Evaluation of the prognostic significance of MSMB and CRISP3 in prostate cancer using automated image analysis

Anna Dahlman,[1,2] Elton Rexhepaj,[3] Donal J Brennan,[3] William M Gallagher,[3] Alexander Gaber,[2] Anna Lindgren,[4] Karin Jirström,[2] and Anders Bjartell[1,2]

1 Department of Clinical Sciences, Division of Urological Cancers, Lund University, Skåne University Hospital Malmö, Sweden
2 Department of Laboratory Medicine, Center of Molecular Pathology, Lund University, Skåne University Hospital Malmö, Sweden
3 UCD School of Biomolecular and Biomedical Science, UCD Conway Institute, University College Dublin, Belfield, Dublin 4, Ireland
4 Center for Mathematical Sciences, Department of Mathematical Statistics, Lund University, Sweden

CORRESPONDENCE TO:

Anders Bjartell, MD, PhD
Department of Clinical Sciences
Division of Urological Cancers
Skåne University Hospital
205 02 Malmö, Sweden
Tel. +46 40 331000 (direct 332685)
Fax +46 40 337049
E-mail: anders.bjartell@med.lu.se

RUNNING TITLE: MSMB expression predicts outcome

ABSTRACT

Despite prostate cancer being the most frequent cancer in males in the Western world, tissue biomarkers for predicting disease recurrence after surgery have not been incorporated into clinical practice. Our group has previously identified β-microseminoprotein (MSMB) and cysteine-rich secretory protein-3 (CRISP3) as independent predictors of biochemical recurrence after radical prostatectomy. The purpose of the present study was to use automated image analysis enabling quantitative determination of MSMB and CRISP3 expression in a large cohort, and to validate the previous findings. MSMB and CRISP3 protein expression was assessed on tissue microarrays constructed from 3,268 radical prostatectomy specimens. Whole-slide digital images were captured, and a novel cytoplasmic algorithm was used to develop a quantitative scoring model for cytoplasmic staining. Classification regression tree analysis was used to group patients with different risk for biochemical recurrence depending on level of protein expression. Patients with tumors expressing high levels of MSMB had a significantly reduced risk for biochemical recurrence after radical prostatectomy (HR= 0.468; 95% CI 0.394-0.556; $p<0.001$). Multivariate analysis adjusted for clinicopathological parameters revealed that MSMB expression was an independent predictor of decreased risk of recurrence (HR= 0.710; 95% CI 0.578-0.872; $p<0.001$). We found no correlation between CRISP3 expression and biochemical recurrence. In the current study, we applied a novel image analysis on a large independent cohort and successfully verified that MSMB is a strong independent factor predicting favorable outcome after radical prostatectomy for localized prostate cancer.

Keywords (3-6): tissue biomarker, PSP94, outcome prediction

INTRODUCTION

Although cancer of the prostate is the most frequent solid cancer in males in the Western countries, and the second leading cause of cancer death,[1] reliable prognostic and treatment predictive tissue biomarkers to guide clinical decision making have yet to be incorporated into clinical practice. Patient prognosis will vary from a rapidly progressing disease with a high probability of death in a minority of patients, to a relatively indolent course that can be controlled requiring little intervention. Accurate risk assessments of clinically significant cancer, stage, and treatment success are pivotal for informed decision making and patient counseling. The postoperative nomogram originally developed by Kattan et al[2] over a decade ago is considered an accurate predictive tool, and is widely used by clinicians to predict five years of freedom from disease recurrence for patients who have undergone radical prostatectomy. However, despite a number of available predictive tools, foretelling patient prognosis has proven abstruse, leading to overtreatment of a large number of men for the benefit of the few who need it.[3]

Tissue microarrays are a highly useful tool in biomarker discovery and validation. Many groups have described tissue biomarkers of prognostic value in prostatic tumours (reviwed in[4]), but the use of such biomarkers to improve the predictive accuracy of existing nomograms have been largely unsuccessful.[5] Limiting factors include the lack of standardized processing procedures for radical prostatectomy tissue specimens, and the lack of methods to reliably quantify immunohistochemical staining. Manual interpretation is highly subjective and hampered by inter- and intra-observer variations. Several groups have targeted this dilemma by developing automated image analysis techniques, with excellent correlations to manual scoring.[6-9]

Our group has previously reported that β-microseminoprotein (MSMB) and the MSMB-binding protein cysteine-rich secretory protein-3 (CRISP3) are independent predictors of patient outcome after radical prostatectomy.[10] MSMB (also known as prostate specific protein of 94 amino acids (PSP94)), is expressed in benign and malignant prostatic epithelium and is along with prostate specific antigen (PSA) and prostate acidic phosphatase (PAP) one of the three most abundant proteins in human seminal plasma.[11, 12] As described in an recent review,[13] several groups have employed a variety of approaches in both tissue and serum samples to demonstrate decreasing levels of MSMB in prostate cancer compared to normal controls, prompting the suggestion that MSMB may be a promising biomarker for prostate cancer.[14-19] Additionally, MSMB, located at chromosome 10q11.2, has recently been reported as an important candidate gene for prostate cancer susceptibility.[20, 21] Several causal risk alleles affecting the level of gene transcription have been identified in the region upstream of the coding sequence.[22-25]

In contrast to MSMB, CRISP3 (also known as specific granule protein of 28 kDa (SGP28)) levels are low in normal and benign prostatic tissue and seminal plasma, but often increased in prostate cancer,[26-28] in turn rendering the proposition that increasing levels of CRISP3 may be a biomarker for prostate cancer.[29] Interestingly, CRISP3 will readily form a complex with MSMB in the seminal plasma,[30] and it has been speculated that this binding may hinder any putative MSMB action. However, to date very little is known about the functions of both MSMB, CRISP3, and the complex they form.

The aim of the present study was to use automated image analysis that enables accurate, quantitative determination of biomarker expression, and by using this method investigate whether previous findings could be validated.[10] Using automated image analysis, we were

able todevelop a quantitative scoring model for MSMB and CRISP3 immunihistochemical staining which could accurately assess protein expression levels in an extensive cohort of 3,268 radical prostatectomy specimens. This automated quantitative scoring model should allow for high-throughput, scrupoulous assessment of immunohistochemical staining in tissue samples and the application of more robust statistical analysis in contrast to qualitative assesment.

MATERIALS AND METHODS

Patients

Tissue specimens were available from 3,268 patients undergoing radical prostatectomy at the Department of Urology, University Medical Center Hamburg-Eppendorf between 1992 and 2005, as previously described.[31] Clinical data included preoperative *PSA* values, clinical and pathologic TNM classification, Gleason score of the preoperative biopsy and prostatectomy specimens, tumor localization, margin status of the prostatectomy specimen, and biochemical recurrence, if available (Table 1). Full follow-up data were available for 2,460 patients, of which 620 developed biochemical recurrence. Mean follow-up time was 34 months (1-144 months). None of the patients received neoadjuvant or adjuvant therapy, and additional therapy was only initiated in the case of biochemical recurrence of the tumor. After surgery, PSA was measured every 3 months in the first year, every 6 months in the second year, and yearly after the third year. Biochemical recurrence was defined as postoperative PSA values of ≥0.1 ng/mL with a confirmatory value. Patients without evidence of tumor recurrence were censored at last follow-up.

Tissue microarrays

The tissue was organized in tissue micro array blocks as previously described.[32] In brief, all cases were histopathologically re-evaluated on hematoxylin-eosin stained whole-mount sections by a pathologist. The index tumor, as defined by the area with the largest tumor focus and/or worst Gleason score, was identified and one 0.6 mm tissue core was punched out and arranged in a tissue micro array format. The 3,268 cores were distributed in 7 blocks. Each block also included various control tissues including normal prostate tissue.

Immunohistochemistry

Tissue micro array sections (4 μm thickness) were mounted, deparaffinised, and processed using the PT Link System (DAKO Cytomation A/S, Glostrup, Denmark), including a heat mediated antigen retieval for 20 min at 97°C in EnvisionFlex Target Retrieval Solution High pH (pH 9, code K8024, DAKO). Serial sections were stained with rabbit polyclonal anti-MSMB or anti-CRISP3 antibodies in an AutostainerPlus staining machine (DAKO). Polyclonal rabbit anti-human CRISP3 was a kind gift from Dr Lene Udby, Copenhagen University Hospital, Denmark,[33] and polyclonal rabbit anti-human MSMB was provided by Dr Per Fernlund, Skåne University Hospital Malmö, Sweden.[34] The final working concentration for anti-MSMB and anti-CRISP3 were 0.3 μg/mL and 3.5 μg/mL, respectively. Slides were processed by the EnVision Flex/HPR Rabbit/Mouse (code K5007, DAKO), with peroxidase-coupled secondary antibodies goat anti-rabbit and goat anti-mouse IgGs (DAKO). Finally, sections were counterstained with Meyer's hematoxylin solution, and coverslips were mounted in Cytoseal XYL (Richard-Allan Scientific, Kalamazoo, MI).

Manual scoring

Scoring of the first 494 cores (one tissue micro array block), was performed by two of the authors using an open discussion procedure. Both expression intensity and fraction of positive

tumor cells were compared. Expression intensity was scored as 0-3: 0 representing the negative cores, and 3 representing the strongly stained cores. The highest intensity observed in the core was the reported value. The fraction was set as the total percentage of immunostained tumor cells in the core. Only cytoplasmic staining was considered. Manual scoring data was compared with automated analysis for MSMB and CRISP3.

Image acquisition and analysis

The Aperio ScanScope XT Slide Scanner (Aperio Technologies, Vista, CA) system was used to capture whole slide digital images with a 20x objective. Slides were de-arrayed to visualize individual cores, using Spectrum software (Aperio). A tumor specific cytoplasmic algorithm (*IHC-MARK*) was developed in house to quantify MSMB and CRISP3 protein expression. *IHC-MARK* was designed to identify tumor cells on the basis on nuclear morphology and disregard non-tumor cells such as normal epithelial or stromal cells, or invading leukocytes as previously described [8]. The algorithm calculated the percentage of positive tumor cells, as well as staining intensity ranging from 0 to 255.

Specificity of automated image analysis

To ascertain the specificity of the *IHC-MARK* algorithm to detect tumor cells, and disregard non-tumor cells, we performed a manual classification of cells as being malignant or benign and compared the manual assessment to the automated image analysis results. Nine tissue microarray cores stained with CRISP3 was selected by an investigator (ER), based on the presence of both non-tumor and tumor cells. Photographs of the immunohistochemically stained cores were compared side-by-side to the corresponding markup images generated by the *IHC-MARK* by an independent investigator (AD). The markup images were then manually assessed, and all cells were categorized. The cells recognized as tumor cells by *IHC-MARK*

were manually categorized as either true positive (true tumor cells) or false positive (non-tumor cells). The cells recognized as non-tumor cells by *IHC-MARK* were categorized as true negative (non-tumor cells) or false negative (tumor cells). For each core, sensitivity of the algorithm was calculated as the number of true positives cells divided by the total number of true positive and false negative cells. Specificity was calculated as 1 minus the number of true negative cells divided by the sum of true negatives and the number of false positives cells in the core. On the basis of the 9 raw data points, with the addition of two points for starting value (0) and ending value (100), a receiver operating characteristics curve was generated based on polynomial interpolation. Additionally, Spearman's non-parametric rank correlation coefficient was computed between the manually and automated detected number of tumor cells.

Statistical analysis

Bivariate correlation analysis (Spearman's Rho) was used to estimate the relationship between manual and automated image analysis. To compare manual and automated classification of cells into malignant or benign, a receiver operating characteristics curve was generated from 9 cores, using polynomial interpolation to predict the receiver operating characteristics response. Prognostic decision rule classification regression tree analysis was used to find subgroups with prognostic value based on fraction of positive tumor cells and staining intensity, and using biochemical recurrence or mortality as dependent variables. The automated scoring results were first separated into two groups depending on the level of staining intensity, and then further subdivided by fraction of positive tumor cells. To determine whether MSMB and CRISP3 expression followed normal distribution, we used Kolmogorov-Smirnov analysis. Significant correlations between expression levels and clinicopathological parameters were established using Mann-Whitney U-test. Cox regression

univariate analysis estimated hazard ratio, and subsequent multivariate analysis included any variable that displayed a significant association with outcome in the univariate analysis. Kaplan-Meier curves and log rank test were used to illustrate the relationship between biochemical recurrence and MSMB. Concordance index was used to determine the predictive accuracy of MSMB and CRISP3 as biomarkers. All tests were two-tailed and a p-value of 0.05 was considered significant. All calculations were performed using SPSS version 17 (SPSS Inc., Chicago, IL) and MatLab 7 (MathWorks, Apple Hill Drive, MA). To compensate for multiple testing in the correlation analyses the conservative Bonferroni adjustment could be considered and a p-value less than 0.007 would then be statistically significant. The p-values presented here have not been adjusted for multiple testing.

RESULTS

Evaluation of immunohistochemical staining

MSMB and CRISP3 protein expression was evaluated using immunohistochemistry on a tissue micro array constructed from 3,268 prostate cancer specimens. Typical staining patterns are presented in Fig. 1. Cytoplasmic MSMB expression was very strong in a vast majority of both benign epithelial cells and in cells of high-grade prostate intraepithelial neoplasia, whereas expression in cancerous tissue was decreased. CRISP3 was usually not expressed in benign prostatic epithelia, but occasionally present in high-grade prostate intraepithelial neoplasia. In general, CRISP3 was localized to the cytoplasm of subsets of tumor cells showing high expression levels.

Quantitative analysis of MSMB and CRISP3 expression

MSMB and CRISP3 protein expression were quantitatively determined using image analysis as described above. Representative markup images generated by the *IHC-MARK* algorithm

are presented in Fig. 1. Thirty eight cores with less than 50 tumor cells were excluded from analysis, leaving 3,230 interpretable cores for both MSMB and CRISP3 staining.

To evaluate the impact of benign epithelium and high-grade prostate intraepithelial neoplasia on the automated annotation, a direct comparison of manual and automated scoring for 494 cores (one tissue micro array block) was performed. Manual annotations were performed (AD and AB) for MSMB and CRISP3 staining intensity and percentage of positive cells, and results were compared to the automated image analysis. Bivariate correlation analysis demonstrated substantial agreement between manual and automated scoring; for MSMB fraction (Spearman's Rho=0.728, P<0.001); for MSMB intensity (Spearman's Rho=0.450, P<0.001); for CRISP3 fraction (Spearman's Rho=0.680, P<0.001); and for intensity (Spearman's Rho=0.546, P<0.001). Repeat analyses following exclusion of cores with high-grade prostate intraepithelial neoplastic lesions demonstrated no significant improvement in the correlation coefficient (data not shown). Based on this, and on relatively low prevalence of intraepithelial neoplastic lesions in the tissue cores, we concluded that contamination with such lesions did not significantly affect the outcome of the automated image analyses.

To ascertain the sensitivity and selectivity of the *IHC-MARK* analysis to recognize malignant cells, we compared the automated data to manual cell-by-cell assessment in 9 randomly selected cores. Subsequent receiver operating characteristics analysis generated a very good prediction response as defined as an area under the curve of 0.727 (Supplementary Figure 1). In addition, we performed a direct comparison between the number of tumor cells detected by manual and automatic assessment, and found an excellent correlation between them (Rho=0.669, p-value<0.001).

The distribution of MSMB and CRISP3 expression obtained by automated analysis is depicted in histograms in Fig 2. A majority of tumors showed a high fraction of MSMB positive tumor cells and high intensity. The fraction of CRISP3 positive tumor cells was smaller but the intensity was generally stronger.

Correlation between MSMB expression and biochemical recurrence

For subsequent statistical analyses, classification regression tree analysis was used to establish thresholds for both staining intensity and percentage of positive tumor cells. Using biochemical recurrence as the dependent outcome variable, four subgroups with different recurrence probabilities were established for MSMB. The subgroups are described in detail in Fig. 3. Compared to the cohort as a whole, patients with low MSMB staining intensity had significantly increased biochemical recurrence rates compared to those with high MSMB staining intensity (29.2% compared to18.8%; Chi square test, p= 0.001). Further subset analysis within the groups defined according to staining intensity revealed that within the low intensity subgroup, patients with low fraction of MSMB positive tumor cells had considerably more biochemical recurrence compared to those with a high fraction of MSMB positive tumor cells (43.2% versus 24.9%; Chi square test, p= 0.001). Likewise in the high intensity subgroup, patients with a low fraction of MSMB positive tumor cells had a increased biochemical recurrence rate compared to those with high fraction of MSMB positive tumor cells (36.6% versus 16.7%; Chi square test, p= 0.001).

Using classification regression tree analysis, we were unable to identify any groups whereby CRISP3 expression correlated with biochemical recurrence (data not shown).

Using MSMB expression level to predict outcome

Further exploring the apparent link between MSMB expression and biochemical recurrence, Kaplan-Meier estimates of recurrence free survival were generated (Fig. 4A-B). Patients with both high intensity and high fraction of positive tumor cells had the best prognosis, whilst those with low fraction of positive tumor cells regardless of intensity had the worst prognosis (Fig. 4A; Log Rank Mantel Cox, $p<0.001$). As previously observed in the classification regression tree analysis, patients with a low fraction of MSMB positive tumor cells had a higher rate of biochemical recurrence regardless of staining intensity. This was also evident following Kaplan-Meier analysis whereby groups clustered together based in fraction of positive tumor cells clustered (Fig. 4A). Therefore, to make subsequent analyses more practical, the groups were dichotomized into two groups based on the fraction of positive tumor cells irrespective of staining intensity (Fig. 4B). Kaplan-Meier analysis on the two MSMB groups demonstrated that increased MSMB expression was associated with a decreased recurrence risk (Log Rank Mantel Cox, $p<0.001$). To analyze whether MSMB levels had any significant predictive accuracy, we calculated the c-index and found that MSMB expression level was significantly associated with biochemical recurrence (c index= 0.604; $p<0.001$).

Examination of the relationship between MSMB protein expression and well established clinicopathological parameters demonstrated that increased MSMB expression correlated significantly with all parameters associated with improved prognosis, including low preoperative PSA serum levels, low clinical stage, low Gleason score, low pathological grade, negative surgical margin, and absence of extraprostatic extension, seminal vesicle invasion, and lymph node involvement (Table 2).

To determine if it could aid in clinical decision making, MSMB expression was compared to well-established predictors of biochemical recurrence in a multivariate model. The clinicopathological predictors included were preoperative serum PSA levels, high Gleason score, extraprostatic extension, seminal vesicle invasion, positive surgical margin, and lymph node involvement. The analysis was restricted to patients for whom complete follow-up data were available (n=2,460). Cox regression univariate analysis demonstrated that increased MSMB expression levels with a decreased biochemical recurrence free survival (HR=0.468; 95% CI 0.394-0.556; p<0.001; Table 3). Cox multivariate analysis confirmed that MSMB is an independent predictor of biochemical recurrence (HR=0.710; 95% CI 0.578-0.872; p=0.001) after controlling for the variables described above. Subsequent inclusion of MSMB in a receiver operating characteristics analysis on the base model generated a modest increase in predictive accuracy as indicated by an increased area under the curve (0.846 compared to 0.839; Supplementary fig. 2).

Level of CRISP3 expression and correlation to prognosis

Although the classification regression tree analysis did not find a significant correlation between CRISP3 levels and biochemical recurrence, we wanted to examine any other relationship CRISP3 and prognosis. Classification regression tree analysis was thus performed using overall survival as the dependent variable and identified three subgroups with different probabilities for survival (Fig. 5A). Due to the very small number of patients in the subgroup of patients with low CRISP3 staining intensity, this group was merged with the group stratified by high intensity and low fraction of positive tumor cells. The resulting Kaplan-Meier curve relating CRISP3 expression to survival was not significant (Fig. 5B; Log Rank Mantel Cox, p=0.707), yet there was a trend that patients with high CRISP3 expression had higher risk of biochemical recurrence (Fig. 5C; Log Rank Mantel Cox, p=0.085).

Furthermore, according to the Cox regression univariate analysis, CRISP3 expression was not significantly associated with biochemical recurrence or mortality (HR=1.270; 95% CI 0.967-1.668; p=0.086 and HR=1.192; 95% CI 0.476-2.988; p=0.708, respectively). However, there was an association between high MSMB expression and decreased risk of mortality (HR=0.522; 95% CI 0.292-0.932; p=0.028).

DISCUSSION

Despite the risk of overtreatment of a large number of patients with relatively indolent prostate cancer, the number of clinically applicable predictive and prognostic biomarkers is disappointingly low. Currently only serum PSA levels are included in clinical assessments, despite the low specificity of this test in localized prostate cancer.[35] Here, we emphasize the role of MSMB as a prognostic biomarker for prostate cancer.

The current study was performed on a large independent cohort of prostate cancer patients, in a high quality tissue micro array that had little sample loss and long follow-up. Lack of full follow-up data decreased the total study to 2,460 patients, but nonetheless this is a considerable number of patients. The tissue micro array contained only one sample from each patient, it was however obtained from the index tumor, thus representing the most appropriate material.

We used classification regression tree analysis to identify subgroups of MSMB expression that correlated differently to biochemical recurrence. This analysis is recognized as a robust and accurate way to predict outcome in that it is not sensitive to background noise, such as missing cases, and readily illustrates the analysis. In concordance with a study previously performed by our group, we find that the most beneficial prognostic factor for recurrence-free

survival is the fraction of MSMB-positive tumor cells.[10] Interestingly, it appears that an MSMB-positive tumor cell fraction as small as 8-10% greatly reduces the risk of recurrence. Furthermore, this remains significant in the multivariate model. MSMB intensity appears to be of less significance. Interestingly, the cut-off values we find to optimally define MSMB high and low expression in the current cohort is very similar to the cut-off values found in our previous study of an independent cohort.[10]

In the current study, we found no significant correlation between CRISP3 expression and biochemical recurrence, neither regarding intensity nor regarding fraction of CRISP3 positive tumor cells. However, similar to our previous findings, there was a trend suggesting that patients with high CRISP3 expression had increased risk for recurrence. Although we considered MSMB to be the stronger marker, we were surprised CRISP3 did not correlate significantly to outcome in this large cohort. Additional studies on long term survival are required to evaluate whether MSMB and/or CRISP3 will be of use in the clinic as prognostic tissue biomarkers for prostate cancer.

Automated annotations are becoming more prevalent as a tool for histopathological assessments since they offer a sensitive and reliable system and remove inherent inter- and intraobserver variability associated with manual assessment. An automated approach provides accurate high throughput analysis as demonstrated in this study of over 3,000 cases. Automated analysis also provides a quantitative assessment enabling robust statistical analyses as demonstrated by the use of decision tree analysis to identify prognostic subgroups in this study. The automated image analysis approach used here was able to identify and selectively evaluate tumor cells based on nuclear morphology.[8] This method is “learning-based”, meaning that a technician has to “teach” the algorithm to differ between the

morphology of a tumor cell and any other cell types that are present in the tissue. However, the impact of heterogenous morphology often seen among prostate cancer cells remains to be fully evaluated, although we demonstrated a high correlation between manual and automated analysis in this study.

In normal prostate and benign prostatic hyperplasia, virtually all epithelial cells express MSMB. It was surprising therefore, to find that the fraction of MSMB-positive cells fell to 10% before the expression correlated with a negative outcome, suggesting that there is a redundancy in protein expression and a fraction of MSMB-expressing cells may be sufficient to maintain any potential tumor suppressing effect(s) that MSMB may exert on tumor cells. Using exogenous MSMB, *in vitro* and *in vivo* studies indicate that MSMB may have several anti-tumor effects on prostate tumor cells, such as suppressing growth and inducing apoptosis;[36-38] decreasing metastatic disease;[39-41] and inhibiting angiogenesis.[42]

Recently MSMB has gained further attention as a candidate prostate cancer susceptibility gene,[20, 21] and several causal risk alleles affecting the level of gene transcription have been identified in the region upstream of the coding sequence.[22-25] The rs10993994 risk allele has the strongest association with prostate cancer risk, and was recently shown to decrease *MSMB* gene expression as well as MSMB protein levels in serum.[43] Additionally, *MSMB* is subject to epigentic silencing by enhancer of zeste homologue-2 (EZH2),[44] a mediator of histone methylation and subsequent repression of target genes that is linked to progression of prostate cancer.[45] We have previously seen that MSMB expression is downregulated by long-term androgen deprivation therapy, and that expression is reduced in castration resistant prostate cancer.[46] Taken together, several lines of evidence support the view that MSMB appears to be important in the transformation and progression of prostate cancer cells.

We are currently performing *in vitro* studies to elucidate whether MSMB may possess anti-tumor mechanisms, as this has not previously been described in detail. Also, studies investigating the effect of an increased CRISP3 expression on MSMB function are completely lacking. One may speculate that increased CRISP3 can bind to MSMB in the extracellular space and thus prevent MSMB from exerting anti-tumor effects. On the other hand, it may be that decreased MSMB expression simply reflects the phenotypic difference of a less differentiated prostate cancer cell compared to a fully mature columnar epithelial cell, rather than being associated with any true tumor suppressor function. Indeed, the *MSMB* gene is one of the most down-regulated genes in less differentiated prostate cancer cells compared to committed epithelial cells.[47]

In conclusion, using an automated image analysis approach and a large independent cohort of prostate cancer patients, we have verified previous findings that MSMB is a strong independent tissue marker for prostate cancer recurrence after radical prostatectomy.

DISCLOSURE/CONFLICT OF INTERESTS

Dr Rexhepaj, Dr Brennan and Prof Gallagher are inventors of a pending patent application in relation to the development of novel automated image analysis approaches in histopathology.

ACKNOWLEDGEMENTS

We are grateful to Dr Thorsten Schlomm and Professor Guido Sauter at the Martini-Clinic, Prostate Cancer Center, University Medical Center, Hamburg, Germany for their generosity in sharing their prostate cancer tissue micro array; to Elise Nilsson for her excellent technical skills in the field of immunohistochemistry; to Dr Lene Udby for providing the CRISP3

antibody; and to Dr Per Fernlund for the MSMB antibody. The UCD Conway Institute is funded by the Programme for Third Level Institutions (PRTLI), as administered by the Higher Education Authority (HEA) of Ireland. Funding is acknowledged from Enterprise Ireland (WG, DB and ER), and from European Union 6th Framework (P-Mark) [grant number LSHC-CT-2004-503011], Swedish Cancer Society, Swedish Research Council (Medicine), Cancer and Research Foundation at University Hospital Malmö, Gunnar Nilsson Cancer Foundation, and the Crafoord Foundation (AB and AD).

Supplementary information is available at *Modern Pathology*'s website.

REFERENCES

1. Jemal A, Siegel R, Ward E*, et al.* Cancer statistics, 2009. CA Cancer J Clin 2009;59:225-49.

2. Kattan MW, Eastham JA, Stapleton AM, Wheeler TM, Scardino PT. A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. J Natl Cancer Inst 1998;90:766-71.

3. Shariat SF, Karakiewicz PI, Roehrborn CG, Kattan MW. An updated catalog of prostate cancer predictive tools. Cancer 2008;113:3075-99.

4. Fiorentino M, Capizzi E, Loda M. Blood and tissue biomarkers in prostate cancer: state of the art. Urol Clin North Am 2010;37:131-41.

5. Shariat SF, Karakiewicz PI, Suardi N, Kattan MW. Comparison of nomograms with other methods for predicting outcomes in prostate cancer: a critical analysis of the literature. Clin Cancer Res 2008;14:4400-7.

6. Faratian D, Kay C, Robson T*, et al.* Automated image analysis for high-throughput quantitative detection of ER and PR expression levels in large-scale clinical studies: the TEAM Trial Experience. Histopathology 2009;55:587-93.

7. Turbin DA, Leung S, Cheang MC*, et al.* Automated quantitative analysis of estrogen receptor expression in breast carcinoma does not differ from expert pathologist scoring: a tissue microarray study of 3,484 cases. Breast Cancer Res Treat 2008;110:417-26.

8. Rexhepaj E, Brennan DJ, Holloway P*, et al.* Novel image analysis approach for quantifying expression of nuclear proteins assessed by immunohistochemistry:

application to measurement of oestrogen and progesterone receptor levels in breast cancer. Breast Cancer Res 2008;10:R89.

9. Brennan DJ, Rexhepaj E, O'Brien SL*, et al.* Altered cytoplasmic-to-nuclear ratio of survivin is a prognostic indicator in breast cancer. Clin Cancer Res 2008;14:2681-9.

10. Bjartell AS, Al-Ahmadie H, Serio AM*, et al.* Association of cysteine-rich secretory protein 3 and beta-microseminoprotein with outcome after radical prostatectomy. Clin Cancer Res 2007;13:4130-8.

11. Lilja H, Abrahamsson P-A. Three predominant proteins secreted by the human prostate gland. Prostate 1988;12:29-38.

12. Abrahamsson PA, Lilja H, Falkmer S, Wadstrom LB. Immunohistochemical distribution of the three predominant secretory proteins in the parenchyma of hyperplastic and neoplastic prostate glands. Prostate 1988;12:39-46.

13. Whitaker HC, Warren AY, Eeles R, Kote-Jarai Z, Neal DE. The potential value of microseminoprotein-beta as a prostate cancer biomarker and therapeutic target. Prostate 2009.

14. Doctor VM, Sheth AR, Simha MM*, et al.* Studies on immunocytochemical localization of inhibin-like material in human prostatic tissue: comparison of its distribution in normal, benign and malignant prostates. Br J Cancer 1986;53:547-54.

15. Tsurusaki T, Koji T, Sakai H*, et al.* Cellular expression of beta-microseminoprotein (beta-MSP) mRNA and its protein in untreated prostate cancer. Prostate 1998;35:109-16.

16. Chan PS, Chan LW, Xuan JW*, et al.* In situ hybridization study of PSP94 (prostatic secretory protein of 94 amino acids) expression in human prostates. Prostate 1999;41:99-109.

17. Sakai H, Tsurusaki T, Kanda S*, et al.* Prognostic significance of beta-microseminoprotein mRNA expression in prostate cancer. Prostate 1999;38:278-84.

18. Vanaja DK, Cheville JC, Iturria SJ, Young CY. Transcriptional silencing of zinc finger protein 185 identified by expression profiling is associated with prostate cancer progression. Cancer Res 2003;63:3877-82.

19. Reeves JR, Dulude H, Panchal C, Daigneault L, Ramnani DM. Prognostic value of prostate secretory protein of 94 amino acids and its binding protein after radical prostatectomy. Clin Cancer Res 2006;12:6018-22.

20. Thomas G, Jacobs KB, Yeager M*, et al.* Multiple loci identified in a genome-wide association study of prostate cancer. Nat Genet 2008;40:310-5.

21. Eeles RA, Kote-Jarai Z, Giles GG*, et al.* Multiple newly identified loci associated with prostate cancer susceptibility. Nat Genet 2008;40:316-21.

22. Chang BL, Cramer SD, Wiklund F*, et al.* Fine mapping association study and functional analysis implicate a SNP in MSMB at 10q11 as a causal variant for prostate cancer risk. Hum Mol Genet 2009;18:1368-75.

23. Lou H, Yeager M, Li H*, et al.* Fine mapping and functional analysis of a common variant in MSMB on chromosome 10q11.2 associated with prostate cancer susceptibility. Proc Natl Acad Sci U S A 2009;106:7933-8.

24. Kote-Jarai Z, Leongamornlert D, Tymrakiewicz M*, et al.* Mutation analysis of the MSMB gene in familial prostate cancer. Br J Cancer 2009.

25. Yeager M, Deng Z, Boland J*, et al.* Comprehensive resequence analysis of a 97 kb region of chromosome 10q11.2 containing the MSMB gene associated with prostate cancer. Hum Genet 2009.

26. Asmann YW, Kosari F, Wang K, Cheville JC, Vasmatzis G. Identification of differentially expressed genes in normal and malignant prostate by electronic profiling of expressed sequence tags. Cancer Res 2002;62:3308-14.

27. Ernst T, Hergenhahn M, Kenzelmann M*, et al.* Decrease and gain of gene expression are equally discriminatory markers for prostate carcinoma: a gene expression analysis on total and microdissected prostate tissue. Am J Pathol 2002;160:2169-80.

28. Bjartell A, Johansson R, Bjork T*, et al.* Immunohistochemical detection of cysteine-rich secretory protein 3 in tissue and in serum from men with cancer or benign enlargement of the prostate gland. Prostate 2006;66:591-603.

29. Kosari F, Asmann YW, Cheville JC, Vasmatzis G. Cysteine-rich secretory protein-3: a potential biomarker for prostate cancer. Cancer Epidemiol Biomarkers Prev 2002;11:1419-26.

30. Udby L, Lundwall A, Johnsen AH*, et al.* beta-Microseminoprotein binds CRISP-3 in human seminal plasma. Biochem Biophys Res Commun 2005;333:555-61.

31. Erbersdobler A, Isbarn H, Steiner I*, et al.* Predictive value of prostate-specific antigen expression in prostate cancer: a tissue microarray study. Urology 2009;74:1169-73.

32. Fleischmann A, Schlomm T, Huland H*, et al.* Distinct subcellular expression patterns of neutral endopeptidase (CD10) in prostate cancer predict diverging clinical courses in surgically treated patients. Clin Cancer Res 2008;14:7838-42.

33. Udby L, Bjartell A, Malm J*, et al.* Characterization and localization of cysteine-rich secretory protein 3 (CRISP-3) in the human male reproductive tract. J Androl 2005;26:333-42.

34. Abrahamsson PA, Andersson C, Bjork T*, et al.* Radioimmunoassay of beta-microseminoprotein, a prostatic-secreted protein present in sera of both men and women. Clin Chem 1989;35:1497-503.

35. Ulmert D, O'Brien MF, Bjartell AS, Lilja H. Prostate kallikrein markers in diagnosis, risk stratification and prognosis. Nat Rev Urol 2009;6:384-91.

36. Garde SV, Basrur VS, Li L*, et al.* Prostate secretory protein (PSP94) suppresses the growth of androgen-independent prostate cancer cell line (PC3) and xenografts by inducing apoptosis. Prostate 1999;38:118-25.

37. Shukeir N, Arakelian A, Kadhim S, Garde S, Rabbani SA. Prostate secretory protein PSP-94 decreases tumor growth and hypercalcemia of malignancy in a syngenic in vivo model of prostate cancer. Cancer Res 2003;63:2072-8.

38. Garde S, Fraser JE, Nematpoor N*, et al.* Cloning, expression, purification and functional characterization of recombinant human PSP94. Protein Expr Purif 2007;54:193-203.

39. Shukeir N, Arakelian A, Chen G*, et al.* A synthetic 15-mer peptide (PCK3145) derived from prostate secretory protein can reduce tumor growth, experimental skeletal metastases, and malignancy-associated hypercalcemia. Cancer Res 2004;64:5370-7.

40. Annabi B, Bouzeghrane M, Currie JC*, et al.* A PSP94-derived peptide PCK3145 inhibits MMP-9 secretion and triggers CD44 cell surface shedding: implication in tumor metastasis. Clin Exp Metastasis 2005;22:429-39.

41. Annabi B, Currie JC, Bouzeghrane M*, et al.* Contribution of the 37-kDa laminin receptor precursor in the anti-metastatic PSP94-derived peptide PCK3145 cell surface binding. Biochem Biophys Res Commun 2006;346:358-66.

42. Lamy S, Ruiz MT, Wisniewski J*, et al.* A prostate secretory protein94-derived synthetic peptide PCK3145 inhibits VEGF signalling in endothelial cells: implication in tumor angiogenesis. Int J Cancer 2006;118:2350-8.

43. Xu B, Wang J, Tong N*, et al.* A functional polymorphism in MSMB gene promoter is associated with prostate cancer risk and serum MSMB expression. Prostate 2010.

44. Beke L, Nuytten M, Van Eynde A, Beullens M, Bollen M. The gene encoding the prostatic tumor suppressor PSP94 is a target for repression by the Polycomb group protein EZH2. Oncogene 2007;26:4590-5.

45. Varambally S, Dhanasekaran SM, Zhou M*, et al.* The polycomb group protein EZH2 is involved in progression of prostate cancer. Nature 2002;419:624-9.

46. Dahlman A, Edsjo A, Hallden C*, et al.* Effect of androgen deprivation therapy on the expression of prostate cancer biomarkers MSMB and MSMB-binding protein CRISP3. Prostate Cancer Prostatic Dis 2010;13:369-75.

47. Birnie R, Bryce SD, Roome C*, et al.* Gene expression profiling of human prostate cancer stem cells reveals a pro-inflammatory phenotype and the importance of extracellular matrix interactions. Genome Biol 2008;9:R83.

TITLES AND LEGENDS TO FIGURES

Fig. 1. Representative photographs depicting prostate cancer tissue immunohistochemically stained for MSMB and CRISP3 protein expression (left), with corresponding markup images generated by the *IHC-MARK* algorithm (right). The algorithm recognizes cancer cells by their morphological properties, and marks it with a corresponding color. The nuclei of *IHC-MARK* negative tumor cells are seen in blue; nuclei of *IHC-MARK* positive tumor cells in red; and cytoplasm of tumor cells in orange. Tumor areas are defined by green lines. MSMB and CRISP3 expression were subsequently stratified into high and low subgroups using classification regression tree analyses (Fig 3 and Fig 5A). Photographs were generated at an original magnification of 20x.

Fig. 2. Histograms depicting distribution of staining intensity and fraction of positive tumor cells based on automated image analysis.

Fig. 3. Flow chart depicting the subdivision of patients into relevant prognostic groups by classification regression tree analysis. MSMB expression was subdivided by intensity and subsequently by fraction of positive tumor cells into four groups with different risk for biochemical recurrence. For CRISP3 expression, no subgroups were significantly linked to outcome. Only patients with complete follow-up data were included (n=2,460).

Fig. 4. Kaplan-Meier curves plotting recurrence-free probability. Patients were stratified by level of MSMB expression into four subgroups (A), with different risk for biochemical recurrence. Black line: high intensity and high fraction of positive tumor cells; dashed line: low intensity and high fraction; grey line: high intensity and low fraction; dotted line: low

intensity and low fraction. The four subgroups were merged into two groups (B), either having high or low MSMB expression. Black line: high fraction of positive tumor cells; grey line: low fraction. Follow-up time ranged from 1-144 months. Only patients with complete follow-up data were included (n=2,460).

Fig. 5. Subgroups of CRISP3 expression may correlate with disease-free survival. Performing classification regression tree analysis on CRISP3 expression, three subgroups were found to correlate differently with disease-free survival (A). The three subgroups were merged into two groups, with high or low CRISP3 expression. Subsequent Kaplan-Meier curves were drawn, plotting disease-free survival (B) and recurrence free survival (C). Black line: low CRISP3 expression; grey line: high expression.

Table 1. Patient characteristics

Variable	***Number of patients (%)***
Age at time of surgery (years)	
Mean (median)	62.1 (63)
<50	83 (2.7%)
50-59	834 (27.0%)
60-69	1,901 (61.5%)
>70	275 (8.9%)
Pre-operative PSA (ng/mL)	
Mean (median)	9.3 (6.8)
<3.0	266 (8.7%)
3.0-10.0	1,922 (63.0%)
10.1-20.0	641 (21.0%)
>20.0	223 (7.3%)
Clinical stage	
T1c	2,026 (65.8%)
T2	1,009 (32.8%)
T3	44 (1.4%)
Biopsy Gleason score	
≤6	2,045 (68.5%)
3+4	662 (22.2%)
4+3	184 (6.2%)
≥8	90 (3.0%)
Prostatectomy Gleason score	
≤6	1,426 (45.9%)
3+4	1,311 (42.2%)
4+3	313 (10.1%)
≥8	54 (1.7%)
Pathological stage	
pT2	2,079 (67.0%)
pT3	980 (31.6%)
pT4	42 (1.4%)
Extraprostatic extension	967 (31.3%)
Positive surgical margins	627 (20.2%)
Lymph node involvement[1]	95 (5.8%)
Follow-up (months)	
Mean (median)	44.5 (34.3)
Range	0.0 – 192.6
Overall no. of biochemical recurrences	620 (20%)

[1]Lymph node status was only available in 1639 patients

Table 2. Correlation of MSMB and CRISP3 expression levels to clinicopatological parameters

	Total no patients (%)	***Patients with low MSMB (%)***	***Patients with high MSMB (%)***	***p-value***
PSA				*<0.001***
<3.0	261(8.6)	51(10.0)	210(8.4)	
3.1-10.0	1,904(63.0)	271(53.0)	1,633(65.0)	
10.1-20.0	634(21.0)	122(23.9)	512(20.4)	
>20.1	223(7.4)	67(13.1)	156(6.2)	
Clinical stage				*<0.001***
T1c	2,000(65.6)	263(50.8)	1,737(68.7)	
T2	1,003(32.9)	236(45.6)	767(30.3)	
T3	44(1.5)	19(3.6)	25(1.0)	
Gleason score				*<0.001***
≤6	1,408(45.8)	171(32.6)	1,237(48.5)	
3+4	1,299(42.3)	227(43.3)	1,072(42.1)	
4+3	311(10.1)	103(19.7)	208(8.2)	
≥8	54(1.8)	23(4.4)	31(1.2)	
Pathological stage				*<0.001***
pT2	2,053(66.9)	262(50.0)	1,791(70.4)	
pT3	974(31.7)	240(45.8)	734(28.8)	
pT4	42(1.4)	22(4.2)	20(0.8)	
Extraprostatic extension				*<0.001***
No	2,092(68.5)	266(51.5)	1,826(72.0)	
Yes	962(31.5)	251(48.5)	711(28.0)	
Positive surgical margin				*0.003**
No	2,445(79.6)	392(75.0)	2,053(80.6)	
Yes	625(20.4)	131(25.0)	494(19.4)	
Lymph node involvement[1]				*0.005**
No	1,529(94.2)	303(90.0)	1,226(95.0)	
Yes	94(5.8)	30(9.0)	64(5.0)	

[1]Lymph node status was available for 1,639 patients

The Mann-Whitney U-test was used for all correlation analyses.

Table 3. Relative risk for biochemical recurrence in patients after radical prostatectomy

	Univariate		***Multivariate***	
	HR (95% CI)	**p-value**	**HR (95% CI)**	**p-value**
PSA	1.036(1.031-1.040)	*<0.001***	1.014(1.008-1.020)	*<0.001***
Clinical stage	1.953(1.691-2.257)	*<0.001***	1.054(0.888-1.252)	0.545
Gleason score	2.909(2.656-3.187)	*<0.001***	1.627(1.430-1.852)	*<0.001***
Pathological stage	4.598(3.975-5.317)	*<0.001***	2.311(1.683-3.174)	*<0.001***
Extraprostatic extension	5.004(4.183-5.987)	*<0.001***	1.243(0.857-1.804)	0.251
Positive surgical margin	2.915(2.475-3.432)	*<0.001***	1.627(1.329-1.991)	*<0.001***
Lymph node involvement[1]	6.463(5.086-8.214)	*<0.001***	2.199(1.658-2.915)	*<0.001***
MSMB level	0.468(0.394-0.556)	*<0.001***	0.710(0.578-0.872)	*0.001***

[1]Lymph node status was available in 1,545 patients

Figure 1

MSMB

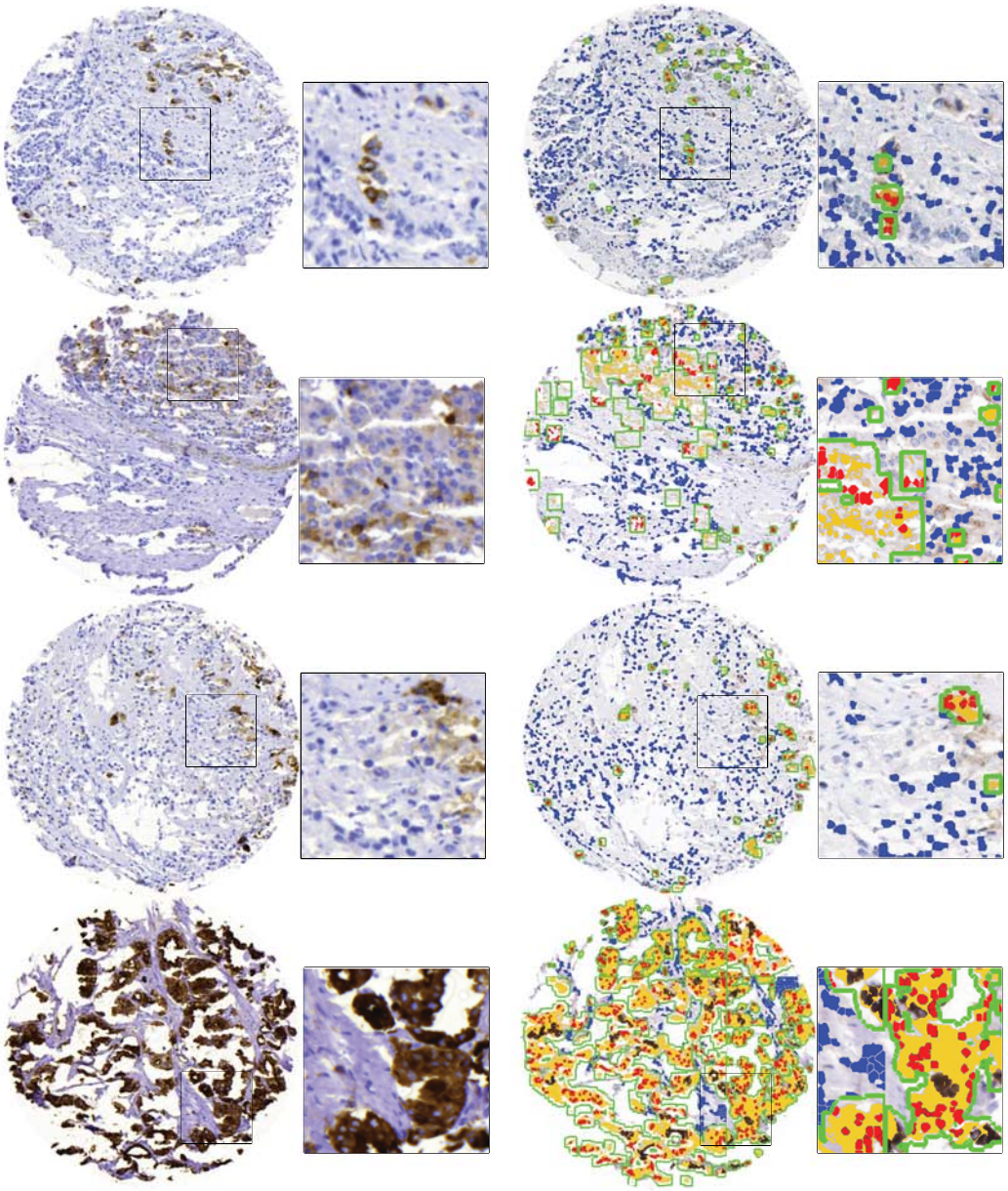


Figure 1 continued

CRISP3

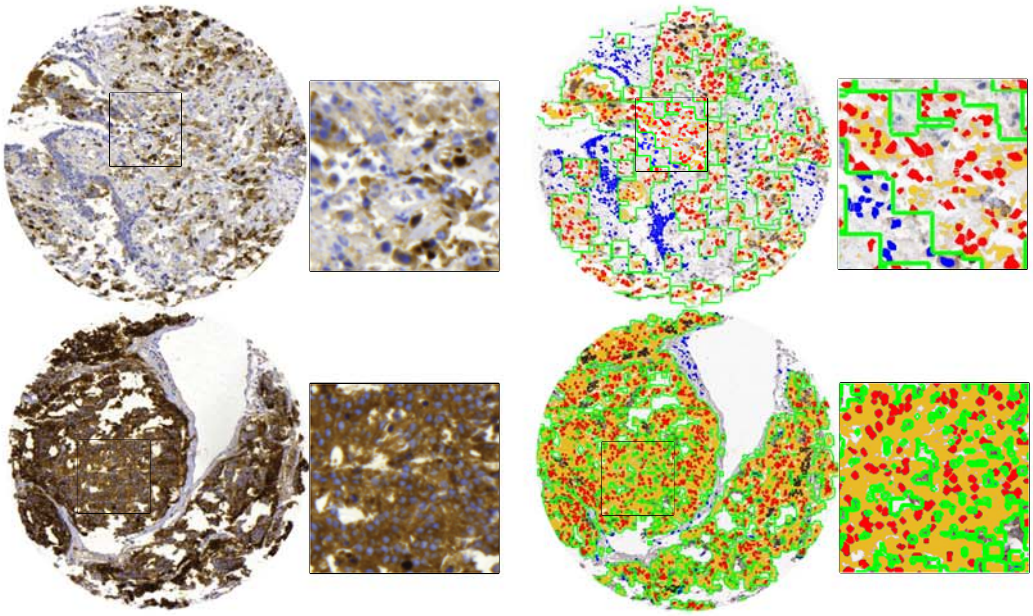


Figure 2. Histograms representing distribution of staning intensity and fraction of positive tumor cells, based on automated image analysis

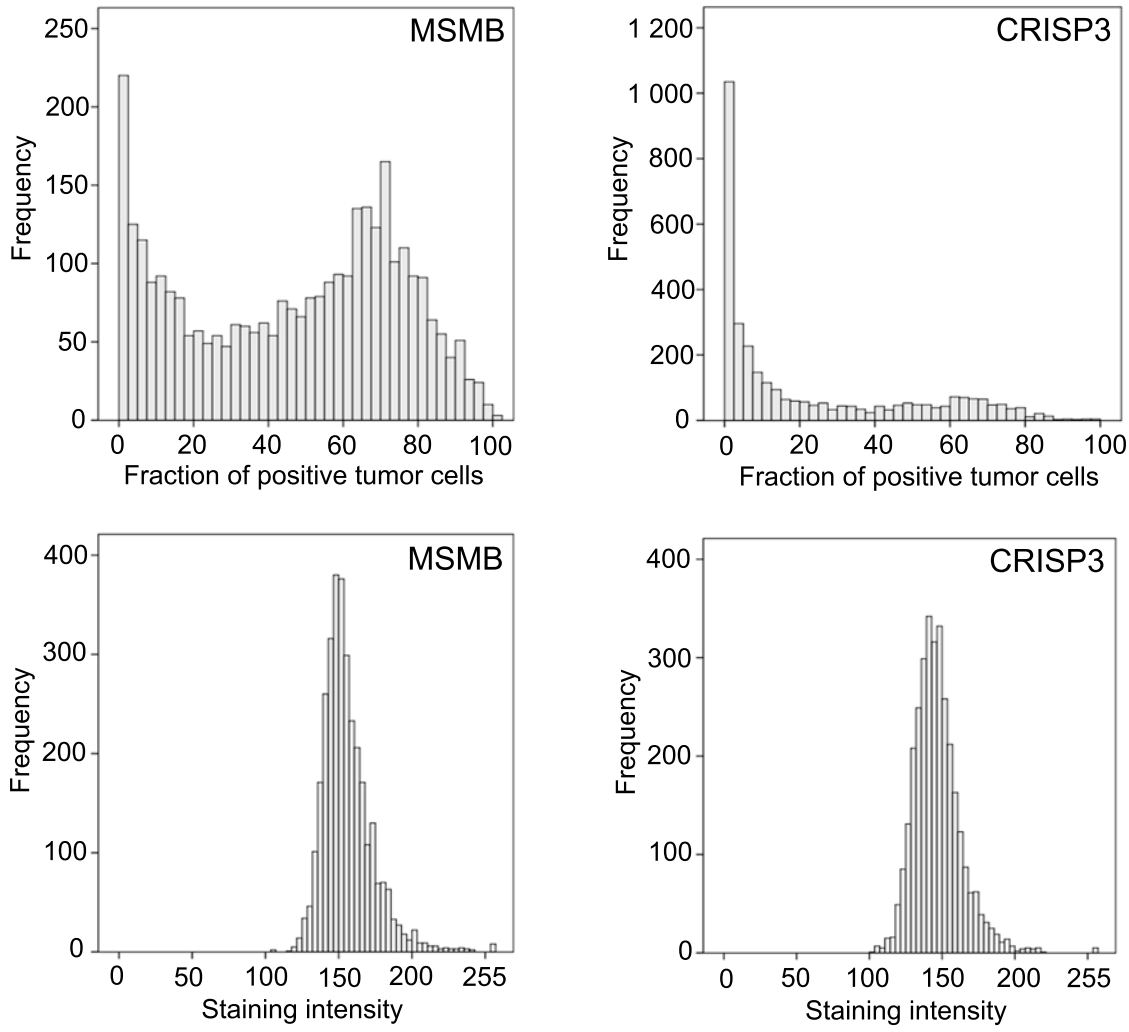


Figure 3 Identification of MSMB-expressing groups of prognostic value

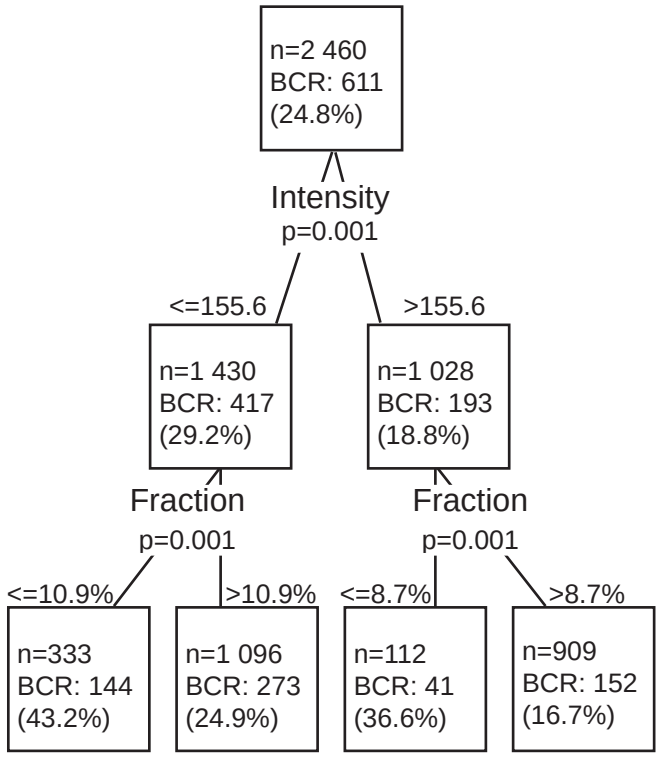


Figure 4

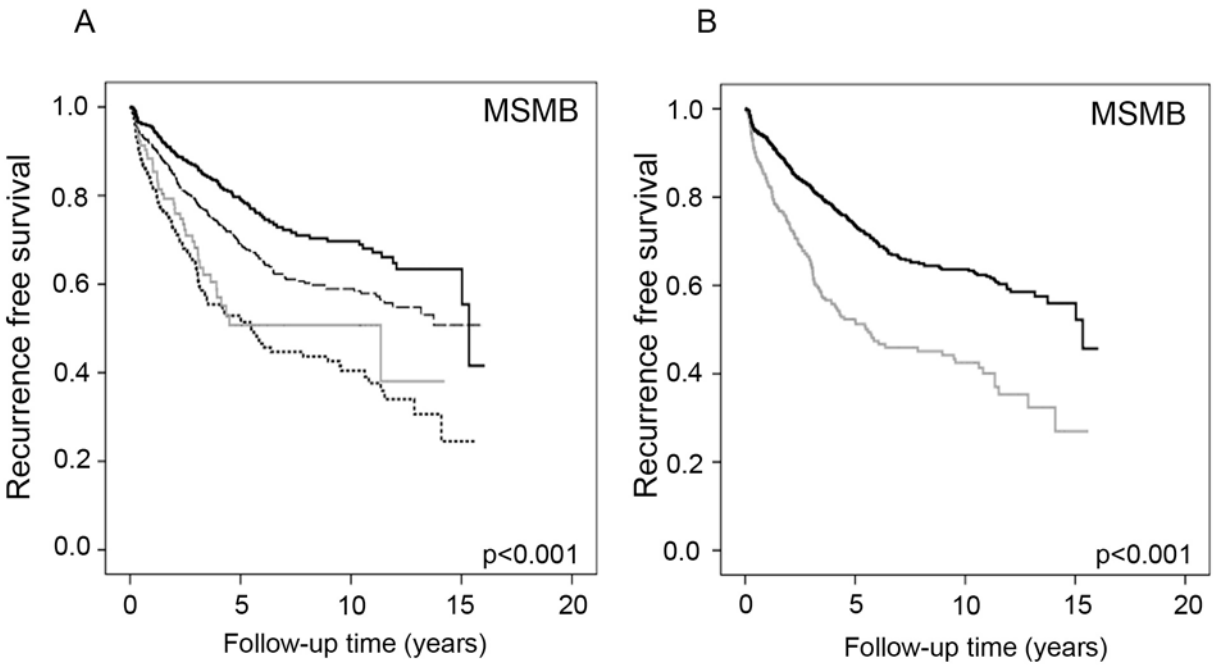
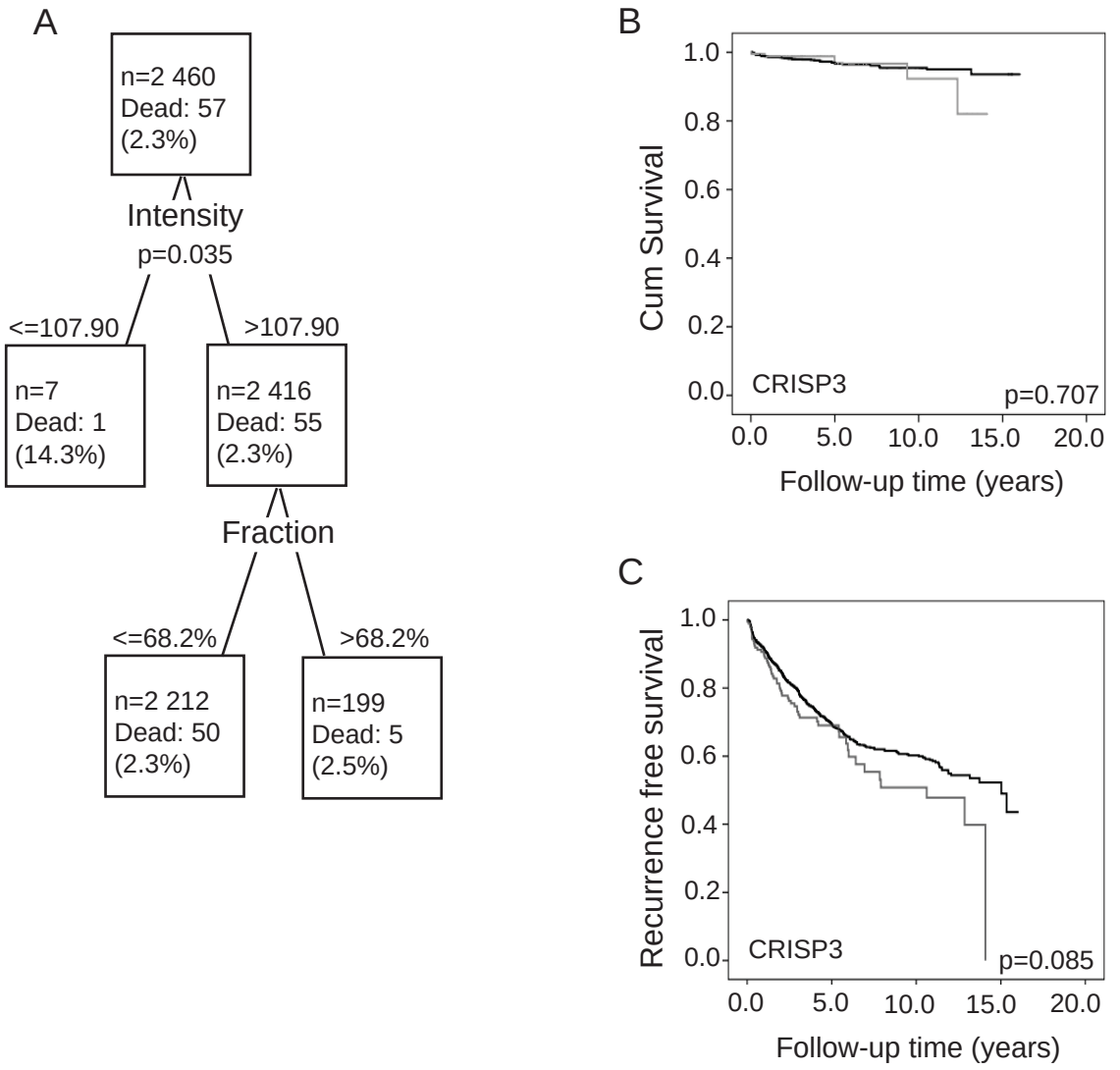
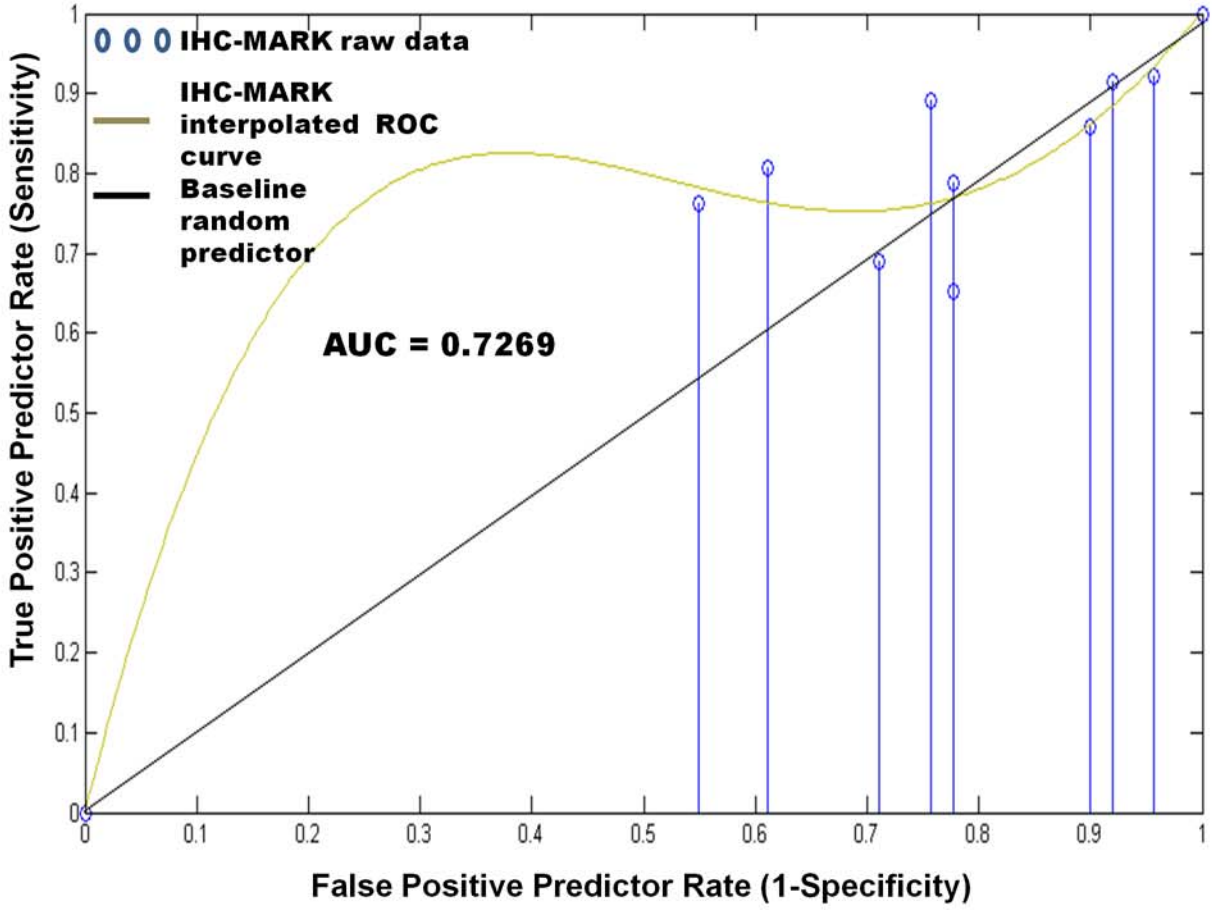

A
MSMB
Recurrence free survival
1.0
0.8
0.6
0.4
0.2
0.0
p<0.001
0
5
10
15
20
Follow-up time (years)
B
MSMB
Recurrence free survival
1.0
0.8
0.6
0.4
0.2
0.0
p<0.001
0
5
10
15
20
Follow-up time (years)

Figure 5 CRISP3 levels and association to overall survival

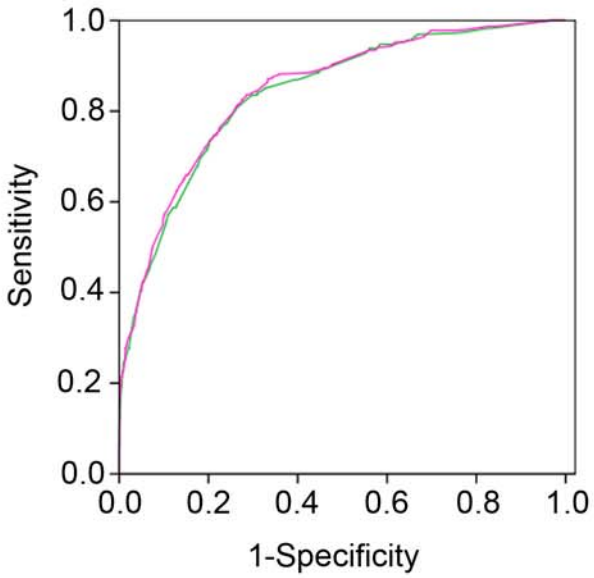


Supplementary figure 1



Supplementary Figure 1. Testing of IHC-MARK using Receiver Operatory Curve analysis. Sensitivity and specificity of IHC-MARK detection platform was tested on a random set of 9 images (i.e. blue circles) stained with CRISP3. A polynomial ROC curve was fitted to the model (i.e. gray continuous line) and compared to the baseline, random predictor (i.e. black straight continuous line).

Supplementary figure 2



				Asyptotic 95% Confidence Interval	
Test result variables:	Area:	Std. error:	Asymptotic sig:	Lower Bound:	Upper Bound:
Clinical parameters	0.839	0.011	0.000	0.817	0.861
Clinical parameters + MSMB	0.846	0.011	0.000	0.824	0.867

Supplementary figure 2: Receiver operating charachteristics analysis was performed to ascertain whether MSMB adds predictive accuracy to the clinical base model used today. The clinical parameters included were the same as for the multivariate analysis (Table 3), and the analysis was performed with (pink line) or without MSMB (green line).