution:	Johns Hopkins University
Nearest person month worked:	1
Contribution to Project:	Dr. Anne Jedlicka coordinated the experiments with CAGE
Funding Support:	NA

Name:	Amanda Dziedzic
Project role:	Research specialist
Researcher Identifier:	NA
Institution:	Johns Hopkins University
Nearest person month worked:	1
Contribution to Project:	Ms. Amanda Dziedzic performed the experiments with CAGE

Name:	Wikum Dinalankara
Project role:	Post-doctoral fellow
Researcher Identifier:	NA
Institution:	Johns Hopkins University
Nearest person month worked:	5 (rounded to 5)
Johns **Contribution to Project:**	Dr. Dinalankara performed bioinformatics and statistical analyses under Dr. Marchionni supervision
Funding Support:	NA

Name:	Eddie Luidy-Imada
Project role:	Post-doctoral fellow
Researcher Identifier:	NA
Institution:	Johns Hopkins University
Nearest person month worked:	12
Contribution to Project:	Dr. Luidy-Imada performed bioinformatics and statistical analyses under Dr. Marchionni supervision
Funding Support:	NA

Name:	Diego Sanchez Martinez
Project role:	Graduate Student
Researcher Identifier:	NA
Institution:	Johns Hopkins University
Nearest person month worked:	8
Contribution to Project:	Dr. Sanchez Martinez performed bioinformatics and statistical analyses under Dr. Marchionni supervision for developing the prognostic signature.
Funding Support:	NA

Name:	Lotte Mulder
Project role:	Student
Researcher Identifier:	NA
Institution:	Johns Hopkins University
Nearest person month worked:	2
Contribution to Project:	Ms. Mulder performed bioinformatics and statistical analyses under Dr. Marchionni supervision for developing the prognostic signature.

Funding Support:	NA
Name:	Ericka M. Ebot
Project role:	Co-investigator
Researcher Identifier:	NA
Institution:	Harvard T.H. Chan School of Public Health
Nearest person month worked:	1 (rounded to 1)
Contribution to Project:	Dr. Ebot provided analytical support for the PHS/HPHS cohorts
Funding Support:	NA
Name:	Anne Jedlicka
Project role:	Co-investigator
Researcher Identifier:	NA

Change in active other support

Dr. Marchionni:

- No longer supported by 1U54RR023561-01A1 (Ford)
- No longer supported by KKESH (Eberhart) – this award is completed
- No longer supported by W81XWH-12-PCRP-TIA – this award is completed
- No longer supported by R01CA163594 (Sidransky) – this award is completed
- PC141474 (Tomlins and Schaeffer) – this award is completed
- R21 AI124776-01 (Romerio) – this award is completed
- 1R01CA211695-01A1 (Hurley) – this award is completed
- R01 PA-13-302 (Marchionni) – this award is now active and moved from pending
- R01CA206027 (Sidransky/Hoque) – this award is now active and moved from pending
- R01CA208709 (Sidransky/Hoque) – this award is now active and moved from pending
- W81XWH-16-PCRP-IDA (Lupold) – this award is now active and moved from pending
- U01CA231776 (Marchionni/Tran/Hann) – this award is now active and moved from pending
- R01CA235681 (Hahn) – this award is now active and moved from pending
- W81XWH-19-1-0292 (Lotan) – this award is now active and moved from pending

Dr. Lotan:

- New Award W81XWH-19-1-0292 (Lotan[PI], Title: Epigenomic Landscape of Primary Prostate Cancer in African-American Men, 10% effort)
- New Award W81-XWH-19-1-0781 (Asrani[PI], Title: mTORC1 Regulates MiTF Expression and Lysosomal Biogenesis, 2% effort)
- New Award R01 CA238218 (Pienta [PI], Title: Dissecting the prostate cancer diaspora, 1% effort)
- New Award PC180810 (Luo [PI], Title: Genetic and genomic determinants of homologous recombination repair deficiency as treatment selection markers for lethal prostate cancer, 5% effort)
- New Award PC180375 (Isaacs[PI], Title: Discovery and Functional Analyses of Susceptibility Genes for Lethal Prostate Cancer, 5% effort)
- No longer supported by W81XWH-15-1-0661, R01CA211695, RSG-17-160-01-CSM.

Other organizations were involved

Organization Name: Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA
Organization Name: Memorial Sloan Kettering Cancer Center, New York, NY, USA

Partner's contribution to the project

Collaboration: Dr. Ericka Ebot (Harvard) provided analytical support for the PHS/HPHS cohorts (<1 person/month effort).

Dr. Anu Gopalan (MSKCC) is a pathologist who created the MSKCC TMAs described above and she has participated in Ki-67 scoring and data analysis of these materials after providing them to us (<1 person/month effort).

8. SPECIAL REPORTING REQUIREMENTS

This project (W81XWH-16-1-0739) is a collaborative award with Dr. Tamara Lotan (Partnering PI, award W81XWH-16-1-0737).

9. APPENDICES

None