Award Number: W81XWH-11-1-0379

TITLE: A Systems Biology Approach to Link Nuclear Factor Kappa B Activation with Lethal Prostate Cancer

PRINCIPAL INVESTIGATOR: Christopher Sweeney

CONTRACTING ORGANIZATION: Dana Farber Cancer Institute
Boston, MA 02115

REPORT DATE: May 2014

TYPE OF REPORT: Annual

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

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
May 2014	Annual	1May2013-30Apr2014

4. TITLE AND SUBTITLE
A Systems Biology Approach to Link Nuclear Factor Kappa B Activation with Lethal Prostate Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-11-1-0379

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Christopher Sweeney

E-Mail: Christopher_sweeney@dfci.harvard.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

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

Dana Farber Cancer Institute

Boston, MA 02115

8. PERFORMING ORGANIZATION REPORT

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: It is appreciated that NFκB activation can either promote cancer cell death or cancer cell survival – the outcome being dependent on the context of parallel biological processes. Notably, the presence of tumor suppressors influences the outcome. Our bioinformatic approach to define a cancer promoting NFκB gene activation signature is well suited for accounting for the varied and opposing roles of NFκB activation. Our prostate cancer specific work has identified absence of a unique set of tumor suppressors that leads to cancer cell survival and ultimately lethal prostate cancer. It is also now appreciated that indiscriminate inhibition of NFκB activation may be problematic as this may block the anti-cancer effect of NFκB activation. As such the increased understanding of NFκB activation's "context dependency" adds further support for the work we are doing. In year one and two of the project we have found elevated cytokines and presence of T. Vaginalis at time of diagnosis of prostate cancer are not associated with higher grade disease nor risk of relapse after prostatectomy. In Year 3 of this grant we have curated the publically available gene expression profile (GEP) data of lethal vs non-lethal prostate cancer and also completed the preparation of the 26,000 GEP data from the Harvard School of Public Health cohorts (relapsed vs not relapsed post prostatectomy). We have also identified a SNP that was associated with lethal disease using a SNP selection process called "dense module GWAS". We have also been collecting TRUS biopsies of patients who progressed on surveillance vs no progression and relapse vs not relapse post prostatectomy to test our SNP and gene sets. In the year 4 no cost extension we will assess whether our candidate SNP and gene set associated with NFκB can reliably identify patients with Gleason 6 low volume prostate cancer not in need of therapy. This work will distinguish itself from the Polaris and Genomic Health gene signatures by being able to select patients for pharmacologic intervention with NFκB inhibitors under development for those with Gleason 6 cancer but with a cancer promoting NFκB activation signature and hopefully obviate the need for radiation of prostatectomy in those without evidence of cancer promoting NFκB activation signature.

15. SUBJECT TERMS
Prostate Cancer, Inflammation, Lethal Prostate Cancer

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
U	U	U	UU	23	19b. TELEPHONE NUMBER *(include area code)*

TABLE OF CONTENTS

INTRODUCTION:

The 4% per year decline in the age specific prostate cancer mortality in the USA has come at the cost of treating a significant number of men who did not require therapy and are living with adverse events that diminish their quality of life. To identify patients with lethal prostate cancer we are deploying a "systems biology approach" to develop a risk scoring system. The systems biology approach is making use of the epidemiological, clinical, pathological and biological data that has implicated Nuclear factor kappa B (NFκB) activation in the development of lethal prostate cancer. Specifically, we hypothesize that lethal prostate cancer results from exogenous insults causing NFκB activation that sets up a vicious cycle with further inflammatory insults and culminates in sustained NFκB activation and carcinogenic changes in the microenvironment. This persistent activation results in progression of prostate cancer to a lethal disease. We aim to identify patients with lethal prostate cancer using a systems biology approach focused on the NFκB pathway which will enable the construction of a risk scoring system to identify patients with localized but potentially lethal prostate cancer in need of therapy and patient who can safely avoid therapy.

BODY:

New data since the original submission of this grant adds further support to the role for inflammation and more specifically nuclear factor kappa B (NFκB) activation in the development of cancer in general and prostate cancer specifically. More importantly it has been realized that NFκB activation can either promote cancer cell death or cancer cell survival – the outcome being dependent on the context of parallel biological processes. Specifically, our data to date has found that absence of a unique set of tumor suppressors in the context of other NFκB genes is associated with lethal outcome[1]. Notably, absence of PTEN was not one of the tumor suppressors. Moreover, drugging NFκB activation by blocking IκK (upstream of NFκB) is problematic due to non-NFκB effects[1]. It is also now appreciated that indiscriminate inhibition of NFκB activation may be problematic as this may block the anti-cancer effect of NFκB activation. As such the increased understanding of NFκB activation's "context dependency" adds further support for the work we are doing.

Brief Summary:

Aim 1A: In years 1, 2 and 3 we have developed a robust network of the NFκB pathway as perturbed in prostate cancer. We leveraged recent advances in Bayesian data integration to simultaneously provide information specific to biological contexts and individual biomolecular mechanisms. We developed this method to identify mechanisms of interaction surrounding NFκB during its activity in cell death, inflammation, adhesion and differentiation. We integrated 18 gene expression datasets chosen to be particularly informative for functional relationships in prostate cancer and 860 non-disease datasets from gene expression, physical interactions, and functional associations. Prior knowledge was further included to inferred genome-wide networks for 442 biological processes (Gene Ontology), including 7 mechanisms of interaction ranging from general functional relationships through specific physical and regulatory activities. Among all inferred networks we focused on 9 biological context networks most informative in prostate cancer. We have investigated NFκB pathways in prostate cancer by identifying 8 genes that were highly functionally related with NFκB in these contexts and highly differentially expressed in prostate cancer microarray experiments. The simultaneous prediction of biomolecular mechanisms finally led to a newly predicted NFκB pathway specific to prostate cancer. Having created the network, we can now assess whether a parsimonious gene set can identify men with prostate cancer destined to progress versus not progress.

Aim 1B and C: In year one and two of the project we have found elevated cytokines and presence of T. Vaginalis at time of diagnosis of prostate cancer are not associated with higher grade disease nor risk of relapse after prostatectomy.

Aim 1D: We have identified a SNP that was associated with lethal disease using a SNP selection process called "dense module GWAS".

Aim 2 and 3: In the year 4 no cost extension we will assess whether our candidate SNP and gene set associated with NFκB can reliably identify patients with Gleason 6 low volume prostate cancer not in need of therapy. This work will distinguish itself from the Polaris and Genomic Health gene signatures by being able to select patients for pharmacologic intervention with NFκB inhibitors under development for those with Gleason 6 cancer but with a cancer promoting NFκB activation signature and hopefully obviate the need for radiation of prostatectomy in those without evidence of cancer promoting NFκB activation signature. We have also been collecting TRUS biopsies of patients who progressed on surveillance vs no progression and relapse vs not relapse post prostatectomy to test our SNP and gene sets. In the year 4 no cost extension we will assess whether our candidate SNP and gene set associated with NFκB can reliably identify patients with Gleason 6 low volume prostate cancer not in need of therapy. This work will distinguish itself from the Polaris and Genomic Health

gene signatures by being able to select patients for pharmacologic intervention with NFκB inhibitors under development for those with Gleason 6 cancer but with a cancer promoting NFκB activation signature and hopefully obviate the need for radiation of prostatectomy in those without evidence of cancer promoting NFκB activation signature.

Task 1. **Identify individual features of NFκB activation which are associated with lethal disease. (Months 1 to 18)**

Task 1A: *Perform gene profiling of tumors and determine whether a set of genes and/or proteins indicative of NFκB activation are associated with lethal prostate cancer. Data will be available at time of commencing the project on 350 patients and we will generate new data on 154 more patients. Data mining and analysis of existing data will be performed to define the 40 gene panel to be assessed for correlation with lethal disease. (Month 1 to 18)*

Accomplishments: In years 1 to 3 we have (i) developed putative gene sets from the initial data-mining efforts of publically available data-sets which are serving as our "discovery gene sets"; (ii) successfully deployed the "Nugen-Affy" assay as a reliable approach for whole genome expression analysis of lethal versus non-lethal prostate cancer – the "discovery gene-sets" are now being inputted into this data-set to define the "training set" and (iii) commenced creation of the Tissue Micro-Arrays (TMA) and extraction of nucleic acids for creation of an independent "validation set" of lethal versus non-lethal prostate cancer.

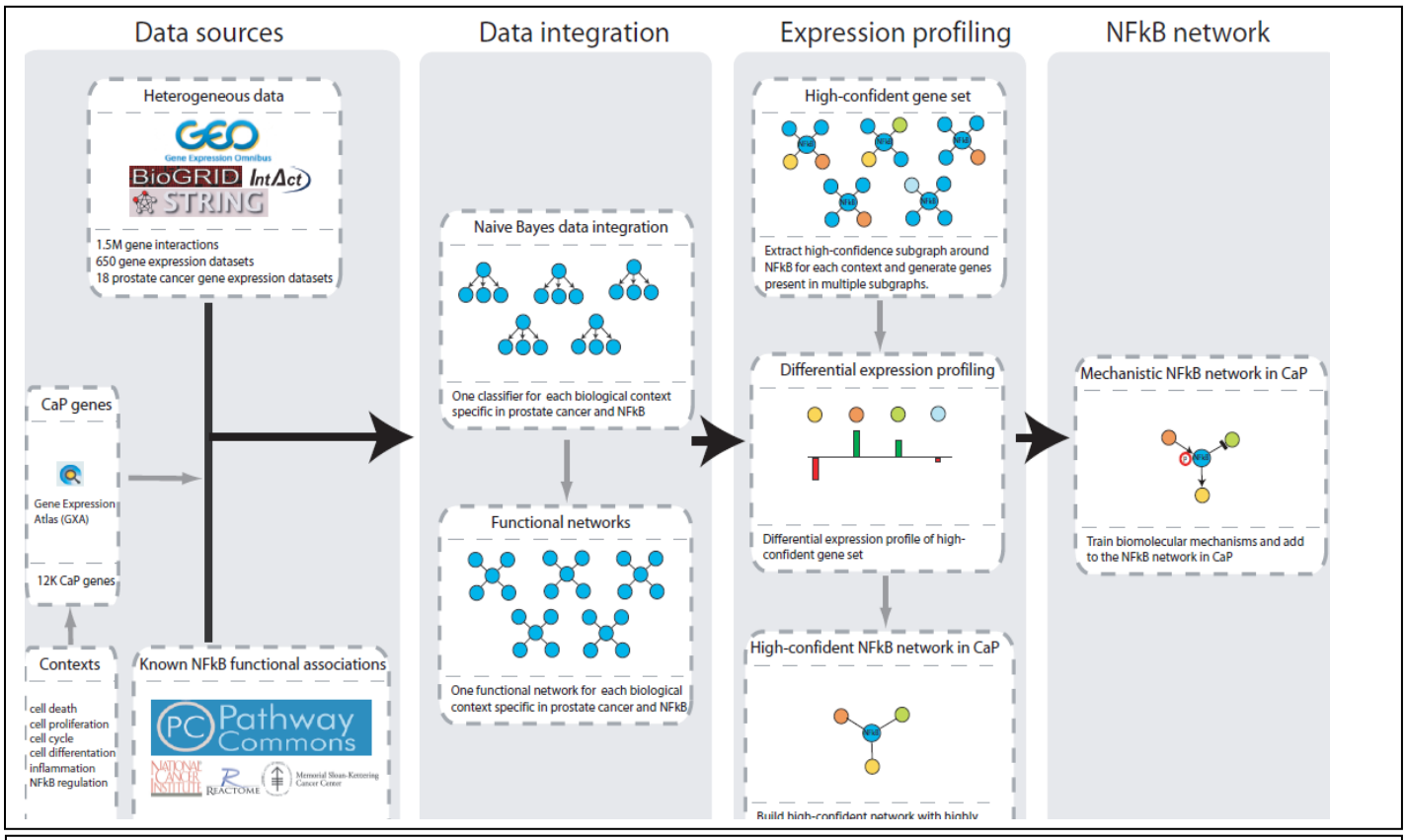


Figure 1: Schematic Summary of Data-mining Process:

- The left columns depict the mining of the 878 publically available databases which lead to the creation of the biological context specific networks.
- High-confidence subgraphs around the NFκB gene were identified in each context specific network and assessed by an additional set of hierarchical mechanism-specific learners to create a complete inferred biomolecular pathway. This identified both characterized and novel NFκB interactors in prostate cancer.
- The 351 selected NFκB related genes created the High-confidence NFκB network and 271 of these were in the 6096 gene DASL gene expression database annotated with lethal versus not lethal outcome from Physicians Health Study cohort of patients (middle figure of third panel). This was used to refine the gene set to those associated with lethal prostate cancer (31 genes).
- The 31 genes were then used to refine the network to define a network of NFκB cancer promoting genes in prostate cancer (fourth panel).

(i) Data-Mining: The research team has completed analyses of total of 878 expression and interaction datasets using context-specific Bayesian learning (Park el al 2010). These datasets included integration of 18 curated prostate cancer expression datasets and GEO[2] and ArrayExpress[3] and cancer-specific arrays from these repositories, and non-condition-specific genomic data such as physical and genetic interactions from BioGRID[4] and IntAct [5], transcriptional regulatory relationships from Transfac[6] and cisRED[7], and miRNA data from miRBase[8]. High-confidence subgraphs around the NFκB gene were identified in each and assessed by an additional set of hierarchical mechanism-specific learners to create a complete inferred biomolecular pathway. These resources provided literally billions of datapoints that have been integrated with our prostate cancer-specific clinical and genomic data, using methodologies developed by members of our research team[9-12]. Such data is key both for enriching detailed mechanistic models of prostate cancer development at the molecular level and, as has been done previously for genetic data[13,14] differentiating common functional variation in the general population from causal variation specific to lethal prostate cancer.

Using this approach 351 genes were identified for further exploration: Specifically, we focused on transcripts (i) correlated with NFκB activation in external expression data from GEO and ArrayExpress. Similar systems approaches have been highly successful in illuminating the entirety of the biomolecular pathways contributing to basic biology phenotypes in model organisms [15,16]. We contend that this integrative modeling is critical to understanding and detecting the development of lethal prostate cancer and will define a critical set of genes ("gene-panel") indicative of NFκB activation and in turn lethal prostate cancer.

The second step in this process was to define NFκB networks that correspond to distinct biological processes. We chose to develop 11 networks from a total of 442 different biological or biochemical processes. The key determinant was whether the biological context was related to cancer biology and associated with the hallmarks of cancer and in turn lethal prostate cancer or NFκB related biology. The 11 contexts analyzed were (i) cell death; (ii) cell migration; (iii) cytokine metabolic process; (iv) mesencyhmal cell differentiation; (v) positive regulation of NFκB, (vi) regulation of cell cycle (vii) regulation of cell differentiation; (viii) regulation of cell motion; (ix) regulation of cell proliferation; (x) stem cell maintenance; (xi) vasculature development. Four of the 11 networks are depicted below.

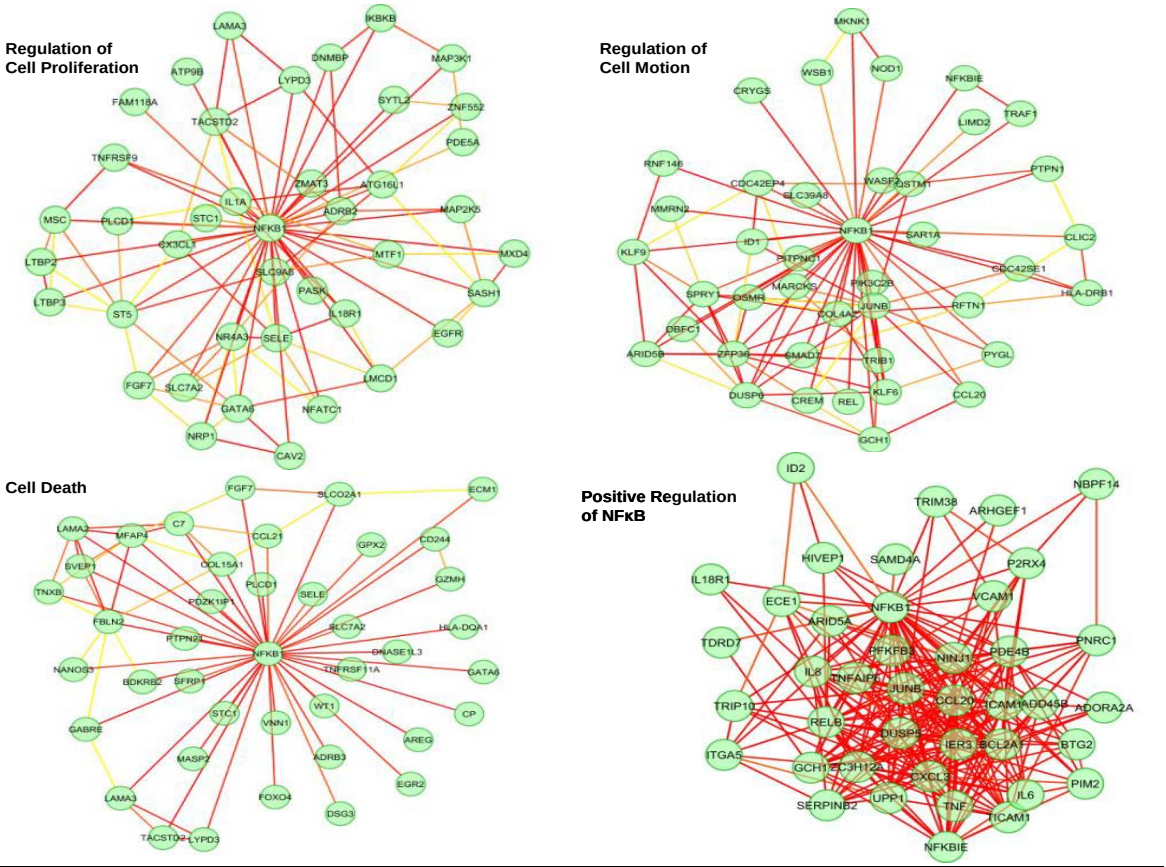
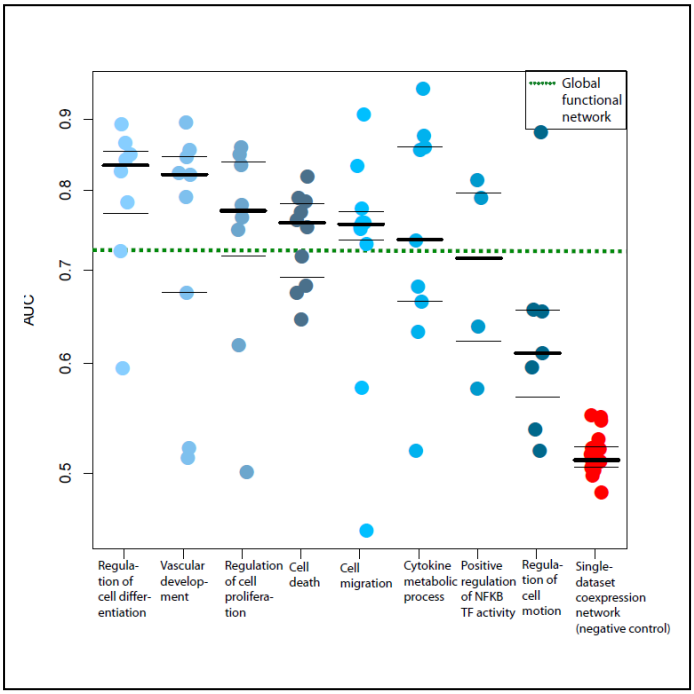


Figure 2: Representative NFκB networks

The notable findings from this work are that there are distinct networks for the unique biological processes. The team then computed for each gene the frequency of co-occurrence among all genes from these 11 networks. It was noted that NFκB1 co-occurs in all 11 context-specific subnetworks (as expected), NFκBIE in 5, CCL20 in 4. Besides NFκB1 there is no other gene that co-occurs in all 11 context-specific sub-networks.



The ability of the NFκB context specific networks to predict biomolecular mechanisms in pathway-specific functional relationship networks in prostate cancer was assessed using receiver operator curves. Figure 3 shows the performance of 8 context-specific inferred networks compared with the performance of a non-context-specific global inferred network (green dotted horizontal line) and a control of coexpression networks from 18 curated prostate cancer specific single expression datasets. 7 of 8 predictions all achieved AUCs over 0.7 for all specific mechanistic interaction types and over 0.75 for general functional associations, both using 10-fold gene-holdout-based cross-validations. The cell death network has so far included several of the highest-confidence links between NFκB1 and characterized examples such as CCL2[17] (regulatory). HDAC1[18] (phosphorylation) and IKBKB[19] (physical). This computational method easily scales to integrate thousands of experimental results and to identify those data most informative regarding specific putative mechanisms of interaction in pathways surrounding genes of interest in cancer.

Identifying NFκB related genes associated with lethal prostate cancer.
Having defined NFκB related genes in prostate cancer, we then sort to determine which genes are associated with lethal prostate cancer in a data-base of patients from the Physicians Health Study with clinical annotation connected to gene expression profile data developed from the 6,096 gene DASL platform. 217 of the 351 genes from the context specific networks were on the 6,096 gene DASL.

An assessment of the differential gene expression was done of the 217 NFκB related genes to identify genes associated with lethal prostate cancer. There were 115 PHS patients (83 indolent, 32 lethal). The DASL data is expressed as difference between the mean log expression in the lethal group and in the indolent group. When the expression is larger in the indolent group, the result has minus indicating the gene is decreased in the lethal group. Differential gene expression was inferred using package *limma*. An analysis was performed with and without Gleason grade as an additional covariate, Gleason scores are recoded as "low" (<7), "med" (=7), and "high" (>7).

Analysis **without** Gleason grade as a covariate:

- PHS subset - 186 of the 6096 genes on the DASL platform were differentially expressed between men who underwent a prostatectomy and relapsed and died of prostate cancer (N=32 lethal disease) compared with those who did not relapse after a prostatectomy (N=83: cured by prostatectomy or had disease that was indolent and did not require a prostatectomy).
- FDR correction was performed using Benjamini-Hochberg method.

	logFC	t	adj P Val	rank	observed score
FOSB	-1.22	-5.53	0.00	11.00	-1.00
ZFP36	-1.07	-5.46	0.00	12.00	-1.00
ATF3	-0.96	-5.22	0.00	15.00	-1.00
EGR2	-1.11	-5.13	0.00	18.00	-1.00
JUNB	-0.81	-4.98	0.00	25.00	-1.00
NR4A3	-1.19	-4.65	0.00	35.00	-1.00
SELE	-0.83	-4.41	0.00	46.00	-1.00
FOSL2	-0.39	-4.25	0.00	52.00	-1.00
BTG2	-0.39	-4.20	0.01	58.00	-1.00
HBEGF	-0.65	-4.11	0.01	66.00	-1.00
SFRP1	-0.79	-4.01	0.01	73.00	-1.00
NEDD9	-0.48	-3.86	0.01	85.00	-1.00
CXCL2	-0.81	-3.80	0.01	96.00	-1.00
DUSP5	-0.76	-3.72	0.02	112.00	-1.00
TRIB1	-0.35	-3.64	0.02	124.00	-1.00
CX3CL1	-0.60	-3.61	0.02	127.00	-1.00
IL1B	-0.57	-3.48	0.03	143.00	-1.00
NPR3	-0.80	-3.47	0.03	145.00	-1.00
CEBPD	-0.53	-3.46	0.03	148.00	-1.00

- 19 of these genes were associated with the NFκB network.

- Swedish subset: 129 genes are differentially expressed at 0.05 FDR level between lethal and indolent subgroups. FDR correction was performed using Benjamini-Hochberg method. 3 genes found to be significant in our analysis belong to the 271 NFκB gene set.

	logFC	t	adj P Val	rank	observed score
SLC39A8	-0.35	-4.20	0.01	28.00	-1.00
SLCO2A1	0.30	4.13	0.01	30.00	1.00
KLF10	0.22	3.96	0.01	41.00	1.00

- It is of note, the tissue analyzed from this analysis was obtained by TURP and is disease arising from the central gland with a different biology to disease from peripheral portions of the prostate gland.
- Patients in this data-set were also managed with "watchful waiting" and did not undergo a prostatectomy.

Given, the lack of uniformity in tissue collection, management and biology between the PHS and Swedish subsets, we elected to focus on the PHS cohort for creation of a discovery gene set of cancer promoting NFκB gene activation. The PHS represent the more commonly ascertained tissue of prostatectomy and TRUS biopsy.

Analysis **with** Gleason Grade as a covariate: On account of trying to improve on the prognostic ability of Gleason Score, we performed analysis with Gleason Grade as a covariate in the PHS prostatectomy series. 0 out of 6096 genes represented on a DASL platform are differentially expressed at 0.05 FDR level between lethal and indolent subgroups in this analysis. (FDR correction was peformed using Benjamini-Hochberg method). When there is no multiple testing correction, there are 384 differentially expressed genes with p-values below 0.05. 31 of these genes belong to the NFKB 271 gene set.

Gleason Score	Non-lethal	Lethal
6	15	0
7	58	9
8	7	11
9	3	9
10	0	3

	logFC	t	P Value	adj P Val	rank	observed scc
ZFP36	-0.82	-3.36	0.00	0.31	20.00	-1.00
JUNB	-0.68	-3.34	0.00	0.31	22.00	-1.00
IL1B	-0.68	-3.31	0.00	0.33	23.00	-1.00
FOSB	-0.88	-3.23	0.00	0.38	24.00	-1.00
ATF3	-0.74	-3.20	0.00	0.38	26.00	-1.00
KLF6	-0.76	-3.17	0.00	0.38	31.00	-1.00
GPX2	0.54	3.11	0.00	0.44	33.00	1.00
EGR2	-0.82	-3.06	0.00	0.45	37.00	-1.00
NR4A3	-0.95	-2.96	0.00	0.48	46.00	-1.00
OAS2	0.62	2.93	0.00	0.50	49.00	1.00
CEBPD	-0.54	-2.78	0.01	0.55	70.00	-1.00
ENPP1	-0.54	-2.74	0.01	0.57	76.00	-1.00
HBEGF	-0.52	-2.64	0.01	0.61	89.00	-1.00
CDC42EP4	-0.53	-2.58	0.01	0.61	104.00	-1.00
DUSP6	-0.55	-2.57	0.01	0.61	110.00	-1.00
ITGA5	-0.69	-2.55	0.01	0.61	115.00	-1.00
BTG2	-0.30	-2.54	0.01	0.61	122.00	-1.00
SFRP1	-0.61	-2.50	0.01	0.62	134.00	-1.00
DUSP5	-0.64	-2.50	0.01	0.62	136.00	-1.00
FOSL2	-0.28	-2.48	0.01	0.62	144.00	-1.00
SELE	-0.57	-2.46	0.02	0.62	149.00	-1.00
MKNK1	0.53	2.43	0.02	0.64	156.00	1.00
CX3CL1	-0.50	-2.43	0.02	0.65	157.00	-1.00
ARHGEF7	-0.38	-2.30	0.02	0.68	209.00	-1.00
BCL6	0.62	2.29	0.02	0.68	216.00	1.00
TRIB1	-0.26	-2.22	0.03	0.70	245.00	-1.00
IER3	-0.42	-2.11	0.04	0.72	314.00	-1.00
BDKRB2	0.43	2.05	0.04	0.74	346.00	1.00
NEDD9	-0.32	-2.04	0.04	0.76	352.00	-1.00
CXCL2	-0.53	-2.00	0.05	0.77	376.00	-1.00
FOXJ1	0.56	1.98	0.05	0.79	384.00	1.00

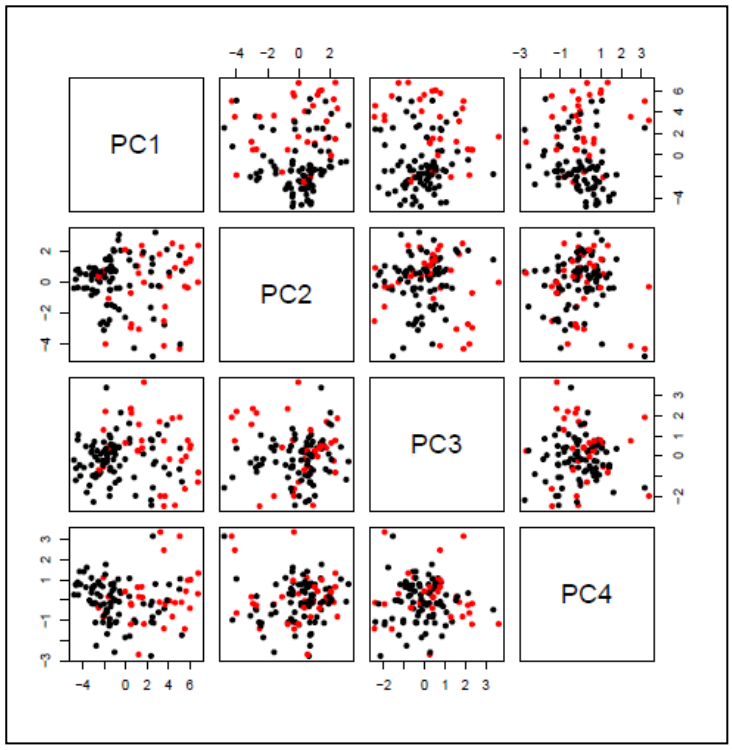
Of the 19 genes differentially expressed with FDR correction and identified without Gleason in the covariate, 18 overlap with the 31 differentially expressed genes with p value of 0.05 (but testing for multiple corrections). Notably a consistent finding is there is down regulation of tumor suppressors in lethal tumors when both types of analyses are used. The tumor suppressors identified to be lost in lethal prostate cancer are:

- *CEBPD:* CCAAT/enhancer-binding protein delta
- DUSP5: Dual specificity protein phosphatase 5
- SFRP1:Secreted frizzled-related protein 1
- NR4A3: neuron-derived orphan receptor 1 (NOR1) also known as NR4A3
- ZFP36: Tristetraprolin (TTP), also known as zinc finger protein 36 homolog

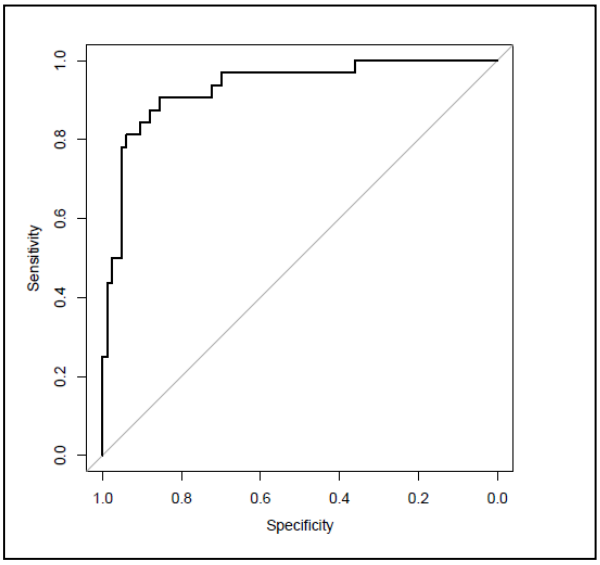
Of the 31 genes identified, 6 are increased in lethal vs non-lethal and

independent of Gleason (GPX2, OAS2, MKNK1, BCL6, BDKRB2, FOXJ1). These were not found when Gleason was not in the model and FDR correction performed. 7 genes were down regulated and independent of Gleason in lethal disease but not found with FDR when Gleason was in the model.

Principal Component Analysis PCA was performed on PHS cases using significant genes form NFκB set. PCA 1, 2 3 and 4 were developed. Logistic regression of the 4 different PCA found PCA1 and PCA3 were significantly associated with lethal cancer.



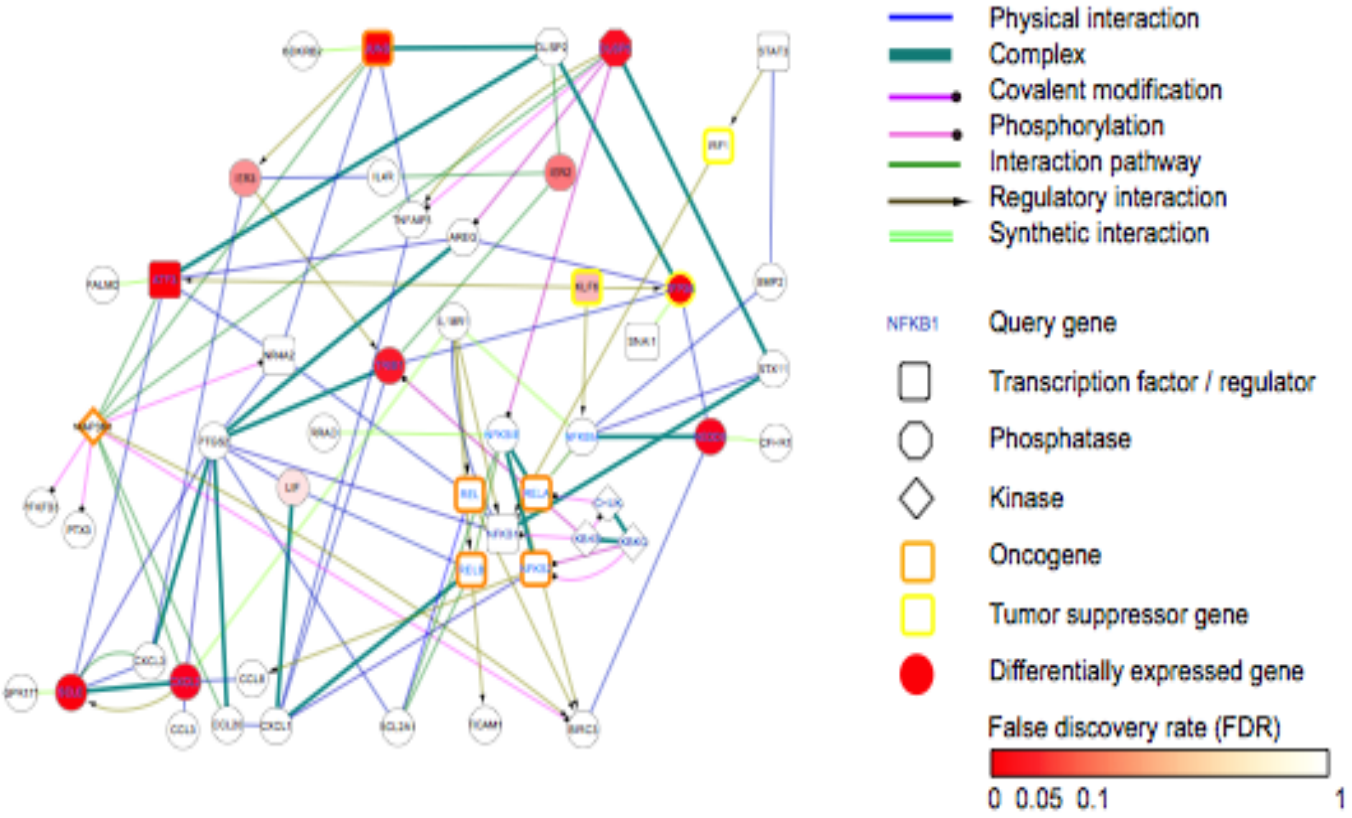
Logistics Regression Deviance Residuals				
Min	1Q	Median	3Q	Max
-2.30	-0.41	-0.20	0.31	2.77
Coefficients				
	Estimate	Std Error	z value	Pr (>|z|)
Intercept	0.47	0.48	0.97	0.32
PC1	0.41	0.12	3.55	0.0004
PC2	-0.26	0.17	-1.51	0.13
PC3	0.64	0.25	2.56	0.01
Gl Low	-18.77	1511	-0.01	0.99
Gl Med	-2.92	0.71	-.41	4.22e-05



ROC were generated and AUC of 0.9313 was found for predicting lethal vs indolent. We note that this is subject to over-fitting because the data used to generate PCA were not subject to correction for multiple testing AND the data inputted was selected to be differentially expressed. These "discovery sets" will next be tested in publically available data sets for association with Gleason score and lethality after prostatectomy [see flowchart below in Section 1 (ii)]

SUMMARY OF PROCESS FOR CREATION OF NFκB PROSTATE CANCER PROMOTING PATHWAY

By the end of year 3, we have completed the computational recovery of mechanistic pathway components specific to the NFκB pathway as perturbed in prostate cancer. In summary, we leveraged recent advances in Bayesian data integration to simultaneously provide information specific to biological contexts and individual biomolecular mechanisms. We developed this method to identify mechanisms of interaction surrounding NFκB during its activity in cell death, inflammation, adhesion and differentiation. We integrated 18 gene expression datasets chosen to be particularly informative for functional relationships in prostate cancer and 860 non-disease datasets from gene expression, physical interactions, and functional associations. Prior knowledge was further included to inferred genome-wide networks for 442 biological processes (Gene Ontology), including 7 mechanisms of interaction ranging from general functional relationships through specific physical and regulatory activities. Among all inferred networks we focused on 9 biological context networks most informative in prostate cancer, as summarized above. We have investigated NFκB pathways in prostate cancer by identifying 8 genes that were highly functionally related with NFκB in these contexts and highly differentially expressed in lethal vs non-lethal prostate cancer microarray experiments. The simultaneous prediction of biomolecular mechanisms finally led to this newly predicted NFκB pathway and associated parallel pathways that interact with it and is specific to prostate cancer.



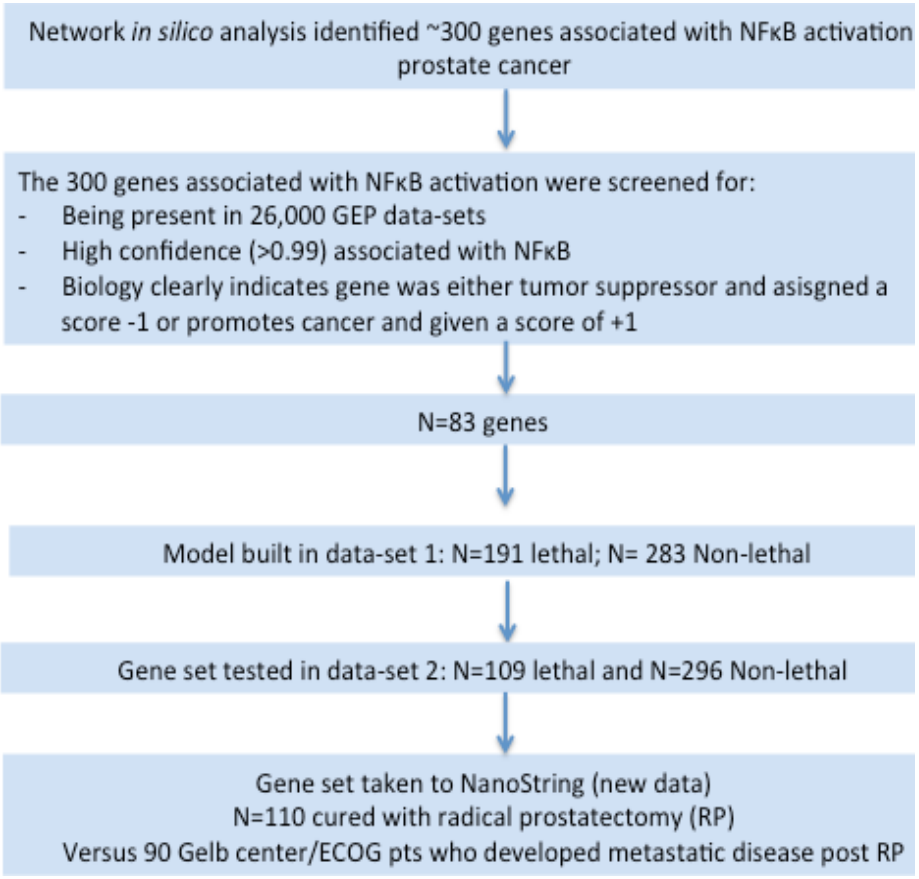
(ii) Assessment of candidate genes in Training Set:

Details of "Nugen-Affy": In the original application, the Harvard School of Public Health collaborators had planned to generate gene-expression profile data using the 24,000 DASL platform from prostatectomy specimens from the Health Professionals Follow-up Study and the Physicians Health Study (to be referred to as HPFS/PHS cohort). However, at time of commencing the assays it became apparent the quality of the assay had declined and failed our pilot study. To address this concern, the HSPH have since successfully established the "Nugen-Affy" platform. This work was not being paid for by this DOD contract and there was no change to the budget or scope of work. The change resulted in access to high quality data and ensures we adhere to the original SOW. Specifically, to conduct transcript profiling in FFPE prostate cancer tissues, whole transcriptome amplification is being paired with microarray technologies. Briefly, RNA extracted from FFPE prostate cancer samples has been amplified using the WT-Ovation FFPE System V2 (Nugen, San Carlos, CA), a whole transcriptome amplification system that allows for complete gene expression analysis from archived FFPE samples known to harbor small and degraded RNA. Using a combination 5 and random primer, reverse transcription creates a cDNA/mRNA hybrid. The mRNA is subsequently fragmented, creating binding sites for DNA polymerase. Isothermal strand-displacement, using a proprietary DNA/RNA chimeric SPIA primer, then amplifies the cDNA. To prepare the amplified DNA for microarray hybridization, the cDNA is fragmented and then labeled with a terminal deoxynucleotidyl transferase that is covalently linked to biotin. The labeled cDNA is then hybridized to a GeneChip Human Exon 1.0 ST microarray (Affymetrics, Santa Clara, CA). This array contains roughly four probes per exon and roughly 40 probes per gene, assessing the expression of roughly 28,000 unique genes. The analysis plan is as follows and detailed in flow chart.

- **Assess for assay reliability.** A pilot study using 11-21 year old prostate tumor specimens and found excellent ($r > 0.95$) concordance of technical replicates and no influence of block age on expression profiles **(Completed)**
- **Assess "discovery set" in Nugen-Affy genomic data to define "training set".** Using the genes associated with lethal prostate from the data-mining efforts described above (discovery set) we will then assess them using genomic "Nugen-Affy" data using SAM and limma. If necessary, additional QC filtering will be applied and the genes re-queried until convergence on a consistent set of markers reproducibly predictive of lethal prostate cancer. We hope that standard sparse regression (lasso) or feature select will

narrow this without loss of predictive accuracy to ~20 genes for application directly to the validation set (Gelb Center samples described below).

- **Our model will be built and tested** as detailed in the flow chart**.** The gene-set chosen from the "Nugen-Affy" work will then be subjected to a multivariate analysis and we will determine whether this gene-set/biological variable has a strong enough association with the lethal prostate cancer to be taken to the validation set. We will estimate Cox proportional hazards model with hazard ratios (HR) and the corresponding 95% confidence intervals (CIs) using, both without (unadjusted) or with (adjusted) the adjustment of factors known to be associated with lethal prostate cancer. The variables will include standard criteria – PSA level, Gleason score and pathological stage.



(iii) Validation Sets: In parallel with the above work we have created a unique cohort to validate whether the outcomes associated with the training set can be reproduced in an independent data-set.

Status of TMA Creation and Extraction of Nucleic Acids from Independent Cohort: Currently we have obtained the following samples from DFCI and ECOG repositories: Blocks from 90 unique patients with metastases post prostatectomy (lethal) and 110 patients without metastases post-prostatectomy (non-lethal/long-term survivors). When we perform the analysis we will use definitions that are harmonized with the HPFS/PHS cohort. Currently, all the TMAs have been created with each TMA laid out with 3 cores of tumor and 2 cores of benign/normal for each case. The TMAs have been made in duplicate and the cores of tumor and of non-cancer tissue have been obtained for nucleic acid studies. This was completed in year 3. Having defined the training set, we will have the genes to guide the custom ordered Nanostring assay. It is anticipated we will perform the nanostring work in July 2014.

In short, we will apply the signature score of cancer promoting NFκB activation directly to the validation set of 77 lethal:77 non-lethal samples. We have previously calculated a C-statistic for Gleason score of 0.86. Compared to a model with Gleason alone, we estimate that we will have 80% power and type 1 error of 0.05 to detect an improvement in the ROC curve of 6% for the signature of activation (i.e increase to 92%) with a rank correlation between models for both lethal and indolent of 0.8. A multivariate analysis will be used in determining whether a biological variable has a strong enough association with the lethal prostate cancer. We will estimate Cox proportional hazards model with hazard ratios (HR) and the corresponding 95% confidence intervals (CIs) using, both without (unadjusted) or with (adjusted) the adjustment of factors known to be associated with lethal prostate cancer. The variables will include standard criteria – PSA level, Gleason score and pathological stage. The gene-set that meets this criteria will be chosen for the risk assessment tool to be assessed in the biopsy cohorts in Year 3 (Aim 3). It is of note that we will also be choosing the 4 most predictive genes which have robust antibodies available for immunohistochemistry studies on the TMAs. We will also make use of Nanostring's nCounter platform which has the potential to be developed as a routine clinical assay.

Task 1B: Perform protein profiling of circulating blood proteins and determine whether a protein or set of proteins indicative of NFκB activation are associated with lethal prostate cancer. Circulating proteins will be assessed in two cohorts of 312 patients. Samples have already been assembled and tied to clinical outcomes. (Month 1 to 18).

We have performed the following assays GROα (CXCL-1), IL-1α, IL-1β, IL-2, IL-6, IL-8, MCP-1 (CCL-2) and TNFα.

	GROα pg/ml	IL-1α pg/ml	IL-1β pg/ml	IL-2 pg/ml	IL-6 pg/ml	IL-8 pg/ml	MCP-1 pg/ml	TNF-α pg/ml
Mean	111.16	0.90	1.06	19.99	8.99	4.79	261.27	5.01
Median	83.90	0.09	0.69	14.06	5.19	3.85	243.72	4.02
Max	994.29	15.30	18.28	270.59	320.41	31.01	1957.95	113.27
Min	0.00	0.00	0.00	0.00	0.00	0.11	1.99	0.00

Study population:

Using Gelb Center Samples we assessed whether elevated levels of serologic cytokines were associated with risk of post-treatment relapse in patients treated with radical prostatectomy or radiation therapy. This allowed us to determine the prognostic value of pre-diagnostic cytokine levels for predicting post-treatment relapse (PTR). We utilized 190 patients from the Gelb Center cohort who received either radical prostatectomy or radiation therapy between 1997 and 2012 and who could be classified according to development of PTR. PTR (N=93) was defined as development of metastases or biochemical recurrence. Biochemical recurrence after radiation therapy was defined as occurrence of nadir+2ng/mL or initiation of salvage therapy if PSA failure had not occurred. Biochemical recurrence after prostatectomy was defined as the observation of two consecutive PSA values ≥0.2 ng/mL, at least 60 days after RP; or if the treatment was given when the PSA was less than 0.2 ng/mL but greater than 0.1 ng/mL. Metastatic disease was defined as metastases to lymph nodes, bone, or distance organs. Patients classified as being without PTR (N=96) were required to have at least 3 years of follow up.

Using EDRN samples we assessed whether serologic cytokines could distinguish between men with an elevated PSA and biopsy revealing clinically significant prostate cancer or indolent cancer or no cancer. 'Clinically Significant' cases are those with a biopsy Gleason score of 3+4 or higher. The desire was for 153 'clinically significant' cases; however, due to limited plasma only 86 'clinically significant' cases were obtained. The criteria for 'Indolent' cases are: GS 6, PSA < 10, no perineural invasion, no more than 50% involvement in a core, PSA density < 0.15, and 3 or less cores. Again, the goal was 153 indolent cases though, with limited plasma, 49 were obtained. The EDRN team provided 79 control cases (with no suspicion for atypia) to make up for the limited clinically significant and indolent cases.

Evaluation of serum markers:

Using pre-diagnostic blood samples, cytokines IL-1α, IL-1α, IL-2, IL-6, IL-8, GROα, MCP-1, and TNFα were assayed on the Meso Scale Discovery (MSD, Gaithersburg, MD) electrochemiluminescence (ECL) platform and MSD Sector Imager 2400 in the laboratory of Raina Fichorova at Brigham and Women's Hospital. The MSD ESL assays have been validated by comparison with traditional ELISA in a multicenter study led by Dr. Fichorova. For the ECL assay 96-well plates were designed specifically for use in the Sector Imager 2400 and supplied by MSD. The plates were first coated with a MSD blocking solution, then washed with a PBS-based buffer prepared in Fichorova Lab, and subsequently incubated with test samples and calibrators (25 µl/well) for 2 hours with shaking. At the end of this incubation period, the plates were washed, a detection antibody added for 1-2 h with shaking, and then washed again before adding MSD Read Buffer (150 µL/well) followed by reading in the MSD Sector 2400 Imager.

Statistical analysis:

Gelb Center: Patient clinical characteristics at diagnosis were summarized as numbers and percentages for categorical variables and as median and interquartile ranges for continuous variables. Serum marker levels were summarized as median and interquartile ranges according to development of lethal prostate cancer. The associations of serum markers with PTR were assessed using Fisher's exact tests using quartiles of each serum marker. Cochran-Armitage tests were used to summarize the trend of serum marker levels and PTR. We also conducted multivariable analyses controlling for age and D'Amico risk score.

EDRN cohort: The association of serum cytokines with clinically significant participant status (as compared to indolent) was graphically explored for each cytokine by boxplots. Cytokines were not log-transformed because each prostate cancer status group had ample sample size (in the least 49 indolent

participants). Formally, the cytokine and clinical significance association was tested using logistic regression. The odds of being clinically significant were predicted separately by each cytokine. Clinically significant participants were compared to indolent participants to see if they had generally 'higher' cytokine expressions. The aim was to determine if there exists a protein profile amongst clinically significant participants—where clinically significant participants have more 'high' cytokines than indolent participants. For each cytokine, the threshold for being 'high' was determined as being above the third quartile of the combined indolent and clinically significant expressions. Ultimately each participant had a count of number of 'high' cytokines, and the distributions of each group were compared using the Chi-Square Test of Independence.

Results: Gelb Center: Selected clinical characteristics are presented in Tables 1 and 2 for Gelb Center and EDRN respectively

In the Gelb Center. The mean age of patients in both groups was 62 years. As expected, men in the PTR group had a higher mean PSA level at diagnosis (7.3 vs. 5.6) and were more likely to have Gleason 8-10 tumors (45% vs. 5%), at least 50% of biopsy cores positive for tumor, and high D'Amico risk score categorization (47% vs. 11%). There were few men with T3/T4 tumors in both groups (3% in PTR and 1% in no PTR group). The associations between each cytokine with respect to lethal prostate cancer are presented in Table 2. In univariate analyses, IL-2 was associated with PTR (p=0.03), but no clear trend in levels was apparent (p-trend=0.49). None of the cytokines were statistically significantly associated with outcome. In multivariable analyses, no statistically significant trends were apparent (data not shown).

In the EDRN cohort: The boxplots convincingly show no difference between indolent and clinically significant prostate cancer for any of the cytokines. The marginal logistic regressions failed to find evidence of cytokine differences between the groups (each p > 0.05). Further stratification of the clinically significant group into Gleason score 3+4 and >/= 4+3 did not help separate the clinically significant group(s) from the indolent group. The median IL 6 appears somewhat higher among GS 3+4 than among GS >= 4+3; however, there is not the expected gradient of IL 6 moving from indolent to GS 3+4 to GS >= 4+3. Also an ANOVA test for a difference in mean IL 6 among these three groups was non-significant (p = 0.075). The best fitting model from backward selection was fit with GRO-a and TNF-a, though neither p-value is less than 0.10.

Conclusions:

With only 93 cases with documented PTR, our power is limited to detect modest associations between levels of cytokines with respect to disease outcomes in the Gelb Center cohort. Although IL-2 levels were found to be statistically significantly associated with PTR, there was no apparent trend for levels being higher or lower in patients with PTR, thus providing limited prognostic utility. While we cannot rule out that pre-diagnostic cytokine levels may be important biologically for prostate cancer development and progression as suggested by previous studies, our investigation does indicate that these markers are inadequate for determining which prostate cancer patients will develop relapse after primary treatment. The EDRN cohort shows no association between any cytokine and clinically significant prostate cancer in men undergoing a TRUS biopsy for an elevated PSA. As such, the cumulative data from the Gelb Center and EDRN cohorts do not provide any evidence that cytokine levels at the time of biopsy or prostatectomy can help identify men with indolent disease and can be spared surgery or intervention. Conversely, none of these cytokines can be used to help identify men with low risk disease who are in need of an intervention to cure them of a potentially lethal cancer. Cytokines are not going to be used as part of the risk scoring system.

SPECIFIC PROGRESS IN YEAR 3: A manuscript has been drafted to detail the null results / no association with lethal disease.

Table 1. Clinical characteristics of patients in the Gelb Center cohort selected according to development of post-treatment recurrence

	N with % or median (q1, q3)	
	No Post-Treatment Relapse N=96	Post-Treatment Relapse N=93
Age at diagnosis	62 (56, 70)	62 (55, 67)
PSA at diagnosis	5.55 (4.40, 7.35)	7.30 (5.05, 12.00)
Gleason at diagnosis		
6 or less	46 (48)	15 (16)
7	44 (46)	36 (39)
8 or higher	5 (5)	42 (45)

Missing	1 (1)	(0)
T stage		
T1	65 (68)	59 (63)
T2	22 (23)	20 (22)
T3/T4	1 (1)	3 (3)
Tx/Unknown	8 (8)	11 (12)
N stage		
N0	30 (31)	49 (53)
N1	0 (0)	3 (3)
Nx/Unknown	66 (69)	41 (44)
M stage		
M0	40 (42)	54 (58)
M1	0 (0)	1 (1)
Mx/Unknown	56 (58)	38 (41)
% of positive cores		
<33%	58 (60)	35 (38)
33%-50%	11 (11)	16 (17)
>50%	18 (19)	35 (38)
Unknown	9 (9)	7 (8)
D'Amico risk group		
High	11 (11)	44 (47)
Intermediate	44 (46)	40 (43)
Low	40 (42)	8 (9)
Unknown	1 (1)	1 (1)

Table 2 Distribution of age and covariates which define Indolent and Clinically Significant groups in EDRN cohorts (in addition assays were performed on 79 controls with elevated PSA and no cancer on biopsy).

	N	**Median (Q1, Q3)**
Age	**135**	**64 (59, 69)**
	N	**%**
Biopsy Gleason	NA	NA
<= 6	49	36
7	61	45
>= 8	25	18
PSA (ng/mL)	NA	NA
< 10	116	86
>= 10	19	14
Perineural Invasion	35	26
Any cores with > 50%	55	41
PSA Density < 0.15	84	62
< 4 Cancerous cores	83	61

Control samples are not listed in the Demographic Table. Also, the core percentage positive for cancer is unknown for 1 Clinically Significant case. Four Clinically Significant and one Indolent case have unknown PSA Density.

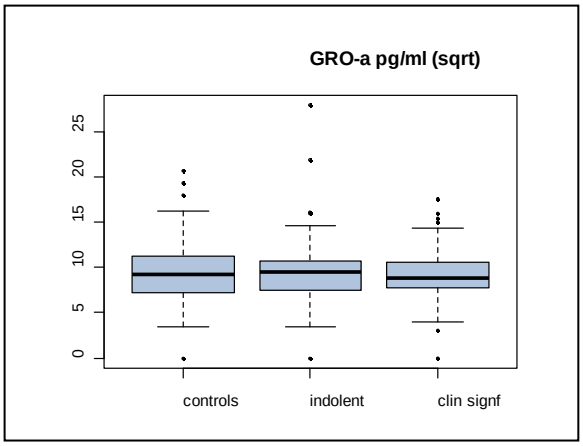
Table 3. The association between pre-diagnosticserological markers with respect to lethal prostate cancer, N=189: Gelb Center Cohort

		Post-treatment Recurrence		Univariate P-value (Fisher's Exact/Trend)
		No	Yes	
IL-1α				0.09/0.22
	Tertile 1	52 (54)	53 (57)	
	Tertile 2	15 (16)	23 (25)	
	Tertile 3	29 (30)	17 (18)	
IL-1β				0.42/0.87
	Q1	28 (29)	22 (24)	
	Q2	19 (20)	25 (27)	
	Q3	22 (23)	26 (28)	
	Q4	27 (28)	20 (22)	
IL-2				0.03/0.49
	Q1	18 (19)	9 (32)	
	Q2	31 (32)	5 (16)	
	Q3	23 (24)	9 (27)	
	Q4	24 (25)	6 (25)	
IL-6				0.34/0.67
	Q1	22 (23)	25 (27)	
	Q2	20 (21)	27 (29)	
	Q3	29 (30)	19 (20)	
	Q4	25 (26)	22(24)	
IL-8				0.13/0.35
	Q1	25 (26)	23 (25)	
	Q2	29 (30)	17 (18)	
	Q3	19 (20)	30 (32)	
	Q4	23 (24)	23 (25)	
MCP-1				0.42/0.72
	Q1	20 (21)	27 (29)	
	Q2	28 (29)	19(20)	
	Q3	25 (26)	23 (25)	
	Q4	23 (24)	24 (26)	
TNF-α				0.23/0.14
	Q1	23 (24)	25 (27)	
	Q2	22 (23)	26 (28)	
	Q3	21 (22)	25 (27)	
	Q4	30 (31)	17(18)	
GRO-α				0.42/0.49
	Q1	28 (29)	19 (20)	
	Q2	20 (21)	27 (29)	
	Q3	25 (26)	23 (25)	
	Q4	23 (24)	24 (26)	

Distribution of cytokines in EDRN cohort at time of biopsy by risk group

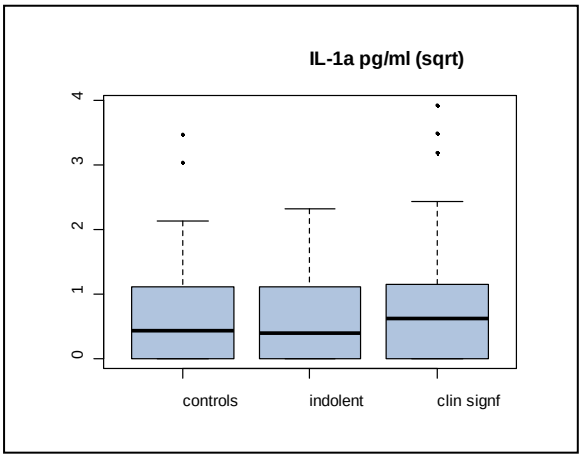
Distribution of GRO-α pg/ml

	controls	indolent	clin signf	indolent / clin signf
Min.	0	0	0	0
1st Qu.	52.52	56.79	59.9	59.24
Median	84.19	89.22	76.98	80.67
Mean	101	115.3	91.35	100
3rd Qu.	125.2	116.2	111.1	114.3
Max.	426.8	777.2	310.9	777.2



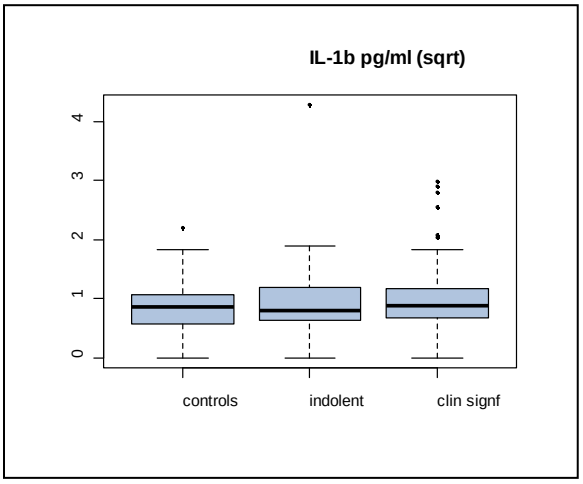
Distribution of IL-1a pg/ml

	controls	indolent	clin signf	indolent / clin signf
Min.	0	0	0	0
1st Qu.	0	0	0	0
Median	0.18	0.15	0.38	0.38
Mean	0.9885	0.9398	1.14	1.067
3rd Qu.	1.23	1.23	1.265	1.26
Max.	11.95	5.38	15.3	15.3



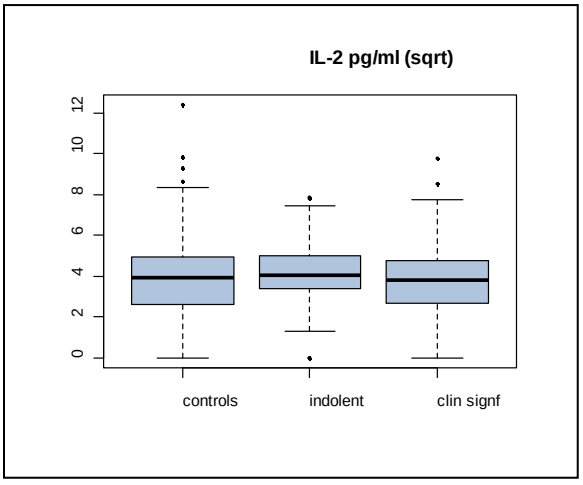
Distribution of IL-1b pg/ml

	controls	indolent	clin signf	indolent / clin signf
Min.	0	0	0	0
1st Qu.	0.33	0.39	0.4575	0.41
Median	0.75	0.62	0.775	0.77
Mean	0.9084	1.245	1.326	1.296
3rd Qu.	1.155	1.43	1.375	1.38
Max.	4.88	18.28	8.93	18.28



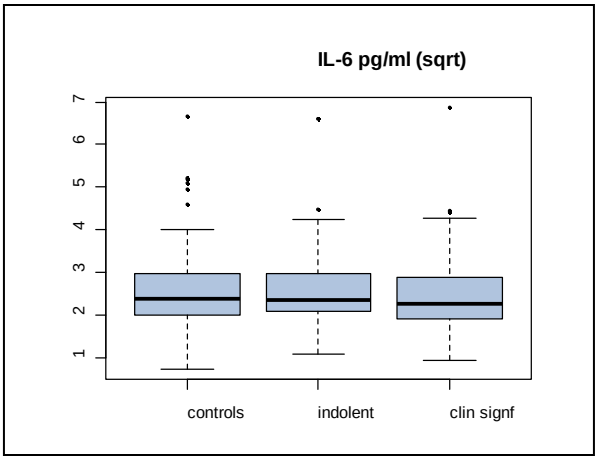
Distribution of IL-2 pg/ml

	controls	indolent	clin signf	indolent / clin signf
Min.	0	0	0	0
1st Qu.	6.82	11.32	7.22	8.935
Median	15.18	16.15	14.29	14.86
Mean	21.27	20.09	18.23	18.91
3rd Qu.	24.62	24.99	22.36	24.16
Max.	153.1	62.01	95.04	95.04



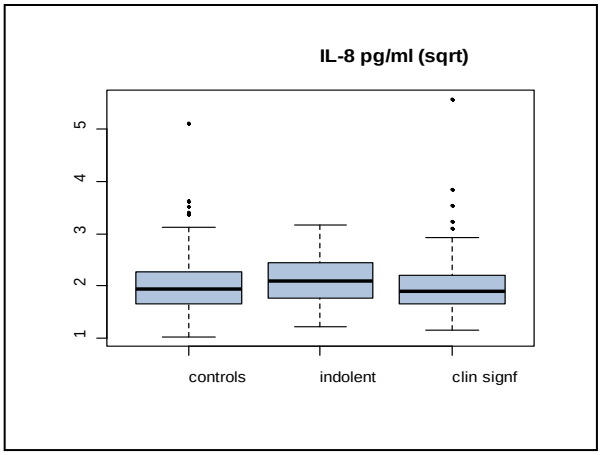
Distribution of IL-6 pg/ml

	controls	indolent	clin signf	indolent / clin signf
Min.	0.56	1.17	0.88	0.88
1st Qu.	4.05	4.43	3.66	3.94
Median	5.68	5.55	5.215	5.37
Mean	7.981	7.58	6.912	7.154
3rd Qu.	8.885	8.82	8.25	8.365
Max.	44.18	43.42	47.01	47.01



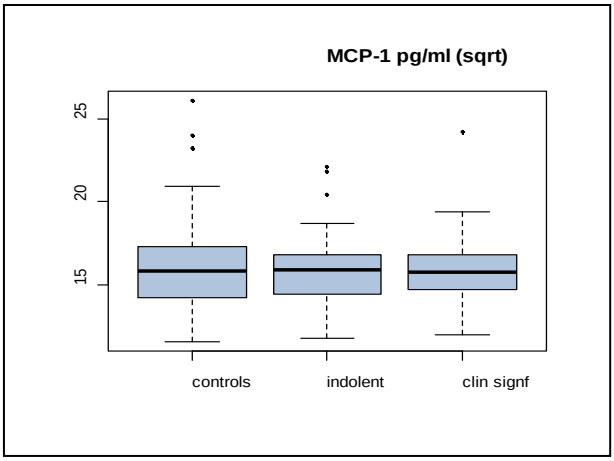
Distribution of IL-8 pg/ml

	controls	indolent	clin signf	indolent / clin signf
Min.	1.07	1.47	1.32	1.32
1st Qu.	2.785	3.1	2.752	2.875
Median	3.81	4.38	3.61	3.74
Mean	4.736	4.619	4.48	4.531
3rd Qu.	5.125	6.03	4.825	5.155
Max.	26.2	10.06	31.01	31.01



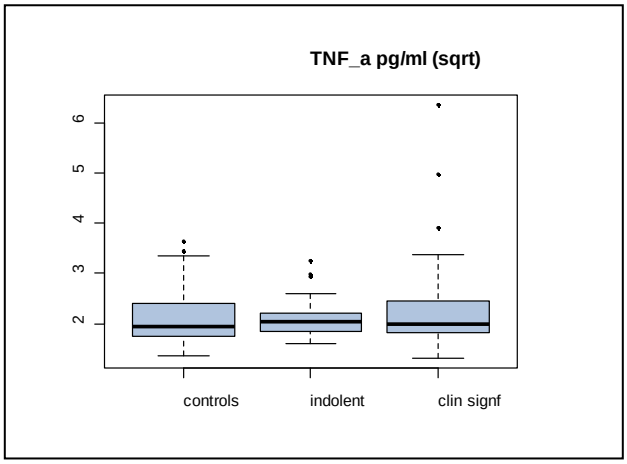
Distribution of MCP-1 pg/ml

	controls	indolent	clin signf	indolent / clin signf
Min.	134.1	138.1	143	138.1
1st Qu.	202.1	208	215.9	209.6
Median	249.2	253	247.1	250.2
Mean	264.5	255.6	252.8	253.8
3rd Qu.	299.7	283.2	282.1	282.7
Max.	680	488.7	587.2	587.2



Distribution of TNF-α pg/ml

	controls	indolent	clin signf	indolent / clin signf
Min.	1.86	2.57	1.75	1.75
1st Qu.	3.045	3.38	3.295	3.315
Median	3.76	4.21	4.02	4.05
Mean	4.744	4.487	5.49	5.126
3rd Qu.	5.825	4.91	5.932	5.195
Max.	13.32	10.59	40.34	40.34



Task 1C: *Assess whether seropositivity for Trichimonas vaginalis correlates with NFκB activation and lethal prostate cancer. We will assess the seropositivity from 111 patients with non-lethal (low risk EDRN samples) and 111 patients with lethal (metastatic, ECOG samples) prostate cancer. Samples have already been obtained and correlated with clinical outcomes. (Months 1 to 18)*

T-Vag was assayed by ELISA and provided by John Alderete of Washington State University. The assay detects IgG antibodies against recombinant T. vaginalis alpha-actinin, one of the most immunogentic trichonomad proteins. Scores of 0 (zero), 1+ and 2+ are negative. A 3+ and 4+ score are considered to be positive. Using non-DOD funds we expanded the analysis to include additional cohorts and the prevalence of seropositivity in the groups analyzed was the following

Gelb Center: N= 96 sample at radiation or surgery and no post treatment relapse: 15%

Gelb Center: N = 93 sample at time of radiation or surgery with post treatment relapse: 14%

EDRN: N=49 samples at time of biopsy and clinically indolent cancer: 12%

EDRN: N=86 samples at time of biopsy and clinically significant cancer: 15%

ECOG: N=111 samples at time of starting hormonal therapy for metastatic disease:13%

Table 1C-1. The association between pre-diagnostic serological markers with respect to lethal prostate cancer, N=189: Gelb Center Cohort

	Post-treatment Recurrence		**Univariate P-value (Fisher's Exact/Trend)**
	No	**Yes**	
***T. vaginalis* score**			0.84/0.77
0-2	82 (85%)	78 (84%)	
3-4	14 (15%)	15 (16%)	

Table 1C-2	EDRN Cohort	
	clin signf	indolent
0	73 (85%)	43 (88%)
1	13 (15%)	6 12%)

Gelb Center T. vag positivity is in table 1C-1 and EDRN is in Table 1C-2. In neither cohort was T. vag status associated with poor risk disease. Tvag positivity did not help distinguish the association of cytokine with clinical significance status (boxplots and logistic regressions $p > 0.05$ – see figures on next page – Fig 1C-1).

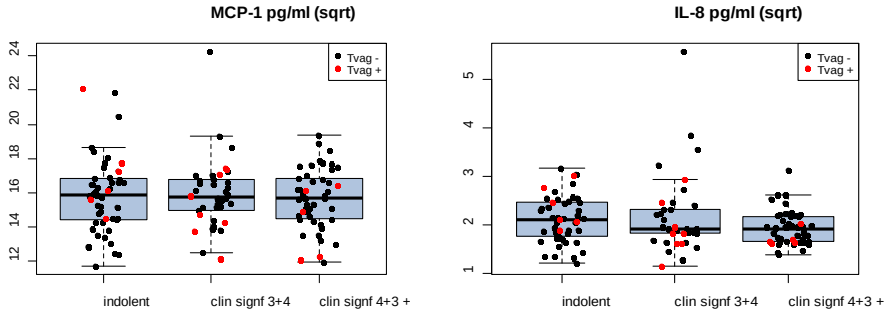
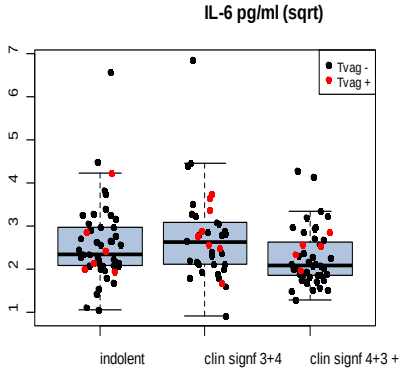
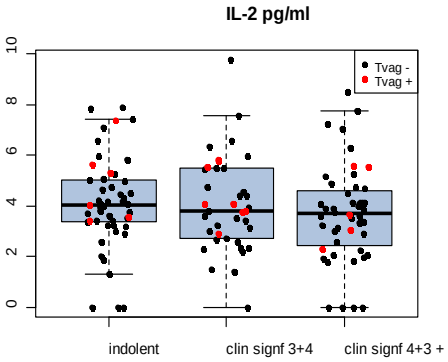
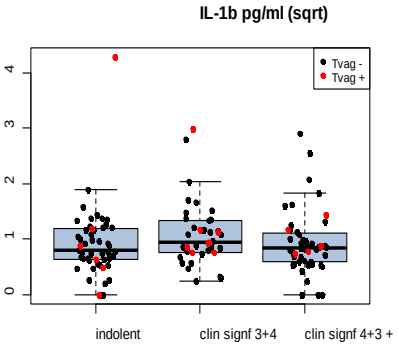
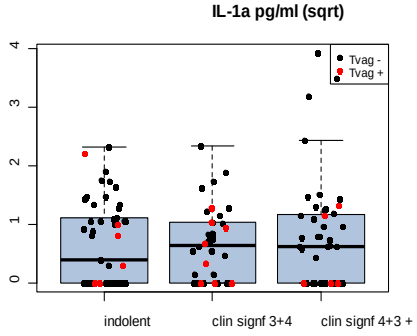


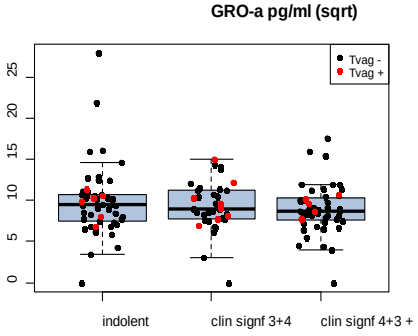
Figure 1C-1
In an attempt to see if T.Vag status was associated with elevation of a given cytokine, the following box plots were constructed from the EDRN cohort

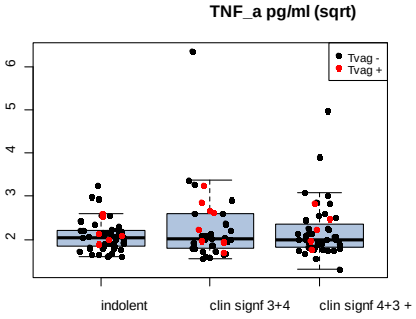












Conclusions: While the overall prevalence of *T. vaginalis* seropostivity is comparable or slightly lower than in previous studies of prostate cancer and the same laboratory was utilized for all of the published studies, only internal comparisons can be directly interpreted, as subtle changes in sample quality or laboratory conditions can impact the overall rates from one study to the next. With only 93 cases with documented PTR, our power is limited to detect modest associations between *T. vaginalis* serostatus with respect to disease outcomes. Moreover, the failure to find an association could be due to a prevalence of infection may be higher in cases without PTR than in disease-free controls, which were used as the comparison group in previous studies. There was no apparent trend for levels T. Vag levels being higher or lower in patients with PTR, thus providing limited prognostic utility. While we cannot rule out that pre-diagnostic cytokine levels and *T. vaginalis* serostatus may be important biologically for prostate cancer development and progression as suggested by previous studies, our investigation does indicate that T Vag status is inadequate for (i)

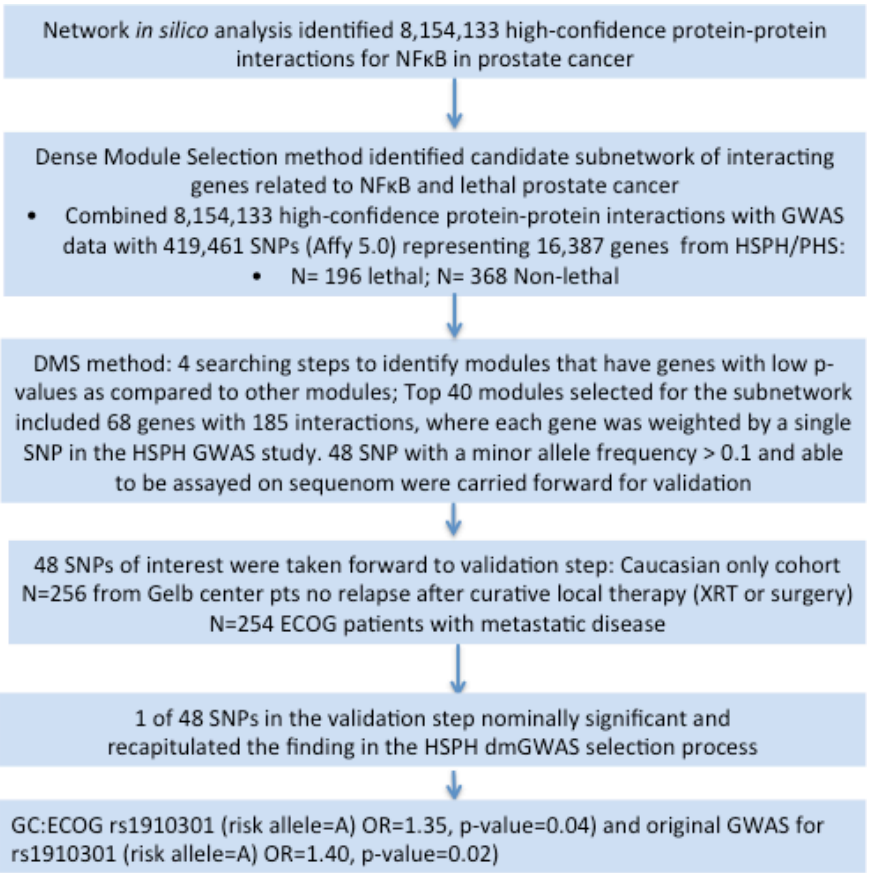
determining which prostate cancer patients will develop relapse after primary treatment or (ii) help in identifying patients with Gl 6 low risk cancer who harbor more aggressive disease and require an intervention. T. vag status will not be put into the risk scoring system.

SPECIFIC PROGRESS IN YEAR 3: A manuscript has been drafted to detail the null results / no association with lethal disease.

Task 1D: Assess whether gene variants associated with NFκB activation are associated with lethal prostate cancer. We will mine existing data sets to define the panel of gene variants to be correlated with lethal disease and then analyze 306 patients. The samples and clinical outcomes have already been assembled. (Months 1 to 18).

Using the network approach detailed in Aim 1A, we had a genome-wide functional association network specific to prostate cancer and the NFκB pathway consisting of 351 genes and 8,154,133 high-confidence functional associations. The dense module searching (DMS) method described by Jia et al[20]. was used to identify a candidate subnetwork of interacting genes related to both (i) the NFκB pathway and (ii) to lethal prostate cancer. The method combines a genome-wide association study conducted by the Harvard School of Public Health (HSPH) with a protein-protein interaction network developed by the team members who completed Aim 1A. The HSPS GWAS was conducted on 196 lethal and 368 indolent cases in the HPFS and PHS for 419,461 SNPs, and the p-value results were used to assign gene weights in the current study. After using the annotation file to assign each SNP on the Affymetrix 5.0 chip to a gene, a single SNP with the lowest p-value is selected to represent each of 16,387 genes. For each SNP-gene pair, the GWAS p-value is used as the gene's weight during DMS. The protein-protein interaction (PPI) network was constructed based on NFκB-based interactions likely to be functionally related to prostate cancer, and includes 8,154,133 high-confidence interactions. These interactions provide the connections between genes that are used to form modules.

As a brief summary, the DMS method then iteratively proceeded through four searching steps to identify modules that have genes with low p-values as compared to other modules. Each gene in the PPI is initially assigned as a seed gene and then genes that interact with the current module gene(s), with an interaction path less than or equal to 2, are identified. For each neighborhood interactor, Z_{m+1} is calculated after including it in the seed module and it is permanently added to the seed module if $Z_{m+1}>1.1*Z_m$, where Z_m is defined as $(\sum z_i)/k$, k is the number of genes in the module, and $z_i=\Phi^{-1}(1-P_i)$. These steps are repeated until no more neighborhood nodes can be added to the module. Normalized Z_m values, called values, are then calculated by comparing the module to 100,000 modules of the same size that were created by randomly selecting genes. After conducting dense module selection, the modules are ranked and the top 40 modules with the highest Z_N values were selected for the subnetwork. The selected subnetwork includes 68 genes with 185 interactions, where each gene was weighted by a single SNP in the GWAS study. Of the 68 SNPs used to represent the genes in the selected subnetwork, the top 40 SNPs with a minor allele frequency > 0.1 were carried forward for subsequent validation.



Samples were pulled and DNA isolated from 256 Gelb Center patients [low risk/non-lethal prostate cancer cohort] and compared with samples from patients with metastatic disease – 254 ECOG patients]. Only one of 40 SNPs was nominally significantly associated with lethal disease in the same direction. In the original HSPH GWAS the rs1910301 SNP with the risk allele being A had an OR of 1.40 (p-value = 0.02) for lethal disease. In the GC/ECOG cohorts, the rs1910301 risk allele A had an OR of 1.35 (p-value=0.04) for lethal disease. It is recognized there is a chance of false discovery given only 1 of 40 SNPs tested was nominally significant. However, it is also a finding that has been reproduced in 2 independent cohorts as the risk allele was in the same direction and the magnitude of the OR for lethal was similar between cohorts and the cohorts had the same ethnic background (Caucasians only). It was also identified via the dense module selection process using NFκB pathway as the underlying biology. We will now test the SNP using a third cohort – the NCI Breast and Prostate Cancer Cohort Consortium with 2,782 aggressive prostate cancer cases and 4,458 controls from the NCI Breast and Prostate Cancer Cohort Consortium which can make use of GWAS data to infer 5.8 million well-imputed autosomal single nucleotide polymorphisms. If it is associated with aggressive prostate cancer in this third cohort, we will assess its utility in the risk prediction model of Aims 2 and 3.

Brief background to rs1910301 SNP: This SNP is in MRPL1 gene that encodes the 39S subunit protein that belongs to the L1 ribosomal protein family. Mammalian mitochondrial ribosomal proteins (mitoribosomes) are involved in protein synthesis within the mitochondrion. Mitochondrial ribosomes consist of a small 28S subunit and a large 39S subunit. The gene for MRPL1 is on chromosome 4 and has a minor allele frequency of 0.323 and is in the Fraser syndrome 1 locus. The Fraser syndrome is multisystem malformation disorder usually comprising cryptophthalmos, syndactyly and renal defects and is due to mutations in the FRAS1 gene that encodes an extracellular matrix protein.

Task 2. Development of a "Risk Scoring System" for Lethal Prostate Using Multiple Factors (Months 18 to 24)

Task 2A Develop a risk scoring system from a composite of factors from Task 1 using prostatectomy cohorts to accurately identify patients with prostate cancer with a lethal potential than individual factors alone. This will be accomplished by developing a model that iteratively correlates various combinations of the different individual factors with lethal disease. We will make use of samples and clinical data from 146 patients which have already been assembled. (Months 12 – 24)

This task has been delayed due to the extra work needed to perform the Nugen-Affy analysis. Nonetheless, we have made substantial progress on a very complicated grant that has laid the foundation for the validation of the putative findings. We have been rigorous in our approach and had to be iterative to ensure reliable and reproducible data. This has led to the delay and the need for the extension (which has been granted). This extra time will allow us to take the project and the encouraging findings to completion. We are set up to complete the project with the 12 month extension as many of the tasks are being done in parallel. This most notably includes the sample collection that will allow us to transition from Aims 1 and 2 to Aim 3 without nay delays. Once the data has been established from Aim 1, a focused effort on creation of the risk scoring system over 3 months will result on completion of Task 2.

Task 3. Determine the predictive capability of "risk scoring system" using biopsy specimens. (Months 25 to 36).

Task 3A Validate the efficacy of the risk scoring system in two independent cohorts of biopsy specimens under the hypothesis that the risk scoring system will accurately identify patients with localized disease with lethal potential when using only biopsy specimens (TRUS or TURP sampling of the prostate). We will make use of samples and clinical data contained in our assembled data-bases. We will assess two cohorts of 146 patients each (total of 292 patients). Samples will be analyzed and correlated with outcome in Months 25 to 31. In months 31 to 36 the risk scoring system will be refined for maximal performance and clinical usability and manuscript written.

This task is contingent on the creation of the risk scoring system in Task 2 and collection of the biopsy specimens. We are actively collecting the Gelb Center lethal and non-lethal specimens. We have had to adopt our plan and not analyze the Swedish TURP specimens due to the inconsistency in the genomic data from TRUS and TURP specimens (secondary to different biology of centrally derived tumors and tissue processing). We have identified a cohort in the HSPH cohorts where TRUS biopsy cancer tissue and specimen for germline genomic analysis is available. We will make sure these 154 patients will be distinct from the patients analyzed

as part of the prostatectomy series in Task 1. We will not require blood given we will not be analyzing blood borne proteins.

KEY RESEARCH ACCOMPLISHMENTS:

- Development of NFκB networks associated with different biological context and an NFκB network associated with lethal prostate cancer
- Curated publically available prostate cancer data-bases for interrogation and are currently being analyzed and will define a gene set associated with NFκB activation.
- Developed "Nugen-Affy" assay at HSPH to a point where it can be reliably analyzed.
- Created informative TMAs and nucleic acid resources for validation of work from HSPH cohorts
- Discovered that T. vaginalis status and inflammation related proteins cannot be used to distinguish patients with low grade disease who harbor a disease that is actually higher risk disease
- Commenced sample collection of TRUS biopsy cohorts: (surveillance: progressed vs not progressed in 5 years and relapsed post prostatectomy vs not relapsed.

REPORTABLE OUTCOMES:

- Manuscripts, abstracts, presentations:
  - Symposium on Systems Biology of Diversity in Cancer at MSKCC, October 18-19 2012, Memorial Sloan-Kettering Cancer Center, New York, USA."Predicting biomolecular mechanisms in complex specific functional relationship networks in prostate cancer"
  - ISMB/ECCB 2013, NetBio Satellite Meeting, July 19 2013, Berlin, Germany. "Predicting biomolecular mechanisms in complex specific functional relationship networks in prostate cancer"
  - Computational reconstruction of NFkB pathway interaction mechanisms during prostate cancer (manuscript in preparation)
  - Seropositivity for Trichinomas Vaginalis and elevated cytokine levels at time of prostate cancer diagnosis do not identify patients with clinically significant localized disease. (manuscript in preparation).
- Licenses applied for and/or issued:
  - None
- Degrees obtained that are supported by this award;
  - None
- Development of cell lines, tissue or serum repositories:
  - Creation TMA and nucleic acid collection from prostatectomy and TRUS biopsies annotated with clinical outcome in collaboration with Gelb Center
- Informatics such as databases and animal models, etc.:
  - None
- Funding applied for based on work supported by this award:
  - None
- Employment or research opportunities applied for and/or received based on experience/training supported by this award:
  - None

CONCLUSION:

As detailed above, we have generated a significant amount of data and are now performing the robust statistical analyses which will lead to reliable new findings. We anticipate having findings either supporting or refuting the hypothesis that tumor and/or germline genetic profiles of inflammation/NFκB activation are associated with lethal prostate cancer. This data will be able to tell us whether germline SNP and/or tumor gene

expression profiling focused on inflammation can be used as a prognostic factor in patients diagnosed with prostate cancer. Moreover, it will set the stage for applying these findings to assess whether one or more of the findings can identify patients with clinically localized disease and is suitable for surveillance as well identify a biology (inflammation) or target (NFκB) to abrogate and prevent progression on surveillance and/or eradication of micrometastatic disease post definitive local therapy.

REFERENCES:

1. Perkins ND. The diverse and complex roles of NF-kappaB subunits in cancer. Nat Rev Cancer 2012;12:121-32.
2. Barrett T, Troup DB, Wilhite SE, et al. NCBI GEO: archive for high-throughput functional genomic data. Nucleic acids research 2009;37:D885-90.
3. Parkinson H, Kapushesky M, Kolesnikov N, et al. ArrayExpress update--from an archive of functional genomics experiments to the atlas of gene expression. Nucleic acids research 2009;37:D868-72.
4. Breitkreutz BJ, Stark C, Reguly T, et al. The BioGRID Interaction Database: 2008 update. Nucleic acids research 2008;36:D637-40.
5. Kerrien S, Alam-Faruque Y, Aranda B, et al. IntAct--open source resource for molecular interaction data. Nucleic acids research 2007;35:D561-5.
6. Matys V, Kel-Margoulis OV, Fricke E, et al. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic acids research 2006;34:D108-10.
7. Robertson G, Bilenky M, Lin K, et al. cisRED: a database system for genome-scale computational discovery of regulatory elements. Nucleic acids research 2006;34:D68-73.
8. Griffiths-Jones S. miRBase: the microRNA sequence database. Methods Mol Biol 2006;342:129-38.
9. Huttenhower C, Haley EM, Hibbs MA, et al. Exploring the human genome with functional maps. Genome research 2009;19:1093-106.
10. Huttenhower C, Mutungu KT, Indik N, et al. Detailing regulatory networks through large scale data integration. Bioinformatics 2009;25:3267-74.
11. Huttenhower C, Troyanskaya OG. Bayesian data integration: a functional perspective. Comput Syst Bioinformatics Conf 2006:341-51.
12. Huttenhower C, Troyanskaya OG. Assessing the functional structure of genomic data. Bioinformatics 2008;24:i330-8.
13. Tomlins SA, Mehra R, Rhodes DR, et al. Integrative molecular concept modeling of prostate cancer progression. Nature genetics 2007;39:41-51.
14. Ergun A, Lawrence CA, Kohanski MA, Brennan TA, Collins JJ. A network biology approach to prostate cancer. Mol Syst Biol 2007;3:82.
15. Workman CT, Mak HC, McCuine S, et al. A systems approach to mapping DNA damage response pathways. Science 2006;312:1054-9.
16. Lu R, Markowetz F, Unwin RD, et al. Systems-level dynamic analyses of fate change in murine embryonic stem cells. Nature 2009;462:358-62.
17. Lawrence DM, Seth P, Durham L, et al. Astrocyte differentiation selectively upregulates CCL2/monocyte chemoattractant protein-1 in cultured human brain-derived progenitor cells. Glia 2006;53:81-91.
18. Zhong H, May MJ, Jimi E, Ghosh S. The phosphorylation status of nuclear NF-kappa B determines its association with CBP/p300 or HDAC-1. Molecular cell 2002;9:625-36.
19. Bouwmeester T, Bauch A, Ruffner H, et al. A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. Nature cell biology 2004;6:97-105.
20. Jia P, Zheng S, Long J, Zheng W, Zhao Z. dmGWAS: dense module searching for genome-wide association studies in protein-protein interaction networks. Bioinformatics 2011;27:95-102.

APPENDICES:
None

SUPPORTING DATA:
None