Research Article

RNA-Binding Motif 3 Protein Expression and Nuclear Architecture Changes as a Combined Biomarker to Predict Aggressive and Recurrent Prostate Cancer

Neil M. Carleton[1*], Guangjing Zhu[2], M. Craig Miller[3], Christine Davis[2], Robert W. Veltri[2]

1. Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213
2. The James Buchanan Brady Urological Institute, Department of Urology, The Johns Hopkins University School of Medicine, Baltimore, MD 21287
3. Statistical Consultant, Quakertown, PA

***Corresponding author**: Neil M. Carleton. Carnegie Mellon University, Department of Biomedical Engineering, 5000 Forbes Ave, Pittsburgh, PA 15213. Tel: 412-266-1991; Email: ncarleto@andrew.cmu.edu; ncarleton42@gmail.com

Robert W. Veltri, PhD. James Buchanan Brady Urological Institute, The Johns Hopkins University School of Medicine, 600 N. Wolfe Street, Baltimore, MD 21287. Tel: 410-952-5411; Email: rveltri11@comcast.net

Citation: Neil M. Carleton, et al. RNA-Binding Motif 3 Protein Expression and Nuclear Architecture Changes as a Combined Biomarker to Predict Aggressive and Recurrent Prostate Cancer. *Cancer Research Frontiers*. 2017; 3(1): 192-203. doi: 10.17980/2017.192

Competing Interests: The authors declare no competing financial interests.

Received May 17, 2018; Revised June 25, 2018; Accepted July 9, 2018. Published July 30, 2018

ABSTRACT

Purpose: The RNA-binding motif protein 3 (RBM3) has been shown to be up-regulated in several types of cancer, including prostate cancer (PCa). Increased RBM3 nuclear expression has been linked to improved clinical outcomes. Given this, we examined RBM3 expression and its relation to nuclear morphometric changes in PCa cells.

Methods: This study utilized two tissue microarrays (TMAs) stained for RBM3 that included 80 total cases of PCa stratified by Gleason score. A software-mediated image processing algorithm identified RBM3-positive cancerous nuclei in the TMA samples and calculated twenty-two features quantifying RBM3 expression and nuclear architecture. Multivariate logistic regression (MLR) modeling was performed to determine if RBM3 expression and nuclear structural changes could predict PCa aggressiveness and biochemical recurrence (BCR). A leave-one-out cross validation (LOOCV) was used to estimate predictive performance of the models.

Results: RBM3 expression was found to be significantly downregulated in highly aggressive GS ≥ 8 PCa samples compared to other Gleason scores ($P < 0.0001$) and significantly down-regulated in recurrent PCa samples compared to non-recurrent samples ($P = 0.0377$). An eleven-feature nuclear morphometric MLR model accurately identified aggressive PCa, yielding a ROC-AUC of 0.90 ($P < 0.0001$) in the raw data set and 0.77 (95% CI: 0.83-0.97) for LOOCV testing. The same eleven-feature model was then used to predict recurrence, yielding a ROC-AUC of 0.92 ($P = 0.0004$) in the raw data set and 0.76 (95% CI: 0.64-0.87) for LOOCV testing.

Significant Conclusions: The RBM3 biomarker alone is a strong prognostic marker for the prediction of aggressive PCa and biochemical recurrence.

Key words: prostate cancer, RBM3, nuclear morphometry, biochemical recurrence, aggressiveness

INTRODUCTION

Prostate cancer (PCa) is the most common cancer in males in the United States and the second leading cause of cancer-related deaths in men (1). Due to the heterogeneous nature of PCa progression, there is a growing need to identify biomarkers to aid in early prognostic risk stratification. While many men now present to the clinic with indolent, non-aggressive tumors due to the widespread screening of prostate-specific

antigen (PSA) levels, identifying patients whose tumors will progress into aggressive phenotypes and whose tumors may recur following radical prostatectomy both remain critical concerns in PCa care (2).

Temperature can be an important element of the physical microenvironment that can regulate cellular stem cell-like properties, especially in the context of cancer (3). Temperature effects are modulated at the cellular level by stress-response pathways that include heat-shock and cold-shock proteins (4-6). In contrast to heat-shock proteins (HSP), which are induced by increased temperatures, cold-shock proteins (CSP) are induced by low temperatures but are downregulated when temperatures are elevated (5,7). The RNA-binding motif protein 3 (RBM3) is a glycine rich protein with an RNA recognition motif capable of binding to both RNA and DNA and is one of the earliest proteins synthesized in response to cold shock (4). The role of RBM3 as a putative cancer biomarker was originally unraveled using an antibody-based discovery approach (www.proteinatlas.org) (8-10). RBM3 is an evolutionarily conserved CSP that has been shown to regulate the translation machinery and facilitate protein synthesis during hypothermic stress and in brain development, where it functions as an RNA chaperone to maintain RNA stability (11,12). Recent studies have suggested that RBM3 has a potential proto-oncogenic function and has been found to be up-regulated in various types of human malignancies, including breast, prostate, ovarian, pancreatic, colorectal, and skin (13-16). Further, RBM3 has been specifically shown to have elevated nuclear expression in many cancer types and accumulating evidence has associated this expression with prolonged times to biochemical recurrence (BCR) and clinical tumor progression in the PCa setting (17). In a recent study that used a gene set enrichment analysis to elucidate the role of RBM3 in underlying biological processes, increased RBM3 expression was associated with DNA-dependent replication, chromatin remodeling, and DNA damage response mechanisms (18).

It is well established that nuclear alterations are a hallmark of many types of cancers, including PCa (19). During cancer progression, alterations to the genetic, epigenetic, and tumor microenvironmental landscape result in significant changes to the nucleus of cancer cells, including size and shape changes, textural changes, and spatial changes in the context of the surrounding cells. Genetic and epigenetic variations during cancer are thought to drive large-scale nuclear alterations during cancer progression. Specifically, it is believed that three-dimensional chromatin organizational alterations and altered methylation patterns have important ramifications for altered gene expression profiles and DNA stability in cancer cell nuclei. Through connections mediated by the nuclear lamina and nuclear matrix, three-dimensional chromatin changes can contribute to nuclear pleomorphism in cancer cells.

Given that RBM3 has been hypothesized to play a role in critical nuclear functions such as chromatin remodeling, DNA damage response, and other post-transcriptional processes, we sought to: (1) further elucidate RBM3 expression in archival PCa samples; (2) develop a nuclear morphometric model to see if measures of RBM3 expression levels and nuclear features could be used to predict disease aggressiveness and BCR; and (3) assess the correlation between RBM3 expression and nuclear architectural features, thus providing an insight into how RBM3 is involved in nuclear alterations.

MATERIALS AND METHODS

Immunohistochemical Tissue Microarray Preparation

Two tissue microarrays (TMAs) were obtained from the Prostate Cancer Biorepository Network (PCBN) at the Johns Hopkins Hospital (JHH; Baltimore, MD). Samples for TMAs were obtained from biopsy samples obtained at JHH from 2002-2011 following IRB approval. Each TMA sample includes four replicates of 0.60mm punched cores. The TMAs were prepared using a Beecher MT1 manual arrayer (Beecher Instruments, Silver Spring, MD) in the TMAJ pathology core facility at the PCBN and processed as previously described (20,21). In brief, TMA samples were incubated with anti-RBM3 antibody produced in rabbit (Sigma; Product Number HPA003624) at a 1:200 dilution for 1 hour at room temperature followed by an overnight incubation at 4°C. Secondary anti-rabbit antibody conjugated with horseradish peroxidase (KPL/SeraCare, Milford, MA) was incubated with the samples at a 1:200 dilution. The slides were subsequently stained with hematoxylin and cover-slipped with Vector Mount.

Prostate Cancer Cohort

Table 1: Summary of the clinicopathologic data for the PCa cases in the tissue microarray (TMA) cohort.

Variable	N (total 80)	%
Gleason Score		
3+3	10	12.5
3+4	40	50.0
4+3	20	25.0
≥ 8	10	12.5
TNM Staging		
2	28	35.0
3A	31	38.8
3B	8	10.0
Unknown	13	16.2
Biochemical Recurrence		
No recurrence	49	61.3
Biochemical recurrence	20	25.0
Increase in PSA > 0.2 ng/mL	12	15.0
Local recurrence of PCa	1	1.3
Distant metastasis	2	2.5
Both local recurrence and distant metastasis	2	2.5
Increase in PSA > 0.2 ng/mL after surgery, decrease in PSA	3	3.7
Death from non-PCa related cause	1	1.3
No follow-up data	10	12.5

The stained slides from each TMA were reviewed and graded by an expert uropathologist. Among the 80 cases in this cohort (40 cases from each TMA), 10 were scored as Gleason sum (GS) 3+3 disease, 20 were scored as GS 3+4, 20 were scored as GS 4+3 disease, and 30 were scored as GS ≥ 8 disease. Further, 20 out of 80 total cases exhibited BCR, which was defined as having a prostate-specific antigen (PSA) level increase greater than 0.2 ng/mL following surgery. Patients with BCR had a mean follow-up time of 8.9 years (median: 10 years; range: 3-13 years). Patients that did not experience BCR had a mean follow-up time of 5.3 years (median: 3 years; range: 1-12 years). The complete and detailed clinicopathologic demographics of the TMA are found in Table 1.

Nuclear Morphometric Analysis

A customized image processing algorithm was created using semi-quantitative methods supported by the ImagePro Premier 9.1 software package (Media Cybernetics, Rockville, MD) to segment and quantify RBM3-stained nuclei in each TMA core image. The algorithm was trained to identify nuclear

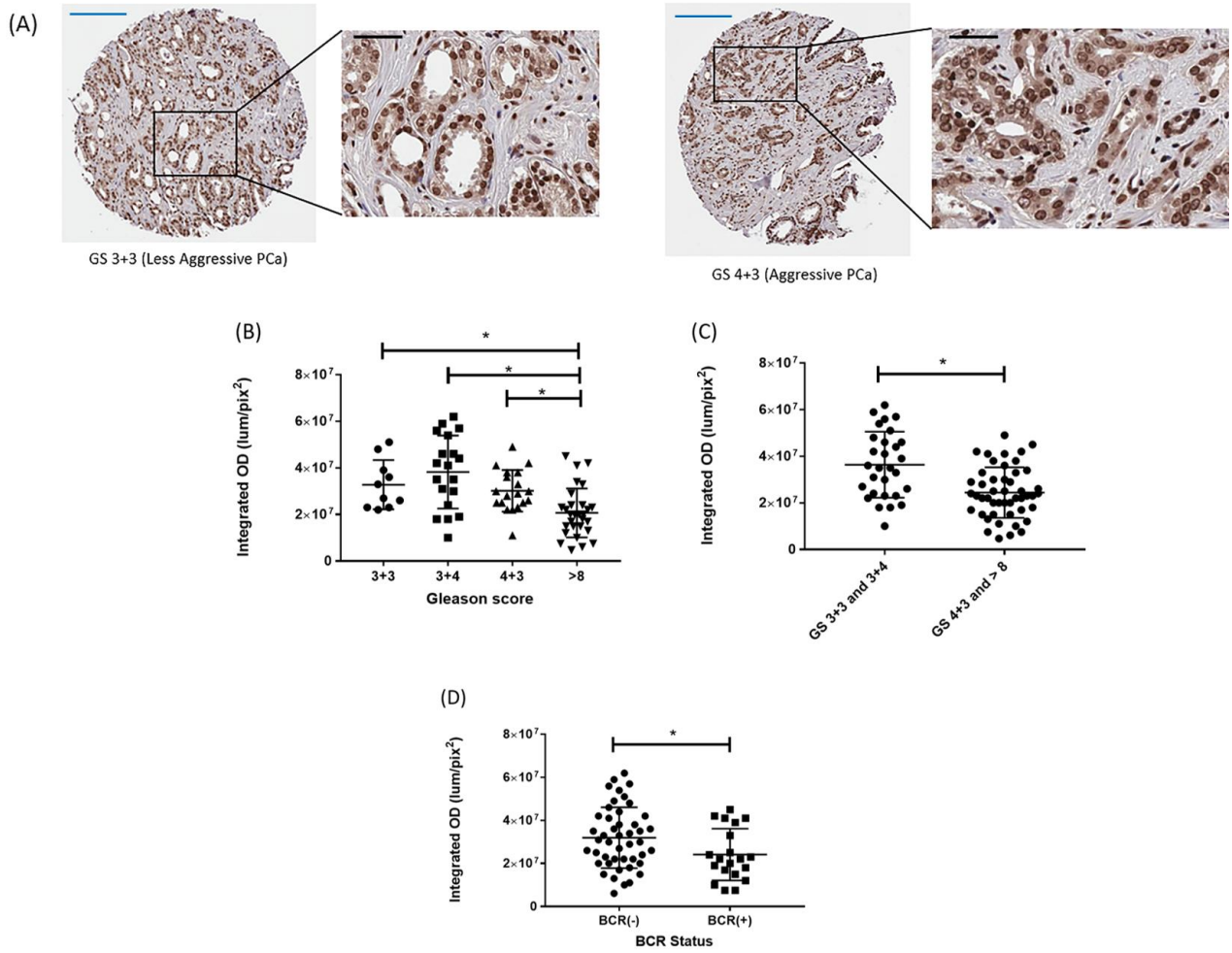

Figure 1: RBM3 is downregulated in aggressive and recurrent prostate cancer. Because RBM3 was previously found to be localized to the nucleus, we examined RBM3 expression in the nucleus of PCa from TMA samples. To quantify RBM3 expression, a measure of staining intensity called Integrated Optical Density (IOD) was used. (A) Representative images of RBM3 staining in TMA images. Scale bar (blue) = 0.25 mm; Scale bar (black) = 0.15 mm. (B) RBM3 expression as represented by IOD varies significantly across Gleason scores. (C) Represents the same data as in (B) but with GS 3+3 and GS 3+4 cases and GS 4+3 and GS ≥ 8 cases grouped together. (D) RBM3 expression is significantly downregulated in biochemical recurrent PCa cases, BCR(+), against cases that did not exhibit biochemical recurrence, BCR(-); * $P < 0.05$.

features from RBM3-positive glandular epithelial cell nuclei by using band-pass filtering for size, color, and intensity. In RBM3-positive nuclei, twenty-two different nuclear features were determined and recorded, including measures of nuclear size and shape and RBM3 staining intensity. One software-derived feature, Integrated Optical Density (IOD), was used as a measure of antibody staining intensity and was therefore used to estimate RBM3 protein content as previously performed (21,22). Using a subset of the most prognostic features (as determined using univariate logistic regression modeling), a multivariate logistic regression (MLR) model was generated to determine if measures of RBM3 staining intensity and nuclear structural changes could be used to distinguish more aggressive (GS 4+3 and GS ≥ 8 disease) PCa from less aggressive (GS 3+3 and GS 3+4) PCa. The same set of significant features was then used in a separate MLR model to determine if they could be used to identify patients who exhibited BCR against those that did not.

Statistical Analysis

Prior to generating the MLR models to identify aggressive PCa and BCR cases, univariate logistic

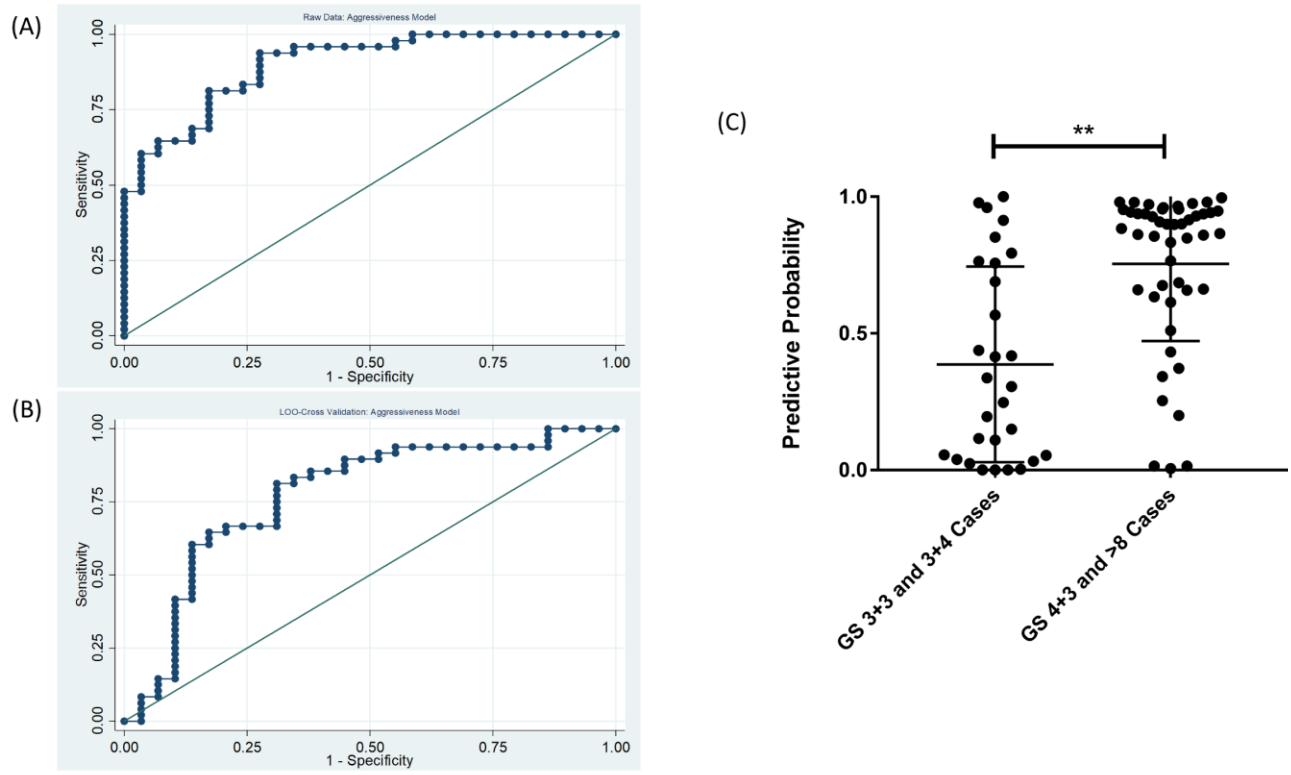

Figure 2: RBM3 and nuclear architectural features predict aggressive prostate cancer. An eleven-feature MLR model was used to predict aggressive (GS 4+3 and GS ≥ 8) PCa cases from less aggressive (GS 3+3 and GS 3+4) PCa cases. (A) Receiver operating characteristic area under the curve (ROC-AUC) = 0.90 in the raw data set ($P < 0.0001$, 95% CI: [0.83-0.97]). (B) ROC-AUC from the leave-one-out cross-validation (LOOCV) testing shows the strength of the model for predicting aggressiveness; ROC-AUC = 0.77 (95% CI: [0.65-0.89]). (C) Plot of the predicted probability from the case in the ROC curve from LOOCV testing in (B). ** $P < 0.0001$.

regression was first performed to identify which features were individually prognostic of aggressive PCa. Using the features that showed univariate significant prognostic capabilities ($P < 0.05$), the MLR models for aggressive PCa prediction and BCR prediction were then generated. Due to the small sample sizes, leave-one-out cross validation (LOOCV) was used to provide insight on how the predictive capabilities of the feature set might behave with respect to an independent patient cohort to address issues such as model overfitting.

Statistical significance was defined as $P < 0.05$. Non-parametric Mann-Whitney U tests were used for all group comparison analyses. Correlations of RBM3 staining intensity with the measures of nuclear structural changes that were in the MLR model were evaluated using Spearman's rank correlation coefficients. GraphPad Prism 7 (GraphPad Software, La Jolla, CA) was used to generate all graphs. STATA 13.0 (StataCorp LLC, College Station, TX) was used for all univariate and multivariate modeling and analysis as well as for the LOOCV approach.

RESULTS

RBM3 is down regulated in aggressive PCa and biochemical recurrent patient samples

Two TMAs containing a total of 80 PCa patient samples were stained for the RBM3 protein. Nuclear features were quantified using the ImagePro Premier 9.1 software. Images of the TMA cores confirmed the nuclear localization of the RBM3 protein (Figure 1A). In order to quantify RBM3 staining intensity in TMA images, the IOD measurement was identified as the most favorable due to its quantification of staining luminance over a given pixel area, an approach previously used by our group (20). The IOD measure of staining intensity from TMA images was used as a measure of RBM3 protein expression. Interestingly, when comparing

Table 2: Summary of the ten features included in the multivariate logistic regression (MLR) model for predicting aggressiveness and biochemical recurrence (BCR).

Feature	Category	Description
Axis Major	Nuclear Size, Shape	Length of major axis of ellipse
Axis Minor	Nuclear Size, Shape	Length of minor axis of ellipse
Bound Box Height	Nuclear Size, Shape	Height of the smallest possible rectangular that can fully enclose the nucleus
Bound Box Width	Nuclear Size, Shape	Width of the smallest possible rectangular that can fully enclose the nucleus
Clumpiness	Texture	Pixel level measurement that measures the texture of the nucleus by quantifying the fraction of pixels deviating from the average
Diameter Max	Nuclear Size, Shape	Length of longest line adjoining two points of region's outline and passing through the centroid
Diameter Mean	Nuclear Size, Shape	Average length of all diameters measured at various intervals around boundary
Diameter Min	Nuclear Size, Shape	Length of shortest line adjoining two points of region's outline and passing through the centroid
Heterogeneity	Texture	Pixel-level measure of nuclear disorder
Integrated Optical Density	Staining Luminance	Quantifies staining intensity due to RBM3 proteins in cancer nuclei

IOD levels across samples of various Gleason scoring, RBM3 showed a significant downregulation in samples with a GS ≥ 8 compared to GS 3+3 samples (P = 0.0018), GS 3+4 samples (P = 0.0001), and GS 4+3 samples (P = 0.0006) (Figure 1B). Similarly, significantly lower levels of RBM3 were observed in BCR cases compared to those that showed no disease recurrence (P = 0.0377).

MLR model distinguishes aggressive PCa phenotype

From the original set of twenty-two nuclear morphometric features that were extracted from the RBM3-positive cancer nuclei, univariate logistic regression revealed that in addition to the IOD for the quantification of the RBM3 staining intensity, ten of the nuclear architectural features were individually significant (P < 0.05) for predicting an aggressive PCa phenotype (GS 4+3 and GS ≥ 8 cases). The full results from the univariate logistic regression for each feature can be found in Supplementary File 1. Table 2 highlights the significant features and their descriptions with respect to quantifying a measure of nuclear architecture or staining intensity. These eleven features were then combined to generate a MLR model for the prediction of aggressive PCa. In the raw data set, this eleven-feature MLR model was significantly predictive for distinguishing aggressive PCa cases from less aggressive cases and yielded a receiver operating characteristic area under the curve (ROC-AUC) of 0.90 (P < 0.0001; 95% CI: 0.83-0.97) (Figure 2A). LOOCV using the eleven features for prediction of aggressive PCa confirmed the predictive capability of the feature set, yielding a

ROC-AUC of 0.77 (95% CI: 0.65-0.89) (Figure 2B), showing that the eleven features generate a robust model for predicting aggressiveness. A plot of the predictive probabilities from the LOOCV testing for each case showed significant separations between the less aggressive versus aggressive PCa cases (P < 0.0001) (Figure 2C).

A file showing the STATA code and STATA outputs from the generation of the MLR model and LOOCV testing to predict aggressiveness can be found in Supplementary File 2.

The eleven-feature MLR model further distinguishes BCR patient samples

We then sought to determine if the same set of eleven features could be used to identify patients with BCR. MLR modeling for prediction of BCR yielded a ROC-AUC = 0.92 (P = 0.0004; 95% CI: 0.85-0.98) (Figure 3A) in the raw data set. LOOCV using the eleven features for prediction of BCR confirmed the predictive capability of the feature set, yielding a ROC-AUC of 0.76 (95% CI: 0.64-0.87). The predictive probability plot of the LOOCV results further shows the significant separation in the BCR-negative and BCR-positive cases (P = 0.0008) (Figure 3C). Additionally, we evaluated the ability of the eleven-feature model to predict BCR compared to the currently used Gleason scoring system (Figure 3D). The ROC-AUC of the eleven-feature model showed comparable performance to the Gleason sum (ROC-AUC = 0.73) for predicting BCR cases (P = 0.70) (Figure 3D).

A file showing the STATA code and STATA outputs from the generation of the MLR model and LOOCV testing to predict BCR can be found in Supplementary File 3.

RBM3 protein staining intensity correlates significantly with features from the MLR model

We evaluated the correlation between RBM3 staining intensity (as represented by the IOD feature) and the other ten univariately significant features from the aggressive and BCR MLR models. We found that RBM3 expression exhibited significant associations with the features with varying degrees of strength of correlation relationships. This included area (Spearman rho = 0.92; P < 0.0001), axis major (Spearman rho = 0.82; P < 0.0001) (Figure 4A), axis minor (Spearman rho = 0.91; P < 0.0001) (Figure 4B), bound box height (Spearman rho = 0.82; P < 0.0001) (Figure 4C), bound box width (Spearman rho = 0.86; P < 0.0001) (Figure 4D), clumpiness (Spearman rho = 0.64; P < 0.0001) (Figure 4E), maximum diameter (Spearman rho = 0.85; P < 0.0001) (Figure 4F), mean diameter (Spearman rho = 0.92; P < 0.0001) (Figure 4G), minimum diameter (Spearman rho = 0.88; P < 0.0001) (Figure 4H), and heterogeneity (Spearman rho = 0.40; P = 0.0003) (Figure 4I). With the strong positive correlations observed here, similar relationships between RBM3 staining and the model parameters were observed as seen in Figure 1B (i.e. as RBM3 expression decreased in higher Gleason score tumors, similar decreases in maximum diameter, for example, also occurred).

DISCUSSION

The RBM3 protein has previously been shown to be upregulated in a wide range of cancer types, including prostate, melanoma, colorectal, breast, and ovarian (14, 16-18, 23, 24) as compared to benign tissues.Overexpression of the nuclear RBM3 protein has been associated with significantly improved survival (16,17). In particular, in the context of PCa, patients with tumors expressing high nuclear levels of RBM3 have a significantly prolonged time to BCR and clinical progression (17). Our study attempts to further clarify the role of RBM3 in PCa by examining its protein expression in a cohort of archival PCa samples with varying Gleason scores and BCR statuses. Further, given the high levels of nuclear localization for RBM3, we sought to determine if RBM3 protein levels in conjunction with other nuclear features could stratify patients that had aggressive (GS 4+3 and GS ≥ 8) PCa and BCR based on changes in RBM3 levels and altered nuclear architectural features. While previous studies have focused on RBM3 expression using only staining intensity by immunohistochemical studies, this study provides the first evaluation of RBM3 in the context of nuclear morphometric changes within human PCa cells. We show that these features can be combined to predict PCa aggressiveness and BCR status.

Previous immunohistochemical analysis of RBM3 first uncovered upregulated RBM3 expression in prostatic intraepithelial neoplasia and invasive prostate tumor samples (17) as compared to benign samples. While this effect was not quantified across various Gleason scores, the same study showed high nuclear expression was specifically associated with improved clinical outcomes (i.e. increased time to

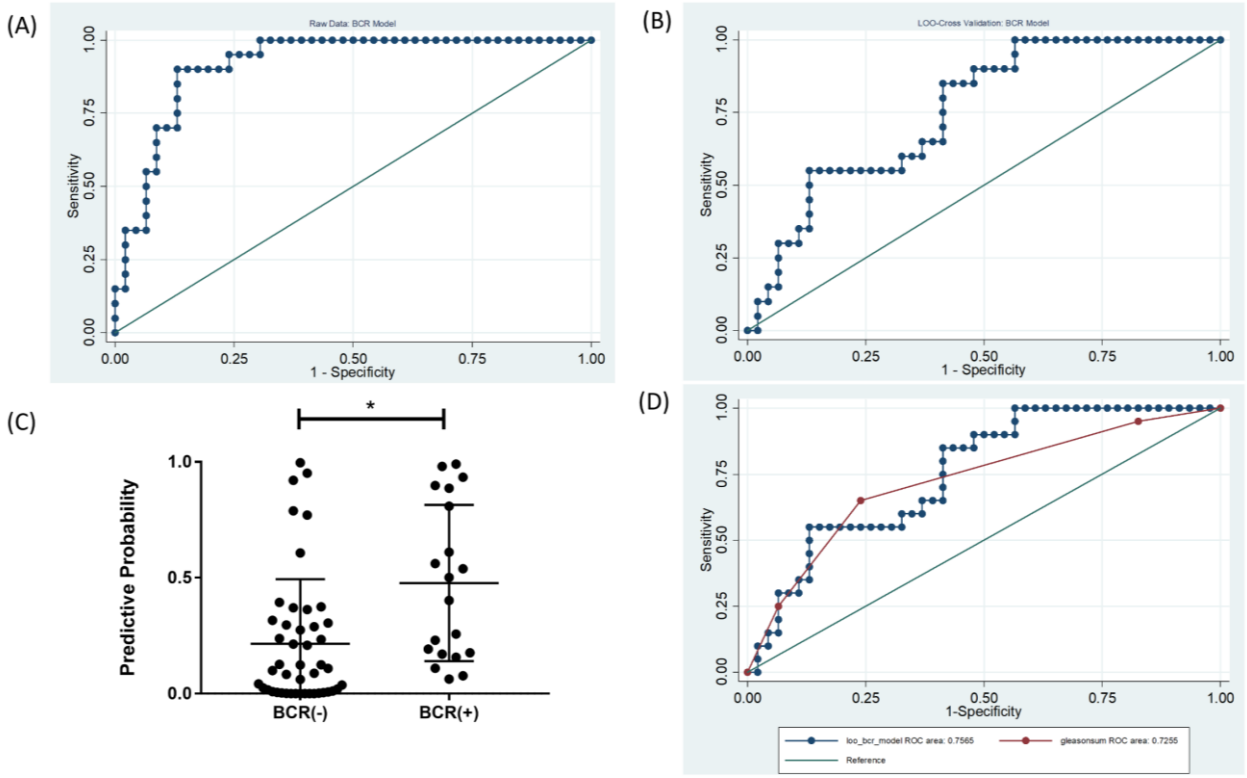

Figure 3: The eleven-feature MLR model further predicts biochemical recurrence. The eleven-feature MLR model was then used to predict biochemical recurrence (BCR). (A) ROC-AUC = 0.92 from the raw data set (*P* = 0.0004; 95% CI: [0.85-0.98]). (B) ROC-AUC from LOOCV testing shows the strength of the model for predicting BCR; ROC-AUC = 0.76 (95% CI: [0.64-0.87]). (C) Plot of the predicted probability from the cases in the ROC curve from the LOOCV set in (B). * *P* = 0.0008. (D) Comparison of the LOOCV model against the currently used Gleason sum shows comparable results for predicting BCR in this sample set (*P* = 0.70).

BCR and disease progression) compared to a low-expressing RBM3 group. Our study examined RBM3 levels across PCa samples of varying Gleason scores and interestingly found that significantly lower RBM3 levels were discovered in higher Gleason score tumors. Further, RBM3 levels were significantly downregulated in PCa that displayed BCR versus those that didn't. In contrast with our results is a study that showed high RBM3 expression was associated with high Gleason score and advanced tumor stage (23). Notably, RBM3 expression was also related to early BCR. The discrepancy in these results may be attributable to the fact that these previous studies used a categorical descriptor of RBM3 staining intensity. Our study used a continuous variable derived from the image processing software, Integrated Optical Density, which may substantiate our results in comparison to categorical measures of staining intensity from the human eye. While our PCa cohort lacked longitudinal survival data, our results suggest that the high RBM3 nuclear expression in low Gleason score tumors, which are typically less aggressive and associated with longer times to BCR, have validity. This result is further substantiated by a 2013 study that showed decreased RBM3 expression in metastatic PCa samples, a finding that suggests that RBM3 upregulation may be important for the early oncogenic events leading to PCa while RBM3 downregulation may have ramifications for progression and metastasis (25). Given that our study further clarifies that RBM3 has a tumor inhibitory role with lower expression levelsin high Gleason score PCa, there now seems to be a growing consensus that RBM3 could be a putative biomarker for PCa. While RBM3 has not previously been shown to be present in the serum or urine of PCa patients, perhaps RBM3 staining of a diagnostic biopsy or a

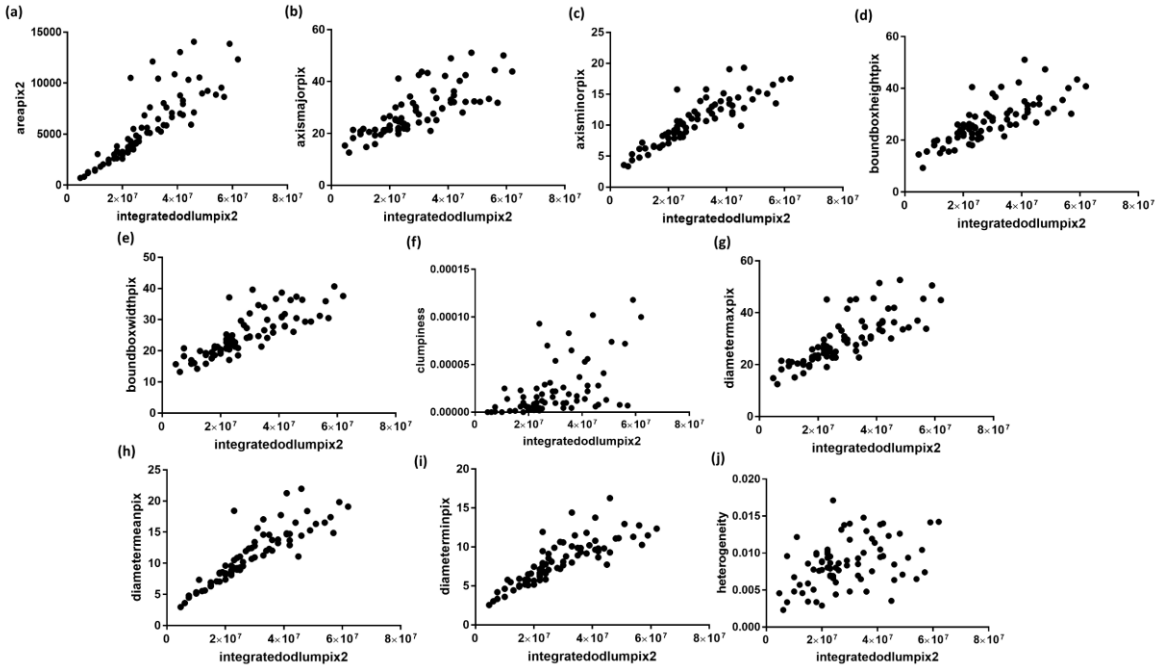

Figure 4: RBM3 staining intensity correlates significantly with nuclear architectural features. Integrated Optical Density (IOD), a measure to quantify the staining intensity of the RBM3 protein expression, correlates significantly with the other ten features included in the eleven-feature model for predicting aggressiveness and BCR:

(a) IOD vs. Area: Spearman rho = 0.92; $P < 0.0001$
(b) IOD vs. Axis Major: Spearman rho = 0.82; $P < 0.0001$
(c) IOD vs. Axis Minor: Spearman rho = 0.91; $P < 0.0001$
(d) IOD vs. Bound Box Height: Spearman rho = 0.82; $P < 0.0001$
(e) IOD vs. Bound Box Width: Spearman rho = 0.86; $P < 0.0001$
(f) IOD vs. Clumpiness: Spearman rho = 0.64; $P < 0.0001$
(g) IOD vs. Diameter Max: Spearman rho = 0.85; $P < 0.0001$
(h) IOD vs. Diameter Mean: Spearman rho = 0.92; $P < 0.0001$
(i) IOD vs. Diameter Min: Spearman rho = 0.88; $P < 0.0001$
(j) IOD vs. Heterogeneity: Spearman rho = 0.40; $P = 0.0003$

surgery sample could provide insight into tumor behavior and aid clinicians into predicting tumor aggressiveness or potential for BCR.

Our data shows that RBM3 staining, in addition to nuclear morphometric features, is predictive of aggressive prostate cancer and biochemical recurrence. The prediction of biochemical recurrence using our novel 11-feature model is slightly better (though not statistically significant) to the currently used Gleason score. We believe a potential application for nuclear morphometry modeling is to aid pathologists in their grading of prostate cancer. Currently, Gleason scoring, which is based on a pathologist's subjective grading, is the best way to predict prostate cancer progression. Even among specialists and expert uropathologists, discrepancies in determining tumor grade and Gleason scoring are significant as prostate cancer is known to have high glandular and nuclear heterogeneity (26). Manual grading remains a laborious process, often resulting in poor score reproducibility between pathologists. Computer-aided approaches can ameliorate many of these challenges by providing an objective algorithm that can track both biomarker staining (not limited to RBM3) as well as sub-visual nuclear changes that a pathologist may not recognize at eye level (27). Further, our eleven-feature can add clinical value by providing a faster, more efficient method to predict tumor progression (both aggressiveness and potential biochemical recurrence) and aid pathologists in their assessment of tissue samples.

Our data shows that it compares favorably to the currently used method and could be feasible method to use clinically.

Our study was limited in that our overall sample size (N = 80) and the number of samples that exhibited BCR (N = 20) was small. This prevents large-scale applicability of the results beyond the samples included in the study. Given the lack of an independent dataset to test our model, this study warrants further external validation in anlarger TMA set. However, we employed leave-one-out cross-validation to address the potential overfitting that may have occurred due to the small sample sizes and to provide a measure of internal validation. Our LOOCV results showed the strength of our eleven-feature model for predicting both aggressiveness and BCR. A decrease was observed in the ROC-AUC from the raw data to the LOOCV results for both the aggressiveness and BCR models. We attribute this to the small sample sizes, but the LOOCV tests still exhibit strong predictive ability. In the case of the BCR model specifically, the model still compared favorably with the currently-used Gleason sum for predicting which patients will experience BCR and adds the efficiency that a computational approach adds over the currently used manual grading system. In addition to the sample sizes, the lack of survival data with our samples limits comparisons with existing studies which examine RBM3 in the context of time to BCR or clinical progression. Further studies to assess our eleven-feature model predicting overall survival or time to BCR are warranted. We believe that having a single, eleven-feature model for predicting both aggressiveness and BCR is a strength of the study. While there are a multitude of potential nuclear features to track in a MLR model, having the same eleven features in both models strengthens the case for testing in future studies in independent samples to further assess applicability.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the late Donald Coffey, PhD for his initial conversations to spark this study. The authors would also like to thank Dr. Prakash Kulkarni (Institute for Bioscience and Biotechnology Research, University of Maryland) for his insightful feedback and comments of the manuscript drafts.

The project described was supported by National Center for Research Resources and National Cancer Institute grant U54CA143803 (RWV and Donald Coffey, Co-PIs) and the Prostate Cancer Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health. The funders had no role in the study design, the collection, analysis, and interpretation of data, involvement in drafting the manuscript, or decision to submit the manuscript for publication.

Abbreviations:

BCR: Biochemical recurrence
CSP: Cold-shock protein
GS: Gleason score
IOD: Integrated Optical Density
LOOCV: Leave-one-out cross validation
MLR: Multivariate logistic regression
PCa: Prostate cancer
PSA: Prostate-specific antigen
RBM3: RNA-binding motif protein 3
ROC-AUC: Receiver-operating characteristic area under the curve
TMA: Tissue microarray

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018;68(1):7-30. DOI: 10.3322/caac.21442.
2. Trewartha D, Carter K. Advances in prostate cancer treatment. Nat Rev Drug Discov. 2013;12(11):823-4. DOI: 10.1038/nrd4068.
3. Phadtare S, Severinov K. RNA remodeling and gene regulation by cold shock proteins. RNA Biol. 2010;7(6):788-95. DOI: 10.4161/rna.7.6.13482.
4. Danno S, Nishiyama H, Higashitsuji H, Yokoi H, Xue JH, Itoh K, et al. Increased transcript level of RBM3, a member of the glycine-rich RNA-binding protein family, in human cells in response to cold stress. Biochem Biophys Res Commun. 1997;236(3):804-7. DOI: 10.1006/bbrc.1997.7059.

5. Nishiyama H, Danno S, Kaneko Y, Itoh K, Yokoi H, Fukumoto M, et al. Decreased expression of cold-inducible RNA-binding protein (CIRP) in male germ cells at elevated temperature. Am J Pathol. 1998;152(1):289-96.
6. Coffey DS, Getzenberg RH, Deweese TL. Hyperthermic biology and cancer therapies: a hypothesis for the "Lance Armstrong effect". JAMA. 2006;296(4):445-8. DOI: 10.1001/jama.296.4.445.
7. Lleonart ME. A new generation of proto-oncogenes: cold-inducible RNA binding proteins. Biochim Biophys Acta. 2010;1805(1):43-52. DOI: 10.1016/j.bbcan.2009.11.001.
8. Uhlén M, Björling E, Agaton C, Szigyarto CA, Amini B, Andersen E, et al. A human protein atlas for normal and cancer tissues based on antibody proteomics. Mol Cell Proteomics. 2005;4(12):1920-32. DOI: 10.1074/mcp.M500279-MCP200.
9. Björling E, Lindskog C, Oksvold P, Linne J, Kampf C, Hober S, et al. A web-based tool for in silico biomarker discovery based on tissue-specific protein profiles in normal and cancer tissues. Mol Cell Proteomics. 2008;7(5):825-44. DOI: 10.1074/mcp.M700411-MCP200.
10. Pontén F, Jirström K, Uhlen M. The Human Protein Atlas--a tool for pathology. J Pathol. 2008;216(4):387-93. DOI: 10.1002/path.2440.
11. Wellmann S, Truss M, Bruder E, Tornillo L, Zelmer A, Seeger K, et al. The RNA-binding protein RBM3 is required for cell proliferation and protects against serum deprivation-induced cell death. Pediatr Res. 2010;67(1):35-41. DOI: 10.1203/PDR.0b013e3181c13326.
12. Sutherland LC, Rintala-maki ND, White RD, Morin CD. RNA binding motif (RBM) proteins: a novel family of apoptosis modulators?. J Cell Biochem. 2005;94(1):5-24. DOI: 10.1002/jcb.20204.
13. Sureban SM, Ramalingam S, Natarajan G, May R, Subramaniam D, Bishnupuri KS, et al. Translation regulatory factor RBM3 is a proto-oncogene that prevents mitotic catastrophe. Oncogene. 2008;27(33):4544-56. DOI: 10.1038/onc.2008.97.
14. Hjelm B, Brennan DJ, Zendehrokh N, Eberhard J, Nodin B, Gaber A, et al. High nuclear RBM3 expression is associated with an improved prognosis in colorectal cancer. Proteomics Clin Appl. 2011;5(11-12):624-35. DOI: 10.1002/prca.201100020.
15. Florianova L, Xu B, Traboulsi S, Elmansi H, Tanguay S, Aprikian A, et al. Evaluation of RNA-binding motif protein 3 expression in urothelial carcinoma of the bladder: an immunohistochemical study. World J Surg Oncol. 2015;13:317. DOI: 10.1186/s12957-015-0730-3.
16. Jögi A, Brennan DJ, Rydén L, Magnusson K, Fernö M, Stål O, et al. Nuclear expression of the RNA-binding protein RBM3 is associated with an improved clinical outcome in breast cancer. Mod Pathol. 2009 Dec;22(12):1564. DOI: 10.1038/modpathol.2009.124
17. Jonsson L, Gaber A, Ulmert D, Uhlén M, Bjartell A, Jirström K. High RBM3 expression in prostate cancer independently predicts a reduced risk of biochemical recurrence and disease progression. Diagn Pathol. 2011 Dec;6(1):91. DOI: 10.1186/1746-1596-6-91
18. Ehlén Å, Nodin B, Rexhepaj E, Brandstedt J, Uhlen M, Alvarado-Kristensson M, et al. RBM3-regulated genes promote DNA integrity and affect clinical outcome in epithelial ovarian cancer. Transl Oncol. 2011;4(4):212-21. DOI: 10.1593/tlo.11106.
19. Zink D, Fischer AH, Nickerson JA. Nuclear structure in cancer cells. Nat Rev Cancer. 2004 Sep;4(9):677. DOI: 10.1038/nrc1430.
20. Zhu G, Liu Z, Epstein JI, Davis C, Christudass CS, Carter HB, et al. A Novel Quantitative Multiplex Tissue Immunoblotting for Biomarkers Predicts a Prostate Cancer Aggressive Phenotype. Cancer Epidemiol Biomarkers Prev. 2015;24(12):1864-72. DOI: 10.1158/1055-9965.EPI-15-0496.
21. Carleton NM, Zhu G, Gorbounov M, Miller MC, Pienta KJ, Resar LMS, et al. PBOV1 as a potential biomarker for more advanced prostate cancer based on protein and digital histomorphometric analysis. Prostate. 2018 May;78(7):547-59. DOI: 10.1002/pros.23499.
22. Veltri RW, Khan MA, Marlow C, Miller MC, Mikolajczyk SD, Kojima M, et al. Alterations in nuclear structure and expression of proPSA predict differences between native Japanese and Japanese-American prostate cancer. Urology. 2006;68(4):898-904. DOI: 10.1016/j.urology.2006.05.008.

23. Grupp K, Wilking J, Prien K, Hube-Magg C, Sirma H, Simon R, et al. High RNA-binding motif protein 3 expression is an independent prognostic marker in operated prostate cancer and tightly linked to ERG activation and PTEN deletions. Eur J Cancer. 2014;50(4):852-61. DOI: 10.1016/j.ejca.2013.12.003.
24. Nodin B, Fridberg M, Jonsson L, Bergman J, Uhlén M, Jirström K. High MCM3 expression is an independent biomarker of poor prognosis and correlates with reduced RBM3 expression in a prospective cohort of malignant melanoma. Diagn Pathol. 2012;7:82. DOI: 10.1186/1746-1596-7-82.
25. Zeng Y, Wodzenski D, Gao D, Shiraishi T, Terada N, Li Y, et al. Stress response protein RBM3 attenuates the stem-like properties of prostate cancer cells by interfering with CD44 variant splicing. Cancer Res. 2013 May 10:1343. DOI: 10.1158/0008-5472.CAN-12-1343.
26. Hamilton PW, Bankhead P, Wang Y, Hutchinson R, Kieran D, McArt DG, et al. Digital pathology and image analysis in tissue biomarker research. Methods. 2014;70(1):59-73. DOI: 10.1016/j.ymeth.2014.06.015.
27. Carleton NM, Lee G, Madabhushi A, Veltri RW. Advances in the computational and molecular understanding of the prostate cancer cell nucleus. J Cell Biochem. 2018;1-16. DOI: 10.1002/jcb.27156.

Statistics/Data Analysis

```
      name:  <unnamed>
       log:  C:\Users\ncarlet4.WIN\Desktop\RBM3\ULR_DataAnalysis.smcl
  log type:  smcl
 opened on:  11 May 2018, 13:12:13

1 . * Univariate Logistic Regression: Identify which features are significant (P < 0.05)
>  for predicting PCa aggressiveness

2 . logistic agg areapix2

Logistic regression                               Number of obs   =         77
                                                  LR chi2(1)      =      20.50
                                                  Prob > chi2     =     0.0000
Log likelihood = -40.755457                       Pseudo R2       =     0.2009

------------------------------------------------------------------------------
         agg | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     areapix2 |   .9996218   .0000983    -3.85   0.000     .9994293    .9998145
       _cons |   15.31591   9.845163     4.25   0.000     4.344955     53.9884
------------------------------------------------------------------------------

3 .
4 . logistic agg aspectratio

Logistic regression                               Number of obs   =         77
                                                  LR chi2(1)      =       0.00
                                                  Prob > chi2     =     0.9944
Log likelihood = -51.003765                       Pseudo R2       =     0.0000

------------------------------------------------------------------------------
         agg | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  aspectratio |    .994633   .7570162    -0.01   0.994     .2237755    4.420926
       _cons |   1.662808    1.15089     0.73   0.463     .4282467    6.456393
------------------------------------------------------------------------------

5 .
6 . logistic agg axismajorpix

Logistic regression                               Number of obs   =         77
                                                  LR chi2(1)      =      21.33
                                                  Prob > chi2     =     0.0000
Log likelihood = -40.337795                       Pseudo R2       =     0.2091

------------------------------------------------------------------------------
         agg | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 axismajorpix |   .8646924   .0325709    -3.86   0.000     .8031541    .9309457
       _cons |   109.7202   122.8227     4.20   0.000     12.23032     984.319
------------------------------------------------------------------------------

7 .
8 . logistic agg axisminorpix

Logistic regression                               Number of obs   =         77
                                                  LR chi2(1)      =      18.54
                                                  Prob > chi2     =     0.0000
Log likelihood = -41.734181                       Pseudo R2       =     0.1817

------------------------------------------------------------------------------
         agg | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 axisminorpix |   .7253965   .0627091    -3.71   0.000     .6123374    .8593304
       _cons |   57.08471   57.78214     4.00   0.000     7.850889    415.0695
------------------------------------------------------------------------------
```

```
9 .
10. logistic agg boundboxheightpix

Logistic regression                               Number of obs   =         77
                                                  LR chi2(1)      =      19.17
                                                  Prob > chi2     =     0.0000
Log likelihood = -41.418193                       Pseudo R2       =     0.1879

------------------------------------------------------------------------------
         agg | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
boundboxheig~x |   .8615584   .0349632    -3.67   0.000     .7956861    .9328841
       _cons |   99.52717   115.1027     3.98   0.000     10.31648    960.1786
------------------------------------------------------------------------------

11.
12. logistic agg boundboxwidthpix

Logistic regression                               Number of obs   =         77
                                                  LR chi2(1)      =      26.27
                                                  Prob > chi2     =     0.0000
Log likelihood = -37.867737                       Pseudo R2       =     0.2576

------------------------------------------------------------------------------
         agg | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
boundboxwidt~x |   .8075284   .0411145    -4.20   0.000     .7308357    .8922691
       _cons |    418.197   567.7396     4.45   0.000       29.228    5983.602
------------------------------------------------------------------------------

13.
14. logistic agg boxareapix2

Logistic regression                               Number of obs   =         77
                                                  LR chi2(1)      =       2.91
                                                  Prob > chi2     =     0.0879
Log likelihood = -49.547687                       Pseudo R2       =     0.0285

------------------------------------------------------------------------------
         agg | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
boxareapix2 |   4.7e+115   7.5e+117     1.65   0.099     2.73e-22    7.9e+252
       _cons |   .0337065   .0794153    -1.44   0.150     .0003328    3.413623
------------------------------------------------------------------------------

15.
16. logistic agg boxxy

Logistic regression                               Number of obs   =         77
                                                  LR chi2(1)      =       0.28
                                                  Prob > chi2     =     0.5954
Log likelihood = -50.862814                       Pseudo R2       =     0.0028

------------------------------------------------------------------------------
         agg | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       boxxy |   7.082719   26.60302     0.52   0.602     .0044983    11151.99
       _cons |    1.09036   .9052896     0.10   0.917     .2142127    5.550016
------------------------------------------------------------------------------
```

```
17.
18. logistic agg circularity

Logistic regression                               Number of obs   =         77
                                                  LR chi2(1)      =       0.26
                                                  Prob > chi2     =     0.6106
Log likelihood = -50.874146                       Pseudo R2       =     0.0025

------------------------------------------------------------------------------
         agg | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 circularity |   3.88e+17   3.11e+19     0.50   0.614     1.98e-51    7.61e+85
       _cons |   .5959826   1.211734    -0.25   0.799     .0110814    32.05319
------------------------------------------------------------------------------

19.
20. logistic agg clumpiness

Logistic regression                               Number of obs   =         77
                                                  LR chi2(1)      =      25.77
                                                  Prob > chi2     =     0.0000
Log likelihood = -38.117978                       Pseudo R2       =     0.2526

------------------------------------------------------------------------------
         agg | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  clumpiness |          0          0    -3.56   0.000            0           0
       _cons |   5.731654   2.303947     4.34   0.000     2.606894    12.60192
------------------------------------------------------------------------------
Note: 1 failure and 0 successes completely determined.

21.
22. logistic agg diametermaxpix

Logistic regression                               Number of obs   =         77
                                                  LR chi2(1)      =      23.19
                                                  Prob > chi2     =     0.0000
Log likelihood = -39.409109                       Pseudo R2       =     0.2273

------------------------------------------------------------------------------
         agg | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
diametermaxpix|   .8621897   .0322622    -3.96   0.000       .80122    .927799
       _cons |   142.3203   165.1697     4.27   0.000     14.63544    1383.974
------------------------------------------------------------------------------

23.
24. logistic agg diametermeanpix

Logistic regression                               Number of obs   =         77
                                                  LR chi2(1)      =      19.33
                                                  Prob > chi2     =     0.0000
Log likelihood = -41.338093                       Pseudo R2       =     0.1895

------------------------------------------------------------------------------
         agg | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
diametermeanpix|   .7557951   .0561175    -3.77   0.000     .6534354    .8741894
       _cons |   44.16861   41.04417     4.08   0.000     7.147028    272.9619
------------------------------------------------------------------------------
```

```
25.
26. logistic agg diameterminpix

Logistic regression                               Number of obs   =         77
                                                  LR chi2(1)      =      10.91
                                                  Prob > chi2     =     0.0010
Log likelihood = -45.548939                       Pseudo R2       =     0.1069

------------------------------------------------------------------------------
         agg | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
diameterminpix |   .7381489   .0744496    -3.01   0.003     .6057479    .8994893
       _cons |   21.29322   19.19771     3.39   0.001     3.637541    124.6449
------------------------------------------------------------------------------

27.
28. logistic agg fourierdescriptormean

Logistic regression                               Number of obs   =         77
                                                  LR chi2(1)      =       3.93
                                                  Prob > chi2     =     0.0473
Log likelihood = -49.037175                       Pseudo R2       =     0.0386

------------------------------------------------------------------------------
                 agg | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval
                     > ]
-------------------- +----------------------------------------------------------
fourierdescriptormean |   337.4626   1105.964     1.78   0.076      .547756    207904.
                     > 6
               _cons |   .5828089   .3573623    -0.88   0.379     .1752237    1.93847
                     > 1
------------------------------------------------------------------------------

29.
30. logistic agg heterogeneity

Logistic regression                               Number of obs   =         77
                                                  LR chi2(1)      =      16.27
                                                  Prob > chi2     =     0.0001
Log likelihood =   -42.8697                       Pseudo R2       =     0.1595

------------------------------------------------------------------------------
         agg | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
heterogeneity |   1.0e-138   9.5e-137    -3.37   0.001     6.1e-219    1.66e-58
       _cons |   29.51295    26.6637     3.75   0.000     5.023288    173.3952
------------------------------------------------------------------------------

31.
32. logistic agg illformedboundary

Logistic regression                               Number of obs   =         77
                                                  LR chi2(1)      =       3.39
                                                  Prob > chi2     =     0.0657
Log likelihood = -49.310453                       Pseudo R2       =     0.0332

------------------------------------------------------------------------------
         agg | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
illformedboundary |   6.120686   6.242337     1.78   0.076     .8292407    45.17723
       _cons |   .0171568   .0440829    -1.58   0.114     .0001115    2.639512
------------------------------------------------------------------------------
```

```
33.
34. logistic agg integratedodlumpix2

Logistic regression                               Number of obs   =         77
                                                  LR chi2(1)      =      14.50
                                                  Prob > chi2     =     0.0001
Log likelihood = -43.755009                       Pseudo R2       =     0.1421

------------------------------------------------------------------------------
         agg | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
integratedo~2|   .9999999   1.83e-08    -3.44   0.001     .9999999           1
       _cons |    12.1958   7.866442     3.88   0.000     3.444885    43.17634
------------------------------------------------------------------------------

35.
36. logistic agg intensitymaxlum

Logistic regression                               Number of obs   =         77
                                                  LR chi2(1)      =       0.07
                                                  Prob > chi2     =     0.7915
Log likelihood = -50.968843                       Pseudo R2       =     0.0007

------------------------------------------------------------------------------
         agg | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
intensityma~m|   .9936337   .0239199    -0.27   0.791     .9478403    1.04164
       _cons |   1.921836   1.175242     1.07   0.285     .5796816    6.371525
------------------------------------------------------------------------------

37.
38. logistic agg intensitymeanlum

Logistic regression                               Number of obs   =         77
                                                  LR chi2(1)      =       4.09
                                                  Prob > chi2     =     0.0432
Log likelihood = -48.960881                       Pseudo R2       =     0.0401

------------------------------------------------------------------------------
         agg | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
intensityme~m|    .978764   .0108506    -1.94   0.053     .9577265    1.000264
       _cons |   26.36446    38.6383     2.23   0.026     1.491231    466.1149
------------------------------------------------------------------------------

39.
40. logistic agg intensityminlum

Logistic regression                               Number of obs   =         77
                                                  LR chi2(1)      =       1.83
                                                  Prob > chi2     =     0.1756
Log likelihood = -50.086436                       Pseudo R2       =     0.0180

------------------------------------------------------------------------------
         agg | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
intensitymi~m|   1.010554    .007955     1.33   0.182     .9950819    1.026266
       _cons |   .1957601   .3151598    -1.01   0.311     .0083435    4.593014
------------------------------------------------------------------------------
```

```
41.
42. logistic agg intensitystddevlum

Logistic regression                               Number of obs   =         77
                                                  LR chi2(1)      =       4.01
                                                  Prob > chi2     =     0.0451
Log likelihood = -48.996716                       Pseudo R2       =     0.0394

------------------------------------------------------------------------------
         agg | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
intensitystd~m |   .8550645   .0695488    -1.93   0.054     .7290616    1.002844
       _cons |   18.49511    23.7625     2.27   0.023     1.490849    229.4457
------------------------------------------------------------------------------

43.
44. logistic agg roundness

Logistic regression                               Number of obs   =         77
                                                  LR chi2(1)      =       4.30
                                                  Prob > chi2     =     0.0381
Log likelihood = -48.852975                       Pseudo R2       =     0.0422

------------------------------------------------------------------------------
         agg | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   roundness |   1.834856   .6206844     1.79   0.073     .9455019     3.56075
       _cons |   .7463464   .3587189    -0.61   0.543     .2909542    1.914504
------------------------------------------------------------------------------

45. log off
       name:  <unnamed>
        log:  C:\Users\ncarlet4.WIN\Desktop\RBM3\ULR_DataAnalysis.smcl
   log type:  smcl
  paused on:  11 May 2018, 13:13:36
-------------------------------------------------------------------------------
```

Statistics/Data Analysis

```
      name:  <unnamed>
       log:  C:\Users\ncarlet4.WIN\Desktop\RBM3\MLR_AggressivenessModel.smcl
  log type:  smcl
 opened on:  11 May 2018, 13:15:14

1 . * Multivariate Logistic Regression: Aggressiveness Model

2 . * Use 11 significant features from univariate logistic regression to predict PCa agg
  > ressiveness in multivariate model.

3 . logistic agg area axismajorpix axisminorpix boundboxheightpix boundboxwidthpix clump
  > iness diametermaxpix diametermeanpix diameterminpix integratedodlumpix2 heterogeneit
  > y

Logistic regression                               Number of obs   =         77
                                                  LR chi2(11)     =      44.41
                                                  Prob > chi2     =     0.0000
Log likelihood = -28.801066                       Pseudo R2       =     0.4353
```

agg	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
areapix2	.9994181	.0012869	-0.45	0.651	.996899	1.001943
axismajorpix	2.603779	1.209992	2.06	0.039	1.047239	6.47385
axisminorpix	.0765185	.1048966	-1.87	0.061	.0052105	1.123698
boundboxheightpix	1.809437	.6979985	1.54	0.124	.8495498	3.853881
boundboxwidthpix	.9220026	.3570531	-0.21	0.834	.4316158	1.96955
clumpiness	0	0	-1.76	0.079	0	.
diametermaxpix	.2637657	.1799907	-1.95	0.051	.0692418	1.004774
diametermeanpix	2.349376	3.883542	0.52	0.605	.0920254	59.97873
diameterminpix	5.930943	6.481499	1.63	0.103	.6964777	50.50569
integratedodlumpix2	1	1.05e-07	0.96	0.336	.9999999	1
heterogeneity	8.65e+67	2.02e+70	0.67	0.503	9.8e-132	7.6e+266
_cons	6.750551	25.36283	0.51	0.611	.0042781	10652

```
Note: 1 failure and 0 successes completely determined.

4 .
5 . logit

Logistic regression                               Number of obs   =         77
                                                  LR chi2(11)     =      44.41
                                                  Prob > chi2     =     0.0000
Log likelihood = -28.801066                       Pseudo R2       =     0.4353
```

agg	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
areapix2	-.0005821	.0012876	-0.45	0.651	-.0031058	.0019416
axismajorpix	.9569639	.4647061	2.06	0.039	.0461568	1.867771
axisminorpix	-2.570223	1.370866	-1.87	0.061	-5.257071	.1166253
boundboxheightpix	.593016	.3857544	1.54	0.124	-.1630487	1.349081
boundboxwidthpix	-.0812073	.3872582	-0.21	0.834	-.8402194	.6778049
clumpiness	-55881.43	31779.76	-1.76	0.079	-118168.6	6405.756
diametermaxpix	-1.332694	.6823885	-1.95	0.051	-2.670151	.004763
diametermeanpix	.8541498	1.65301	0.52	0.605	-2.38569	4.09399
diameterminpix	1.780183	1.092828	1.63	0.103	-.3617196	3.922086
integratedodlumpix2	1.01e-07	1.05e-07	0.96	0.336	-1.05e-07	3.08e-07
heterogeneity	156.4304	233.722	0.67	0.503	-301.6562	614.517
_cons	1.909624	3.75715	0.51	0.611	-5.454255	9.273503

```
Note: 1 failure and 0 successes completely determined.
```

```
6 .
7 . qui predict agg_model, pr

8 .
9 . roctab agg agg_model
```

Obs	ROC Area	Std. Err.	Asymptotic Normal [95% Conf.	Interval]
77	**0.9009**	**0.0343**	**0.83360**	**0.96812**

```
10. roctab agg agg_model, graph title("Raw Data: Aggressiveness Model", size(vsmall))

11. lsens, xlab(0.00(.05)1.00, labs(vsmall) grid) ylab(0.00(.05)1.00, labs(vsmall) angle
  > (horizontal) grid) title("Raw Data: Sensitivity and Specificity Plot", size(vsmall))

12.
13.
14.
15. dotplot agg_model, over(agg) center msize(small) title("Raw Data for AGG", size(vsma
  > ll)) ytitle("Predicted Probability", size(medium)) ylab(0.00(0.10)1.00, angle(horizo
  > ntal) labs(small)) xlab(, labs(small))

16. * Now that raw data model has been created, use leave-one-out cross validation metho
  > d to evaluate model.

17. xtile loo_agg_group = uniform() if agg~=., nq(50)

18.
19. gen loo_agg_model = .
  (80 missing values generated)

20.
21. forvalues i = 1/50 {
    2.
22. qui logistic agg area axismajorpix axisminorpix boundboxheightpix boundboxwidthpix c
  > lumpiness diametermaxpix diametermeanpix diameterminpix integratedodlumpix2 heteroge
  > neity if loo_agg_group~=`i'
    3.
23. qui predict loo_agg_model`i', pr
    4.
24. qui replace loo_agg_model = loo_agg_model`i' if loo_agg_group==`i'
    5.
25. drop loo_agg_model`i'
    6.
26. }

27.
28. roctab agg loo_agg_model
```

Obs	ROC Area	Std. Err.	Asymptotic Normal [95% Conf.	Interval]
77	**0.7723**	**0.0602**	**0.65432**	**0.89022**

```
29. roctab agg loo_agg_model, graph title("LOO-Cross Validation: Aggressiveness Model",
  > size(vsmall))

30.
31.
```

```
32.
33. lsens, xlab(0.00(.05)1.00, labs(vsmall) grid) ylab(0.00(.05)1.00, labs(vsmall) angle
  > (horizontal) grid) title("LOO-Cross Validation: for agg", size(vsmall))

34.
35.
36.
37. dotplot loo_agg_model, over(agg) center msize(small) title("LOO-Cross Validation: Ag
  > gressiveness Model", size(vsmall)) ytitle("Predicted Probability", size(medium)) yla
  > b(0.00(0.10)1.00, angle(horizontal) labs(small)) xlab(, labs(small))

38. log off
       name:  <unnamed>
        log:  C:\Users\ncarlet4.WIN\Desktop\RBM3\MLR_AggressivenessModel.smcl
   log type:  smcl
  paused on:  11 May 2018, 13:21:13
-------------------------------------------------------------------------------
```

Statistics/Data Analysis

```
      name:  <unnamed>
       log:  C:\Users\ncarlet4.WIN\Desktop\RBM3\MLR_BCR_Model.smcl
  log type:  smcl
 opened on:  11 May 2018, 13:22:24

1 . * BCR Model: Apply the 11-feature model to see if it can predict BCR cases

2 . logistic bcr area axismajorpix axisminorpix boundboxheightpix boundboxwidthpix clump
  > iness diametermaxpix diametermeanpix diameterminpix integratedodlumpix2 heterogeneit
  > y

Logistic regression                               Number of obs   =         66
                                                  LR chi2(11)     =      33.95
                                                  Prob > chi2     =     0.0004
Log likelihood = -23.510199                       Pseudo R2       =     0.4193

------------------------------------------------------------------------------
         bcr | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     areapix2 |   1.004058   .0019859     2.05   0.041     1.000173    1.007958
 axismajorpix |   .1726408   .1634401    -1.86   0.064     .0269964     1.10403
 axisminorpix |   .1754509    .262404    -1.16   0.245     .0093562    3.290111
boundboxheightpix |   4.000533   2.225608     2.49   0.013     1.344531    11.90323
 boundboxwidthpix |   2.914428   1.394366     2.24   0.025      1.14106    7.443863
    clumpiness |          0          0    -1.11   0.267            0           .
 diametermaxpix |   1.532945   1.413965     0.46   0.643     .2514149    9.346781
diametermeanpix |   .0157737   .0294069    -2.23   0.026     .0004084    .6092926
 diameterminpix |   2.110992   2.522785     0.63   0.532     .2028807    21.96507
integratedodlumpix2 |          1   1.30e-07     0.54   0.591     .9999998           1
  heterogeneity |   4.2e+189   1.4e+192     1.35   0.178     6.60e-87           .
          _cons |   9.479703   55.39723     0.38   0.700     .0001006    893352.5
------------------------------------------------------------------------------

3 .
4 . logit

Logistic regression                               Number of obs   =         66
                                                  LR chi2(11)     =      33.95
                                                  Prob > chi2     =     0.0004
Log likelihood = -23.510199                       Pseudo R2       =     0.4193

------------------------------------------------------------------------------
         bcr |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     areapix2 |     .00405   .0019779     2.05   0.041     .0001733    .0079266
 axismajorpix |  -1.756542   .9467059    -1.86   0.064    -3.612051    .0989675
 axisminorpix |  -1.740396   1.495598    -1.16   0.245    -4.671713    1.190921
boundboxheightpix |   1.386428   .5563279     2.49   0.013      .296045    2.47681
 boundboxwidthpix |   1.069674   .4784355     2.24   0.025     .1319574    2.00739
    clumpiness |  -35391.39   31872.71    -1.11   0.267    -97860.75    27077.97
 diametermaxpix |   .4271907    .922385     0.46   0.643    -1.380651    2.235032
diametermeanpix |   -4.14941   1.864296    -2.23   0.026    -7.803364   -.4954567
 diameterminpix |   .7471581   1.195071     0.63   0.532    -1.595137    3.089453
integratedodlumpix2 |   6.99e-08   1.30e-07     0.54   0.591    -1.85e-07    3.25e-07
  heterogeneity |   436.6306   324.0202     1.35   0.178    -198.4372    1071.698
          _cons |   2.249153   5.843772     0.38   0.700     -9.20443    13.70274
------------------------------------------------------------------------------

5 .
```

```
6 . qui predict bcr_model, pr

7 .
8 . roctab bcr bcr_model

                    ROC                    —Asymptotic Normal——
          Obs       Area     Std. Err.      [95% Conf. Interval]
——————————————————————————————————————————————————————————————
           66     0.9163       0.0336        0.85047     0.98214

9 . roctab bcr bcr_model, graph title("Raw Data: BCR Model", size(vsmall))

10.
11.
12.
13. lsens, xlab(0.00(.05)1.00, labs(vsmall) grid) ylab(0.00(.05)1.00, labs(vsmall) angle
  > (horizontal) grid) title("Raw Data: Sensitivity and Specificity Plot", size(vsmall))

14.
15.
16.
17. dotplot bcr_model, over(bcr) center msize(small) title("Raw Data: BCR Model", size(v
  > small)) ytitle("Predicted Probability", size(medium)) ylab(0.00(0.10)1.00, angle(hor
  > izontal) labs(small)) xlab(, labs(small))

18. * Apply leave-one-out cross validation

19. xtile loo_bcr_group = uniform() if bcr~=., nq(50)

20.
21. gen loo_bcr_model = .
  (80 missing values generated)

22.
23. forvalues i = 1/50 {
    2.
24. qui logistic bcr area axismajorpix axisminorpix boundboxheightpix boundboxwidthpix c
  > lumpiness diametermaxpix diametermeanpix diameterminpix integratedodlumpix2 heteroge
  > neity if loo_bcr_group~=`i'
    3.
25. qui predict loo_bcr_model`i', pr
    4.
26. qui replace loo_bcr_model = loo_bcr_model`i' if loo_bcr_group==`i'
    5.
27. drop loo_bcr_model`i'
    6.
28. }

29.
30. roctab bcr loo_bcr_model

                    ROC                    —Asymptotic Normal——
          Obs       Area     Std. Err.      [95% Conf. Interval]
——————————————————————————————————————————————————————————————
           66     0.7565       0.0603        0.63830     0.87475

31. roctab bcr loo_bcr_model, graph title("LOO-Cross Validation: BCR Model", size(vsmall
  > ))

32.
```

```
33.
34.
35. lsens, xlab(0.00(.05)1.00, labs(vsmall) grid) ylab(0.00(.05)1.00, labs(vsmall) angle
  > (horizontal) grid) title("LOO-Cross Validation: Sensitivity and Specificity Plot", s
  > ize(vsmall))

36.
37.
38.
39. dotplot loo_bcr_model, over(bcr) center msize(small) title("LOO-Cross Validation: BC
  > R Model", size(vsmall)) ytitle("Predicted Probability", size(medium)) ylab(0.00(0.10
  > )1.00, angle(horizontal) labs(small)) xlab(, labs(small))

40. * Compare LOO_BCR_Model to Gleason score for predicting BCR

41. roccomp bcr loo_bcr_model gleasonsum

                              ROC                    —Asymptotic Normal——
                    Obs       Area     Std. Err.      [95% Conf. Interval]
---------------------------------------------------------------------------
loo_bcr_mo~l         66     0.7565       0.0603        0.63830     0.87475
gleasonsum           66     0.7255       0.0659        0.59644     0.85464
---------------------------------------------------------------------------
  Ho: area(loo_bcr_mo~l) = area(gleasonsum)
      chi2(1) =     0.14       Prob>chi2 =   0.7040

42. roccomp bcr loo_bcr_model gleasonsum, graph

43. log off
        name:  <unnamed>
         log:  C:\Users\ncarlet4.WIN\Desktop\RBM3\MLR_BCR_Model.smcl
    log type:  smcl
   paused on:  11 May 2018, 13:27:06
-------------------------------------------------------------------------------
```