ARTICLES

AQCH ANALYTICAL AND QUANTITATIVE CYTOPATHOLOGY AND HISTOPATHOLOGY®

An Official Periodical of **The International Academy of Cytology** *and the* **Italian Group of Uropathology**

Lymph Node Metastasis Status in Breast Carcinoma Can Be Predicted via Image Analysis of Tumor Histology

Mark D. Zarella, Ph.D., **David E. Breen,** Ph.D., **Md. Alimoor Reza,** M.S., **Aladin Milutinovic,** M.E., and **Fernando U. Garcia,** M.D.

OBJECTIVE: To develop a method whereby axillary lymph node (ALN) metastasis can be predicted without ALN dissection, via computational image analysis of routinely acquired tumor histology.
STUDY DESIGN: We employed digital image processing to stratify patients based on the histological attributes of the primary tumor. We extracted image features that capture the nuclear and architectural properties of the specimen. We then used a novel machine learning algorithm to transform image features into a scalar score that provided not only a metastasis prediction but also the certainty of classification.
RESULTS: We applied this procedure to 101 patients with a ground truth established by histological examination of the lymph nodes and found that 68.3% of the cohort could be classified, exhibiting a correct prediction rate of 88.4%.
CONCLUSION: These results demonstrate a technique that potentially can be used to supplant existing surgical methods to determine ALN metastasis status, thereby reducing patient morbidity associated with overtreatment. (Anal Quant Cytopathol Histpathol 2015; 37:273–285)

Keywords: computer-assisted image processing, diagnostic imaging, digital pathology, image processing, informatics, lymph nodes, machine learning, metastasis, morphometry, neoplasm metastasis.

From the Department of Pathology and Laboratory Medicine, Drexel University College of Medicine, and the Department of Computer Science, College of Computing and Informatics, Drexel University, Philadelphia, Pennsylvania, USA.

Dr. Zarella is Assistant Professor, Department of Pathology and Laboratory Medicine, Drexel University College of Medicine.

Dr. Breen is Associate Professor, Department of Computer Science, College of Computing and Informatics, Drexel University.

Mr. Reza is Graduate Student, Department of Computer Science, College of Computing and Informatics, Drexel University.

Mr. Milutinovic is Research Assistant, Department of Pathology and Laboratory Medicine, Drexel University College of Medicine.

Dr. Garcia was Professor, Department of Pathology and Laboratory Medicine, Drexel University College of Medicine, and currently is Medical Director, Cancer Treatment Centers of America, Eastern Regional Medical Center, Philadelphia.

Supported by the Pennsylvania Department of Health through the Commonwealth Universal Research Enhancement (CURE) Program grants 4100047631 and 4100062203.

Address correspondence to: Mark D. Zarella, Ph.D., Department of Pathology and Laboratory Medicine, Drexel University College of Medicine, 245 North 15th Street, Philadelphia, PA 19102, USA (mark.zarella@drexelmed.edu).

Financial Disclosure: The authors have no connection to any companies or products mentioned in this article.

0884-6812/15/3705-0273/$18.00/0

Axillary lymph node (ALN) metastasis status remains one of the most critical prognostic variables for breast cancer management decision-making and patient survival. At present, the standard of care for many breast cancer patients includes dissection of the sentinel lymph node (SLN)—the lymph node determined to be the most likely first site of metastasis—using intraoperative tracers or high contrast dyes. The most common treatment strategy for SLN-positive patients is surgical excision of the nonsentinel ALNs to facilitate tumor staging and treatment selection. However, this procedure is often accompanied by chronic lymphedema, seroma, paresthesia due to diminished lymphatic drainage and damage of intercostal nerves, or shoulder and arm dysfunction, as reported in 35–38% of patients.[1,2] To avoid these complications, ALN excision is often advised only for cases exhibiting macrometastases (i.e., SLN tumor sizes >2.0 cm), but the risk of ALN involvement for even small SLN deposit sizes <0.4 cm has been shown to be significant (21%).[3] This suggests that the exclusionary criterion must be refined in order to safely indicate whether patients should undergo ALN dissection.

Additional prognostic factors have been sought in order to reliably predict ALN status and therefore better guide patient selection for ALN surgery. Patient age, primary tumor size, lymphovascular invasion, and extranodal extension[4-7] have been found to be associated with an increased incidence of ALN metastasis. These insights have produced models to predict ALN involvement with overall accuracy exhibiting area under ROC curve (AUC) values that often approach 0.7 or 0.8,[7-12] indicating moderate predictive capacity. For instance, Van Zee and colleagues[13] incorporated histological characteristics of the SLN in addition to primary tumor attributes to predict ALN metastasis and demonstrated predictive success corresponding to an area under the ROC curve of 0.77. Further evaluation of this model on independent data sets have revealed AUC values between 0.63 and 0.74[14-17] and low false-positive and false-negative rates (9.1% and 4.1%, respectively[15]) for the highest scoring cases, although these comprised only 30% of the total number of cases in their cohort. While these approaches have provided clinicians a number of tools with which to assess the likelihood of ALN metastasis in patients, reliability is still limited for widespread applicability in patients. As a result, ALN dissection remains the recommended course of treatment by the American Society of Clinical Oncology[18] even though the majority of patients that undergo this procedure ultimately do not demonstrate ALN involvement.[19]

Microscopic evaluation of tumor histology provides insights about a tumor's attributes, including nuclear grade, which has been shown to be associated with ALN metastasis.[9,11] However, this has not yet produced an objective method by which metastasis status can be reliably predicted but instead relies on subjective scoring and interpretation. Previous analyses of nuclear morphology have revealed a significant correlation between nuclear size and shape and patient outcomes,[20-26] suggesting that the phenotypic variations observed in breast tumors contain information related to patient prognosis. Similarly, Beck et al[27] recently showed that stromal features that play a key role in the architectural underpinnings of breast tumors can be quantitatively evaluated and harnessed to stratify patients. Here we describe a method that utilizes the rich information contained in nuclear morphology and tumor architecture to predict ALN status, relying only on routinely acquired histological images of primary tumor specimens.

Materials and Methods

We developed an algorithm to predict lymph node status from tumor histology using specimens from 101 patients who had undergone lumpectomy as a primary treatment modality to both train the algorithm and measure its performance. The use of these specimens was approved by the Drexel University College of Medicine Institutional Review Board. The nodal status associated with each case was assessed via SLN biopsy and, in SLN-positive cases, confirmed with ALN dissection. Sentinel lymph nodes were sectioned according to a protocol which included the following: cut at 0.3–0.5-cm intervals along the short axis and submit entirely for microscopic examination; section the tissue within the paraffin blocks until a complete face of the lymph node is identified. Two H&E slides and one AE1/AE3 immunostain, 20 mm apart, were studied histologically.

Image Acquisition and Preprocessing

High-resolution (0.5 μm/pixel) whole slide digital images were obtained at 20× magnification using

the Aperio Scanscope XT (Aperio, Vista, California, USA). Image analysis was performed on the raw image data using custom software written in Matlab (MathWorks, Natick, Massachusetts, USA). The slide with the highest proportion of positive staining was selected for analysis from each case. Positive staining was estimated by first training a classifier to distinguish stained pixels from unstained pixels. Training was accomplished by manually selecting positively stained pixels and negatively stained pixels from a set of representative images and optimally dividing them in hue-saturation-value (HSV) space (Supplementary Figure 1) using support vector machine (SVM) learning. Subsequently, a 6000×6000 pixel region of interest (ROI) was defined as the region with the greatest number of positively stained images. These ROIs served as the basis for all further analyses.

Nuclear Segmentation

In order to capture the properties of cell nuclei within the ROIs, an initial step was required to detect candidate nuclei and segment them from the background. We applied the classifier from the previous step to identify candidate pixels that are likely to represent stained nuclei. We then applied the watershed transform[28] to the candidate pixel map to extract image contours and define "objects" bound by these contours. Objects that contained fewer than 17 contiguous pixels were considered artifacts and discarded from the analysis.

Feature Extraction

Feature extraction operated at 2 scales in order to characterize both the fine-scale single nucleus features (e.g., nuclear shape, color) and architectural features (e.g., nuclear density, tubule formation). A total of 9 features were individually extracted from each segmented nucleus. Three color-based features were comprised of the mean hue, saturation, and value representations of all pixels within a nucleus. Six shape-based features included perimeter and area measurements computed as the sum of pixels that comprise the nucleus boundary and the sum of pixels enclosed by the boundary, respectively: aspect-ratio, defined as the ratio of the major and minor axis length of an ellipse fit to the nucleus; circularity, a measure of roundness defined as the ratio of area to perimeter, squared; and 2 values that apply concepts from stochastic geometry[29] to quantify nuclear "shape." These consist of a radial contact measure, which computes the radius of the largest circle enclosed by a given nucleus placed randomly within the nucleus (Supplementary Figure 2A), and a line contact measure, which computes the length of the line subtended by the contours of a nucleus when positioned randomly within the nucleus (Supplementary Figure 2B). In addition, the 2 stochastic measures were repeated after performing a morphological closing of the segmented image ($k=17$ pixels), which produced 2 architectural metrics. Unlike the other 9 features, these do not correspond to individual nuclei. The number of iterations performed to generate the architectural distributions was chosen to match the number of nuclei in the ROI so that all feature distributions were composed of the same number of values.

Dimension Reduction

Each case contained a characteristic set of 11 distributions that quantitatively described the morphological properties of the tumor. These distributions were used to train a classifier, but a more concise quantitative representation of the distributions was necessary in order to ensure that the classifier was trained with the appropriate dimensionality. This is important in order to reduce the likelihood of "overfitting" and therefore increase the generalizability of the classifier to other data sets. We transformed each distribution into a small set of second-order statistics by calculating the first 4 moments of the distribution. As a result, each distribution was reduced to its mean, variance, skew, and kurtosis. We tested the ability of this representation to accurately characterize the properties of a distribution by reconstructing each distribution by randomly sampling from a Beta distribution defined by the 4 statistics. We then estimated the difference between a measured distribution and reconstructed distribution by pairing samples in sorted order and computing the mean-square difference.

Classification

The moments derived from each feature distribution were used to train an SVM classifier. This procedure uses a mathematical construct to separate 2 data classes (N0 and N+) in a multidimensional space.[30] Each data point corresponded to a case and each data dimension corresponded to a moment. After normalization along the different data dimensions, a hyperplane that optimally separated the cases within the space was computed. This was

performed separately for each of the 11 features and was repeated using a range of free parameters that govern the behavior of the SVM. These parameters include the box constraint value, which defines the tradeoff between accuracy on the training data set and complexity of the model, and the kernel function, which defines a nonlinear mapping of the data points into an alternate space. More complex kernel functions allow the model to adapt to nuances in the training set but also introduce the possibility of overfitting. We used box constraint values that spanned 10^{-3} to 10^{3} in log steps and kernel functions that included polynomials of orders 1 through 6 and radial basis functions (RBFs) with σ values between 0.2 and 1.0 applied to the normalized space. Therefore, 77 combinations of box constraint and kernel functions were possible for each model, and we applied each of these to the data sets containing the first 2, 3, or 4 moments, resulting in a total of 231 classifiers for each feature.

Often, models with different parameters produced identical results. To maximize computational efficiency, we discarded classifiers that exhibited high error rates (>0.45) using leave-one-out cross validation, as well as those that generated redundant predictions. When ≥2 models generated identical predictions on our data set, we elected to exclude those built with the more complex kernel functions (i.e., highest polynomial order), box constraint values that deviated the most from 1, and those with the largest number of moments. This reduced the number of classifiers generated for each feature, improving the computational efficiency of the next stage.

The above procedure allowed us to experimentally derive a set of weak classifiers that could be combined to generate a unified prediction of metastasis status. Therefore, we developed a second stage that weighted each weak classifier, c, by a factor of w and summed the weighted binary predictions from leave-one-out cross-validation, P, to produce a scalar metastasis score between 0 and 1 for each case, n:

$$score_n = \sum_c w_c P_{cn}.$$

Weights were initially uniform and iteratively adjusted (100,000 iterations) by increasing the weights of the classifiers that made correct predictions on cases in which the metastasis score was most different from the ground truth, y:

$$w_{c,i+1} = w_{c,i}\left(1 + \sum_n s_{cn}\,|\,score_n - y_n\,|\right)$$

$$s_{cn} = \left\{ \begin{array}{l} 1,\ for\ P_{cn} = y_n \\ -1,\ for\ P_{cn} \neq y_n \end{array} \right\}.$$

Weights were normalized after each iteration to ensure that the possible range of metastasis scores was confined to the interval between 0 and 1.

Validation

The prediction performance of the classifier was measured using leave-one-out cross-validation. In this procedure a single case is held out of the complete data set to test a classifier trained by the remaining cases. This procedure is repeated by holding out a new case during each iteration until all cases in the data set have served as test data. The result of this analysis is a prediction set the same size as the complete data set.

Results

We analyzed high-resolution (0.5 µm/pixel) whole-slide images from 101 primary breast carcinoma specimens stained with a complete prognostic panel. Each specimen was associated with a known metastasis status determined by examination of the lymph nodes at the time of surgical resection of the primary tumor. We used this ground truth to train and validate an algorithm to predict metastasis status from histological images.

Feature Extraction and Representation

The shape and color of nuclei,[20,22,26,31-36] as well as their spatial distributions within a tumor,[27,37] have been shown to be associated with metastasis and patient outcomes. We hypothesized that this information could be harnessed to generate quantitative predictions of nodal status. In order to represent the structure of tumors with an informative set of parameters, we extracted a set of features from the segmented nuclei of each specimen. These were comprised of 3 color-based features, 6 shape-based features, and 2 architectural features described in *Methods* (Table I).

Feature analysis of a 6000×6000 pixel region of interest (ROI) placed within a histological image yielded distributions of values associated with each nuclear and architectural feature (Figure 1A, black line). This strategy allowed us to quantitatively evaluate tumor attributes within individual feature domains. To reduce the likelihood of overfitting

Table I *Optimal SVM Tuning Parameters*

	Kernel function	Box constraint	No. of moments
Color-based features			
Hue	Polynomial 6	3	4
Saturation	RBF 1.0	3	2
Intensity	Polynomial 1	0	2
Shape-based features			
Area	Polynomial 2	3	4
Perimeter	RBF 0.2	3	2
Circularity	Polynomial 5	–2	2
Aspect ratio	RBF 0.4	1	4
Distance transform	RBF 0.2	2	4
Line sweep	Polynomial 5	2	4
Architectural features			
Distance transform (global)	RBF 0.4	2	3
Line sweep (global)	Polynomial 4	2	2

The tuning parameters for the most strongly weighted prediction set for each feature reveal a mixture of properties. Polynomial and radial basis function (RBF) kernels were used, and their corresponding order and standard deviation are listed, respectively. Box constraint values spanned 10^{-3} to 10^{3} and the exponent is listed. The number of moments used often impacted the predictions produced by the classifier; 2 signifies that only the mean and variance of the distributions were used, and 3 signifies that only the mean, variance, and skewness were used.

a model to the training data, we reduced each distribution to a small set of descriptors: the mean, variance, skew, and kurtosis. To demonstrate that these statistics were successful in describing the properties of the distribution, we reconstructed each distribution by resampling from a Beta distri-

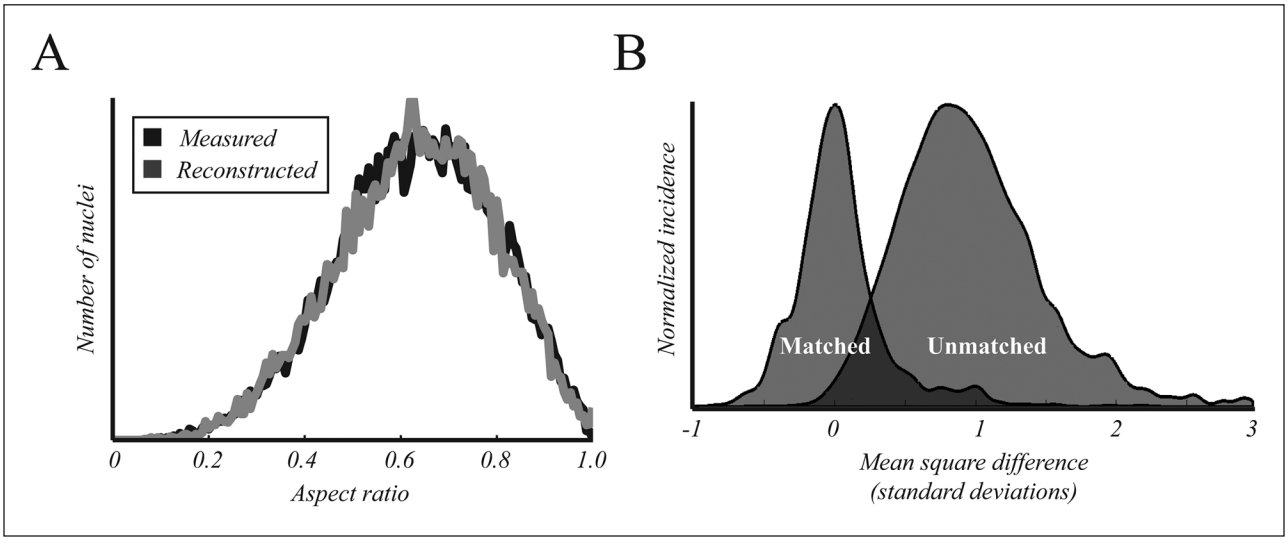


Figure 1 Representation of feature distributions by second-order statistics. (A) A histogram of aspect ratio values from a representative case reveals the shape of the distribution (black line). The mean, variance, skewness, and kurtosis of the distribution were measured and used to generate a histogram from an equal number of random samples drawn from a 4-parameter Beta distribution (gray line). The selected case is a representative example of the agreement between the distributions, as it had the median mean square difference between the measured and reconstructed distributions. (B) The mean square difference between the measured and reconstructed distributions was measured within (matched comparison) and across (unmatched comparison) all cases to assess whether the 4 second-order statistics produced reliable reconstructions. Matched comparisons showed significantly smaller mean square differences than did unmatched comparisons. To account for the difference in the number of matched versus unmatched samples, a kernel density estimator was used and then normalized for the purpose of facilitating a comparison between the 2 distributions. Only moderate overlap between the distributions is observed.

bution defined by all 4 statistics (Figure 1A, gray line) and measured the error between the measured and reconstructed distributions by computing the mean-square difference between them (*matched* comparison). To show that this representation uniquely characterized a given distribution, we also measured the mean-square differences between each recomputed distribution and the distributions of all other cases in the cohort for that feature (*unmatched* comparison). Figure 1B shows that the *matched* mean-square differences were significantly smaller than the *unmatched* mean-square differences (Wilcoxon signed-rank test, $p << 0.01$), demonstrating that the transformation of these distributions to their second-order statistics did not destroy the characteristics that made them different from one another.

Feature-based Predictions

We gathered the 4 distribution statistics for each feature for all cases in our cohort and used these values as the inputs to 11 SVM classifiers. In this way we used leave-one-out cross-validation to produce 11 individual feature-specific predictions of metastasis status (N0 and N+) for each left-out case. Since SVM has been shown to be highly dependent on the tuning parameters provided,[38] we performed this procedure using an exhaustive set of tuning parameters comprised of 11 different SVM kernels (polynomial and radial basis functions with a range of orders and variances, respectively) and 7 box constraint values. This ensured that we considered a broad range of tuning parameters that systematically probed the classification performance of SVM by manipulating the trade-off between model simplicity and accuracy on the training data set. In addition, we performed classification based on the first 2, 3, or 4 statistics of the distributions in order to maximize the available information to the classifier while also avoiding overfitting. Together, this meant that for each feature we generated a total of 231 weak classifiers for every left-out case.

Consistent with previous reports, the predictions generated by this procedure indeed differed depending on the choice of input parameters. We discarded classifiers that produced overall prediction error rates that exceeded 0.45 and found that the remaining number of classifiers that produced unique predictions across the entire data set ranged from 1 to 39 per feature (Figure 2A, inset). Despite the large number of predictions that could be produced merely by adjusting the input parameters, the prediction variability tended to be much smaller within features than across features. The predictions produced by leave-one-out cross-validation were evaluated individually for every weak classifier. We mapped these predictions into a 101-dimension state space, in which each dimension represented a single case in the cohort. Therefore, the position of a point within the space characterized the overall prediction "pattern" produced by a weak classifier, and the distance between any two points in this space represented the difference between the prediction patterns produced by the two classifiers. To demonstrate this visually, we used multidimensional scaling to collapse these points into a two-dimensional plane, preserving their positions relative to one another. In Figure 2A we show that predictions arising from the same features tended to be in close proximity, suggesting that each feature had a characteristic prediction pattern. This is confirmed by applying hierarchical clustering to the nontransformed data, which shows that points within the same feature domains tended to cluster together, forming distinct groupings (Figure 2B). Furthermore, points within the same feature class (architectural, color-based, and nuclear shape-based) exhibited very little overlap with other feature classes. For example, although architectural features measured using radial contact overlap with those measured using the line contact procedure (Figure 2A, triangles), they do not overlap with color-based (Figure 2A, circles) or nuclear shape-based prediction patterns (Figure 2A, squares). These results imply that the predictive information from each of the 3 feature classes is not redundant and demonstrates the importance of incorporating information from all 3 feature classes into a unified prediction in which each feature complements the others.

Metastasis Score

To form a unified prediction from the predictions of the constituent features, we incorporated a second stage that performed a weighted sum of the predictions produced by the weak classifiers individually. This sum represented a "metastasis score" which can be used to assess the likelihood that a tumor has metastasized to the axillary lymph nodes. Scores ranged from 0 to 1, where values close to 0 indicated strong lymph node–negative predictions, values close to 1 indicated strong

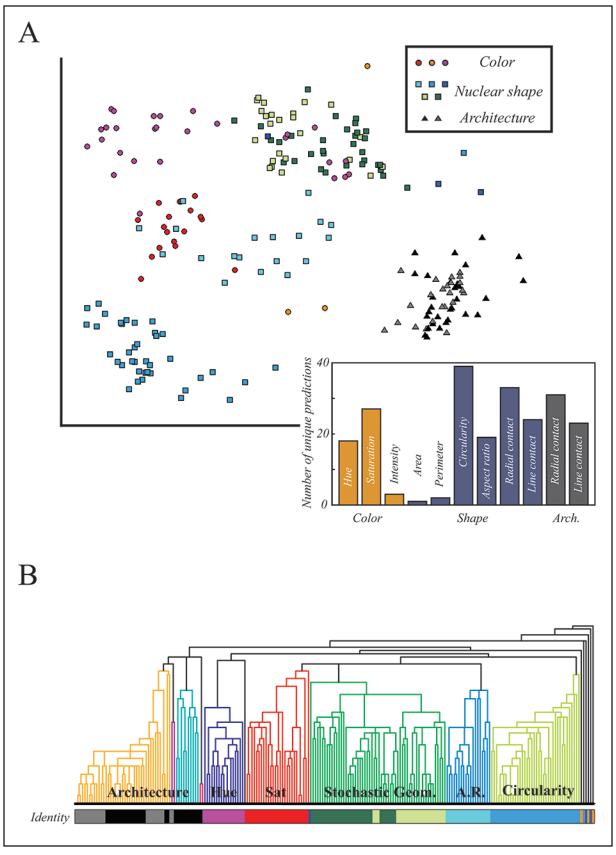


Figure 2 Similarity of feature predictions. (A) Multidimensional scaling was applied to the 101 predictions generated by each SVM classifier applied to each feature using leave-one-out cross-validation. Only classifiers that produced an overall success rate >0.55 were included. Each data point represents a 101-case set of predictions collapsed onto the two-dimensional plane shown. The distance between points represents the similarity of their predictions. The number of unique points per feature is shown in the inset. (B) Hierarchical clustering demonstrates that the most similar prediction sets typically are derived from the same features. The dendrogram is colored for clarity, showing major branch points. The identity of each element is indicated below by the color bar, which corresponds to the data point colors in (A). Intensity, area, and perimeter are not labeled and did not form any major clusters (evident on the right hand side of the dendrogram). AR = aspect ratio.

lymph node–positive predictions, and values close to 0.5 indicated prediction uncertainty. In Figure 3 we show the metastasis scores for all 101 cases. Cases that resulted in metastasis scores close to 0 or 1 were usually correct predictions. Importantly, the scalar nature of the output of this algorithm allowed us to derive diagnostic thresholds that assess the appropriateness of this technique for a given case. Here we found that a correct prediction rate of 88.4% can be achieved for scores >0.606 and <0.460, producing an AUC of 0.87. Of those scores 25 of the 29 (86%) node-positive predictions were indeed node-positive, and 36 of the 40 (90%) node-negative predictions were node-negative. Therefore, using these metastasis score thresholds, the technique characterized 68.3% of the cohort with a false negative rate of 10% and a false positive rate of 14%. Although future investigation evaluating larger data sets could help improve the accuracy of this important diagnostic criterion, it is noteworthy that these thresholds are unbalanced around the neutral value of 0.5, emphasizing the algorithm's sensitivity of identifying tumors as node-negative.

Feature Weights

To produce metastasis scores, a two-stage learning algorithm was adopted in which SVM predictions were individually computed for unique features and tuning parameters, and the predictions that the classifiers produced were linearly combined using a set of weights that were determined from a secondary training stage (Figure 4). The weights derived from the second stage can be interpreted as the contribution of each weak classifier to the final metastasis score. We found that all 3 feature classes strongly contributed to the unified score, with the shape-based features contributing the most strongly (Figure 5A). Notably, the circularity measure, which was among the most nonoverlap-

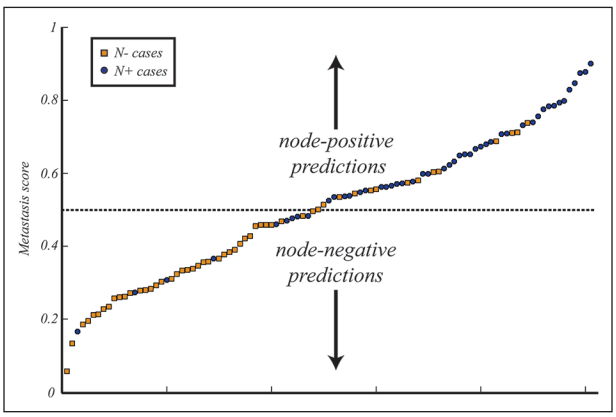


Figure 3 Metastasis scores. Final metastasis scores for all 101 cases using leave-one-out cross-validation are shown in sorted order. Orange points represent true node-negative cases, while blue points represent true node-positive cases. A robust segregation is observed at the two ends of the score, with a mixture of node-positive and node-negative cases near 0.5.

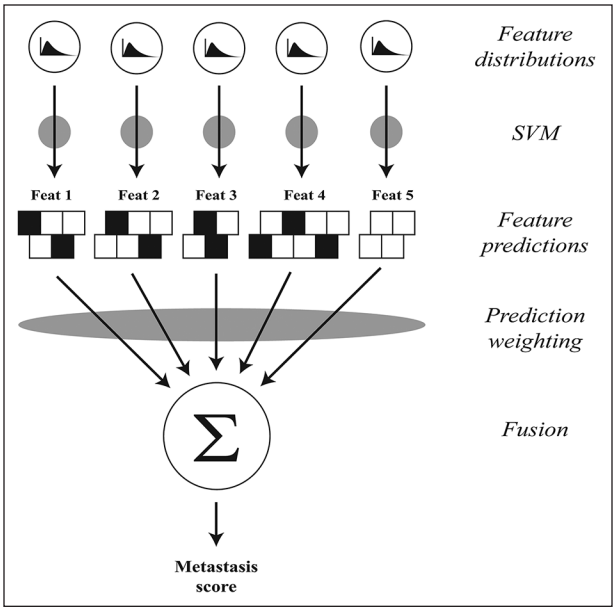


Figure 4 Metastasis score generation from histologic images. Feature distributions are formed from the analysis of nuclear and architectural properties. Four representative statistics are extracted from these distributions and used to form individual metastasis predictions within segregated feature domains via classifiers trained using SVM. Each feature potentially generates multiple predictions depending on how many of the 4 statistics are used and the SVM tuning properties used to form the classifier. All predictions are then multiplied by trained weights and linearly combined to produce a scalar metastasis score.

ping of the predictors (Figure 2), was the most strongly weighted single feature. Likewise, the architectural radial contact measure was strongly weighted, considerably more so than the architectural line contact measure. Given the substantial prediction overlap between these two features, it is likely that the architectural line contact measure would have been much more strongly weighted in the absence of architectural radial contact. The area measure did not contribute to the metastasis score.

Generally, very few weak classifiers within each feature domain were weighted strongly, consistent with the notion that predictions produced within the same feature domain were largely redundant. When we reran the procedure by utilizing only the most strongly weighted tuning parameters within each feature domain, the results were not appreciably different (data not shown). In Figure 5B the strongest weight from each feature domain was selected, and the overall contributions from each feature appear similar to those in Figure 5A, which depicts the sum of weights. The predictive success of SVM using the classifiers with the strongest weights is shown in Figure 5C. It is evident that the linear combination of weighted SVM predictions improved performance beyond that of any single feature domain, emphasizing the cooperative nature of the second stage of the algorithm.

The most strongly weighted tuning parameters were generally those with box constraint values much greater than 1 and nonlinear kernel functions (Table I), indicating that the optimal SVM models were formed by tailoring the model to achieve high

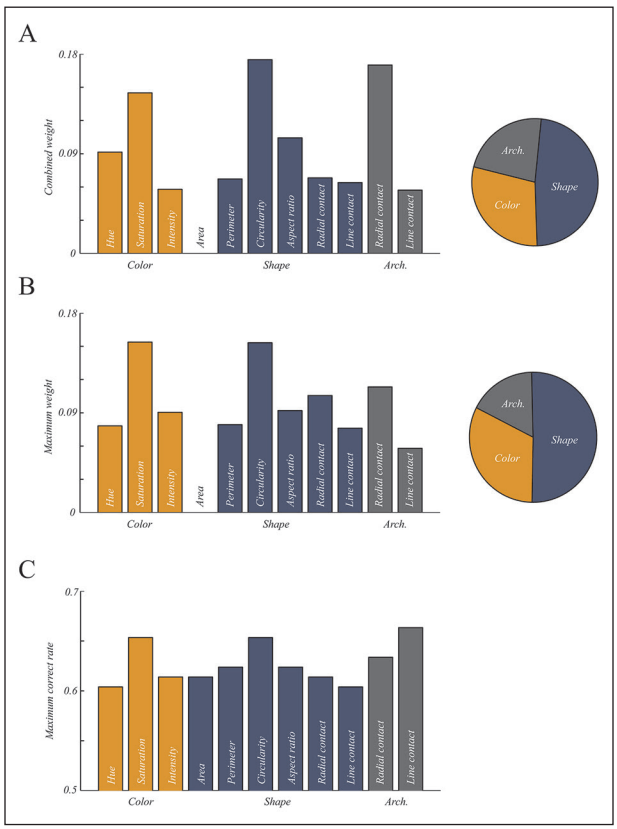


Figure 5 Feature contributions. (A) Contributions of each feature to the metastasis score by summing the weights assigned to all SVM classifiers within each feature dimension. The pie chart illustrates the distribution of shape, color, and architectural weights. (B) The strength of the maximum weight within each feature dimension as a proportion of the total set of maximum weights. The weights in (B) are not markedly different from those in (A). (C) The correct rate associated with the most strongly weighted classifier within each feature domain. A comparison of (B) and (C) reveals that the correct rate alone did not define the strength of weighting.

training data accuracy. Evaluation of the second stage weights also revealed that the addition of the skewness and kurtosis statistics to describe feature distributions often improved classification performance (Table I). Interestingly, addition of the skewness statistic without kurtosis was not usually associated with higher weights. Further study is needed to determine whether this implies that skewness offers little additional information beyond what the other statistics provide and whether further improvements could be achieved by eliminating skewness altogether.

Metastasis Score Correlation with Traditional Prognostic Variables

Although the technique described exhibits a high degree of precision for the most confident predictions, approximately one-third of the cohort generates metastasis scores that are close to the neutral value of 0.5 and that perform near chance as a group (53% correct rate, area under ROC curve: 0.48). In order to understand the limitations of the metastasis score, we asked whether some tumor attributes were correlated with low confidence predictions to help guide the development of exclusion criteria. We evaluated the expression of ER, PR, Ki-67, Her2, p53, and BCL-2 using immunohistochemistry. We found that the *uncertain* prediction group had a greater incidence of hormone receptor–positive cases and a lower incidence of p53-positive cases than did the node-positive and node-negative prediction groups (Table II). Expression of Her2 and BCL-2 exhibited a graded response in which the *uncertain* prediction group had an incidence of expression between that of node-positive and node-negative prediction groups. Surprisingly, Ki-67 expression did not appear to correlate with metastasis score.

Table II *Immunohistochemical Markers Associated with Metastasis Score*

	Prediction type		
Total	**Node-negative 40**	**Uncertain 32**	**Node-positive 29**
ER/PR	40	32	29
positive (%)	*55.0*	*65.6*	*58.6*
Her2	40	32	28
3+ (%)	*7.5*	*12.5*	*17.9*
Ki-67	40	32	29
positive (%)	*85.0*	*84.4*	*86.2*
p53	40	32	28
positive (%)	*35.0*	*25.0*	*46.4*
BCL-2	30	21	22
positive (%)	*43.3*	*38.1*	*22.7*

The 3 prediction groups are as defined as in Figure 6. The percentages of positivity showed a marked dependence on metastasis score. The uncertain group, which consisted of more node-positive than node-negative cases (19 vs. 13), often had immunohistochemical profiles that were distinct from the groups with more confident predictions.

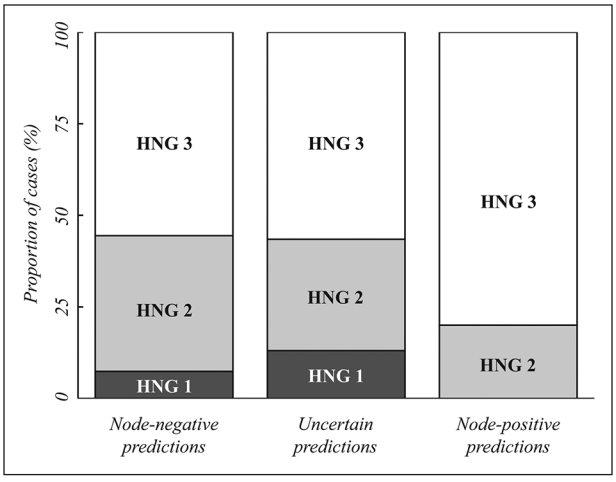


Figure 6 Histologic Nottingham grades are correlated with prediction. Cases were separated into 3 groups based on their metastasis scores; node-negative cases had scores <0.461 (N=40), node-positive cases had scores >0.606 (N=29), and uncertain cases had scores between those values (N=32). Node-positive scores consisted of a higher proportion of HNG 3 cases.

The node-positive prediction group was also associated with a greater number of histologic Nottingham grade (HNG) 3 tumors than was the node-negative prediction group and *uncertain* prediction group, which contained similar HNG distributions (Figure 6). This finding is consistent with previously reported relationships between histologic grade and metastasis[39] but highlights that the *uncertain* prediction group more closely resembles the node-negative prediction group in this regard despite consisting of more node-positive tumors than node-negative tumors (19 of 32). Indeed, even the node-positive cases within the *uncertain* prediction group exhibited a prevalence of HNG 3 similar to that of the node-negative group (57% vs. 56%). Node-positive cases of HNG 1 were always misclassified or fell within the *uncertain* group in our sample.

Discussion

Microscopic evaluation of histopathology has his-

torically been the cornerstone of pathology diagnosis and staging, but the accuracy and reproducibility of this approach can suffer due to its subjective nature.[40,41] Several investigators have attempted to overcome this shortcoming by developing image processing algorithms to produce objective measures from digital histological images. Gurcan et al[42] present a review of image analysis methods for histopathological analysis. Here we have shown that axillary lymph node metastasis status can be accurately predicted using a completely automated image analysis procedure which analyzes routinely acquired histological images of primary tumor samples from definitive surgical treatment at the scale of single nuclei. Using geometric principles and chromatic transformations, the attributes of cell nuclei and their organization within the tumor were quantitatively evaluated, and these properties were used to train a classifier that identified the factors that contribute to the prediction of ALN status. As a result, a metastasis score was generated that was shown to be predictive of a patient's metastasis status.

Of 101 cases 69 had metastasis scores that deviated enough from the neutral value of 0.5 to be characterized as "certain" predictions, together achieving a success rate of 88.4% and AUC of 0.87. This result compares favorably to previous attempts to predict ALN metastasis from clinical, pathological, and histological data.[3,4,7-13,15,43-45] Of the 29 high-scoring node-positive predictions, 4 were false positives. It is noteworthy that the ground truths for all 4 cases were not based on complete axillary dissection but rather an examination of a relatively small number of lymph nodes identified as sentinel nodes (3, 3, 4, and 5—significantly less than the median of 20 from this group). Furthermore, these cases did not undergo complete sectioning and histologic examination of the sentinel lymph nodes, which have been shown to harbor metastases in 11.4% of SLNs initially thought to be negative.[46] Therefore, it remains possible that metastasis may have been present but undetected in these cases. Cases that receive high metastasis scores could potentially be used by both surgeons and pathologists to flag cases that warrant greater scrutiny.

Image Feature Analysis

Many of the morphological features analyzed in this study have been shown previously to have prognostic significance. For example, nuclear aspect ratio and circularity have been shown to be associated with patient outcome.[20] Likewise, in a previous study by our group we showed that the stochastic metrics employed here were capable of predicting histologic grade.[25] In addition, area and the standard deviation of area have been shown to contain prognostic information[21,23,24,26] and can improve the prediction of patient outcomes over ALN status alone.[22] Although our metastasis score ultimately did not utilize nuclear area, it is likely that this feature would have contributed in the absence of the line contact metric, which showed complete overlap with the area prediction. If trained with a larger cohort, it is possible that there are nuances in a subset of tumors that may be captured by area but not by line contact. Application of this technique on additional data sets may reveal an improved set of weights that utilize all of the image features tested.

Comparison to Previous Models

Several models have been developed to predict ALN status, primarily based on primary tumor characteristics, SLN attributes, and clinical data. The Memorial Sloan-Kettering Cancer Center (MSKCC) nomogram[13] is among the most reviewed and relies on 9 variables to produce likelihoods of ALN metastasis. It has been shown to exhibit false negative rates of 4.1%,[15] 14%,[47] and 19%[48] for the most confident N0 predictions. However, the most confident predictions in these studies were present in only 2.6%, 12%, and 25.6% of patients, respectively. This demonstrates that in order to expand the applicability of this technique to larger patient populations, a higher false negative rate must be accepted. Furthermore, this technique has been associated with high false positive rates (62%[47]), implying that it is only appropriate for the purposes of ALN surgery exclusion and not to identify high-risk patients.

In this study we demonstrated a high degree of prediction accuracy (AUC=0.87) in 68.3% of the patient cohort, achieved by utilizing morphological information. This technique was accurate for flagging high-risk cases, while retaining high low-risk specificity. In comparison to the MSKCC nomogram, our model maintained high node-negative prediction accuracy for 39.6% of the 101 cases in the cohort, with a false negative rate of only 10%. These results suggest that morphological information contains useful information for the prediction of ALN metastasis that may have application

broader than that for clinicopathological data. Future investigation is warranted to reveal whether automated histological image analysis, when used in conjunction with clinicopathological data analysis, can provide more informative predictions than either technique alone.

Incorporation of Histological Image Analysis in Patient Care

Metastasis prediction from primary tumor histology adds a valuable tool to breast cancer patient management, but its role within the overall context of patient care remains unclear. Inclusion of primary tumor attributes towards ALN evaluation could reduce the morbidity and costs associated with additional surgeries by avoiding ALN dissection in patients with metastasis scores that meet the appropriate diagnostic criteria. Likewise, strong N+ predictions could eliminate the need to perform SLN biopsy altogether.

The ability to apply this prediction algorithm preoperatively to diagnostic core needle biopsies could aid in surgical planning and help guide neoadjuvant treatment before examination of the primary tumor is performed. In this study we analyzed only complete surgical resections. Validation of this technique with tissue acquired from core needle biopsies is necessary to confirm that this method could in fact be applied preoperatively. Despite core needle biopsy producing less tissue for analysis in comparison to excised specimens, it is often an accurate predictor of primary tumor characteristics, in one study exhibiting high concordance of histologic type (100%), hormone-receptor status (89%), and Her2 expression (92%)[49]—all of which are important variables for developing personalized treatment strategies. We suggest that incorporating image analysis of core needle biopsy material may provide a powerful and cost effective quantitative tool to assess the likelihood of axillary lymph node metastasis prior to surgical intervention.

References

1. Engel J, Kerr J, Schlesinger-Raab A, Sauer H, Holzel D: Axilla surgery severely affects quality of life: Results of a 5-year prospective study in breast cancer patients. Breast Cancer Res Treat 2003;79:47-57
2. Giuliano AE, Haigh PI, Brennan MB, Hansen NM, Kelley MC, Ye W, Glass EC, Turner RR: Prospective observational study of sentinel lymphadenectomy without further axillary dissection in patients with sentinel node-negative breast cancer. J Clin Oncol 2000;18:2553-2559
3. Changsri C, Prakash S, Sandweiss L, Bose S: Prediction of additional axillary metastasis of breast cancer following sentinel lymph node surgery. Breast J 2004;10:392-397
4. Abdessalam SF, Zervos EE, Prasad M, Farrar WB, Yee LD, Walker MJ, Carson WB, Burak WE Jr: Predictors of positive axillary lymph nodes after sentinel lymph node biopsy in breast cancer. Am J Surg 2001;182:316-320
5. Weiser MR, Montgomery LL, Tan LK, Susnik B, Leung DY, Borgen PI, Cody HS 3rd: Lymphovascular invasion enhances the prediction of non-sentinel node metastases in breast cancer patients with positive sentinel nodes. Ann Surg Oncol 2001;8:145-149
6. Farshid G, Pradhan M, Kollias J, PG Gill: A decision aid for predicting non-sentinel node involvement in women with breast cancer and at least one positive sentinel node. Breast 13:494-501
7. Meretoja TJ, Leidenius MH, Heikkila PS, Boross G, Sejben I, Regitnig P, Luschin-Ebengreuth G, Zgajnar J, Perhavec A, Gazic B, Lazar G, Takacs T, Voros A, Saidan ZA, Nadeem RM, Castellano I, Sapino A, Bianchi S, Vezzosi V, Barranger E, Lousquy R, Arisio R, Foschini MP, Imoto S, Kamma H, Tvedskov TF, Kroman N, Jensen MB, Audisio RA, Cserni G: International multicenter tool to predict the risk of non-sentinel node metastases in breast cancer. J Natl Cancer Inst 2012;104:1888-1896
8. De Laurentiis M, Gallo C, De Placido S, Perrone F, Pettinato G, Petrella G, Carlomagno C, Panico L, Delrio P, Bianco AR: A predictive index of axillary nodal involvement in operable breast cancer. Br J Cancer 1996;73:1241-1247
9. Olivotto IA, Jackson JS, Mates D, Andersen S, Davidson W, Bryce CJ, Ragaz J: Prediction of axillary lymph node involvement of women with invasive breast carcinoma: A multivariate analysis. Cancer 1998;83:948-955
10. Anan K, Mitsuyama S, Tamae K, Nishihara K, Iwashita T, Abe Y, Ihara T, Toyoshima S: Axillary lymph node metastases in patients with small carcinomas of the breast: Is accurate prediction possible? Eur J Surg 2000;166:610-615
11. Bevilacqua J, Cody H 3rd, MacDonald KA, Tan LK, Borgen PI, Van Zee KJ: A prospective validated model for predicting axillary node metastases based on 2,000 sentinel node procedures: The role of tumour location [corrected]. Eur J Surg Oncol 2002;28:490-500
12. Takada M, Sugimoto M, Naito Y, Moon HG, Han W, Noh DY, Kondo M, Kuroi K, Sasano H, Inamoto T, Tomita M, Toi M: Prediction of axillary lymph node metastasis in primary breast cancer patients using a decision tree-based model. BMC Med Inform Decis Mak 2012;12:54
13. Van Zee KJ, Manasseh DM, Bevilacqua JL, Boolbol SK, Fey JV, Tan LK, Borgen PI, Cody HS 3rd, Kattan MW: A nomogram for predicting the likelihood of additional nodal metastases in breast cancer patients with a positive sentinel node biopsy. Ann Surg Oncol 2003;10:1140-1151
14. Reyal F, Belichard C, Rouzier R, de Gournay E, Senechal C, Bidard FC, Pierga JY, Cottu P, Lerebours F, Kirova Y, Feron JG, Fourchotte V, Vincent-Salomon A, Guinebretiere JM, Sigal-Zafrani B, Sastre-Garau X, De Rycke Y, Coutant C: Non-sentinel lymph node metastasis prediction in breast cancer with metastatic sentinel lymph node: Impact of molecular subtypes classification. PLoS One 2012;7:e47390
15. Hessman CJ, Naik AM, Kearney NM, Jensen AJ, Diggs BS, Troxell ML, Vetto JT: Comparative validation of online nomograms for predicting nonsentinel lymph node status

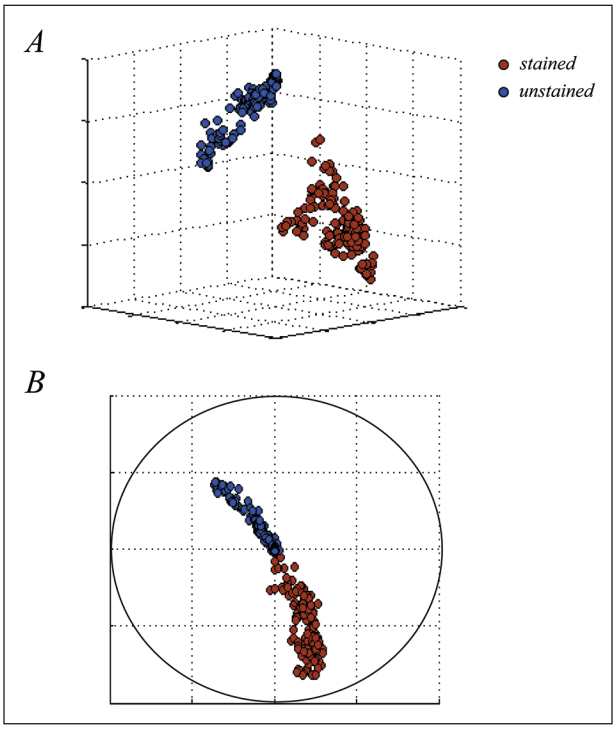
in sentinel lymph node-positive breast cancer. Arch Surg 2011;146:1035-1040

16. Moghaddam Y, Falzon M, Fulford L, Williams NR, Keshtgar MR: Comparison of three mathematical models for predicting the risk of additional axillary nodal metastases after positive sentinel lymph node biopsy in early breast cancer. Br J Surg 2010;97:1646-1652
17. Scow JS, Degnim AC, Hoskin TL, Reynolds C, Boughey JC: Assessment of the performance of the Stanford Online Calculator for the prediction of nonsentinel lymph node metastasis in sentinel lymph node-positive breast cancer patients. Cancer 2009;115:4064-4070
18. Lyman GH, Giuliano AE, Somerfield MR, Benson AB 3rd, Bodurka DC, Burstein HJ, Cochran AJ, Cody HS 3rd, Edge SB, Galper S, Hayman JA, Kim TY, Perkins CL, Podoloff DA, Sivasubramaniam VH, Turner RR, Wahl R, Weaver DL, Wolff AC, Winer EP; American Society of Clinical Oncolory: American Society of Clinical Oncology guideline recommendations for sentinel lymph node biopsy in early-stage breast cancer. J Clin Oncol 2005;23:7703-7720
19. Fournier K, Schiller A, Perry RR, Laronga C: Micrometastasis in the sentinel lymph node of breast cancer does not mandate completion axillary dissection. Ann Surg 2004;239: 859-863; discussion 863-865
20. Baak JP, Kurver PH, De Snoo-Niewlaat AJ, De Graef S, Makkink B, Boon ME: Prognostic indicators in breast cancer--morphometric methods. Histopathology 1982;6:327-339
21. Aaltomaa S, Lipponen P, Eskelinen M, Kosma VM, Marin S, Alhava E, Syrjanen K: The significance of nuclear morphometric variables as prognostic predictors in breast cancer. Anticancer Res 1991;11:1663-1669
22. Baak JP, Van Dop H, Kurver PH, Hermans J: The value of morphometry to classic prognosticators in breast cancer. Cancer 1985;56:374-382
23. Kuenen-Boumeester V, Hop WC, Blonk DI, Boon ME: Prognostic scoring using cytomorphometry and lymph node status of patients with breast carcinoma. Eur J Cancer Clin Oncol 1984;20:337-345
24. Larsimont D, Kiss R, d'Olne D, de Launoit Y, Mattheiem W, Paridaens R, Pasteels JL, Gompel C: Correlation between nuclear cytomorphometric parameters and estrogen receptor levels in breast cancer. Cancer 1989;63:2162-2168
25. Petushi S, Zhang J, Milutinovic A, Breen DE, Garcia FU: Image-based histologic grade estimation using stochastic geometry analysis. Proc. SPIE 7963, Medical Imaging 2011: Computer-Aided Diagnosis, 79633E-79645E
26. Pienta KJ, Coffey DS: Correlation of nuclear morphometry with progression of breast cancer. Cancer 1991;68:2012-2016
27. Beck AH, Sangoi AR, Leung S, Marinelli RJ, Nielsen TO, van de Vijver MJ, West RB, van de Rijn M, Koller D: Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. Sci Transl Med 2011;3:108ra113
28. Vincent L, Soille P: Watersheds in digital spaces: An efficient algorithm based on immersion simulations. IEEE Trans Pattern Anal Mach Intell 1991;13:583-598
29. Stoyan D, Kendall WS, Mecke J: Stochastic geometry and its applications. *In* Wiley Series in Probability and Statistics: Applied Probability and Statistics. Second edition. 1995, Chichester, New York, Wiley, xix, p 436
30. Vapnik VN: An overview of statistical learning theory. IEEE Trans Neural Netw 1999;10:988-999
31. Ladekarl M: Quantitative histopathology in ductal carcinoma of the breast: Prognostic value of mean nuclear size and mitotic counts. Cancer 1995;75:2114-2122
32. Aubele M, Auer G, Voss A, Falkmer U, Rutquist LE, Höfler H: Different risk groups in node-negative breast cancer: Prognostic value of cytophotometrically assessed DNA, morphometry and texture. Int J Cancer 1995;63:7-12
33. Eskelinen M, Lipponen P, Aaltomaa S, Kosma VM, Syrjänen K, Alhava E: Breast cancer in young women: Clinical, histological and morphometric prognostic factors. Anticancer Res 1992;12:1237-1242
34. Umbricht C, Oberholzer M, Gschwind R, Christen H, Torhorst J: Prognostic significance (relapse, non-relapse) of nuclear shape parameters in lymph node negative breast cancer. Anal Cell Pathol 1989;1:11-23
35. Van der Linden HC, Baak JP, Smeulders AW, Lindeman J, Meyer CJ: Morphometry of breast cancer: I. Comparison of the primary tumours and the axillary lymph node metastases. Pathol Res Pract 1986;181:236-242
36. van der Linden HC, Baak JP, Lindeman J, Hermans J, Meyer CJ: Morphometry and breast cancer: II. Characterisation of breast cancer cells with high malignant potential in patients with spread to lymph nodes: Preliminary results. J Clin Pathol 1986;39:603-609
37. Fulford LG, Easton DF, Reis-Filho JS, Sofronis A, Gillett CE, Lakhani SR, Hanby A: Specific morphological features predictive for the basal phenotype in grade 3 invasive ductal carcinoma of breast. Histopathology 2006;49:22-34
38. Chapelle O, Vapnik V, Bousquet O, Mukherjee S: Choosing multiple parameters for support vector machines. Mach Learn 2002;46:131-159
39. Klar M, Foeldi M, Markert S, Gitsch G, Stickeler E, Watermann D: Good prediction of the likelihood for sentinel lymph node metastasis by using the MSKCC nomogram in a German breast cancer population. Ann Surg Oncol 2009; 16:1136-1142
40. Boiesen P, Bendahl PO, Anagnostaki L, Domanski H, Holm E, Idvall I, Johansson S, Ljungberg O, Ringberg A, Ostberg G, Fernö M: Histologic grading in breast cancer--reproducibility between seven pathologic departments. South Sweden Breast Cancer Group. Acta Oncol 2000;39: 41-45
41. Dalton LW, Pinder SE, Elston CE, Ellis IO, Page DL, Dupont WD, Blamey RW: Histologic grading of breast cancer: Linkage of patient outcome with level of pathologist agreement. Mod Pathol 2000;13:730-735
42. Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, Yener B: Histopathological image analysis: A review. IEEE Rev Biomed Eng 2009;2:147-171
43. Noguchi M, Ohta N, Thomas M, Kitagawa H, Earashi M, Miyazaki I, Mizukami Y: A retrospective study on the clinical and biological prediction of axillary lymph node metastasis in breast cancer. Surg Today 1993;23:573-579
44. Noguchi M, Ohta N, Thomas M, Kitagawa H, Earashi M, Miyazaki I, Mizukami Y: Clinical and biological prediction of axillary and internal mammary lymph node metastases in breast cancer. Surg Oncol 1993;2:51-58
45. Guarnieri A, Neri A, Correale PP, Lottini M, Testa M, Mariani

F, Tucci E, Megha T, Cintorino M, Carli A: Prediction of lymph node status by analysis of prognostic factors and possible indications for elective axillary dissection in T1 breast cancers. Eur J Surg 2001;167:255-259

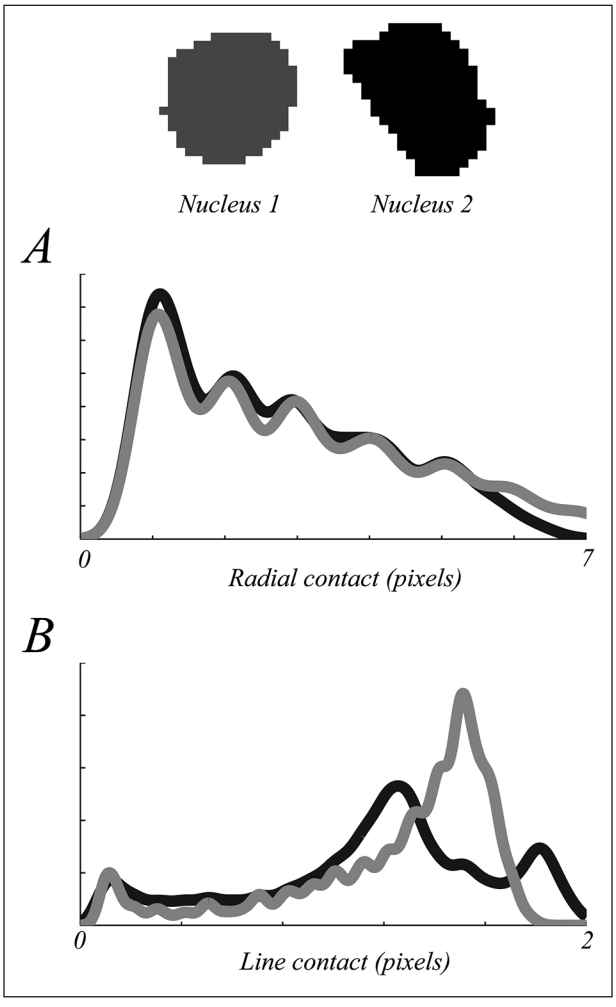
46. Weaver DL, Le UP, Dupuis SL, Weaver KA, Harlow SP, Ashikaga T, Krag DN: Metastasis detection in sentinel lymph nodes: Comparison of a limited widely spaced (NSABP protocol B-32) and a comprehensive narrowly spaced paraffin block sectioning strategy. Am J Surg Pathol 2009;33:1583-1589
47. Berrang TS, Lesperance M, Truong PT, Walter C, Hayashi AH, Olivotto IA: Which prediction models best identify additional axillary disease after a positive sentinel node biopsy for breast cancer? Breast Cancer Res Treat 2012; 133:695-702
48. van den Hoven I, Kuijt GP, Voogd AC, van Beek MW, Roumen RM: Value of Memorial Sloan-Kettering Cancer Center nomogram in clinical decision making for sentinel lymph node-positive breast cancer. Br J Surg 2010;97:1653-1658
49. Greer LT, Rosman M, Mylander WC, Hooke J, Kovatich A, Sawyer K, Buras RR, Shriver CD, Tafra L: Does breast tumor heterogeneity necessitate further immunohistochemical staining on surgical specimens? J Am Coll Surg 2013;216: 239-251

Supplementary Figure 1



Positive staining from HSV pixel values. Positive nuclear staining was defined by manual selection of a set of representative pixels in different regions across 3 exemplary images. Points outside positively stained nuclei were separately selected, which included pixels within unstained nuclei, stroma, and blank regions of the slide. (A) Selected points are shown in HSV cylindrical coordinates, in which the distance from the origin in the horizontal plane represents saturation, the angle in the horizontal plane represents hue, and the vertical axis represents value. Brown points signify stained pixels and blue points signify unstained pixels. The optimal separation between these groups formed a hyperplane defined within all 3 dimensions. (B) Projection of the points to the hue-saturation plane demonstrates the specific hue associated with each group.

Supplementary Figure 2



Stochastic geometry characterization of nuclear shape. Radial contact (A) and line contact (B) distributions were computed for two example nuclei shown in the top panel. Both nuclei had identical areas (99 pixels). Kernel density estimation with $\sigma=0.35$ was performed to generate plots for nucleus 1 (gray lines) and nucleus 2 (black lines). Nucleus 1 produced higher radial contact values and line contact values, as expected from its rounder and less convoluted border.