

Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Wang X, Yu J, Sreekumar A, et al. Autoantibody signatures in prostate cancer. *N Engl J Med* 2005;353:1224-35.

Supplementary Appendix for

Autoantibody Signatures In Prostate Cancer

Xiaoju Wang, Jianjun Yu, Arun Sreekumar, Sooryanarayana Varambally, Ronglai Shen, Donald Giacherio, Rohit Mehra, James E. Montie, Kenneth J. Pienta, Martin G. Sanda, Philip Kantoff, Mark A. Rubin, John T. Wei, Debashis Ghosh, Arul M. Chinnaiyan

CONTENTS

- I. Supplementary Methods**
- II. Randomization and re-analysis of the phage-peptide microarray data**
- III. Clinical associations of the autoantibody signature**
- IV. Descriptions of the 22 phage peptide predictor**
- V. ELISA validation of the immunoreactivity for selected phage clones**
- VI. Immunofluorescence and immunohistochemical analyses of eIF4G1**
- VII. Supplementary Tables**
- VIII. Supplementary Figures**
- IX. References (for Supplementary Appendix)**

I. SUPPLEMENTARY METHODS

Construction of T7 phage display prostate cancer cDNA libraries. Total RNA was isolated separately from six prostate cancer tissue samples (**Supplementary Table 1**) according to a standard Trizol protocol.¹ Equal amounts of total RNA from six tissues were pooled together and poly(A) RNA was purified from the total RNA pool. OrientExpression cDNA Synthesis and Cloning System (Novagen) was used for the construction of the T7 phage prostate cancer cDNA libraries. In order to ensure the representation of both N-terminal and C-terminal amino acid sequences and eliminate the 3' bias inherent from oligo(dT)-primed strands, equal amounts of mRNA from each was used to construct two cDNA libraries using directional oligo(dT) primers and random primers in parallel. After vector ligation and T7 packaging, two cDNA phage display libraries were constructed and the library titers were determined by plaque assay with 4.2×10^6 pfu for the oligo(dT) primer library and 2.2×10^6 pfu for the random primer library. Phage particles from the two libraries were combined to make a phage library pool. After amplification, glycerol was added and the libraries were stored at -80°C . The quality and diversity of the phage library was tested by PCR amplification and gel electrophoresis of 19 randomly selected phage colonies (**Supplementary Figure 1**) and DNA sequencing of 30 randomly selected phage colonies from the library (**Supplementary Table 14**).

Amplification of Libraries. Five milliliters of LB with carbenicillin was inoculated at 37°C overnight with a single colony of BLT5615 from a freshly streaked plate. Overnight culture was added to 100 ml of LB with carbenicillin and grew to an OD_{600} of 0.5. One mM of IPTG was added and the cells were allowed to grow further for 30 min. An appropriate volume of culture was infected with phage library at multiplicity of infection (MOI) of 0.001-0.01 (i.e. 100-1000 cells for each pfu). The infected bacteria were incubated with shaking at 37°C for 1-2 hr until lysis was observed. The phage lysate was then clarified by spinning at 8000 g for 10 min. The supernatant was collected and stored at -80°C .

Enrichment for Prostate Cancer-Specific Phage-Peptide Clones. To enrich for phage clones that bind to IgGs specifically associated with prostate cancer, we performed positive and negative selection. First, a pre-clearing step was used to remove non-specific phage clones by pre-adsorbing the phage libraries onto purified IgG pool from 10 normal sera (**Supplementary Table 2**). Next, the pre-cleared phage libraries were selected onto a pool of IgGs purified from the sera of 19 localized prostate cancer patients (**Supplementary Table 2**). Protein A/G agarose beads (Pierce) were used to purify IgGs from the sera. The bound phages were eluted and amplified in bacteria, thus completing one round of biopanning. Five cycles of affinity selections were carried out for enrichment of prostate cancer-specific phage clones. Briefly, 10 μl protein-A/G agarose beads were placed into 1.5 ml eppendorf tubes and washed two times with $1 \times \text{PBS}$. Washed beads were blocked with 1% BSA at 4°C for 1 hr. The beads were then incubated at 4°C with 15 μl of individual serum from control or prostate cancer patients at 1:50 dilution in 1% BSA. After incubation overnight, the beads were washed with $1 \times \text{PBS}$ by centrifuging at 1000 g for 2 min. After three washes, 10 μl of $1 \times \text{PBS}$ was added to each tube, and 10 tubes of protein A/G-IgG complex from 10 control sera and 19 tubes of prostate cancer sera were combined to make IgG pools of control and prostate cancer respectively. These control and prostate cancer IgG pools associated with protein A/G beads were stored at 4°C as stocks for subsequent biopanning.

Twenty microliters of control IgG pool was incubated with 30 μ l amplified phage library pool diluted at 1:40 with 10% BSA at 4°C. After 2 hrs, the mixture was centrifuged at 1000 g for 2 min. The beads with non-specifically bound phage particles were discarded, and the supernatant was collected. Next, the supernatant was incubated with 30 μ l of the prostate cancer IgG pool at 4°C overnight. The mixture was centrifuged at 1000 g for 2 min and the supernatant was discarded. To elute the bound phage, 100 μ l of 1% SDS was added and incubated at room temperature for 10 min to break up the antibody-antigen reaction without disrupting T7 phage particles. The bound phages were removed from the beads by centrifuging at 5500 g for 8 min. Eluted phages were transferred to 10 ml culture of BLT5615 cells for amplification.

Construction and Screening of Phage-Peptide Protein Microarrays. The phage library ($\sim 10^{10}$ pfu) from the fifth cycle of biopanning was diluted at 1:10⁸ and allowed to grow on LB agar plates with carbenicillin. A total number of 2,304 random phage colonies were picked and amplified in 96-well plates. The phage lysates were spotted onto Nitrocellulose Film Slides (Grace Bio-Labs) to make high density phage-peptide protein microarrays using a GMS 417 printer (Affymetrix). Empty T7 phage without cDNA insert and anti-human IgG at 1:1000 dilution were spotted in triplicate as negative and positive controls, respectively. Before processing, the arrays were rinsed briefly in a 4% nonfat milk/PBS with 0.1% tween-20 to remove unbound phage, and then transferred immediately to 4% nonfat milk/PBS as a blocking solution for 1 hr at room temperature. Without allowing to dry, 2 ml of PBS containing human serum (1:500 dilution) and T7-tag antibody (Novagen) (1:5000) was applied to the surface of the slides. To test the specificity of the immune response, reactive serum was first quenched of non-specific activity by pre-adsorbing with 50 fold higher amount (v/v) of bacterial lysate (OD₆₀₀ of 0.5) and then used for incubation as described below. The arrays were incubated with sera from prostate cancer or control individuals for 1 hour at room temperature and then washed 5 times in PBS/0.1% Tween-20 solution for 5 min each. All washes were performed at room temperature. After washing, the arrays were incubated with 2 ml of PBS containing Cy3-labeled goat anti-mouse antibody and Cy5-labeled goat anti-human antibody (Jackson ImmunoResearch) at a dilution of 1:5000 for both for 1 hr in the dark. Five washes were performed using PBS/0.1% Tween-20 solution with 5 mins each. The arrays were dried by centrifuging at 500 g for 5 min and scanned.

Normalization and Analysis of the Microarray Data. Slides were scanned and quantified using GenePix 4000B scanner (Axon Laboratories). The Cy5/Cy3 ratios were calculated for each phage spot, and values for duplicate spots were averaged. The difference between duplicates was <5% for 98% of the spots. Analyses of repeated experiments using same serum samples revealed that the results were very consistent with correlation coefficient greater than 98%. According to the experimental design, the median of Cy5/Cy3 was utilized so as to control the small variations in the amount of phage particles being spotted. The ratio of Cy5/Cy3 for each spot was subtracted by median of Cy5/Cy3 of the negative T7 empty spots with the observation that the signal for the T7 empty phage on each chip highly correlated with the signal intensity for the whole array. A Z-transformation was applied to data so that the mean of each clone was zero across arrays and the standard deviation was 1.0. Normalized data (without log transformation) was then subjected to two-way clustering analysis with use of Cluster and TreeView.²

Development of Phage-Peptide Predictor. In order to decrease the complexity of subsequent validation studies and develop a focused microarray from the 2,304 element phage-peptide microarray, we evaluated 20 sera from prostate cancer patients and 11 control subjects (chosen randomly from the respective UM cohorts) (**Supplementary Table 3**). Most of the sera from prostate cancer patients contained antibodies that reacted with phage-peptide clones on the microarrays (19/20), while most of the controls did not (1/11) (See **Supplementary Figure 3** for representative images). After normalization of all values obtained by the scanner, we selected phage-peptide clones that yielded a Cy5/Cy3 ratio with an absolute value greater than 1.2 in at least one of the serum samples. This analysis identified 186 phage-peptide clones that reacted with sera from prostate cancer patients, and they, together with negative control phage clones, were used for subsequent screening of sera from patients and controls (training and validation phases).

By employing 186-element phage-peptide microarray platform, 257 sera samples were tested. These samples were divided into training and validation set. Training set was used to build a class prediction model by a leave-one-out-cross-validation (LOOCV) strategy in Genetic Algorithm/K-Nearest Neighbor (GA/KNN) ($k=3$ in this study) method.³ The raw data was normalized as described above. The normalized array data was then applied to GA for selection of the clones and assessment of their relative predictive importance by ranking them based on their frequency of occurrence in GA solutions with the top-most clone assigned a rank of 1. Different numbers of the top-most clones were used to build a different KNN prediction model. Prediction accuracy and error were calculated using LOOCV to evaluate the performance of the models. As few as 10 phage clones performed with similar accuracy, but to maintain a diversity of clones for validation, we used the 22 phage-peptide predictor (**Supplementary Table 10**). Finally, the top-ranked 22 clones were selected based on their best performance on specificity and sensitivity. For the validation sample set, a weighted voting scheme was adopted, similar to that described previously.⁴ Let class 0 and class 1 represent non-cancer and cancer samples, respectively. Each informative phage clone, derived from the training set, casts a weighted vote for a class 0 or 1: $v_x = T_x (e_x - b_x)$ where e_x is the signal value of phage x for each individual validation sample on array images, T_x is the t-statistic for comparing the two class means of phage x in the training set, and b_x is $(\mu_0 + \mu_1)/2$, where μ_0 and μ_1 denote the means of phage x for class 0 or 1 in the training set. A negative v_x indicates a vote for class 0 and a positive value indicates a vote for class 1. The total vote V_0 for class 0 is obtained by summing the absolute values of the negative votes over the informative phage-peptides, while the total vote V_1 for class 1 is obtained by summing the absolute values of the positive votes. The final voting score V_s is $V_1 - V_0$ and the final vote for class 0 or 1 is $\text{sign}(V_s)$ and the *confidence* in the prediction of the winning class is $|V_1 - V_0| / (V_0 + V_1)$, where V_i is the vote for class i .

Sequence Analysis of 22 Phage Clones. The top 22 phage clones were amplified by PCR using T7 capsid forward and reverse primers (Novagen). Briefly, 2 μl of fresh phage lysate with titer of $\sim 10^{10}$ pfu was incubated with 100 μl of 10 mM EDTA, pH 8.0 at 60°C for 10 min. After centrifuging at 14,000 g for 3 min, 2 μl of denatured phage was used for PCR in 100 μl volume of reaction under standard condition. PCR products were confirmed on 1% agarose gel containing ethidium bromide. After purifying with MultiScreen-FB filter plate (Millipore) following manufacturer's protocol, PCR products were sequenced using T7 capsid forward

primer to determine the cDNA inserts. DNA sequence and translated protein sequence were aligned using NCBI BLAST.

ELISA Analysis of Immunoreactivity for Phage Clones. ELISAs were developed for the phage clones to confirm their immunoreactivity with different patient serum. Ninety-six well Max-Sorb™ microtiter plates (NUNC) were coated with 100 µl of diluted T7-tag antibody (Novagen) using 1X PBS at 1:1000 overnight at 4°C on an orbital shaker. All the additions were in 100 µl volumes unless otherwise mentioned. Dilutions of serum and secondary detection reagents were carried out in 1:5 HPE buffer (R&D systems). After washing 5 times with PBS/Tween-20 using an EL404 microplate autowasher (Bio-Tek), the plates were blocked first with 200 µl of 2% BSA/PBS for 2 hrs followed by 200 µl of superbloc (Pierce) for 2 mins, both at room temperature. Phages and the T7 empty phage as negative control were separately diluted at 1:25 to a final titration of $\sim 10^9$ pfu. After washing as above, the plate was incubated with 100 µl of diluted phages for 2 hrs at RT. Serially diluted serum samples (1:500, 1:1000 and 1:2000) were added to each well, and incubated for 1 hr at RT. After washing, the plates were then incubated with 1:10000 diluted HRP-conjugated anti-human IgG for 1 hr at RT. The plates were then developed using 100 µl TMB substrate system (Sigma) for 30 min after final washing. The reaction was stopped using 50 µl of 1.5 M H₂SO₄ and read at 450 nm using an ELx 800 universal microplate reader (Bio-Tek).

Meta-Analysis of Gene Expression. The gene expression level of four genes, namely BRD2, eIF4G1, RPL13a and RPL22, were studied using ONCOMINE (www.oncomine.org).⁵ Briefly, each gene was searched on the database, and the results were filtered by selecting prostate cancer. The data from study classes of benign prostate, prostate cancer and / or metastatic prostate cancer with $p < 0.05$ were used to plot the box plots with SPSS11.5. P values for each group were calculated using student t-test.

Immunoblot Analysis. Tissues were homogenized in NP-40 lysis buffer containing 50 mmol/L Tris-HCl, pH 7.4, 1% Nonidet P-40 (Sigma) and complete protease inhibitor cocktail (Roche). Fifteen µg of protein extracts were mixed with SDS sample buffer and electrophoresed onto a 4-15% linear gradient SDS-polyacrylamide gel under reducing conditions. The separated proteins were transferred onto polyvinyl difluoride membranes (Amersham). The membranes were then incubated for 1 hour in blocking buffer (Tris-buffered saline with 0.1% Tween (TBS-T) and 5% nonfat dry milk). Membranes were incubated with purified eIF4G1 rabbit polyclonal at 1:4000 dilution (Bethyl), RPL22 mouse monoclonal (BD biosciences) at 1:400 dilution, BRD2 rabbit polyclonal (Abgent) diluted at 1:400 and RPL13a rabbit polyclonal (kind gift of Dr. Paul Fox) used at 1:4000 dilution and incubated overnight at 4°C. After washing three times with TBS-T buffer, the membrane was incubated with horseradish peroxidase-linked donkey anti-rabbit IgG or rabbit anti-mouse IgG HRP conjugate (Amersham) at 1:5000 for 1 hour at room temperature. After washing the blots with TBS-T and TBS, the signals were visualized with the ECL detection system (Amersham) and autoradiography. To monitor equal loading, the membranes were incubated with anti-human GAPDH antibody (Abcam) at 1:25,000 dilution for two hours and the signals were visualized.

Immunofluorescence and confocal microscopy. The prostate cancer tissue section slides were soaked in xylene to remove paraffin. Antigen was retrieved by heating the slides in citrate buffer

pH6.0 for 15 minutes in a pressure cooker. The slides were then blocked in PBS-T with 5% normal donkey serum for 1 hour. A mixture of rabbit anti-eIF4G1 (Bethyl) antibody and mouse anti-E-cadherin (BD biosciences) antibody was added to the slides at 1:40 and 1:250 dilutions respectively and incubated for 1 hour at room temperature. Slides were then incubated with secondary antibodies (anti-mouse Alexa 488 and anti-rabbit Alexa 555 at 1:1000 dilution) were incubated for 1 hour. After washing the slides with PBS-T and PBS, the slides were mounted using vectashield mounting medium containing DAPI. Confocal images were taken with Zeiss LSM510 META (Carl Zeiss) imaging system using ultraviolet, Argon and Helium Neon 1 light source. The triple color images were exported as TIFF images and color balanced.

Tissue Microarray Analysis. Tissue cores with diameter of 0.6 mm were taken from a spectrum of benign prostate and prostate cancer tissues and represented in triplicate in a high density tissue microarray using a manual tissue arrayer (Beecher Instruments). Three tissue cores were sampled to account for tumor and tissue heterogeneity. To evaluate eIF4G1 protein expression, immunohistochemistry was performed using standard protocol [DAKO EnVision +System, HRP (DAB), K4011]. 5- μ m thick paraffin embedded tissue microarrays were dewaxed and hydrated in xylene and ethanol respectively. Antigen retrieval was performed in citrate buffer, at pH 6.0 for 15 minutes in pressure cooker. The slides were incubated with eIF4G1 (Bethyl) rabbit polyclonal antibody for 1 hour at 1:50 dilution. The secondary antibody was labelled with biotin and applied for 30 min. Standard biotin-avidin complex immunohistochemistry was carried out (DAKO) for 30 min followed by peroxidase/diaminobenzidine substrate/chromagen. The slides were counterstained with hematoxylin. Protein expression was scored as negative (score = 1), weak (score = 2), moderate (score = 3), or strong (score = 4) using a system that has been validated previously.⁶

II. RANDOMIZATION AND RE-ANALYSIS OF THE PHAGE-PEPTIDE MICROARRAY DATA

To control for potential bias introduced by sequential development of a training and validation set, the autoantibody profiles from all 257 samples (**Fig.2**) were randomly divided into a training set and a validation set with the same proportion of control and cancer sera. We evaluated three different classification approaches including PAM (Predictive Analysis of Microarray)⁷, k -nearest neighbor analysis (KNN)⁸ and diagonal linear discriminate analysis (DLDA)⁹. For KNN and DLDA analysis, the top clones ranked by two-sided t-test were selected, ranging from 10 to 100 and then used to build a classifier. Leave-one-out cross validation (LOOCV) was carried out for each combination of clones in the training set. The optimal number of clones which yielded the lowest misclassification error was selected as the best signature and then applied to the validation samples for prediction. For PAM analysis, we performed default 10-fold cross-validation on the training set, selected a threshold value giving the minimum cross-validated misclassification error rate, and then applied the threshold value to predict validation samples. Thus, for KNN ($k=5$) analysis, we found a prediction accuracy of 82.1% (83.0% sensitivity and 81.3% specificity) on the validation set. Similar performance accuracies on the validation set were observed by PAM and DLDA analysis with 77.8% and 80.3% accuracy, respectively.

III. CLINICAL ASSOCIATIONS OF THE AUTOANTIBODY SIGNATURE

Receiver operating curves (ROC) was generated for the immunoreactivity of the 22 phage peptides and PSA levels in the validation sample of 128 sera (**Fig. 2**). ROC curves, generated for the 22 phage peptides, indicated that, at a specificity of 85.7%, the sensitivity was 84.6% (at a cut-off voting score of -3.64, area under the curve [AUC] was 0.927 [95% confidence interval 0.881- 0.973]; $p < 0.0001$) (**Fig. 4C**). With PSA, the AUC was 0.796 ($p < 0.001$, 95% confidence interval 0.713 - 0.879). Among the subjects with intermediate PSA levels of 4–10 ng/ml (the critical diagnostic range) in the 128 sample set, the phage-peptide predictor still had significant discriminatory power ($p < 0.0001$) in discriminating between patients with prostate cancer and control subjects, whereas PSA was not ($p = 0.498$) (**Fig. 4D**). The AUC was 0.931 (95% CI = 0.859 to 1.0) for the phage-peptide predictor and 0.562 (95% CI = 0.382 to 0.742) for PSA along. Thus, at a cut-off voting score of -3.64, the 22 phage-peptide predictor had high sensitivity and specificity of 85.7% and 85.0%, respectively, in detecting patients with prostate cancer. Further, when the lower limit of PSA was decreased to 2.5 ng/ml, the discriminatory power of the phage-peptide predictor was maintained ($p < 0.0001$), with an AUC of 0.939 (95% CI = 0.879 to 0.998), whereas that for PSA along decreased slightly to 0.496 (95% CI = 0.333 to 0.659) (**Fig. 4E**). The sensitivity and specificity for the predictor at a cut-off voting score of -3.64 were 85.2% and 87.0%, respectively.

To compare the ROCs for the 22 phage-peptide predictor and PSA, a permutation test was performed based on the difference between the AUCs for the two diagnostic techniques accounting for the fact that the same samples were used for both assays. For the entire validation set, the difference in AUCs was significant ($p < 0.001$). In addition, the difference was also significant for the subjects with PSA level between 4-10 ng/ml ($p = 0.004$) and with PSA level between 2.5-10 ng/ml ($p < 0.001$).

To evaluate whether the 22 phage-peptides predictor is a useful supplement to PSA, we performed logistic regression on the validation set. We first used disease (cancer/non-cancer) as the response and carried out univariate logistic regression for the normalized weighted voting scores and PSA respectively (see Methods). We found that both tests are statistically significant (odds ratio [OR] for the voting scores = 74.22, 95% CI = 16.17-340.67, $p < 0.001$; OR for PSA = 4.17, 95% CI = 2.05-8.47, $p < 0.001$). Next, we performed multivariate logistic regression with disease as the response and fit both the voting scores and PSA as covariates. We found that the effect of voting scores was strongly significant (OR = 47.69; 95% CI = 9.47-240.21; $p < 0.001$) after adjusting for the effect of PSA (OR = 2.91; 95% CI = 1.29-6.56; $p = 0.01$), indicating that the 22 phage-peptide predictor provides additional predictive value over preoperative PSA level.

IV. DESCRIPTIONS OF THE 22 PHAGE PEPTIDES

The phage protein microarray strategy allows us to easily identify humoral response targets by sequencing. Thus, we sequenced the 22 top discriminating clones identified by supervised analysis. Five out of the 22 clones were found to be in-frame and in known expressed sequences (**Supplementary Table 16**). These five included Bromodomain Containing Protein 2 (BRD2), Eukaryotic Translation Initiation Factor 4 Gamma 1 (eIF4G1), Ribosomal Protein L22 (RPL22), Ribosomal Protein L13a (RPL13a), and hypothetical protein XP_373908. To our knowledge, none of these proteins have been associated with prostate cancer previously as either an over-expressed protein or as a humoral response target. BRD2, also known as RING3, is a nuclear transcription factor kinase known to be up-regulated in human leukemias.^{10,11} BRD2 has been

shown to specifically interact with acetylated lysine 12 on histone H4.¹² Initiation factors of the eIF4 group are important in the recognition of the 5' cap region of messenger RNAs (mRNA) as well as unwinding of the mRNA structure.¹³ Among them, eIF4G1 plays a central role in the assembly of the preinitiation complex.¹⁴ Interestingly, eIF4G1 has been shown to be overexpressed in head and neck squamous cell carcinoma¹⁵ and squamous lung carcinoma patients^{16,17} and produces a humoral immune response.¹⁸ Furthermore, overexpression of eIF4G1 has been shown to transform NIH3T3 cells.¹⁹ RPL22 and RPL13a are cytoplasmic ribosomal proteins that are the components of the 60S subunit.²⁰ RPL22 has been shown to be overexpressed in lung cancer.^{21,22} RPL13a was identified as a candidate interferon-Gamma Activated Inhibitor of Translation (GAIT) and thus mediates transcript-specific translational control.²⁰

Except hypothetical protein XP_373908, all four of the in-frame phage clones were intracellular proteins involved in regulating transcription or translation in rapidly growing cells. The remaining 17 prostate cancer-specific phage clones were either in un-translated regions of expressed genes or out of frame in the coding sequence of known genes (**Supplementary Table 16**). This is intriguing as these clones likely represent epitopes or part of epitopes that are structurally similar to expressed proteins but unrelated or weakly related at the protein sequence level (**Supplementary Table 17**). Of note, three of the remaining 17 discriminating clones represented an epitope encoded by overlapping sequence from the 5' un-translated region (UTR) of the BMI1 gene (5'-UTR_BMI1), which is a Polycomb Group (PcG) protein implicated in various cellular processes including self-renewal.^{23,24} Coincidentally, we have previously reported dysregulation of the PcG protein, EZH2 in prostate cancer²⁵ and PcG proteins in general function as multi-component complexes. Interestingly, protein BLAST analysis of the peptide sequence shared by the three phage clones representing the 5'-UTR_BMI1 identified significant homology (E value = 5×10^{-4}) to a glycine-rich stretch of the androgen receptor (**Supplementary Table 17**). This is especially intriguing as it is well known that androgens play an important role in prostate cancer progression.^{26,27} This was the only phage clone in our 22 phage-peptide predictor that was represented by multiple independent clones. Interestingly, in 1985, Liao and Witte reported that 37% of males and only 3% of females had significant autoantibodies to androgen receptor.²⁸ Although they did not report analysis of androgen receptor autoantibody association with prostate cancer in their study, higher-titer autoantibodies to androgen receptor were more common in males older than 66 (among whom prostate cancer has higher expected prevalence) than in females or younger males.

Many of these markers are represented by novel peptides, the most intriguing of which appears to be a region of the androgen receptor (5'-UTR_BMI1). Interestingly, several of the phage clones identified, represented known proteins including BRD2, eIF4G1, RPL13a, and RPL22, all of which are dysregulated in prostate cancer tissues. The role of these proteins in regulating the transcription and translation of rapidly dividing cancer cells, warrants further investigation as they may also play a role in neoplastic progression. Our approach suggests that analogous autoantibody profiles of other cancers could be developed for screening, diagnosis, and potentially prognosis. As gene expression analyses of tumor specimens have been widely employed for the molecular classification of cancer,^{1,29-31} a similar global approach may be envisioned for autoantibody profiles derived from serum.

V. ELISA VALIDATION OF THE IMMUNOREACTIVITY

To validate the observations we made using phage-peptide protein microarrays, we generated an ELISA using three of the phage clones including the 5'-UTR_BMI1, eIF4G1 and RPL22. Phage particles were purified and coated onto 96-well plates for subsequent incubation with representative sera from prostate cancer patients and controls. As shown in **Supplementary Figure 4A**, prostate cancer patients in general produce a higher immune activity to these phage peptides relative to controls. Titration of the immune activity to the 5'-UTR_BMI1 clone is shown as a representative example in **Supplementary Figure 4B**.

VI. IMMUNOFLUORESCENCE AND IMMUNOHISTOCHEMICAL ANALYSES OF eIF4G1

To assess the protein expression levels of the potential immune response antigens *in situ* we employed immunofluorescence analysis. Only one out of the four antibodies used for immunoblot analysis (**Fig. 5C**) were compatible for tissue staining purposes (data not shown). The antibody that was successful for these applications was directed against the eIF4G1 protein. Importantly, we observed a weak staining in benign epithelia (**Supplementary Figure 5, panel 1, panel 4** (right)) as compared to a strong cytoplasmic staining of eIF4G1 in prostate cancer epithelia (**Supplementary Figure 5, panel 2, panel 4** (left)) and prostate cancer metastasis (**Supplementary Figure 5, panel 3**). Overexpression of eIF4G1 in prostate cancer was further demonstrated by immunohistochemistry on high-density prostate cancer tissue microarrays (**Supplementary Figure 5, panel 7, 8 and 9**). A chi-squared test for association dichotomized the staining intensity with an odds ratio of 4.06 (95% CI = 2.52 to 6.53, $p < 0.001$) (**Fig. 5E**), which indicates the significant over-expression of eIF4G1 staining between cancer and benign tissues.

VII. SUPPLEMENTARY TABLES

Supplementary Table 1. Clinical and pathology information of prostate cancer patients used for construction of phage display cDNA library (NA, not available).

Patient ID	Age	PSA (ng/ml)	Gleason	Gland Weight (g)	Dimension of Maximum Tumor (cm)	Stage	Hormone Treatment	Radiation Treatment
1	60	5.6	3+4	41.5	1.5	T3a+	None	None
2	64	25.0	4+4	60.0	1.0	T3b+	None	None
3	60	23.9	4+5	78.0	1.0	T3a+	None	None
4	57	NA	3+3	38.0	2.2	NA	NA	NA
5	60	4.6	3+3	36.2	1.7	T2	None	None
6	55	4.2	3+3	47.8	0.8	NA	None	None

Supplementary Table 2. Clinical and pathology information for the 29 samples used for biopanning (NA, not available).

Sample ID	Age	PSA (ng/ml)	Gleason	Gland Weight (g)	Dimension of Maximum Tumor (cm)	Stage	Hormone Treatment	Radiation Treatment
PCA-BP-1	56	4.4	8	45.3	1.0	T2a	None	None
PCA-BP-2	66	4.2	8	36.8	3.1	T3b	None	None
PCA-BP-3	74	26.9	7	33.2	1.7	T4	None	None
PCA-BP-4	69	3.1	6	67.6	1.3	T2b	None	None
PCA-BP-5	62	1.5	6	32.6	0.3	T2a	None	None
PCA-BP-6	NA	60.6	7	79.6	2.1	T3b	None	None
PCA-BP-7	70	4.8	6	33.4	2.5	T2b	None	None
PCA-BP-8	63	9.1	8	40.2	1.6	T3a	None	None
PCA-BP-9	66	10.3	7	66.4	0.7	T2a	None	None
PCA-BP-10	69	4.3	7	28.5	1.4	T2b	None	None
PCA-BP-11	73	0.1	8	108.0	1.1	T3a	None	None
PCA-BP-12	59	7.5	6	54.1	0.8	T2b	None	None
PCA-BP-13	65	9.2	6	74.0	1.1	T2b	None	None
PCA-BP-14	48	5.3	6	49.1	2.0	T2b	None	None
PCA-BP-15	74	13.6	7	89.4	2.0	T2b	None	None
PCA-BP-16	54	8.8	6	47.0	0.1	T2a	None	None
PCA-BP-17	58	3.4	7	38.3	1.5	T2b	None	None
PCA-BP-18	63	11.8	7	27.1	1.2	T2b	None	None
PCA-BP-19	50	6.0	7	27.7	1.3	T2b	None	None
Control-BP-1	65	1.0						
Control-BP-2	63	0.7						
Control-BP-3	53	1.5						
Control-BP-4	59	1.1						
Control-BP-5	75	1.5						
Control-BP-6	54	1.4						
Control-BP-7	59	0.9						
Control-BP-8	56	0.8						
Control-BP-9	65	0.9						
Control-BP-10	71	0.9						

Supplementary Table 3. Clinical and pathology information for the 31 samples used with the 2304 element phage-peptide microarray in the discovery phase (NA, not available).

Sample ID	Age	PSA (ng/ml)	Gleason	Gland Weight (g)	Dimension of Maximum Tumor (cm)	Stage	Hormone Treatment	Radiation Treatment
PCA-1	63	1.7	7	NA	NA	T2a	None	None
PCA-2	49	7.4	7	29.7	1.6	T2b	None	None
PCA-3	73	0.1	6	108.0	1.1	T2b	None	None
PCA-4	57	6.2	7	35.0	2.0	T2a	None	None
PCA-5	55	4.5	7	37.1	2.0	T2b	None	None
PCA-6	44	6.0	7	26.7	1.3	T2b	None	None
PCA-7	54	4.3	7	41.2	1.4	T2b	None	None
PCA-8	60	11.8	7	41.3	1.9	T2b	None	None
PCA-9	41	9.0	6	34.3	0.5	T2b	None	None
PCA-10	68	6.8	7	59.9	1.3	T2b	None	None
PCA-11	47	5.3	6	48.6	2.0	T2b	None	None
PCA-12	55	4.4	7	45.0	1.0	T2a	None	None
PCA-13	NA	9.1	8	53.8	1.7	T3b	None	None
PCA-14	46	5.8	7	41.5	1.5	T2b	None	None
PCA-15	58	3.7	7	39.1	1.3	T2a	None	None
PCA-16	59	5.6	9	56.7	4.0	T3a	None	None
PCA-17	51	5.1	6	50.5	0.8	T2a	None	None
PCA-18	65	2.5	7	36.2	2.0	T2a	None	None
PCA-19	57	6.2	7	35.0	2.0	T2a	None	None
PCA-20	55	4.5	7	37.1	2.0	T2b	None	None
Control-1	50	0.3						
Control-2	77	7.6						
Control-3	64	1.0						
Control-4	57	1.7						
Control-5	51	1.5						
Control-6	58	1.1						
Control-7	50	1.4						
Control-8	53	1.7						
Control-9	69	0.9						
Control-10	53	0.7						
Control-11	63	0.6						

Supplementary Table 4. Clinical and pathology information for prostate cancer patients from the Dana Farber Cancer Institute (NA, not available).

Serum ID	PSA (ng/ml)	Age	Gleason	Stage
DFCI_1	11.2	61	3+4=7	T2a
DFCI_2	19.0	57	3+4=7	T3b
DFCI_3	8.4	66	4+5=9	T2a
DFCI_4	14.5	60	3+4=7	T3b
DFCI_5	6.7	49	3+3=6	T2a
DFCI_6	28.6	53	4+5=9	T2b
DFCI_7	NA	75	3+3=6	T2a
DFCI_8	10.1	71	3+4=7	NA
DFCI_9	19.0	68	3+4=7	T2a
DFCI_10	12.0	66	3+4=7	NA
DFCI_11	5.4	63	3+3=6	T2a
DFCI_12	18.9	68	4+4=8	T2a

Supplementary Table 5. Clinical and pathology information of post-prostatectomy patients.

Sample ID	Age	Gleason	PSA (ng/ml) Before Surgery	Stage	Days After Surgery	PSA (ng/ml) After Surgery
Post_1	48	3+4	3.6	T2a	2085	0.1
Post_2	52	3+3	4.5	T2a	2613	0.1
Post_3	65	3+5	7.1	T2b	2800	0.1
Post_4	63	4+5	44.6	T3b	1772	0.8
Post_5	51	3+3	0.5	T2a	2516	0.1
Post_6	55	3+3	21.4	T2b	2285	0.2
Post_7	55	3+3	21.4	T2b	2285	0.4
Post_8	68	3+3	1.4	T2a	1936	0.1
Post_9	49	3+4	7.1	T2b	2645	0.1
Post_10	63	4+5	3.7	T2b	520	0.1
Post_11	58	3+2	8.4	T2b	1835	0.1
Post_12	61	3+3	6.6	T2b	1423	0.1
Post_13	56	3+3	13.0	T2b	1214	0.1
Post_14	55	4+3	11.3	T2b	1537	0.1

Supplementary Table 6. Clinical and pathology information of hormone-refractory prostate cancer patients.

Sample ID	Age	RRP	Gleason	Disease Stage at Diagnosis	PSA (ng/ml)	Current Bone Disease	Current Soft Tissue Disease	Hormone Treatment	Chemotherapy
HR_1	58	Y	5+4=9	T3b	4	N	N	Y	N
HR_2	76	N	N/A	Metastatic	85	Y	Y	Y	N
HR_3	76	Y	N/A	T2	16	Y	Y	Y	N
HR_4	68	N	3+4=7	T1a	15	N	N	Y	N
HR_5	69	N	4+3=7	T2c	239	Y	N	Y	Y
HR_6	76	N	4+4=8	T2	6	Y	N	Y	Y
HR_7	44	N	5+4=9	Metastatic	674	Y	Y	Y	N
HR_8	72	N	1+2=3	Metastatic	7	Y	Y	Y	Y
HR_9	72	N	4+4=8	Metastatic	412	Y	N	Y	Y
HR_10	45	N	4+3=7	Metastatic	0.1	Y	N	Y	Y
HR_11	52	N	4+3=7	Metastatic	183	Y	N	Y	Y

Supplementary Table 7. Clinical and pathology information of lung adenocarcinoma patients (NA, not available).

Sample ID	Age	Gender	Stage	Survival Time	Differentiation
AD10	58	M	Ia	36.0	Well
AD104	63	M	Ia	63.8	Poor
AD108	81	M	Ib	62.0	Moderate
AD109	46	M	Ia	70.8	Moderate
AD111	72	M	Ia	45.8	NA
AD114	73	M	Ia	29.5	Well
AD120	63	M	Ia	40.3	Moderate
AD122	63	M	Ib	38.4	Moderate
AD124	70	M	Ib	21.8	Moderate
AD125	72	M	IIIa	6.3	Poor
AD127	73	M	IIIa	3.2	Poor
AD129	58	M	IIIa	4.9	Poor
AD130	57	M	IIa	24.3	NA
AD136	64	M	IIb	10.3	NA
AD138	67	M	Ia	28.7	NA
AD140	50	M	Ia	25.8	Well
AD141	74	M	Ia	25.2	Moderate
AD142	78	M	Ia	18.9	Moderate
AD143	77	M	Ib	12.1	Poor
AD145	63	M	IIb	22.9	Well
AD147	74	M	Ia	6.2	Poor
AD148	62	M	Ib	23.3	Well
AD149	59	M	Ia	19.7	NA
AD150	85	M	Ib	17.6	NA
AD152	73	M	Ib	17.0	NA
AD155	50	M	Ia	0.3	NA
AD31	56	M	IIIb	47.7	Well
AD32	55	M	IIb	29.3	Well
AD7	46	M	Ib	68.6	Well
AD89	53	M	IV	24.6	NA

Supplementary Table 8. Clinical and pathology information for the 129 samples used in the training set (NA, not available).

Sample ID	Age	PSA (ng/ml)	Gleason	Gland Weight (g)	Dimension of Maximum Tumor (cm)	Stage	Hormone Treatment	Radiation Treatment
Cancer-1	64	7.7	7	NA	NA	T2b	None	None
Cancer-2	63	1.7	7	NA	NA	T2a	None	None
Cancer-3	49	7.4	7	29.7	1.6	T2b	None	None
Cancer-4	55	5.2	7	28.0	1.5	T2b	None	None
Cancer-5	62	10.2	7	45.0	2.0	T2b	None	None
Cancer-6	52	1.3	6	31.0	0.1	T2a	None	None
Cancer-7	59	18.3	7	57.2	2.0	T2b	None	None
Cancer-8	48	3.6	7	47.1	2.0	T2a	None	None
Cancer-9	50	6.7	6	49.4	0.6	T2b	None	None
Cancer-10	65	7.6	6	67.0	1.0	T2b	None	None
Cancer-11	51	1.2	6	30.7	0.1	T2a	None	None
Cancer-12	67	1.4	6	55.9	1.0	T2b	None	None
Cancer-13	59	4.4	6	59.0	2.3	T2a	None	None
Cancer-14	52	3.1	6	38.8	1.3	T2a	None	None
Cancer-15	58	4.4	6	58.9	2.3	T2a	None	None
Cancer-16	56	6.4	7	36.5	2.0	T2a	None	None
Cancer-17	70	11.3	7	80.1	0.9	T2b	None	None
Cancer-18	54	3.7	7	51.0	1.3	T2b	None	None
Cancer-19	54	8.8	6	47.0	0.1	T2a	None	None
Cancer-20	73	0.1	6	108.0	1.1	T2b	None	None
Cancer-21	61	8.2	6	66.0	0.6	T2b	None	None
Cancer-22	67	2.5	7	37.0	1.0	T2a	None	None
Cancer-23	57	6.2	7	35.0	2.0	T2a	None	None
Cancer-24	55	4.5	7	37.1	2.0	T2b	None	None
Cancer-25	44	6.0	7	26.7	1.3	T2b	None	None
Cancer-26	69	2.8	7	28.5	1.4	T2b	None	None
Cancer-27	58	7.3	5	69.5	1.4	T2b	None	None
Cancer-28	59	3.9	7	NA	0.9	T2b	None	None
Cancer-29	58	6.5	5	69.5	1.4	T2b	None	None
Cancer-30	74	13.6	7	89.4	2.0	T2b	None	None
Cancer-31	61	5.3	6	66.0	0.6	T2b	None	None
Cancer-32	69	0.1	7	28.5	1.4	T2b	None	None
Cancer-33	55	0.4	7	48.8	1.8	T3b	None	None
Cancer-34	73	0.1	7	40.5	1.6	T3b	None	None
Cancer-35	65	9.2	6	74.0	1.1	T2b	None	None
Cancer-36	52	0.1	7	42.0	NA	T2b	None	None
Cancer-37	54	4.3	7	41.2	1.4	T2b	None	None
Cancer-38	56	17.0	9	52.0	1.8	T3a	None	None
Cancer-39	61	3.9	6	44.1	0.3	T2b	None	None
Cancer-40	60	11.8	7	41.3	1.9	T2b	None	None
Cancer-41	41	9.0	6	34.3	0.5	T2b	None	None
Cancer-42	59	6.9	7	37.4	1.3	T2b	None	None
Cancer-43	68	6.8	7	59.9	1.3	T2b	None	None
Cancer-44	65	NA	7	NA	NA	T2a	None	None

Cancer-45	71	19.8	7	89.9	2.1	T3b	None	None
Cancer-46	57	1.7	6	43.2	0.5	T2b	None	None
Cancer-47	47	5.3	6	48.6	2.0	T2b	None	None
Cancer-48	56	17.3	7	59.0	1.8	T2a	None	None
Cancer-49	58	3.4	7	38.3	1.5	T2b	None	None
Cancer-50	55	4.4	7	45.0	1.0	T2a	None	None
Cancer-51	66	4.2	NA	36.8	3.1	T3b	None	None
Cancer-52	NA	9.1	NA	NA	NA	NA	None	None
Cancer-53	46	5.8	7	41.5	1.5	T2b	None	None
Cancer-54	58	3.7	7	39.1	1.3	T2a	None	None
Cancer-55	59	5.6	9	56.7	4.0	T3a	None	None
Cancer-56	51	5.1	6	50.5	0.8	T2a	None	None
Cancer-57	51	13	7	48.0	3.0	T2b	None	None
Cancer-58	65	2.5	7	36.2	2.0	T2a	None	None
Cancer-59	48	7.6	7	55.9	1.1	T2a	None	None
Control-1	51	1.2						
Control-2	66	1.7						
Control-3	49	0.9						
Control-4	77	4.2						
Control-5	46	1.4						
Control-6	74	5.6						
Control-7	67	1.1						
Control-8	68	2.1						
Control-9	71	1.8						
Control-10	52	0.1						
Control-11	64	0.7						
Control-12	71	3.4						
Control-13	77	0.2						
Control-14	57	1.1						
Control-15	51	1.2						
Control-16	80	1.4						
Control-17	63	6.4						
Control-18	63	4.3						
Control-19	63	9						
Control-20	67	7.5						
Control-21	62	6.2						
Control-22	68	6.6						
Control-23	60	4.2						
Control-24	64	7.4						
Control-25	65	1.0						
Control-26	72	6.4						
Control-27	55	0.1						
Control-28	66	1.1						
Control-29	63	2.9						
Control-30	61	1.6						
Control-31	83	3.1						
Control-32	52	0.7						
Control-33	77	7.6						
Control-34	69	6.6						
Control-35	50	2.4						

Control-36	63	0.7
Control-37	50	0.3
Control-38	57	1.7
Control-39	63	8.4
Control-40	74	5.3
Control-41	64	1.6
Control-42	66	0.4
Control-43	51	1.5
Control-44	62	1.3
Control-45	59	0.6
Control-46	58	1.1
Control-47	75	1.5
Control-48	80	7.7
Control-49	54	3.5
Control-50	50	1.4
Control-51	63	0.6
Control-52	56	0.7
Control-53	54	1.8
Control-54	62	2.0
Control-55	53	1.7
Control-56	59	0.9
Control-57	56	0.8
Control-58	65	0.9
Control-59	66	5.9
Control-60	71	0.1
Control-61	69	1.0
Control-62	64	0.7
Control-63	60	4.4
Control-64	53	0.7
Control-65	52	0.6
Control-66	71	5.0
Control-67	61	7.3
Control-68	67	7.7
Control-69	NA	5.0
Control-70	62	4.2

Supplementary Table 9. Clinical and pathology information for the 128 samples used in the validation set (NA, not available).

Sample ID	Age	PSA (ng/ml)	Gleason	Gland Weight (g)	Dimension of Maximum Tumor (cm)	Stage	Hormone Treatment	Radiation Treatment
Cancer-1	49	NA	6	NA	NA	NA	None	None
Cancer-2	69	NA	6	NA	NA	NA	None	None
Cancer-3	NA	7.6	7	NA	NA	NA	None	None
Cancer-4	63	44.6	9	53.0	2.3	T3b	None	None
Cancer-5	58	7.1	5	69.5	1.4	T2b	None	None
Cancer-6	48	NA	7	NA	NA	NA	None	None
Cancer-7	53	6.2	6	NA	NA	T2a	None	None
Cancer-8	59	12.9	7	57.2	2.0	T2b	None	None
Cancer-9	75	NA	6	NA	NA	T2a	None	None
Cancer-10	65	7.7	7	NA	NA	T2b	None	None
Cancer-11	52	7.8	7	45.4	1.2	T2b	None	None
Cancer-12	65	3.0	7	78.5	2.1	T2b	None	None
Cancer-13	61	12.3	7	NA	NA	T2a	None	None
Cancer-14	51	2.5	6	28.6	2.2	T2a	None	None
Cancer-15	77	46.3	8	NA	NA	T2a	None	None
Cancer-16	55	11.3	7	37.1	2.0	T2b	None	None
Cancer-17	53	28.6	9	NA	NA	T2b	None	None
Cancer-18	71	10.1	7	NA	NA	NA	None	None
Cancer-19	53	0.1	8	38.0	1.8	T2b	None	None
Cancer-20	75	13.8	8	NA	NA	T2a	None	None
Cancer-21	59	4.7	7	NA	NA	T2a	None	None
Cancer-22	63	5.4	6	NA	NA	T2a	None	None
Cancer-23	53	7.3	5	44.6	0.8	T2b	None	None
Cancer-24	68	18.9	8	NA	NA	T2a	None	None
Cancer-25	51	5.8	7	48.0	3.0	T2b	None	None
Cancer-26	63	NA	8	NA	NA	NA	None	None
Cancer-27	63	0.1	6	40.3	1.2	T2b	None	None
Cancer-28	57	19.0	7	NA	NA	T3b	None	None
Cancer-29	66	8.4	9	NA	NA	T2a	None	None
Cancer-30	56	6.3	7	64.0	1.2	T2a	None	None
Cancer-31	64	10.5	4	62.5	0.2	T2a	None	None
Cancer-32	52	4.7	6	45.0	0.4	T2a	None	None
Cancer-33	46	2.9	6	51.8	0.4	T2b	None	None
Cancer-34	NA	4.8	7	NA	NA	NA	None	None
Cancer-35	69	7.5	7	68.4	1.3	T2b	None	None
Cancer-36	49	7.1	7	29.7	1.6	T2b	None	None
Cancer-37	63	16.3	9	78.0	4.8	T3b	None	None
Cancer-38	53	NA	6	NA	NA	NA	None	None
Cancer-39	77	NA	7	NA	NA	NA	None	None
Cancer-40	68	1.4	6	25.0	1.2	T2a	None	None
Cancer-41	77	20.7	7	120.0	3.1	T3b	None	None
Cancer-42	49	6.7	6	NA	NA	T2a	None	None
Cancer-43	63	6.3	5	45.5	1.3	T2b	None	None
Cancer-44	72	26.3	7	NA	NA	T3b	None	None

Cancer-45	66	12.0	7	NA	NA	NA	None	None
Cancer-46	69	5.6	7	NA	NA	NA	None	None
Cancer-47	68	19.0	7	NA	NA	T2a	None	None
Cancer-48	57	3.7	7	30.4	1.4	T3a	None	None
Cancer-49	60	14.5	7	NA	NA	T3b	None	None
Cancer-50	55	21.4	6	32.4	2.0	T2b	None	None
Cancer-51	47	NA	7	NA	NA	NA	None	None
Cancer-52	62	4.8	7	52.6	1.5	T2b	None	None
Cancer-53	58	8.4	5	69.5	1.4	T2b	None	None
Cancer-54	61	11.2	7	NA	NA	T2a	None	None
Cancer-55	77	2.2	7	NA	NA	T2a	None	None
Cancer-56	69	5.5	6	NA	NA	T2a	None	None
Cancer-57	63	3.7	9	42.5	1.8	T2b	None	None
Cancer-58	71	10.3	6	50.7	2.5	T2b	None	None
Cancer-59	50	5.2	6	50.7	0.8	T2a	None	None
Cancer-60	40	3.1	5	43.0	0.1	T2a	None	None
Control-1	76	4.2						
Control-2	58	5.2						
Control-3	78	4.2						
Control-4	74	5.6						
Control-5	58	4.0						
Control-6	64	9.9						
Control-7	74	8.2						
Control-8	74	5.6						
Control-9	71	6.1						
Control-10	72	4.1						
Control-11	74	1.6						
Control-12	50	1.2						
Control-13	79	1.4						
Control-14	68	18.6						
Control-15	63	24.5						
Control-16	62	12.1						
Control-17	65	0.4						
Control-18	60	7.8						
Control-19	54	NA						
Control-20	63	22.0						
Control-21	53	0.4						
Control-22	76	1.5						
Control-23	50	1.9						
Control-24	77	0.6						
Control-25	67	4.9						
Control-26	51	2.2						
Control-27	60	1.2						
Control-28	50	0.7						
Control-29	54	0.6						
Control-30	69	0.1						
Control-31	52	1.7						
Control-32	55	1.7						
Control-33	75	1.5						
Control-34	61	2.4						

Control-35	61	0.8
Control-36	55	1.9
Control-37	58	0.6
Control-38	68	2.3
Control-39	57	1.3
Control-40	52	0.5
Control-41	77	1.3
Control-42	74	2.3
Control-43	60	1.2
Control-44	75	2.1
Control-45	76	3.6
Control-46	59	0.7
Control-47	60	0.9
Control-48	51	0.8
Control-49	72	4.1
Control-50	74	8.4
Control-51	67	6.6
Control-52	53	0.6
Control-53	67	1.1
Control-54	62	3.0
Control-55	82	3.0
Control-56	53	0.6
Control-57	50	0.1
Control-58	77	7.6
Control-59	68	6.7
Control-60	63	0.7
Control-61	51	0.3
Control-62	57	1.7
Control-63	62	8.4
Control-64	58	0.5
Control-65	65	0.5
Control-66	49	1.4
Control-67	62	7.3
Control-68	66	7.7

Supplementary Table 10. Comparison of different prediction models.

Phage clones were ranked according to their performance in prediction of unknown samples using GA/KNN algorithm. Models were built using the top 10, 20, 21, 22, 23, 24, 25, 25, 30, 50, or 100 clones, and the prediction accuracy and corresponding error for each model were calculated using leave-one-out-cross-validation (LOOCV). Although accuracy of the models was comparable between 10 and 22 clones, the 22 phage-clone model was chosen for further study to keep as many features as possible.

Number of Clones	Prediction Accuracy (%)	Error
10	93	9/129
20	93	9/129
21	93	9/129
22	93	9/129
23	91	12/129
24	90	13/129
25	90	13/129
26	91	12/129
30	91	12/129
50	88	15/129
100	86	18/129

Supplementary Table 11. Summary of class predictions for training sample set.

A prediction model was built using the 22 phage peptides selected from leave-one-out-cross-validation strategy of GA/KNN.³² The column “call” is the prediction of a sample by the model along with the confidence value. Error indicates the misclassified samples.

Sample Name	Call	Confidence	Pathology	Error
Train-1	Normal	0.43	Cancer	*
Train-2	Cancer	0.329	Cancer	
Train-3	Normal	0.326	Cancer	*
Train-4	Normal	0.236	Cancer	*
Train-5	Cancer	0.233	Cancer	
Train-6	Normal	0.227	Cancer	*
Train-7	Cancer	0.226	Cancer	
Train-8	Cancer	0.211	Cancer	
Train-9	Cancer	0.202	Cancer	
Train-10	Cancer	0.195	Cancer	
Train-11	Cancer	0.193	Cancer	
Train-12	Cancer	0.191	Cancer	
Train-13	Cancer	0.18	Cancer	
Train-14	Cancer	0.176	Cancer	
Train-15	Cancer	0.176	Cancer	
Train-16	Cancer	0.168	Cancer	
Train-17	Cancer	0.165	Cancer	
Train-18	Cancer	0.163	Cancer	
Train-19	Cancer	0.156	Cancer	
Train-20	Cancer	0.153	Cancer	
Train-21	Cancer	0.148	Cancer	
Train-22	Cancer	0.146	Cancer	
Train-23	Cancer	0.146	Cancer	
Train-24	Cancer	0.121	Cancer	
Train-25	Cancer	0.116	Cancer	
Train-26	Cancer	0.109	Cancer	
Train-27	Cancer	0.108	Cancer	
Train-28	Cancer	0.105	Cancer	
Train-29	Cancer	0.093	Cancer	
Train-30	Cancer	0.089	Cancer	
Train-31	Cancer	0.088	Cancer	
Train-32	Cancer	0.085	Cancer	
Train-33	Cancer	0.075	Cancer	
Train-34	Cancer	0.074	Cancer	
Train-35	Cancer	0.069	Cancer	
Train-36	Cancer	0.069	Cancer	
Train-37	Cancer	0.067	Cancer	
Train-38	Cancer	0.066	Cancer	
Train-39	Cancer	0.063	Cancer	
Train-40	Cancer	0.056	Cancer	
Train-41	Normal	0.056	Cancer	*
Train-42	Cancer	0.052	Cancer	
Train-43	Cancer	0.05	Cancer	
Train-44	Cancer	0.042	Cancer	
Train-45	Cancer	0.041	Cancer	
Train-46	Cancer	0.04	Cancer	
Train-47	Cancer	0.033	Cancer	
Train-48	Cancer	0.027	Cancer	
Train-49	Cancer	0.024	Cancer	

Train-50	Cancer	0.024	Cancer	
Train-51	Normal	0.021	Cancer	*
Train-52	Cancer	0.019	Cancer	
Train-53	Cancer	0.018	Cancer	
Train-54	Cancer	0.013	Cancer	
Train-55	Normal	0.01	Cancer	*
Train-56	Cancer	0.003	Cancer	
Train-57	Cancer	0.001	Cancer	
Train-58	Cancer	0	Cancer	
Train-59	Cancer	0	Cancer	
Train-60	Normal	0.828	Normal	
Train-61	Normal	0.827	Normal	
Train-62	Normal	0.769	Normal	
Train-63	Normal	0.768	Normal	
Train-64	Normal	0.726	Normal	
Train-65	Normal	0.695	Normal	
Train-66	Normal	0.684	Normal	
Train-67	Normal	0.678	Normal	
Train-68	Normal	0.648	Normal	
Train-69	Normal	0.645	Normal	
Train-70	Normal	0.638	Normal	
Train-71	Normal	0.606	Normal	
Train-72	Normal	0.598	Normal	
Train-73	Normal	0.581	Normal	
Train-74	Normal	0.562	Normal	
Train-75	Normal	0.558	Normal	
Train-76	Normal	0.538	Normal	
Train-77	Normal	0.527	Normal	
Train-78	Normal	0.527	Normal	
Train-79	Normal	0.507	Normal	
Train-80	Normal	0.487	Normal	
Train-81	Normal	0.463	Normal	
Train-82	Normal	0.453	Normal	
Train-83	Normal	0.439	Normal	
Train-84	Normal	0.437	Normal	
Train-85	Normal	0.433	Normal	
Train-86	Normal	0.41	Normal	
Train-87	Normal	0.403	Normal	
Train-88	Normal	0.394	Normal	
Train-89	Normal	0.37	Normal	
Train-90	Normal	0.367	Normal	
Train-91	Normal	0.358	Normal	
Train-92	Normal	0.355	Normal	
Train-93	Normal	0.354	Normal	
Train-94	Normal	0.345	Normal	
Train-95	Normal	0.331	Normal	
Train-96	Normal	0.324	Normal	
Train-97	Normal	0.323	Normal	
Train-98	Normal	0.233	Normal	
Train-99	Normal	0.231	Normal	
Train-100	Normal	0.226	Normal	
Train-101	Normal	0.222	Normal	
Train-102	Normal	0.216	Normal	
Train-103	Normal	0.197	Normal	
Train-104	Normal	0.166	Normal	
Train-105	Normal	0.151	Normal	

Train-106	Normal	0.142	Normal	
Train-107	Normal	0.14	Normal	
Train-108	Normal	0.138	Normal	
Train-109	Normal	0.136	Normal	
Train-110	Normal	0.132	Normal	
Train-111	Normal	0.13	Normal	
Train-112	Normal	0.102	Normal	
Train-113	Normal	0.097	Normal	
Train-114	Normal	0.097	Normal	
Train-115	Normal	0.087	Normal	
Train-116	Normal	0.075	Normal	
Train-117	Normal	0.074	Normal	
Train-118	Normal	0.069	Normal	
Train-119	Normal	0.058	Normal	
Train-120	Normal	0.045	Normal	
Train-121	Normal	0.038	Normal	
Train-122	Cancer	0.016	Normal	*
Train-123	Normal	0.016	Normal	
Train-124	Normal	0.016	Normal	
Train-125	Normal	0.007	Normal	
Train-126	Cancer	0.002	Normal	*
Train-127	Normal	0	Normal	
Train-128	Normal	0	Normal	
Train-129	Normal	0	Normal	

Supplementary Table 12. Class predictions for the independent validation set.

Prediction of the independent validation dataset is based on a “weighted voting” approach of classification.⁴ A 22 phage-peptide predictor was used to assess 128 samples as either “Normal” or “Cancer”. See **Supplementary Table 11** for column descriptions.

Sample Name	Call	Confidence	Pathology	Error
Test-1	Cancer	0.61	Cancer	
Test-2	Cancer	0.332	Cancer	
Test-3	Cancer	0.006	Cancer	
Test-4	Cancer	0.168	Cancer	
Test-5	Cancer	0.3	Cancer	
Test-6	Cancer	0.226	Cancer	
Test-7	Cancer	0.423	Cancer	
Test-8	Normal	0.016	Cancer	*
Test-9	Cancer	0.159	Cancer	
Test-10	Cancer	0.164	Cancer	
Test-11	Cancer	0.429	Cancer	
Test-12	Cancer	0.924	Cancer	
Test-13	Cancer	0.182	Cancer	
Test-14	Cancer	0.958	Cancer	
Test-15	Normal	0.334	Cancer	*
Test-16	Cancer	0.614	Cancer	
Test-17	Cancer	0.253	Cancer	
Test-18	Normal	0.224	Cancer	*
Test-19	Cancer	0.947	Cancer	
Test-20	Cancer	0.002	Cancer	
Test-21	Cancer	0.283	Cancer	
Test-22	Normal	0.484	Cancer	*
Test-23	Cancer	0.28	Cancer	
Test-24	Cancer	0.027	Cancer	
Test-25	Cancer	0.177	Cancer	
Test-26	Cancer	0.823	Cancer	
Test-27	Cancer	0.277	Cancer	
Test-28	Normal	0.031	Cancer	*
Test-29	Normal	0.101	Cancer	*
Test-30	Cancer	0.717	Cancer	
Test-31	Normal	0.314	Cancer	*
Test-32	Cancer	0.365	Cancer	
Test-33	Cancer	0.282	Cancer	
Test-34	Cancer	0.907	Cancer	
Test-35	Cancer	0.567	Cancer	
Test-36	Normal	0.294	Cancer	*
Test-37	Cancer	0.946	Cancer	
Test-38	Cancer	0.528	Cancer	
Test-39	Cancer	0.075	Cancer	
Test-40	Cancer	0.762	Cancer	
Test-41	Cancer	0.926	Cancer	

Test-42	Cancer	0.225	Cancer	
Test-43	Cancer	0.931	Cancer	
Test-44	Cancer	0.319	Cancer	
Test-45	Cancer	0.231	Cancer	
Test-46	Cancer	0.579	Cancer	
Test-47	Cancer	0.816	Cancer	
Test-48	Cancer	0.835	Cancer	
Test-49	Cancer	0.778	Cancer	
Test-50	Cancer	0.679	Cancer	
Test-51	Cancer	0.287	Cancer	
Test-52	Normal	0.364	Cancer	*
Test-53	Normal	0.54	Cancer	*
Test-54	Cancer	0.722	Cancer	
Test-55	Cancer	0.401	Cancer	
Test-56	Cancer	0.682	Cancer	
Test-57	Cancer	0.574	Cancer	
Test-58	Normal	0.771	Cancer	*
Test-59	Cancer	0.53	Cancer	
Test-60	Cancer	0.239	Cancer	
Test-61	Normal	0.646	Normal	
Test-62	Normal	0.366	Normal	
Test-63	Cancer	0.538	Normal	*
Test-64	Normal	0.849	Normal	
Test-65	Normal	0.766	Normal	
Test-66	Normal	0.957	Normal	
Test-67	Normal	0.127	Normal	
Test-68	Normal	0.876	Normal	
Test-69	Normal	0.863	Normal	
Test-70	Normal	1	Normal	
Test-71	Normal	0.044	Normal	
Test-72	Cancer	0.189	Normal	*
Test-73	Normal	0.508	Normal	
Test-74	Normal	0.449	Normal	
Test-75	Normal	0.766	Normal	
Test-76	Normal	0.77	Normal	
Test-77	Normal	0.2	Normal	
Test-78	Normal	0.968	Normal	
Test-79	Normal	0.971	Normal	
Test-80	Normal	0.538	Normal	
Test-81	Normal	0.118	Normal	
Test-82	Normal	0.595	Normal	
Test-83	Normal	0.844	Normal	
Test-84	Normal	0.237	Normal	
Test-85	Normal	0.16	Normal	
Test-86	Cancer	0.07	Normal	*
Test-87	Normal	0.877	Normal	
Test-88	Normal	0.789	Normal	
Test-89	Normal	0.951	Normal	

Test-90	Normal	0.93	Normal	
Test-91	Normal	0.996	Normal	
Test-92	Normal	0.446	Normal	
Test-93	Normal	0.897	Normal	
Test-94	Normal	0.475	Normal	
Test-95	Normal	1	Normal	
Test-96	Normal	1	Normal	
Test-97	Normal	0.864	Normal	
Test-98	Normal	0.504	Normal	
Test-99	Normal	0.906	Normal	
Test-100	Normal	0.456	Normal	
Test-101	Normal	0.902	Normal	
Test-102	Cancer	0.051	Normal	*
Test-103	Normal	0.193	Normal	
Test-104	Normal	0.122	Normal	
Test-105	Normal	0.799	Normal	
Test-106	Normal	0.291	Normal	
Test-107	Normal	0.973	Normal	
Test-108	Normal	0.756	Normal	
Test-109	Cancer	0.503	Normal	*
Test-110	Normal	1	Normal	
Test-111	Cancer	0.746	Normal	*
Test-112	Normal	0.479	Normal	
Test-113	Normal	0.428	Normal	
Test-114	Normal	0.625	Normal	
Test-115	Normal	0.018	Normal	
Test-116	Normal	0.963	Normal	
Test-117	Normal	0.794	Normal	
Test-118	Normal	0.417	Normal	
Test-119	Normal	0.523	Normal	
Test-120	Normal	0.671	Normal	
Test-121	Normal	0.145	Normal	
Test-122	Cancer	0.059	Normal	*
Test-123	Cancer	0.093	Normal	*
Test-124	Normal	0.86	Normal	
Test-125	Normal	0.913	Normal	
Test-126	Normal	0.548	Normal	
Test-127	Normal	0.5	Normal	
Test-128	Normal	0.982	Normal	

Supplementary Table 13. Class predictions for samples in the “other” category (i.e., post-prostatectomy, hormone-refractory prostate cancer, and lung adenocarcinoma patients). Prediction of the independent validation dataset is based on a “weighted voting” approach of classification.⁴ A 22 phage-peptide predictor was used to assess 55 samples as either “Normal” or “Cancer”. See **Supplementary Table 11** for column descriptions.

Sample ID	Call	Confidence	Pathology	Classified as prostate cancer
HR_1	Normal	1	Hormone-refractory PCA	
HR_2	Cancer	0.979	Hormone-refractory PCA	*
HR_3	Normal	0.997	Hormone-refractory PCA	
HR_4	Normal	0.94	Hormone-refractory PCA	
HR_5	Normal	0.738	Hormone-refractory PCA	
HR_6	Normal	0.999	Hormone-refractory PCA	
HR_7	Normal	0.985	Hormone-refractory PCA	
HR_8	Normal	0.896	Hormone-refractory PCA	
HR_9	Cancer	0.092	Hormone-refractory PCA	*
HR_10	Cancer	0.538	Hormone-refractory PCA	*
HR_11	Normal	0.93	Hormone-refractory PCA	
Post_1	Cancer	0.434	Post-prostatectomy PCA	*
Post_2	Normal	0.829	Post-prostatectomy PCA	
Post_3	Normal	0.88	Post-prostatectomy PCA	
Post_4	Cancer	0.414	Post-prostatectomy PCA	*
Post_5	Normal	0.656	Post-prostatectomy PCA	
Post_6	Normal	0.177	Post-prostatectomy PCA	
Post_7	Normal	0.748	Post-prostatectomy PCA	
Post_8	Normal	0.119	Post-prostatectomy PCA	
Post_9	Cancer	0.249	Post-prostatectomy PCA	*
Post_10	Cancer	0.629	Post-prostatectomy PCA	*

Post_11	Normal	0.962	Post-prostatectomy PCA	
Post_12	Cancer	0.175	Post-prostatectomy PCA	*
Post_13	Normal	0.53	Post-prostatectomy PCA	
Post_14	Normal	0.949	Post-prostatectomy PCA	
AD10	Normal	1	Lung Adenocarcinoma	
AD104	Cancer	0.306	Lung Adenocarcinoma	*
AD108	Normal	0.701	Lung Adenocarcinoma	
AD109	Normal	1	Lung Adenocarcinoma	
AD111	Normal	0.919	Lung Adenocarcinoma	
AD114	Cancer	0.532	Lung Adenocarcinoma	*
AD120	Normal	0.138	Lung Adenocarcinoma	
AD122	Normal	0.237	Lung Adenocarcinoma	
AD124	Normal	0.12	Lung Adenocarcinoma	
AD125	Cancer	0.396	Lung Adenocarcinoma	*
AD127	Normal	0.94	Lung Adenocarcinoma	
AD129	Cancer	0.033	Lung Adenocarcinoma	*
AD130	Normal	0.542	Lung Adenocarcinoma	
AD136	Cancer	0.238	Lung Adenocarcinoma	*
AD138	Normal	0.454	Lung Adenocarcinoma	
AD140	Normal	0.608	Lung Adenocarcinoma	
AD141	Cancer	0.561	Lung Adenocarcinoma	*
AD142	Normal	0.165	Lung Adenocarcinoma	
AD143	Cancer	0.565	Lung Adenocarcinoma	*

AD145	Normal	0.704	Lung Adenocarcinoma	
AD147	Cancer	0.312	Lung Adenocarcinoma	*
AD148	Normal	0.753	Lung Adenocarcinoma	
AD149	Cancer	0.113	Lung Adenocarcinoma	*
AD150	Normal	0.461	Lung Adenocarcinoma	
AD152	Normal	0.594	Lung Adenocarcinoma	
AD155	Normal	1	Lung Adenocarcinoma	
AD31	Normal	0.688	Lung Adenocarcinoma	
AD32	Normal	0.728	Lung Adenocarcinoma	
AD7	Normal	1	Lung Adenocarcinoma	
AD89	Normal	0.828	Lung Adenocarcinoma	

Supplementary Table 14. Sequences of 30 randomly selected phage colonies from original phage display library. Of these 30 clones, 28 had unique cDNA inserts with an average length of 442 bp (range= 121-1158 bp). Of the remaining two clones, one had no sequence data and the other was empty clone. Importantly, 6 out of 30 random clones are in-frame expressed, indicating the high-quality of the library. cDNA identity is the original gene assignment of clone sequence in Gene Bank database. Size of inserts is the length, in base pair, of cDNA clone sequence. Translated peptide sequences are the linear amino acids residues translated through reading frame in the phage vector.

Clone No.	cDNA Identity	Size of inserts (bp)	Translated Peptide Sequences (*, stop codon)	Comments
15G3	RAB40C	521	NSCAASGGAAQRPGEQTGPVLPLEAWVTGALAPT GWRVWSCGQSPSSTQERGERRCTWQPTFLLLPKS LPTWG*	3' UTR
2G10	MGST1	384	NSVLPIYNEYFLRF*	3' UTR
2F10	SMYD5	317	LLLRDPREGRCPPHRPFPGPWSQVKPVWPLTHWV LTPSQRSALLSQPPCPLHSSLATG*	3' UTR
6F3	SFPQ	401	WRRSNEHGRSLWFRRPEISTSRWWWHL*	3' UTR
1C6	APEX nuclease	395	RGCSFAHKRGFAGSLNGKLESGPLIKILNWLEGRSR E*	5' UTR
9A10	BIRC3	565	DRCLRLGLSLLVHASSASGQQVYKGGKRLLLDFF FSFFYKSKHLKMET*	5' UTR
19F4	Coagulation factor III precursor	547	NSLVPTKLISFVSANLGQPTIQSFEQVGTKVNVTVED ERTLVRNNTFLSLRDVFGKDLIYTLYYWKSSSSGK VSIF*	in-frame
20D12	NT5C2L1 protein	361	RSQRPEESEPLEKKGKYEKPKAKPLNTSSKKWGSF FIDSVLGLENTEDSLVYTWSCKRISTYSTIAIPSIEAIA ELPLDYKFTRFSSSNSKTAGYYPSLRPHSSN*	in-frame
21E12	CGI-115 protein	577	IKQRDKRLEWEMMCRVKPDVVQDKETERNLQRIAT RGVVQLFNAVQKHQKNVDEKVKEAGSSMRKRAKLI STVSKKDFISVLRGMDGSTNETASSRKKPKAKQTEV KSEEGPGWTILRDDFMMGASMKDWDKESDGPDDS RPESASDSDT*	in-frame
22H2	KIAA0863 protein	530	KRQRNEVMHDSSFSVKKRKLDPDGHGAEDQRHGEE QPPILNADAAPGPEKVTSSVVPFKRQRNESRTEGPIV KDEALQILALDPKYEGRSYEEKKQFLKDYFHKKPY PSKKEIELLSSLFWVWKIDVASFFGKRRYICMKAIKN HKPSVLLGFDMSSELKNVKHRLNFEYEP	in-frame
2D12	P40	425	NGTKLENTLQDIIQENFPNLARQANVQIQEIQRTTPQR YSSRRPTPRHIVRFTKVEMKEKMLRAAREKGRVTL KGKPIRLTADLSAETLQARREWGPIFNILKEKNFQPR ISYPAKLSLRPHSSN*	in-frame

4B10	RPL15	405	NSYWVGEDSTYKFFEVLIDPFHKAIRRNPDQTQWITK PVHKKHREMRGLTSAGRKSRGLGKGHKFHHTIGGSR RAAWRRRNTLQLHRYR*	in-frame
12F5	CTSB (Plus)	619	QPLRQRQPLRCAQAGLQALGCSAGWI*	out-of- frame
11G2	FTL	258	LSSRTSRSQKMSGVKPQTP*	out-of- frame
18H8	MSTP128 mRNA	388	AARAPRPLPTARRDAHSPSTRKEVPSPVTGSCRVQI YTELHNQPKTT*	out-of- frame
19D3	PRDX4	502	NSDL*	out-of- frame
21G2	HSPA8	339	RNAWGISWWWSSSLWWCFLRAHH*	out-of- frame
22D10	RPS27	345	RPTHTRTCLSQRISFIPLQKRRRGNTRRNAWCRAPI PTSWM*	out-of- frame
6A6	Complement component C3	825	VRERSGRGPWSSSRRGTPSSWPSDNPALPLRPS*	out-of- frame
6E7	LTBP1	266	RSQKWQLHPLKRKYLHWIKRKNLSLVNPSCQLQAF PLFICIHFSQ*	out-of- frame
12F7	Mitochondrion	236	PVSGVLALALLAKLFLVNSLCRRYRG*	
15C12	Clone	1158	KWLNGWIDFFFTLHHDIQSDREVFISHGPCCQIAYSI V*	
15H5	mRNA clone	394	GPRPHMPS*	
18B3	Clone (chromosome 10)	376	WICKRFIETIHELNLFNKRKFFKRAGCWLGAHAHAC NPSTLEGRGGQSQGQEFETGLTIPMAKPHLCQHYK NQPGVVAHAGNPSYSGG*	
20E1	TMX2	500	VRGLVKVQGISPPSPTTQRKVSSLYSPLLNPBKIGG CSILSSMIKTLWLKVSP*	
21F7	cDNA clone	121	ITGNLFSCDNNFFKKRIFTFQKYVCLYAH*	
24H1	EEF1A1	250	IWPGWFRIIT*	
3C10	AK2B	395	NSITSSFSFLFSMRSSSISACLTVRGKPSRRKPFLH KGVSKFFSMSSTTISLTSFTASPR*	
6D7	T7 Vector			
13B10	No Sequence Data			

Supplementary Table 15. Sequences of 49 biopanned phage colonies. Of these 49 clones, 47 had unique cDNA inserts with an average length of 438 bp (range= 152-1014 bp), in which two clones (1G4 and 17F8) contained same cDNA sequence. Other two clones had no sequence data. To be noted, 11 out of 49 sequenced clones were in-frame expressed. See **Supplementary Table 14** for column description.

Clone No.	cDNA Identity	Size of inserts (bp)	Translated Peptide Sequences (*, stop codon)	Comments
20C12	ACTB	152	TKHCRVPPKKFVPQLEMYEGFWSPWEWVEAR AYLYTDLRPVE*	3' UTR
20F1	Glutamine synthase	439	WGQGWSQNTCPHRFRDKSWLVNLSMLLTEGV LGLFSGRNSMSLKQAF*	3' UTR
22F8	SMAP-4	606	NSTFSRLINNERVRKLLWNIQVNFVKDSEKWK LELLLR*	3' UTR
3C6	FZD6	406	NSFRQLTRILLNIKKCDSDISANQMEKR*	3' UTR
4B1	Spermidine synthase	231	PRPLQPRAGPPAGSRPPTKCYKPQNAARPALL GGLSVCLSLWRSTSKPIPAVYSAISLPSVAPHSP NTCIYSK*	3' UTR
16F7	BMI1	774	GGGGVGGRRGGGGGGGGRGAGGGRGAGAGG GRPEAA*	5' UTR
2E6	TFIP11	870	RCPAYTGMVKLNLVRWTLGQTARERQARGLTS STPEFEPYEQWNLGTTKDPDIILSSVIEKRIWDL RRYLLIF	5' UTR
9E4	C20orf11	294	QRCLSVHAQCGILRLTALPALPSALGGASKPLP QLLFFSCRNKSE*	5' UTR
1D7	RPL18	590	SSRRTNSTFNQVVLKRLFMSRTNRPPLSLRMI RKMKLPGRENKTAVVVGTTDDVRVQEVPKLV CALRVTSRARSRLRAGGKILTFDQLALDSPKGC GTVLLSGPRKGREVYRHFGKAPGTPHSHTKPY VRSKGRKFERARGRRASRGYKN*	in-frame
20A6	DC24	552	PICFNFNIHSHRTNHILYLLRNHTYPHLGYHHPM RQPARTPERRHILPILHPSRLPSPHRTNLHSQH PRLTKHSTTDSHCPRTIKLLSQQLNMTSLHNSFY SKDTSRLRTPLMTP*	in-frame
20H11	DNA topoisomerase II	689	KKVVEAVNSDSDSEFGIPKKTTPKGKGRGAKK RKASGSENEGDPGRKTSKTTSKPKKTSFD QDSDVDIFPSDFTEPPSLPRTGRARKEVKYFA ESDEEEDDVFAMFN*	in-frame
20H2	mitochondrial RPL32	332	QRGVLARNYWERLLRKLQPGRPFPPWGPAL AVQGPAMFTEPANDTSGSKENSSLLDSIFWMA	in-frame

APKNRRTIEVNRCRRRNPNQKLAAALE*

22F10	RPL23	480	PAFKMSKRGRGGSSGAKFRISLGLPVGAVINCA DNTGAKNLYIISVKGIKGRNLRLPAAGVGMV ATVKKGKPELRKKVHPAVVIRQRKSYRRKDG LYFEDNAGVIVNNKGEMKGSAITGPVAKECADL WPKACGRTRVTS*	in-frame
24D8	NEFH	346	EKTEVAKKEEAEDKKKVPTPEKEAPAKVEVKED AKPKEKTEVAKKEPDDAKAKEPSKPAEKKEAVP EKKDTKEEKAKKPEEKPKTEAKAKLAAALE*	in-frame
2A10	Prostaglandin D2 synthase 21kDa	354	ATHHTLWMGLALLGVLGDLQAAPEAQVSVQPN FQQDKFLGRWFSAGLASNSSWLREKKAALSMC KSVVAPATDGGGLNLTSTFLRKNQCETRSLRPHS SN*	in-frame
2E8	HES1	881	PRAAVRSESRNVLTESARIARGKITDLANLSAAN HDAAIFPGFGAAKNLSTFAVDGKDCKVNEVE RVLKEFHQAGKPIGLCCIAVLAQVLRGVEVTV GHEQEEGKWPYAGTAEAIKALGAKHCVKEVV EAHVDQKNKVTTTAFMCEALHYIHDGIGAMV RKVLELTGK*	in-frame
5A10	BING4 protein	423	QKFCRIDKSRKLPHSKAKTRSRLEVAEAEET SIKAARSELLLAEEPFGFLEGEDGDTAKICQADI VEAVDIASAAKHFDLNLQFGPYRLNYSRTGRH LAFGGRRGHVAALDWVTKKLAAALE*	in-frame
6B11	Serine proteinase inhibitor	337	NSLSEANTKFMFDLFQQFRKSKENNIFYSPI SITSLALGMVLLGAKDNTAQQIKKVLHFDQVT ENTTGKAAATYHVDKSGNVHHQFQKLAAALE*	in-frame
9A8	NIPBL	397	KKYKDRPQIARVVQKTSSGFSVQWMAGSYSGS WTEAKRRDGRKLPVWVDTIKESDIIYKKIAL TSA NKLTNKVVQTLRSLYAAKDGTS*	in-frame
1G4/ 17F8	PTDSS1	635	NSLYWWHSSHSETVLRRLPHRHTVQARRNTML GVWGHWFPGGHCLHKIWTRSL*	out-of- frame (2 clones)
20B3	cDNA FLJ45990	311	RVSRGWGLGTKALPRPYAPWGFWLCKGVKDC PSLPRPLPGFCSFFLAVNYAKGAGKGRVGGG KPGGSLKDWGTGPVRKEPSLRPHSSN*	out-of- frame
20D7	Genomic DNA (chromosome 11)	450	PKEDYLRHLLIKLSKIKNKERILKAARGNK RITYKGT PNCLAVDFSVETLQARREWHDLL KMPREKNP LS*	out-of- frame
21H3	S100A9	423	EKICKIFSRRRIRMKRS*	out-of- frame

4H3	cDNA clone	426	NSWELLTFHRKSASPGDGFVLWYDPFTPRISLF KTCFFKRSLDNESSGRRQQLKTVAFGVLE*
6F4	Clone (chromosome 10)	182	KRSACNLGRFTVLRNLFWHNTFMLQPHRLLR VNL*
8C2	COL10A1	374	KHFVL*
9C10	Fructose-6- phosphate,2-kinase	455	YNWGYSSGVEHLTAEQERGAPGTVAQACNPS PLGGQGGRITRSGDQDHPGQHGETPSIKNTIKTI LANTAKPRLLKTLRSLSWLTWRNPVY*
15E3	No Sequence Data		
2B11	No Sequence Data		

Supplementary Table 16. Sequence identity for top 22 clones.

The average size of in-frame phage clones was 96 amino acids in length with minimum size of 64 residues and maximum of 134 residues, while, for remaining clones, the average size was 41 residues long with minimum and maximum length of 11 and 92, respectively. See **Supplementary Table 14** for column description.

Clone ID	cDNA Identity	Size of inserts (bp)	No. of Clones	Protein Sequences (*, stop codon)	Reading Frame
24E1	eIF4G1	438	1	IRDPNQGGKDITEEIMSGARTAST PTPPQTGGGLEPQANGETPQVAV IVRPDDRSQGAIADRPGLPGPEH SPSESQPSSPSPSPSPVLEPG SEPNLAVLSIPGDTMTTIQMSVEE*	in-frame
8A6	BRD2	504	1	ESRPMSYDEKRQLSLDINKLPGEK LGRVVHIIQAREPSLRDSNPEEIEI DFETLKPSTLRELERYVLSCLRKK PRKPYSTYEMRFISWF*	in-frame
12B6	RPL13a	441	1	RCEGINISGNFYRNKLKYLAFRLK RMNTNPSRGPYHFRAPSRIFWRT VRGMLPHKTKRGQAALDRLKVFD GIPPPYDKKKADGGSCCPQGRAS EAYKKVCLSGAPGSRGWLEVPGS DSHPGGEEEACGRTRVTS*	in-frame
16D12	RPL22	244	1	ITVTSEVPFSKRYLKYLTKKYLKKN NLRDWLRVVANSKESYELRYFQIN QDEEEEESLRPHSSN*	in-frame
21D10	hypothetical protein XP_353238	293	1	PASASILAGVPMYRNEFTAWYRR MSVVYIGITWSVLGSLLYSRTM AKSSVDQKDGSASEVPSELSERP SLRPHSSN*	in-frame
17F10	UREB1	462	1	RMPKEPLKIPVATSRTQASLGKQK CRRRIMMSLRQRWQMGISWMGR LKPTQW*	out-of-frame
24G4	PLS3	537	1	EGSVYQCCEKGKKQVCSQRIFKW MRWLPLRFPKMSLMNSKRPLQKL ISTATDSFVTMFMSSSRKLICHY QDIK*	out-of-frame
8E10	BRMS1L	414	1	APRTRTLRARRSPRMEIAQKMMM KTVKEEEWNVWMKCPILKNSLPIS KINFIKND*	out-of-frame
8D1	5'-UTR_BMI1	774	12	QRSGRDNGDVGAGAPFRLSSTS QPRRIKPIAPPPRAPSPGAGGGG GGGRGGGGGGPGGGGVGGRGG GGGGGGRGAGGGRGAGAGGGGR	

				PEAA*
3C4	5'-UTR_BMI1	728	5	GGGRGAGGGRGAGAGGGRPEAA*
18D2	5'-UTR_BMI1	1105	1	GVGGRGGGGGGGGRGAGGGRGAGAGGGRPEAA*
1B4	cDNA clone	594	1	ILYPETLLKLLISLRRFWAEMMEFSRYTIMSSEN RDNL TSSFPN*
15F1	RP3-323M22	221	1	LVSILLTKTIY*
22B1	cDNA clone	435	1	QSQHGGPENFKI*
2B10	3'-UTR-MEP50	587	1	NSLPLFPPQNSMGPDI FCPGPLSLDVESLNAV FIDF*
21B4	LAMR1	477	1	REMVPRMRRTSRASIIHHPTE*
6E2	SFRS14	316	1	KAECFKNLIVKKQKSLCSGFKEHLNEASILAQVSVSSSKRVWKS WENLISSFMVWNP AHLIISIPNLEKTSDL SMMSKLAAALE*
15H9	cDNA clone	329	1	NNVSALLGWQK*
4B9	cDNA clone	202	1	PFCKFRILSPRCLSDATQWPFKVLFKWDCSSNSFLGPN*
7F8	3'-UTR-EEF2	275	1	PTLFPFLQRETQMSKLILTNALRGLFGYMARSGFCPRKKGKTRG*
20F6	Chromosome 16 clone	448	1	NSDLPGSLVLSSLYDSNVYSESPVFLQAHE*
21H4	cDNA clone	507	1	QKLCQAKEKGMCMKLRMLWECQKLYSLGF*

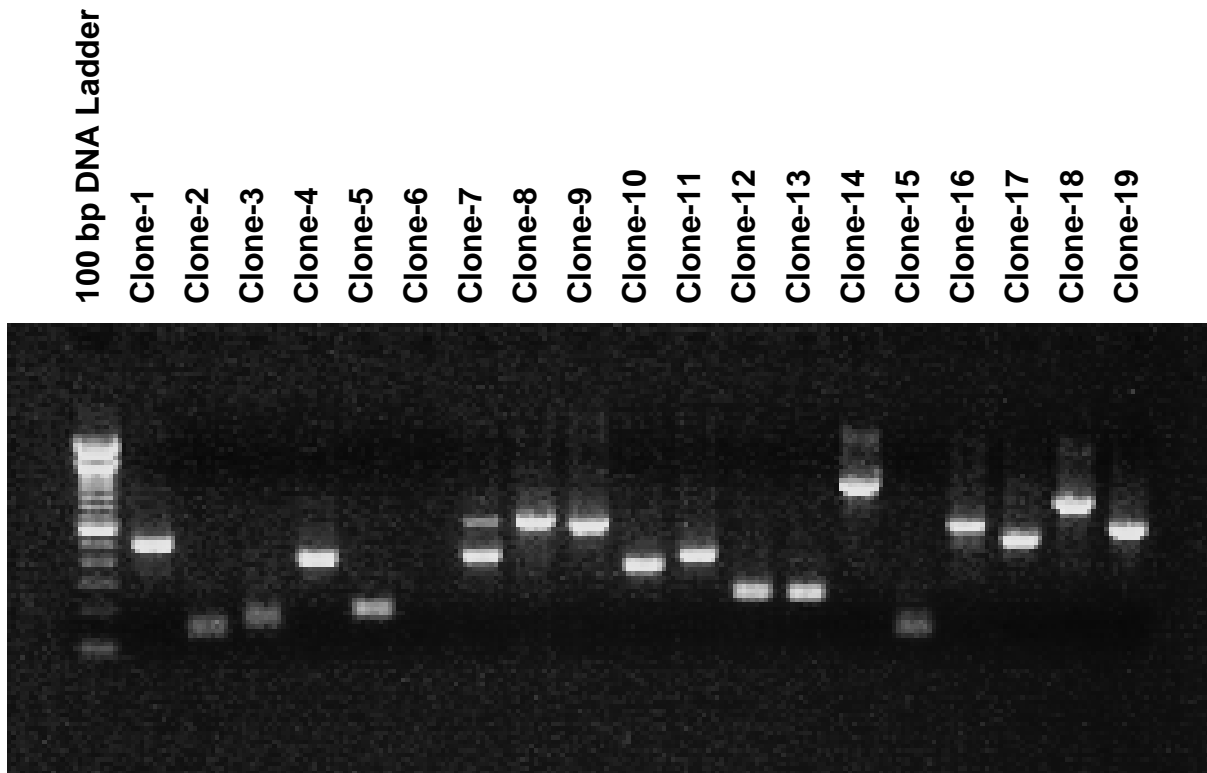
Supplementary Table 17. Protein sequence alignments. Seventeen out-of-frame phage peptides were aligned to protein database in NCBI using NCBI BLAST. Protein identities are known proteins that significantly match to the peptide sequence.

Clone ID	Protein Identity	E value	Sequence Alignment(Query, peptide sequence; Sbjct, protein hit)	Protein Sequence
1B4	unnamed protein product	1.00E-18	Query: ILYPETLLKLLISLRRFWAEMMEFSRYTIMSSENDRDLTSS ILYP+ LLKLLISLR F E MEFSRY IMSS N+DN+TSS Sbjct: ILYPQALLKLLISLRSFCTETMEFSRYRIMSSANKDNMTSS	ILYPETLLKLLISLRRFWAEM MEFSRYTIMSSENDRDLTSS FPN
21B4	hypothetical protein XP_373740	5.00E-05	Query: MVPRMRTSRASIHNIK MVP M R+S+ SIHNIK Sbjct: MVPCMRSSQTSIHNIK	REMVPRMRTSRASIHNIK TE
5'-UTR_BMI1 (8D1 / 3C4 / 18D2)	Androgen Receptor	5.00E-04	Query: GGGRGAGGGRGAGAGGGRPEA GGG G GGG G G GGG EA Sbjct: GGGGGGGGGGGGGGG--EA	GGGRGAGGGRGAGAGGG RPEAA
6E2	Serine/threonine-protein kinase DCAMKL1	0.78	Query: IVKKQKSLCSGFKEHL--NEASIL I+KK K C G KEH+ NE SIL Sbjct: IIKKSK--CRG-KEHMIQNEVSIL	KAECFKNLIVKKQKSLCSGF KEHLNEASILAQVSVSSSKR VWKSWEHLISSFMVWNPAAH LIISIPNLEKTSDLMSMMLIF LLGSRFFRSPRGIF
17F10	hypothetical protein MGC20470	1.8	Query: PLKIPVA----TSRT--QASLGK--QK PLKIP A TSRT ASL K QK Sbjct: PLKIPPARVTLTSRTTAGAASLTKWIQK	RMPKEPLKIPVATSRTQASL GKQKRRRIMMSLRQRWQ MGISWMGRCLKPTQW
24G4	NR1H2	5.0	Query: LRFPMKMLNSKRPLQKLIS LRFPM LM KL+S Sbjct: LRFPRM-LM-----KLVS	EGSVYQCCEKGGKQVCSQ RIFKWMRWLPLRFPKMSLM NSKRPLQKLISATDSFVTM NFMSSSRKLICHYQDIK
20F6	Type XVIII collagen	4.5	Query: YDSNVYSES YDSNV++ES Sbjct: YDSNVFAES	NSDLPGSLVLSLYDSNVY SESPVFLQAAE
21H4	PTPL1-associated RhoGAP 1	3.4	Query: QKLCQAKEKGM +KLCQA E GM Sbjct: EKLCQALENGM	QKLCQAKEKGMCMKLRML WECQKLYSLGF
8E10	Unknown Protein	3.6	Query: PRTRTLRRARRSPRMEI-AQK P TRTLR PRME AQK Sbjct: PTTRTLRL----PRMEAPAQK	APRTRTLRRARRSPRMEIAQK WMMKTVKEEENWVWMMKC PILKNSLPISKINFIKND
7F8	B lymphocyte cell adhesion molecule	2.0	Query: LQRETQMSKLIL---TNALRG L++E QMSKLIL T +RG Sbjct: LEQEQQMSKLILHSVTKDMRG	PTLFPFLQRETQMSKLILTN ALRGLFGYMARSGFCPRKG KGTRG F
4B9	BRF1 protein	7.6	Query: PRCLSDATQWP PRCL A+QWP Sbjct: PRCL-HASQWP	PFCKFRILSPRCLSDATQWP FKVLFKWDCSSNSFLGPN
15H9	MOV10-like 1	28	Query: VSALLGWQK V++LLGW+K Sbjct: VTSLLGWEK	NNVSALLGWQK
22B1	hypothetical protein FLJ40243	28	Query: PENFKI PENFKI Sbjct: PENFKI	QSQHGGPENFKI
15F1	zinc finger protein 292	38	Query: LLTKTIY LLTKT+Y Sbjct: LLTKTVY	LVSILLTKTIY
2B10	BCL9-2	7.2	Query: PPQNSM-----GPDIFCPGPLSLDVESLNA PPQNSM GPD SLNA Sbjct: PPQNSMMMAPGGPD-----SLNA	NSLPLFPQNSMGPDIFCPG PLSLDVESLNAVFIDF

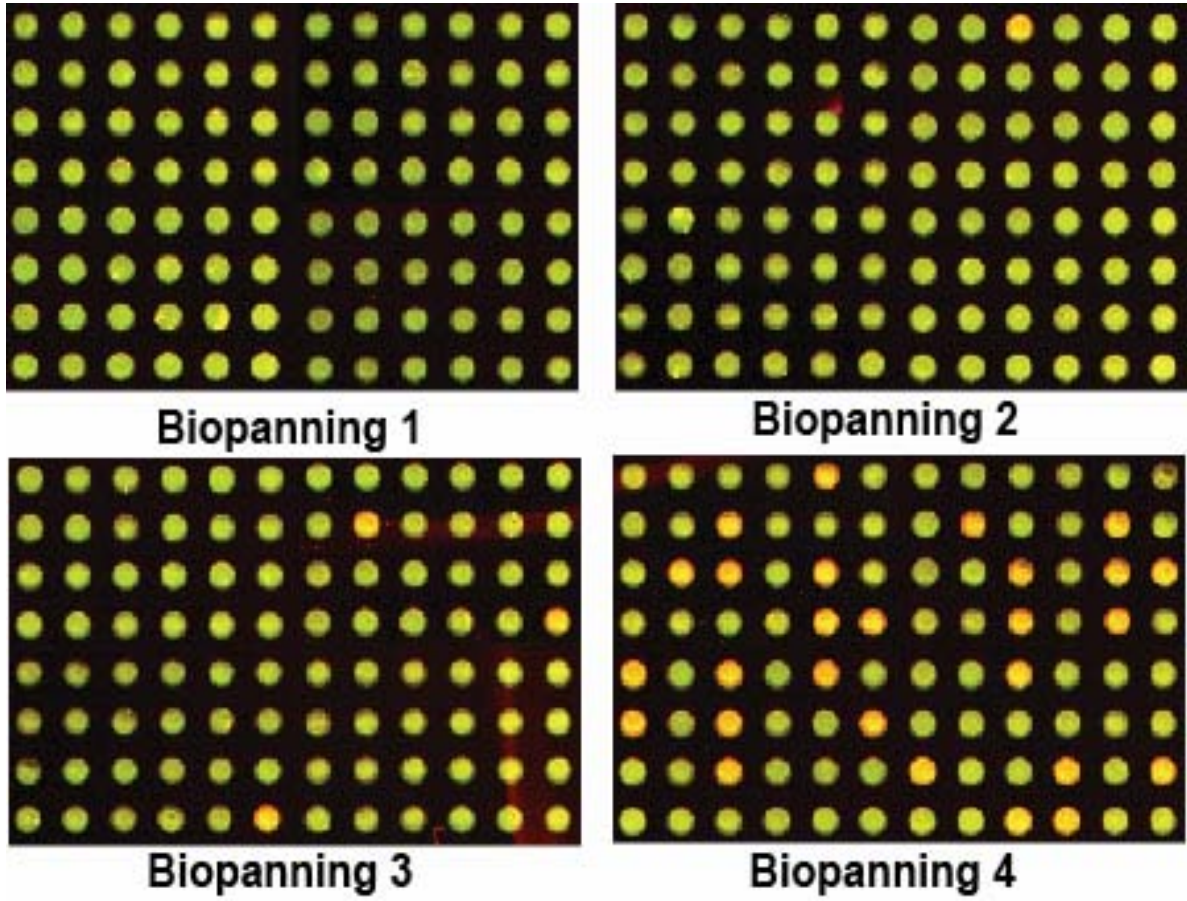
Supplementary Table 18. Association between the 22-phage peptide profile and various clinical and pathological parameters.

Variable	Analysis	PSA	Total Gleason	Major Gleason	Pathological Stage
22-phage profile	Pearson correlation	-.063	-.129	-.014	-.024
	P value	.521	.198	.891	.810

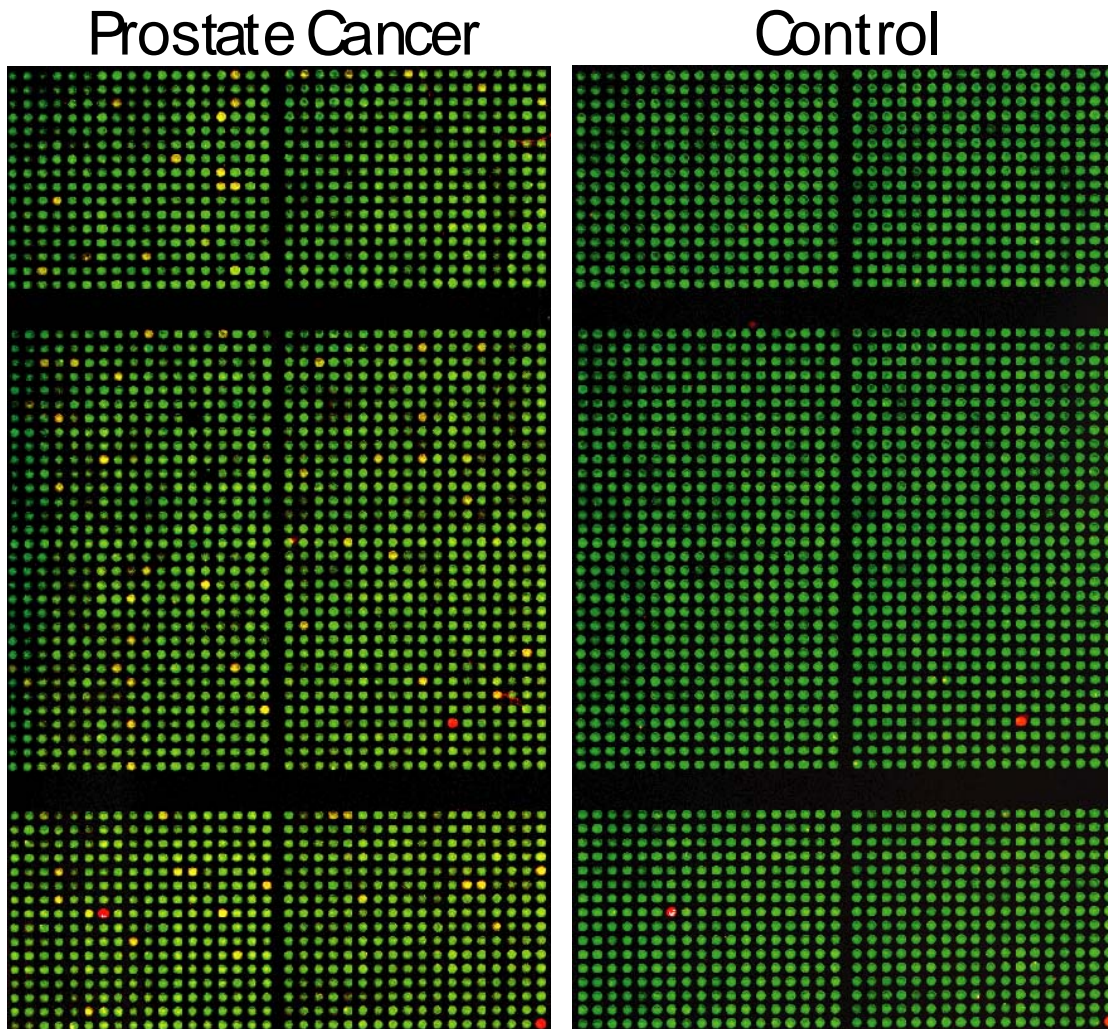
VIII. SUPPLEMENTARY FIGURES



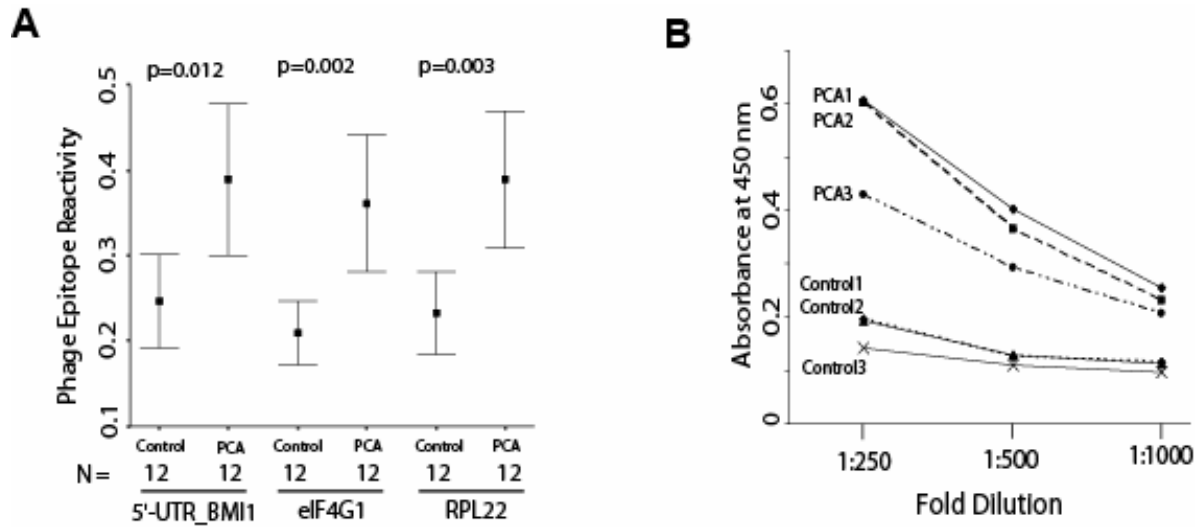
Supplementary Figure 1. Amplification of cDNA inserts by PCR in 19 randomly selected phage colonies from prostate cancer phage display library. Overall, the size of cDNA inserts in this library ranges between 150 bp to 1000 bp in length. Of 19 randomly clones, 95% have cDNA fragment inserts (clone-6 has no band), of which two clones have same bands.



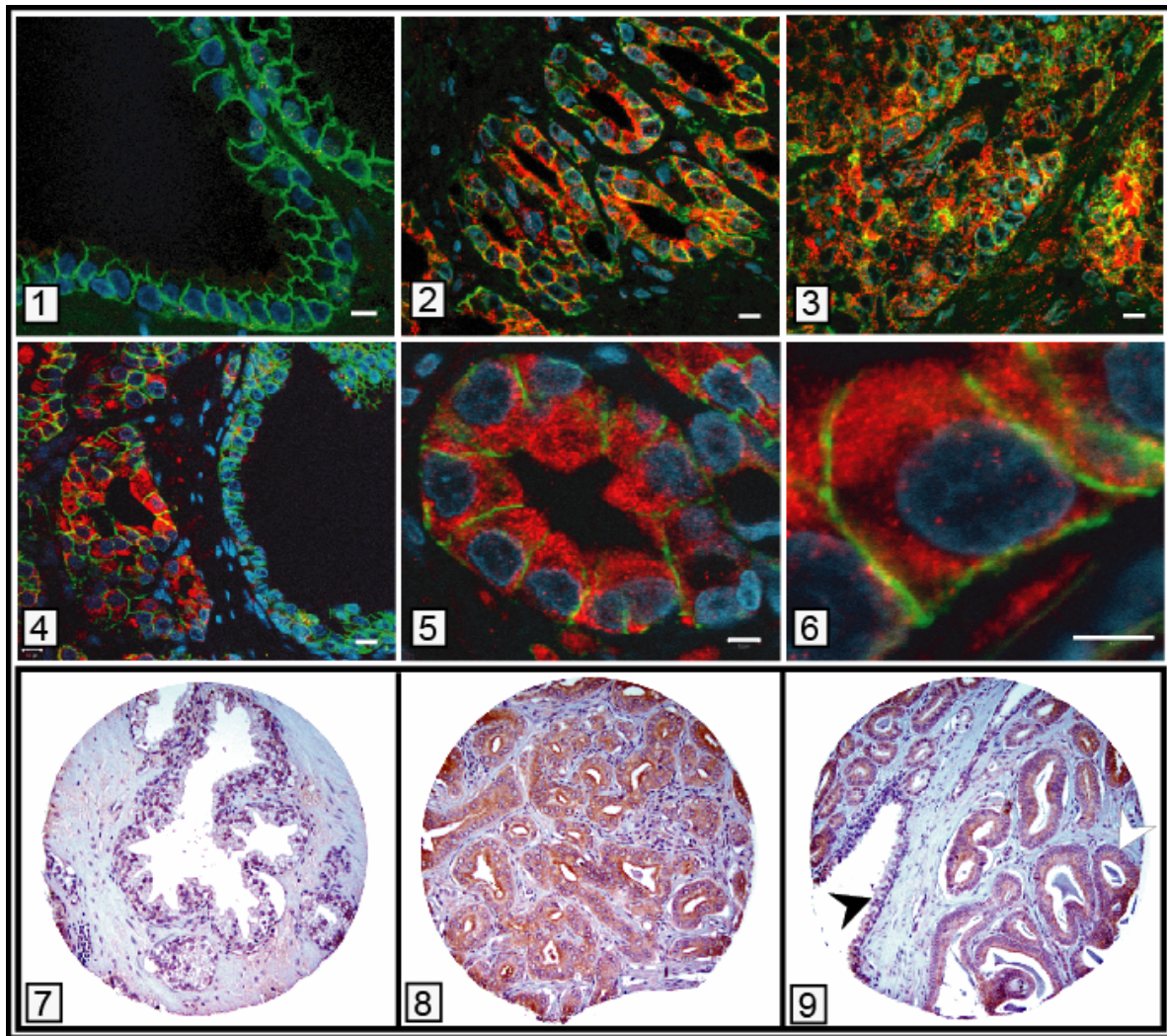
Supplementary Figure 2. Images showing the enrichment of positive clones with progress of different cycles of biopanning.



Supplementary Figure 3. Autoantibody signatures in prostate cancer. Representative images of high-density protein microarray demonstrated the difference of immunoreactivity in prostate cancer patients and controls. Yellow spots represent the immunoreactivity in sera samples while green spots represent no reactivity. The three red spots represent spotted anti-human IgG used as a positive control.



Supplementary Figure 4. ELISA-based validation of the immune response candidates. A, Normalized population mean ratios for the immune activity of 5'-UTR_BMI1, eIF4G1 and RPL22 from 12 control subjects and 12 patients with prostate cancer as assessed by ELISA. Error bars represent 95% confidence intervals. B, Titration curves of the immunoreactivity to a representative phage clone (5'-UTR_BMI1). Titrations of sera from three prostate cancer patients and controls are shown.



Supplementary Figure 5. Expression of the humoral response candidate eIF4G1 in prostate cancer. Immunofluorescence staining of eIF4G1 (panel 1-6). Panel 1 shows benign prostate epithelia, while Panel 2 and 3 displaying clinically localized prostate cancer and metastatic prostate cancer, respectively. Panel 4 shows clinically localized prostate cancer (left) adjacent to a benign gland (right). Panel 5 and 6 demonstrate magnifications of a single prostate cancer gland and single prostate cancer cell. Stains for eIF4G1 (red), E-cadherin (green) and nuclei (blue) were employed. Scale bar represents 5 μ m. Immunohistochemical analyses of human prostate tissues (panel 7-9). Benign prostate glands (panel 7) show weak staining of eIF4G1. Prostate cancer glands (panel 8) reveal strong eIF4G1 expression. Panel 9 shows clinically localized prostate cancer (open arrow) adjacent to a benign gland (black arrow).

IX. REFERENCES

1. Dhanasekaran, S.M. et al. Delineation of prognostic biomarkers in prostate cancer. *Nature* **412**, 822-6 (2001).
2. Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* **95**, 14863-8 (1998).
3. Li, L., Darden, T.A., Weinberg, C.R., Levine, A.J. & Pedersen, L.G. Gene assessment and sample classification for gene expression data using a genetic algorithm/k-nearest neighbor method. *Comb Chem High Throughput Screen* **4**, 727-39 (2001).
4. Golub, T.R. et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531-7 (1999).
5. Rhodes, D.R. et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* **6**, 1-6 (2004).
6. Rubin, M.A. et al. Overexpression, amplification, and androgen regulation of TPD52 in prostate cancer. *Cancer Res* **64**, 3814-22 (2004).
7. Tibshirani, R., Hastie, T., Narasimhan, B. & Chu, G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A* **99**, 6567-72 (2002).
8. Ripley, B.D. *Pattern Recognition and Neural Networks.*, (Cambridge University Press, Cambridge, 1996).
9. Dudoit, S., Fridlyand, J. & Speed, T.P. Comparison of discrimination methods for the classification of tumors using gene expression data. *Journal of the American Statistical Association* **97**, 77-87 (2002).
10. Denis, G.V. & Green, M.R. A novel, mitogen-activated nuclear kinase is related to a *Drosophila* developmental regulator. *Genes Dev* **10**, 261-71 (1996).
11. Denis, G.V., Vaziri, C., Guo, N. & Faller, D.V. RING3 kinase transactivates promoters of cell cycle regulatory genes through E2F. *Cell Growth Differ* **11**, 417-24 (2000).
12. Kanno, T. et al. Selective recognition of acetylated histones by bromodomain proteins visualized in living cells. *Mol Cell* **13**, 33-43 (2004).
13. Gingras, A.C., Raught, B. & Sonenberg, N. Regulation of translation initiation by FRAP/mTOR. *Genes Dev* **15**, 807-26 (2001).
14. Morino, S., Imataka, H., Svitkin, Y.V., Pestova, T.V. & Sonenberg, N. Eukaryotic translation initiation factor 4E (eIF4E) binding site and the middle one-third of eIF4GI constitute the core domain for cap-dependent translation, and the C-terminal one-third functions as a modulatory region. *Mol Cell Biol* **20**, 468-77 (2000).
15. Cromer, A. et al. Identification of genes associated with tumorigenesis and metastatic potential of hypopharyngeal cancer by microarray analysis. *Oncogene* (2003).
16. Bauer, C. et al. Overexpression of the eukaryotic translation initiation factor 4G (eIF4G-1) in squamous cell lung carcinoma. *Int J Cancer* **98**, 181-5 (2002).
17. Bauer, C. et al. Translation initiation factor eIF-4G is immunogenic, overexpressed, and amplified in patients with squamous cell lung carcinoma. *Cancer* **92**, 822-9 (2001).
18. Brass, N. et al. Translation initiation factor eIF-4gamma is encoded by an amplified gene and induces an immune response in squamous cell lung carcinoma. *Hum Mol Genet* **6**, 33-9 (1997).
19. Fukuchi-Shimogori, T. et al. Malignant transformation by overproduction of translation initiation factor eIF4G. *Cancer Res* **57**, 5041-4 (1997).

20. Mazumder, B. et al. Regulated release of L13a from the 60S ribosomal subunit as a mechanism of transcript-specific translational control. *Cell* **115**, 187-98 (2003).
21. Miura, K. et al. Laser capture microdissection and microarray expression analysis of lung adenocarcinoma reveals tobacco smoking- and prognosis-related molecular profiles. *Cancer Res* **62**, 3244-50 (2002).
22. Racz, A. et al. Expression analysis of genes at 3q26-q27 involved in frequent amplification in squamous cell lung carcinoma. *Eur J Cancer* **35**, 641-6 (1999).
23. Park, I.K. et al. Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. *Nature* **423**, 302-5 (2003).
24. Molofsky, A.V. et al. Bmi-1 dependence distinguishes neural stem cell self-renewal from progenitor proliferation. *Nature* **425**, 962-7 (2003).
25. Varambally, S. et al. The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419**, 624-9 (2002).
26. Singh, A. & Figg, W.D. Upregulation of the Androgen Receptor During Prostate Cancer Progression. *Cancer Biol Ther* **3**(2004).
27. Taplin, M.E. & Balk, S.P. Androgen receptor: A key molecule in the progression of prostate cancer to hormone independence. *J Cell Biochem* **91**, 483-190 (2004).
28. Liao, S. & Witte, D. Autoimmune anti-androgen-receptor antibodies in human serum. *Proc Natl Acad Sci U S A* **82**, 8345-8 (1985).
29. Perou, C.M. et al. Molecular portraits of human breast tumours. *Nature* **406**, 747-52 (2000).
30. van 't Veer, L.J. et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530-6. (2002).
31. Bittner, M. et al. Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* **406**, 536-40 (2000).
32. Armstrong, S.A. et al. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat Genet* **30**, 41-7 (2002).