

## IMAGING

# Systematic Analysis of Breast Cancer Morphology Uncovers Stromal Features Associated with Survival

Andrew H. Beck,<sup>1,2\*</sup> Ankur R. Sangoi,<sup>1,3</sup> Samuel Leung,<sup>4</sup> Robert J. Marinelli,<sup>5</sup> Torsten O. Nielsen,<sup>4</sup> Marc J. van de Vijver,<sup>6</sup> Robert B. West,<sup>1</sup> Matt van de Rijn,<sup>1</sup> Daphne Koller<sup>7†</sup>

The morphological interpretation of histologic sections forms the basis of diagnosis and prognostication for cancer. In the diagnosis of carcinomas, pathologists perform a semiquantitative analysis of a small set of morphological features to determine the cancer's histologic grade. Physicians use histologic grade to inform their assessment of a carcinoma's aggressiveness and a patient's prognosis. Nevertheless, the determination of grade in breast cancer examines only a small set of morphological features of breast cancer epithelial cells, which has been largely unchanged since the 1920s. A comprehensive analysis of automatically quantitated morphological features could identify characteristics of prognostic relevance and provide an accurate and reproducible means for assessing prognosis from microscopic image data. We developed the C-Path (Computational Pathologist) system to measure a rich quantitative feature set from the breast cancer epithelium and stroma (6642 features), including both standard morphometric descriptors of image objects and higher-level contextual, relational, and global image features. These measurements were used to construct a prognostic model. We applied the C-Path system to microscopic images from two independent cohorts of breast cancer patients [from the Netherlands Cancer Institute (NKI) cohort,  $n = 248$ , and the Vancouver General Hospital (VGH) cohort,  $n = 328$ ]. The prognostic model score generated by our system was strongly associated with overall survival in both the NKI and the VGH cohorts (both log-rank  $P \leq 0.001$ ). This association was independent of clinical, pathological, and molecular factors. Three stromal features were significantly associated with survival, and this association was stronger than the association of survival with epithelial characteristics in the model. These findings implicate stromal morphologic structure as a previously unrecognized prognostic determinant for breast cancer.

## INTRODUCTION

In the mid-19th century, it was first appreciated that the process of carcinogenesis produces characteristic morphologic changes in cancer cells (1). Patey and Scarff showed in 1928 (2) that three histologic features—tubule formation, epithelial nuclear atypia, and epithelial mitotic activity—could each be scored qualitatively, and the assessments could be combined to stratify breast cancer patients into three groups that showed significant survival differences. This semiquantitative morphological scoring scheme has been refined over the years (3–5) but still remains the standard technique for histologic grading in invasive breast cancer. Although considerable effort has been devoted recently to molecular profiling for assessment of prognosis and prediction of treatment response in cancer (6, 7), microscopic image assessment is still the most commonly available (and in some places in the world, the only) tool that is financially and logistically feasible.

Although the three epithelial features scored in current grading systems are useful in assessing cancer prognosis, valuable prognostic information can also be derived from other factors, including properties of the cancer stroma such as its molecular characteristics (8–15)

and morphological features [such as stromal fibrotic focus, a scar-like area in the center of a carcinoma (16)]. Thus, we sought to develop a high-accuracy, image-based predictor to identify new clinically predictive morphologic phenotypes of breast cancers, thereby providing new insights into the biological factors driving breast cancer progression.

The development of such a system could also address other problems relevant to the clinical treatment of breast cancer. A limitation to the current grading system is that there is considerable variability in histologic grading among pathologists (17), with potentially negative consequences for determining treatment. An automated system could provide an objective method for predicting patient prognosis directly from image data. Moreover, once established, this system could be used in breast cancer clinical trials to provide an accurate, objective means for assessing breast cancer morphologic characteristics, allowing objective stratification of breast cancer patients on the basis of morphologic criteria and facilitating the discovery of morphologic features associated with response to specific therapeutic agents.

## RESULTS

### Experimental design overview

We developed the Computational Pathologist (C-Path), a machine learning-based method for automatically analyzing cancer images and predicting prognosis. To construct and evaluate the model, we acquired hematoxylin and eosin (H&E)-stained histological images from breast cancer tissue microarrays (TMAs) (figs. S4 and S5). The

<sup>1</sup>Department of Pathology, Stanford University School of Medicine, Stanford, CA 94305, USA. <sup>2</sup>Biomedical Informatics Training Program, Stanford University School of Medicine, Stanford, CA 94305, USA. <sup>3</sup>Department of Pathology, El Camino Hospital, Mountain View, CA 94040, USA. <sup>4</sup>Genetic Pathology Evaluation Centre, University of British Columbia, Vancouver, British Columbia V6H 3Z6, Canada. <sup>5</sup>Department of Biochemistry, Stanford University, Stanford, CA 94305, USA. <sup>6</sup>Department of Pathology, Academic Medical Center, Meibergdreef 9, 1105AZ Amsterdam, Netherlands. <sup>7</sup>Department of Computer Science, Stanford University, Stanford, CA 94305, USA.

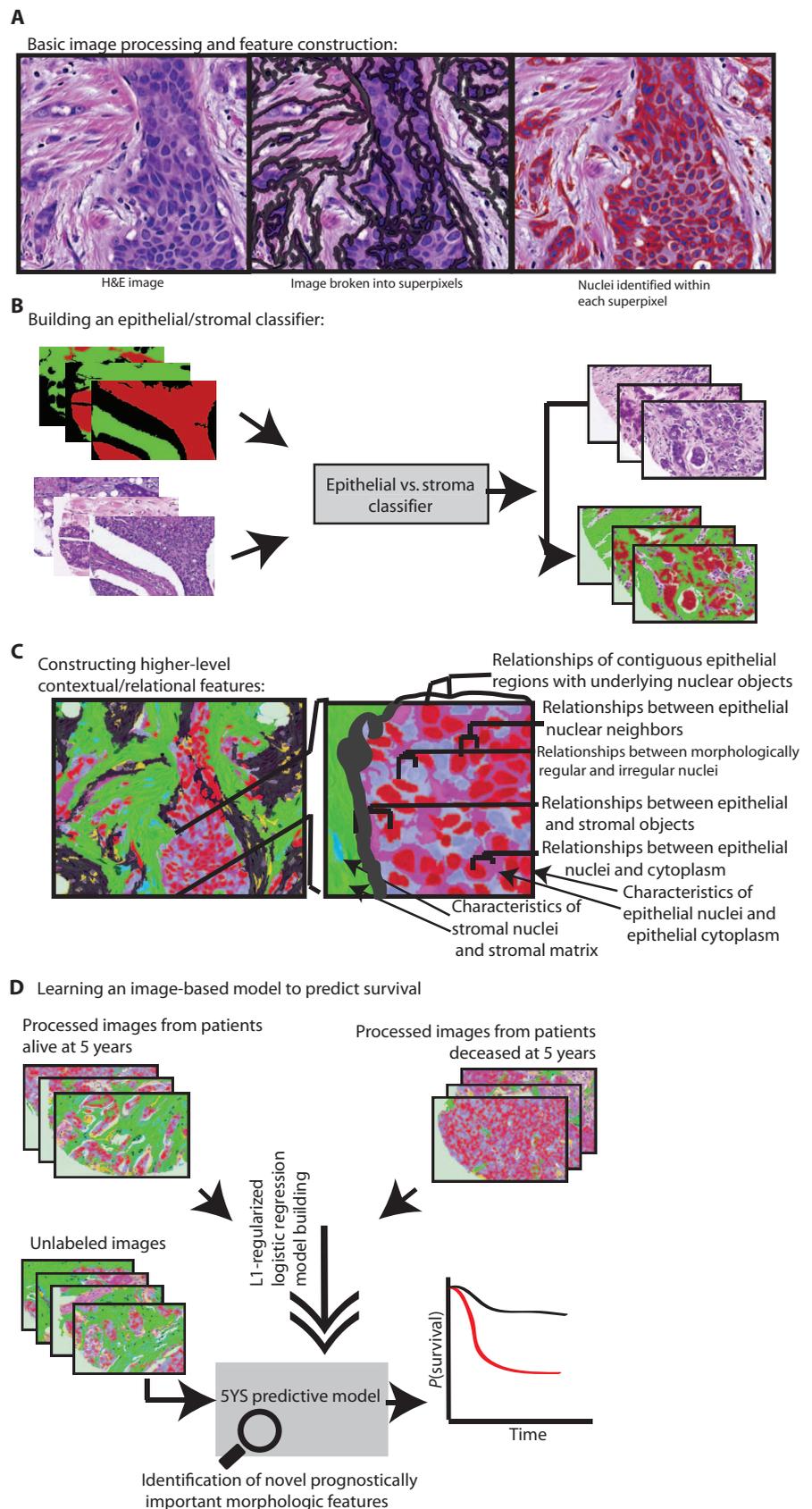
\*Present address: Department of Pathology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02115, USA.

†To whom correspondence should be addressed. E-mail: koller@cs.stanford.edu

**Fig. 1.** Overview of the image processing pipeline and prognostic model building procedure. **(A)** Basic image processing and feature construction. **(B)** Building an epithelial-stromal classifier. The classifier takes as input a set of breast cancer microscopic images that have undergone basic image processing and feature construction and that have had a subset of superpixels hand-labeled by a pathologist as epithelium (red) or stroma (green). The superpixel labels and feature measurements are used as input to a supervised learning algorithm to build an epithelial-stromal classifier. The classifier is then applied to new images to classify superpixels as epithelium or stroma. **(C)** Constructing higher-level contextual/relational features. After application of the epithelial-stromal classifier, all image objects are subclassified and colored on the basis of their tissue region and basic cellular morphologic properties (epithelial regular nuclei = red; epithelial atypical nuclei = pale blue; epithelial cytoplasm = purple; stromal matrix = green; stromal round nuclei = dark green; stromal spindled nuclei = teal blue; unclassified regions = dark gray; spindled nuclei in unclassified regions = yellow; round nuclei in unclassified regions = gray; background = white). (Left panel) After the classification of each image object, a rich feature set is constructed. **(D)** Learning an image-based model to predict survival. Processed images from patients alive at 5 years after surgery and from patients deceased at 5 years after surgery were used to construct an image-based prognostic model. After construction of the model, it was applied to a test set of breast cancer images (not used in model building) to classify patients as high or low risk of death by 5 years.

TMA contains 0.6-mm-diameter cores (median of two cores per case) that represent only a small sample of the full tumor. We acquired data from two separate and independent cohorts: Netherlands Cancer Institute (NKI; 248 patients) and Vancouver General Hospital (VGH; 328 patients).

Unlike previous work in cancer morphometry (18–21), our image analysis pipeline was not limited to a predefined set of morphometric features selected by pathologists. Rather, C-Path measures an extensive, quantitative feature set from the breast cancer epithelium and the stroma (Fig. 1). Our image processing system first performed an automated, hierarchical scene segmentation that generated thousands of measurements, including both standard morphometric descriptors of image objects and higher-level contextual, relational, and global image features. The pipeline consisted of three stages (Fig. 1, A to C, and tables S8 and S9). First, we used a set of processing steps to separate the tissue from the background, partition the image into small regions of coherent appearance known as superpixels, find nuclei within the superpixels, and construct



nuclear and cytoplasmic features within the superpixels (Fig. 1A). Within each superpixel, we measured the intensity, texture, size, and shape of the superpixel and its neighbors. Second, to produce more biologically meaningful features, we classified superpixels as epithelium or stroma (Fig. 1B). We use a machine learning approach ( $L_1$ -regularized logistic regression), in which we hand-labeled superpixels from 158 images (107 NKI and 51 VGH) and used those to train the epithelium/stroma classifier. The resulting classifier comprises 31 features (Supplementary Material, tables S4 and S5, and figs. S6 to S10) and achieves a classification accuracy of 89% on held-out data. To construct our final set of features to be used in the prognostic model, we first recomputed the values of the basic features separately within the epithelium and stroma. We subclassified nuclei as “typical” or “atypical” and obtained object measurements from contiguous epithelial and stromal regions, as well as from epithelial nuclei, epithelial atypical nuclei, epithelial cytoplasm, stromal round nuclei, stromal spindled nuclei, stromal matrix, and unclassified objects. We computed a range of relational features (Fig. 1C) that capture the global structure of the sample and the spatial relationships among its different components, such as mean distance from epithelial nucleus to stromal nucleus, mean distance of atypical epithelial nucleus to typical epithelial nucleus, or distance between stromal regions (Supplementary Material, tables S6 and S7, and figs. S9 and S10). Overall, this resulted in a set of 6642 features per image. For patients with multiple TMA images (208 of 248 NKI patients; 192 of 328 VGH patients), these statistics were summarized as their mean across the images (Supplementary Material).

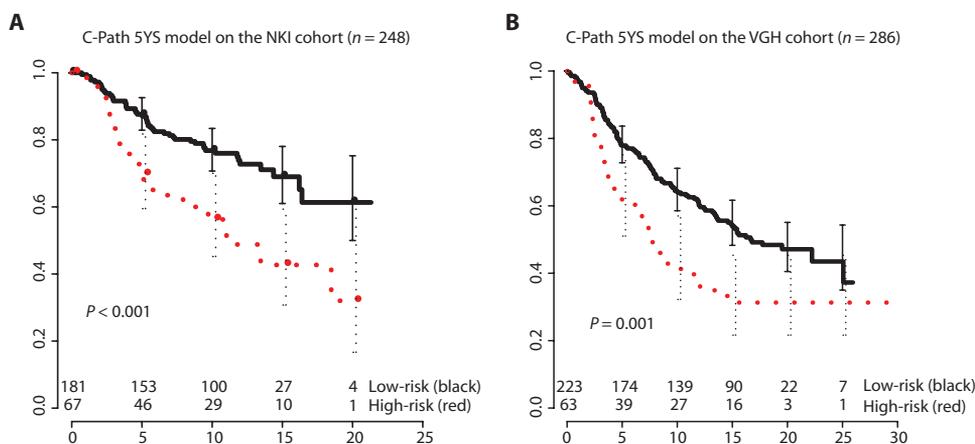
The NKI images were used to build an image feature-based prognostic model to predict the binary outcome of 5-year survival (5YS model) (Fig. 1D) using  $L_1$ -regularized logistic regression, implemented in the R package glmnet (22). Model performance on the NKI data set was assessed by eightfold cross-validation, where the data set is split into

about eight equal folds, and in each fold, the model was built with seven of the folds (up to 217 cases) and evaluated on the held-out fold. For each instance in the data set, we determined the C-Path model result when that instance was held out during cross-validation, which allowed this prediction to be used for evaluating model performance on unseen data. This procedure is known as “prevalidation” (23). To further assess performance of the model, we trained the prognostic model on the full NKI data set and tested the model on the VGH data set (see Materials and Methods and Supplementary Material). We excluded from this analysis the 42 VGH cases that had been used in training the epithelial-stromal classifier.

### Survival analysis on the NKI data set

The prevalidation C-Path 5YS scores were highly associated with overall survival (log-rank  $P < 0.001$ ) (Fig. 2A and table S1). When cases were stratified by their previously assigned grade (see also fig. S1A), the C-Path score was significantly associated with survival within histologic grade 2 tumors (log-rank  $P = 0.004$ ), but not with survival within grade 1 and 3 tumors on the NKI data set (Fig. 3, A to C).

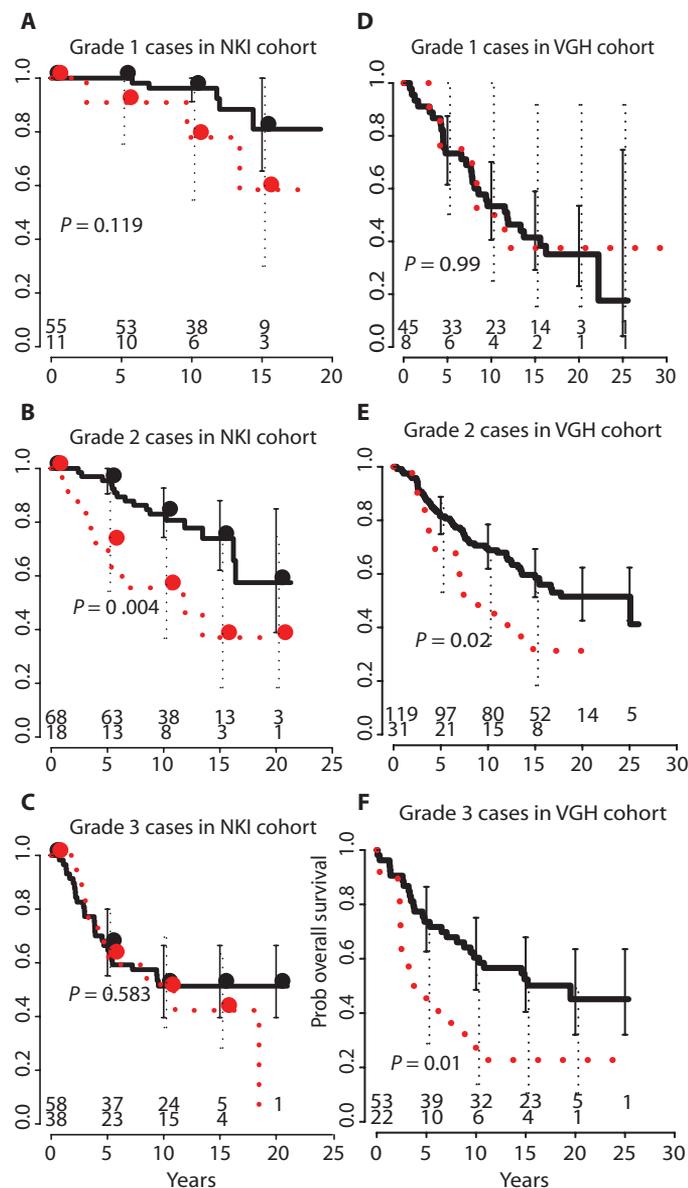
Using a multivariate Cox proportional hazards analysis, we next set out to assess the added prognostic value of the C-Path score in the context of other measured prognostic factors. In addition to standard clinical measurements, the tumors from all patients in the NKI data set had previously undergone expression profiling by microarray, allowing each case to be classified according to several standard breast cancer molecular signatures: the 70-gene prognosis signature score (24), the genomic grade index score (25), the invasiveness gene signature (26), the hypoxia gene signature (27), and intrinsic molecular subtype (28). The subtype classifications used in our analysis come from the original publications or from the supplemental data (29). The prevalidation C-Path scores were significantly associated with 5-year survival independent of any of the other clinical or molecular factors: grade, estrogen receptor (ER) status, age, tumor size, lymph node status, mastectomy, chemotherapy, 70-gene prognosis signature, hypoxia signature, wound response signature, genomic grade index, or intrinsic molecular subtypes ( $P = 0.02$ ) (Table 1A). The only other features significantly associated with survival were the hypoxia signature and age (Table 1A).



**Fig. 2.** Kaplan-Meier survival curves of the 5YS model predictions and overall survival on the NKI and VGH data sets. Cases classified as high risk are plotted on the red dotted line and cases classified as low risk on the black solid line. The error bars represent 95% CIs. The y axis is probability of overall survival, and the x axis is time in years. The numbers of patients at risk in the high- and low-risk groups at 5-year intervals are listed beneath the curves. **(A)** NKI data set. Patients were stratified into low- and high-risk groups based on predictions of the 5YS model on held-out cases during cross-validation. **(B)** VGH data set. VGH patients were stratified into low- and high-risk groups based on predictions of the 5YS model trained on the full NKI data set. In both data sets (A and B), cases predicted to be high risk showed significantly worse overall survival than cases predicted to be low risk (log-rank  $P < 0.001$  in both analyses). VGH cases used to train the epithelial-stromal classifier have been excluded from the analysis.

The previously assigned histologic grading scores used in our comparison came from manual pathologic interpretation of whole-slide microscopic images by a centralized review. To directly compare the performance of the C-Path system to pathological grading on the exact same set of images, we applied standard pathological grading criteria to the TMA images used in the C-Path analysis (mitotic activity, nuclear pleomorphism, and tubule formation were semiquantitatively scored from 1 to 3, and the scores were summed with a sum of less than 6 receiving a grade of 1, sum of 6 to 7 receiving a grade of 2, and sum greater than

7 receiving a grade of 3); the pathologist grading the images was blinded from the survival data. Although the C-Path predictions on the NKI data set were strongly associated with survival, the pathologic grade derived from the same TMA images showed no significant association with survival (log-rank  $P = 0.4$ ), highlighting the difficulty of obtaining accurate prognostic predictions from these small tumor samples.



**Fig. 3.** Kaplan-Meier survival curves of the 5YS model predictions and overall survival on the NKI and VGH data sets, stratified by grade. Cases classified as high risk are plotted on the red dotted line and cases classified as low risk on the black solid line. The error bars represent 95% CIs. The y axis is probability of overall survival, and the x axis is time in years. The numbers of patients at risk in the high- and low-risk groups at 5-year intervals are listed beneath the curves. (A) Grade 1 cases in NKI cohort. (B) Grade 2 cases in NKI cohort. (C) Grade 3 cases in NKI cohort. (D) Grade 1 cases in VGH cohort. (E) Grade 2 cases in VGH cohort. (F) Grade 3 cases in VGH cohort. VGH cases used to train the epithelial-stromal classifier have been excluded from the analysis.

### Survival analysis on the VGH data set

We next tested the prognostic model on the VGH data set, which was not used in constructing the prognostic model. In addition to being an additional data set, the cases from VGH represented a cohort of patients with distinct clinical features. The NKI data set was limited to women younger than 53 years with stage I or II breast cancer. In contrast, the VGH data come from a population-based cohort with a higher proportion of older women and women with more advanced disease. A subset of the VGH cases with survival data (51 images from 42 cases) were used for training of the epithelial-stromal classifier, which was built to classify superpixels as epithelium or stroma and implemented as part of the image processing pipeline. We excluded these 42 cases from our survival analysis.

The C-Path score was significantly associated with overall survival in this independent group of cases (log-rank  $P = 0.001$ ) (Fig. 2B and table S1). Notably, the standard histologic grading scores that had been obtained by routine pathological analysis of whole-slide images with standard grading criteria on the original patient material showed no significant association with survival on this same cohort of patients (fig. S1B; log-rank  $P = 0.29$ ), perhaps due to the greater variability of the grading process, in which grades were assigned independently by individual community pathologists (17). On VGH, significant survival stratification was achieved by the 5YS model within both grade 2 and grade 3 tumors (log-rank  $P = 0.02$  and  $0.01$ , respectively; Fig. 3, D to F). We constructed a multivariate Cox proportional hazards model that considered age, lymph node status, mastectomy, ER status, grade, size, and C-Path 5YS model score. In this multivariate model, the C-Path 5YS model score, age, and lymph node status were independently significantly associated with patient survival (all  $P < 0.05$ ) (Table 1B). Grade, size, and ER status were not significant independent predictors of survival in this multivariate model.

To assess the generalizability of the full image processing pipeline, we have repeated the entire analysis with training of the epithelial-stromal classifier limited exclusively to the 107 NKI images. This pipeline resulted in decreased performance of the prognostic model (fig. S2), with statistically significant (log-rank  $P < 0.05$ ) survival stratification observed only on the NKI data set. These findings suggest that a relatively large, varied set of training images is important for robust performance of the epithelial-stromal classifier and that accurate epithelial-stromal segmentation is important for extracting the most prognostically informative morphological features.

### Assessing significance of features

To identify morphologic features that robustly contribute to the C-Path model, we performed a bootstrap analysis on the NKI data set to generate 95% confidence intervals (CIs) for the coefficient estimates for the image features in the C-Path model. This analysis revealed 11 features with a 95% CI that does not include zero (table S2). These 11 features included 3 stromal features (Fig. 4) and 8 epithelial features (Fig. 5).

We assessed correlation of these features with pathological assessment of epithelial tubule formation, mitotic activity, and nuclear pleomorphism, which are the standard features used in histologic grading. The top associations of the 11 C-Path features with pathological grading features were a negative correlation of tubule formation with “stromal matrix textural variability” (Spearman’s  $\rho = -0.21$ ,  $P = 0.001$ ) and positive correlation of both mitotic activity and nuclear pleo-

**Table 1.** Multivariate Cox proportional hazards model to predict survival in NKI (A) and VGH (B) cohorts.

	Exp(coef)	Exp(-coef)	Lower 95%	Upper 95%	P
A. Multivariate Cox regression (NKI, <i>n</i> = 248)					
C-Path 5YS model score	1.78	0.56	1.11	2.86	0.017
Hypoxia	1.8	0.56	1.09	2.96	0.021
Age	0.79	1.26	0.63	1	0.046
Seventy-gene prognosis signature	2.01	0.5	0.94	4.26	0.070
Size	1.22	0.82	0.94	1.59	0.137
Invasiveness gene signature	1.84	0.54	0.69	4.9	0.226
Mastectomy	1.3	0.77	0.81	2.07	0.276
Wound response signature	1.44	0.69	0.72	2.89	0.299
ERBB2 molecular subtype	1.88	0.53	0.53	6.67	0.326
Grade	1.19	0.84	0.77	1.83	0.431
Basal molecular subtype	0.68	1.47	0.15	3	0.612
ER	0.82	1.22	0.36	1.87	0.634
Lymph node	1.15	0.87	0.55	2.38	0.715
Genomic grade index	1.09	0.91	0.52	2.32	0.814
Luminal A molecular subtype	1.14	0.87	0.36	3.65	0.819
Luminal B molecular subtype	0.93	1.07	0.29	3.03	0.907
Chemotherapy	1.04	0.96	0.49	2.21	0.925
B. Multivariate Cox regression (VGH, <i>n</i> = 286)					
Age	1.83	0.55	1.49	2.24	<0.0001
Lymph node	1.94	0.52	1.32	2.84	0.0007
C-Path 5YS model score	1.54	0.65	1.02	2.32	0.039
ER	0.79	1.27	0.46	1.34	0.376
Mastectomy	0.72	1.39	0.32	1.6	0.417
Size	0.95	1.05	0.85	1.07	0.444
Grade	1.06	0.95	0.74	1.5	0.760

morphism with the C-Path feature “number of epithelial nuclei from unclassified regions” (Spearman’s rho = 0.27 and 0.33, respectively, both  $P < 0.001$ ).

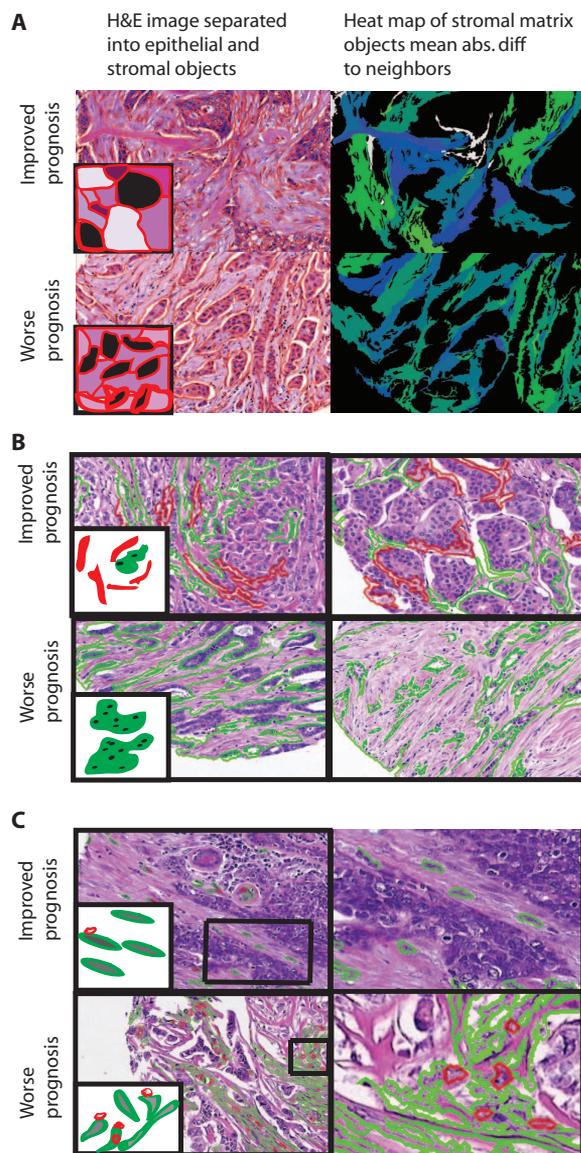
Seven of the top features in the bootstrap analysis were relational features characterizing the contextual relationships of epithelial and stromal objects to their neighbors. Because cancer is a disease of abnormal tumor cell growth and abnormal cellular relationships between tumor cells and stroma (unlimited replicative potential, loss of growth inhibition between neighboring transformed cells, cancer cell invasion of neighboring tissue) (30, 31), it is perhaps not surprising that relational features form key prognostic factors in breast cancer.

To test the prognostic value of the stromal features identified by our analysis, we tested the predictive performance of stromal features and epithelial features separately. The model using only stromal features was highly associated with overall survival in the VGH data set (log-rank  $P = 0.004$ ) (Fig. 6) and showed survival association similar to that of the full C-Path model. For both grade 2 and grade 3 breast cancers, the stromal model predictions were associated with survival (both log-rank  $P < 0.05$ ). The predictions from the model composed solely of epithelial features were associated with survival

overall (log-rank  $P = 0.02$ ), and this association was strongest for stratification within histologic grade 3 tumors (log-rank  $P = 0.002$ ), with no statistically significant stratification observed in grade 1 and 2 tumors (both log-rank  $P > 0.2$ ). Pathologists currently use only epithelial features in the standard grading scheme for breast cancer and other carcinomas. Our findings suggest that evaluation of morphologic features of the tumor stroma may offer significant benefits for assessing prognosis.

The stromal feature with the largest coefficient in the prognostic model was a measure of the variability of the stromal matrix intensity differences with its neighbors (Fig. 4A). High values were associated with improved outcome. Breast cancer tissue that received a high score tended to contain larger contiguous regions of stroma separated from larger contiguous epithelial regions. This pattern of cancer growth more closely approximates epithelial-stromal relationships observed in the normal breast. This pattern results in a high score, because in stroma-rich areas, stromal matrix regions border exclusively other stromal matrix regions, whereas in other areas the stromal matrix directly borders epithelial regions. Cases that receive a low score tend to have relatively uniform distribution of epithelium

and stromal matrix throughout the image, with thin cords of epithelial cells infiltrating through stroma across the image, so that each stromal matrix region borders a relatively constant proportion of ep-



**Fig. 4.** Top stromal features associated with survival. **(A)** Variability in absolute difference in intensity between stromal matrix regions and neighbors. Top panel, high score (24.1); bottom panel, low score (10.5). (Insets) Top panel, high score; bottom panel, low score. Right panels, stromal matrix objects colored blue (low), green (medium), or white (high) according to each object's absolute difference in intensity to neighbors. **(B)** Presence of stromal regions without nuclei. Top panels, high scores; bottom panels, 0 score. Green, stromal contiguous regions with score 0; red, stromal contiguous regions with high score. (Insets) Red stromal regions are thin and do not contain nuclei; green regions are larger with nuclei. **(C)** Average relative border of stromal spindle nuclei to stromal round nuclei. Top panel, low score; bottom panel, high score. (Insets) Stromal spindled nuclear objects are green and stromal round nuclear objects are red. Right panels, higher magnification of a portion of the larger image.

ithelial and stromal regions. The stromal feature with the second largest coefficient (Fig. 4B) was the sum of the minimum green intensity value of stromal-contiguous regions. This feature received a value of zero when stromal regions contained dark pixels (such as inflammatory nuclei). The feature received a positive value when stromal objects were devoid of dark pixels. This feature provided information about the relationship between stromal cellular composition and prognosis and suggested that the presence of inflammatory cells in the stroma is associated with poor prognosis, a finding consistent with previous observations (32). The third most significant stromal feature (Fig. 4C) was a measure of the relative border between spindled stromal nuclei to round stromal nuclei, with an increased relative border of spindled stromal nuclei to round stromal nuclei associated with worse overall survival. Although the biological underpinning of this morphologic feature is currently not known, this analysis suggested that spatial relationships between different populations of stromal cell types are associated with breast cancer progression.

### Reproducibility of C-Path 5YS model predictions on samples with multiple TMA cores

For the C-Path 5YS model (which was trained on the full NKI data set), we assessed the inpatient agreement of model predictions when predictions were made separately on each image contributed by patients in the VGH data set. For the 190 VGH patients who contributed two images with complete image data, the binary predictions (high or low risk) on the individual images agreed with each other for 69% (131 of 190) of the cases and agreed with the prediction on the averaged data for 84% (319 of 380) of the images. Using the continuous prediction score (which ranged from 0 to 100), the median of the absolute difference in prediction score among the patients with replicate images was 5%, and the Spearman correlation among replicates was 0.27 ( $P = 0.0002$ ) (fig. S3). This degree of inpatient agreement is only moderate, and these findings suggest significant inpatient tumor heterogeneity, which is a cardinal feature of breast carcinomas (33–35). Qualitative visual inspection of images receiving discordant scores suggested that inpatient variability in both the epithelial and the stromal components is likely to contribute to discordant scores for the individual images. These differences appeared to relate both to the proportions of the epithelium and stroma and to the appearance of the epithelium and stroma. Last, we sought to analyze whether survival predictions were more accurate on the VGH cases that contributed multiple cores compared to the cases that contributed only a single core. This analysis showed that the C-Path 5YS model showed significantly improved prognostic prediction accuracy on the VGH cases for which we had multiple images compared to the cases that contributed only a single image (Fig. 7). Together, these findings show a significant degree of inpatient variability and indicate that increased tumor sampling is associated with improved model performance.

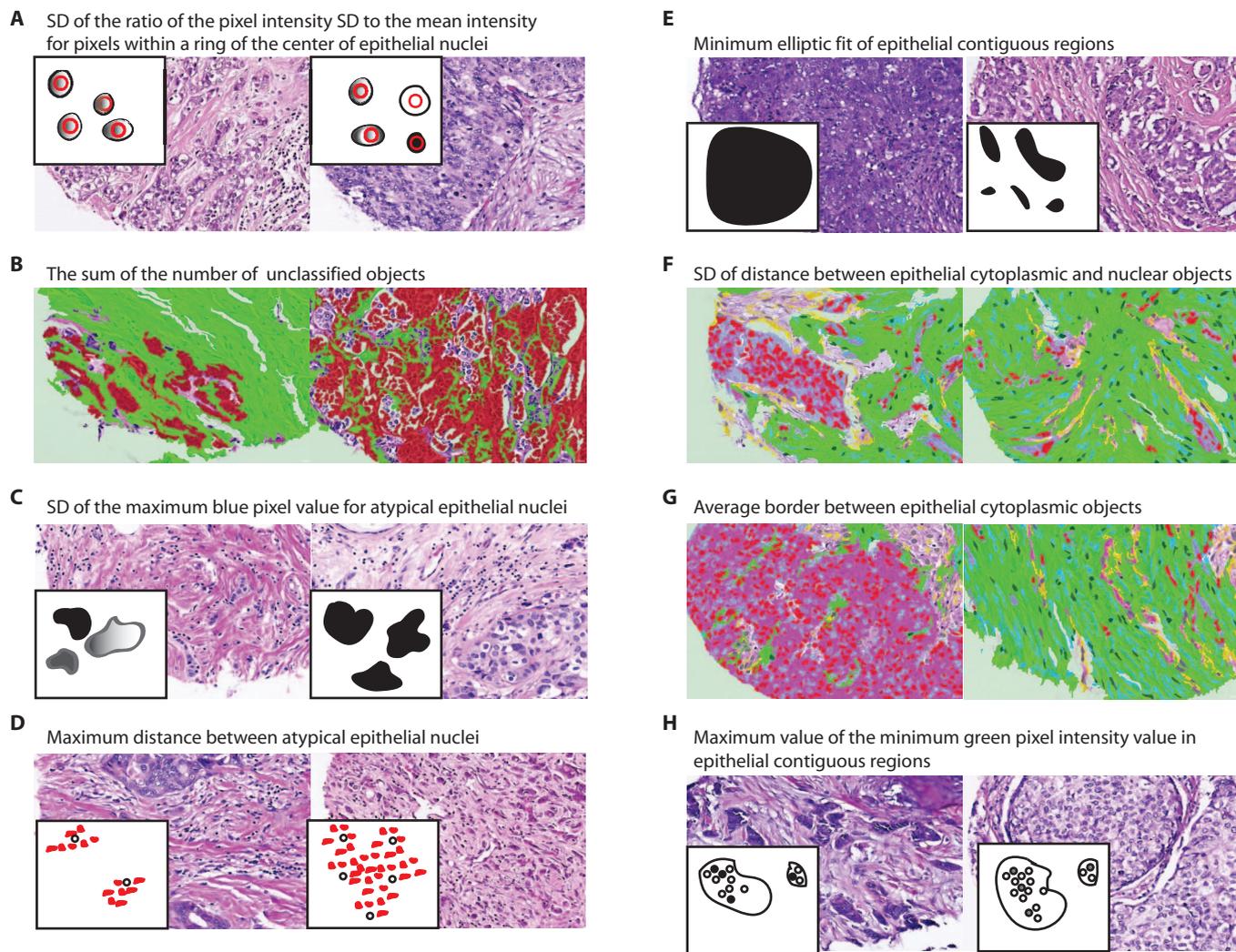
## DISCUSSION

We have developed a system for the automatic hierarchical segmentation of microscopic breast cancer images and the generation of a rich set of quantitative features to characterize the image. On the basis of these features, we built an image-based model to predict patient outcome and to identify clinically significant morphologic features. Most previous work in quantitative pathology has required

laborious image object identification by skilled pathologists, followed by the measurement of a small number of expert predefined features, primarily characterizing epithelial nuclear characteristics, such as size, color, and texture (21, 36). In contrast, after initial filtering of images to ensure high-quality TMA images and training of the C-Path models using expert-derived image annotations (epithelium and stroma labels to build the epithelial-stromal classifier and survival time and survival status to build the prognostic model), our image analysis system is automated with no manual steps, which greatly increases its scalability. Additionally, in contrast to previous approaches, our system measures thousands of morphologic descriptors of diverse

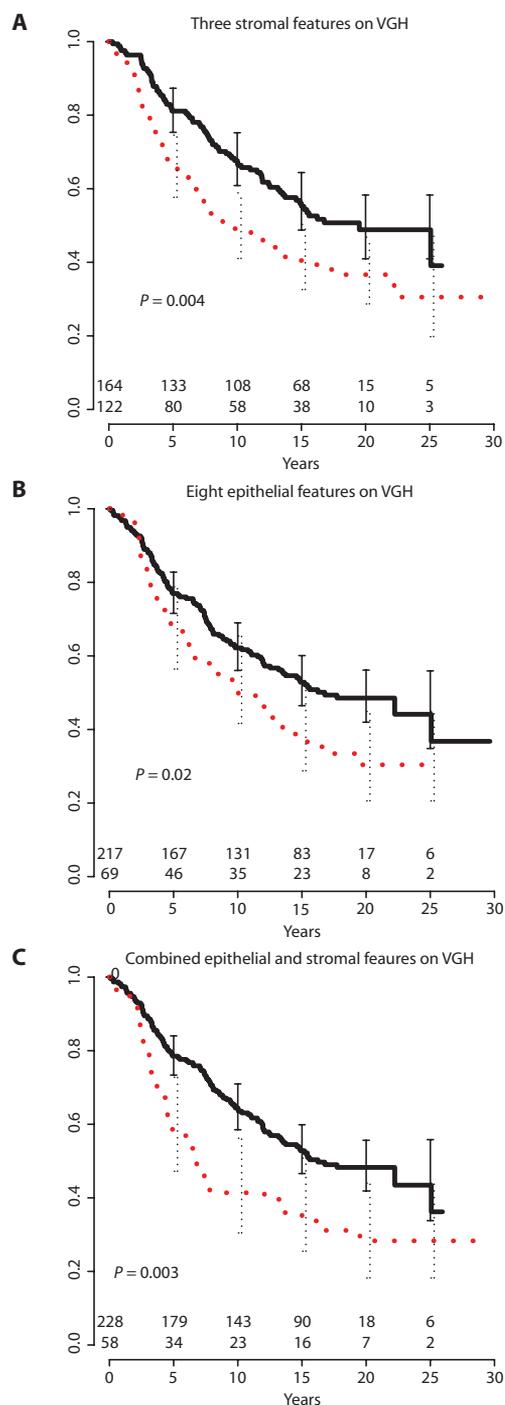
elements of the microscopic cancer image, including many relational features from both the cancer epithelium and the stroma, allowing identification of prognostic features whose significance was not previously recognized.

Using our system, we built an image-based prognostic model on the NKI data set and showed that in this patient cohort the model was a strong predictor of survival and provided significant additional prognostic information to clinical, molecular, and pathological prognostic factors in a multivariate model. We also demonstrated that the image-based prognostic model, built using the NKI data set, is a strong prognostic factor on another, independent data set with very different



**Fig. 5.** Top epithelial features. The eight panels in the figure (A to H) each shows one of the top-ranking epithelial features from the bootstrap analysis. Left panels, improved prognosis; right panels, worse prognosis. **(A)** SD of the (SD of intensity/mean intensity) for pixels within a ring of the center of epithelial nuclei. Left, relatively consistent nuclear intensity pattern (low score); right, great nuclear intensity diversity (high score). **(B)** Sum of the number of unclassified objects. Red, epithelial regions; green, stromal regions; no overlaid color, unclassified region. Left, few unclassified objects (low score); right, higher number of unclassified objects (high score). **(C)** SD of the maximum blue pixel value for atypical epithelial nuclei. Left, high

score; right, low score. **(D)** Maximum distance between atypical epithelial nuclei. Left, high score; right, low score. (Insets) Red, atypical epithelial nuclei; black, typical epithelial nuclei. **(E)** Minimum elliptic fit of epithelial contiguous regions. Left, high score; right, low score. **(F)** SD of distance between epithelial cytoplasmic and nuclear objects. Left, high score; right, low score. **(G)** Average border between epithelial cytoplasmic objects. Left, high score; right, low score. **(H)** Maximum value of the minimum green pixel intensity value in epithelial contiguous regions. Left, low score indicating black pixels within epithelial region; right, higher score indicating presence of epithelial regions lacking black pixels.



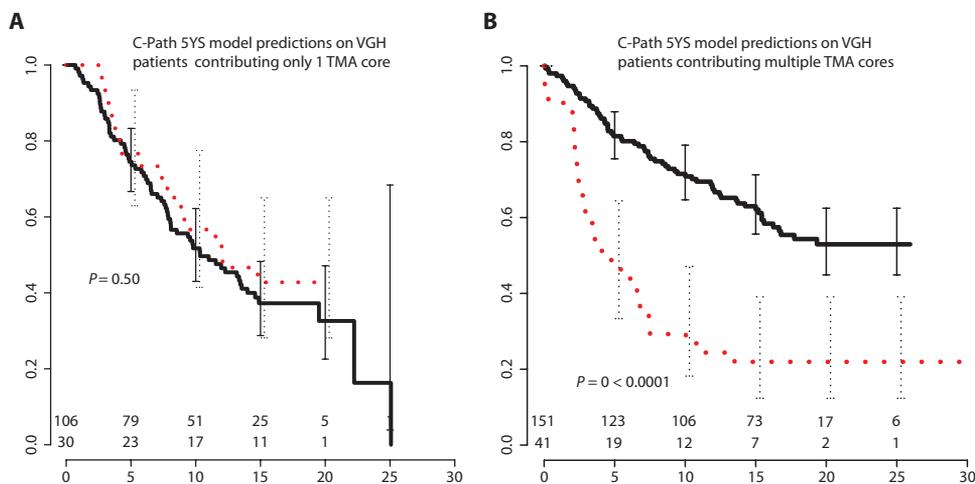
**Fig. 6.** Kaplan-Meier survival curves of prognostic models built in analysis on VGH data set limited to top features identified on NKI data set. Cases classified as high risk are plotted on the red dotted line, and cases classified as low risk on the black solid line. The error bars represent 95% CIs. The y axis is probability of overall survival, and the x axis is time in years. The numbers of patients at risk in the high- and low-risk groups at 5-year intervals are listed beneath the curves. **(A)** Prognostic model built with three top stromal features. **(B)** Prognostic model built with eight top epithelial features. **(C)** Prognostic model built with top epithelial and stromal features.

characteristics (VGH). These findings suggest that the C-Path model might be adapted to provide an objective, quantitative tool for histologic grading of invasive breast cancer in clinical practice.

A key goal of our project was to use an unbiased data-driven approach to discover prognostically significant morphologic features in breast cancer. This discovery-based approach has been widely used in analysis of genomic data, but not yet in the study of cancer morphology from microscopic images of patient samples. Microscopic images of cancer samples represent a rich source of biological information, because this level of resolution facilitates the detailed quantitative assessment of cancer cells' relationships with each other, with normal cells, and with the tumor microenvironment, all of which represent key "hallmarks of cancer" (31).

Of the top 11 features that were most robustly associated with survival in a bootstrap analysis, 8 were from the epithelium and 3 were from the stroma. A prognostic model built on only the three stromal features was a stronger predictor of patient outcome than one built from the epithelial features and is equally as predictive as the model built from all features. These stromal features included a measure of stromal inflammation, a process that has previously been implicated in breast cancer progression (32), as well as several stromal morphologic features whose prognostic significance in breast cancer has not previously been studied. Despite the growing recognition of stromal molecular characteristics and the tumor microenvironment in the regulation of carcinogenesis (8–14), since the grading of breast cancer began in the early 20th century, grading criteria have consisted entirely of epithelial features. Our analysis suggests that stromal morphologic structure is an important prognostic factor in breast cancer. Understanding the molecular basis for the prognostically significant stromal morphologic phenotypes uncovered in our analysis will be informative.

Our study has several limitations, which will need to be addressed before translation of the C-Path system for use in clinical medicine. First, it will be necessary to establish the effectiveness of the system on whole-slide images. All images used in our study came from breast cancer TMA images. Each TMA image captures only a minute portion of the full tumor volume, which is much smaller than the multiple whole-slide images used in routine diagnostic pathology. This fact is both a strength and a limitation of this study. On the one hand, our work demonstrates the ability to apply image analysis tools within a machine learning framework to build a powerful microscopic image-based prognostic model from very small samples of a tumor. This suggests that C-Path may prove useful for deriving prognostically important information from small tumor biopsy specimens. On the other hand, it is likely that we could have derived a more powerful prognostic model by analyzing whole-slide images, because these might allow the generation of additional higher-level features (such as measurements of tumor heterogeneity) and might facilitate more robust model performance because we would be summarizing our features over a much larger area of the tumor. Our image processing and machine learning pipeline is not specific to the use of TMA images and could be adapted and retrained with a data set of whole-slide images. Whole-slide images will require either manual or automated identification of breast cancer, because these larger images typically contain regions of both cancer and normal surrounding breast tissue. The TMA-based system did not require this step, because the TMA cores tend to sample exclusively areas of breast cancer. Nevertheless, the method's higher performance on patients



**Fig. 7.** Kaplan-Meier survival curves of the 5YS model predictions and overall survival on cases from the VGH cohort, stratified according to whether the case contributed one or multiple TMA cores. Cases classified as high risk are plotted on the red dotted line and cases classified as low risk on the black solid line. The error bars represent 95% CIs. The y axis is probability of overall survival, and the x axis is time in years. The numbers of patients at risk in the high- and low-risk groups at 5-year intervals are listed beneath the curves. **(A)** C-Path 5YS model predictions on VGH patients contributing only one TMA core. **(B)** C-Path 5YS model predictions on VGH patients contributing multiple TMA cores.

contributing multiple TMA cores suggests that, once this challenge is addressed, performance of the prognostic model is likely to improve with whole-slide samples.

Our study was limited to two large breast cancer patient cohorts. An important future direction for research will be to test the model on additional independent cohorts of breast cancer patients to evaluate more fully the model's generalizability. Specifically, the C-Path system must be systematically evaluated on a diverse set of images from different institutions where samples are handled in different ways. As part of this evaluation, the robustness of the epithelial-stromal classifier and the prognostic model must be evaluated separately to determine the robustness of each component of the C-Path system. Our results suggest that before applying C-Path to images from a new institution that uses a different slide processing regimen, it may be useful to train the epithelial-stromal classifier on a subset of images from the new institution. This process is likely to require labeling of 50 to 60 images, which can be performed by a trained pathologist in about 1 hour. This retraining process is analogous to the standard pathological evaluation process of histologic images from diverse institutions, in which pathologists use the visual characteristics of known morphologic structures (nuclei, cytoplasm, epithelium, stroma) from images acquired from a new institution to recalibrate their visual interpretations before applying fixed histologic grading criteria. Given the ability of our model to generalize across two diverse cohorts, it seems plausible that only a retraining of the epithelial-stromal classifier will be needed and that the prognostic features and relative weights in the prognostic model will be robust across data sets.

A final critical step for the translation of C-Path to clinical medicine will be the increased utilization of digital images in routine diagnostic pathology. Even today, most surgical pathology diagnoses are made from images viewed directly on a light microscope, and digital slide scanners are not routinely used in diagnostic surgical pathology. Beyond the technical challenges, innovative leadership among pa-

thologists will be critical for facilitating widespread implementation of quantitative, digital systems in surgical pathology laboratories (37). However, the availability of a high-accuracy, robust, automated predictor of cancer prognosis has significant promise to improve the clinical practice of pathology, especially in parts of the world where expert pathologists may be in short supply (38).

Although the work reported here has focused on predicting survival for patients with invasive breast cancer and on discovering morphologic features associated with prognosis, our unbiased methods are not specific to this setting. Hence, they can be applied much more broadly. We believe that the flexible architecture of the C-Path system—consisting of the construction of a comprehensive feature set within a machine learning framework—will enable the application of C-Path to build a library of image-based models in multiple cancer types, each optimized to predict

a specific clinical outcome, including response to particular pharmacologic agents, thereby allowing this approach to be used to directly guide treatment decisions.

## MATERIALS AND METHODS

### Patient samples

We acquired H&E-stained histological images from breast cancer tissue TMAs from two independent institutions: NKI (248 patients represented in TA110 to TA116) (26) and VGH (328 patients represented in TA268, TA274, and TA280) (38). All images are available at [tma.stanford.edu/tma\\_portal/C-path/](http://tma.stanford.edu/tma_portal/C-path/). Images were manually reviewed. Images that contained out-of-focus areas, less than 10% of tissue from the TMA core, or folded-over areas of tissue were removed. About 8% of the image files were removed, leaving a total of 671 NKI and 615 VGH images in the analysis (figs. S4 and S5).

### Image processing pipeline

We developed a customized image processing pipeline within the Definiens Developer XD image analysis environment (see Supplementary Methods and tables S8 and S9). The pipeline consists of three stages: basic image processing and feature construction, training and application of the epithelium/stroma classifier, and construction of higher-level features. Per image, we computed the mean, SD, min, and max of each feature and ultimately generated a set of 6642 features per image. For patients with multiple images (208 of 248 NKI patients; 192 of 328 VGH patients), these statistics were summarized by their mean across the images (Supplementary Material).

### Learning a prognostic model

The NKI images were used to build an image feature-based prognostic model to predict the binary outcome of 5-year survival (5YS model)

(table S3). To focus the model on the most relevant features, we used  $L_1$ -regularized logistic regression, implemented in the R package glmnet (22). Model performance on the NKI data set was assessed by eight-fold cross-validation; in each fold, the model was built using up to 217 cases of the NKI data set and evaluated on the held-out set of 31 cases. If a case from the training set was censored before 5 years (7 of 248 cases), the case was excluded from the training set. The  $\lambda$  parameter that controls the sparsity of the model was tuned at each fold by leave-one-out cross-validation on the training cases for that fold. During each fold, the value of  $\lambda$  was chosen that minimized the binomial deviance on the held-out training cases. The logistic regression model computes a probability of 5-year survival. To stratify patients into low- and high-risk groups, we selected the cut point whose stratification maximized the statistical significance of the difference in overall survival between the high- and the low-risk groups on the training cases, as indicated by the log-rank test statistic. The model and cut point were then applied to the held-out cases, so that all held-out cases received a binary classification. To assess the statistical significance of the survival stratification observed between cases predicted to be low risk versus high risk, we computed a log-rank  $P$  value using the survdiff function in the R package survival. To assess the statistical significance of feature coefficients in multivariate Cox proportional models, we assessed each feature's Wald statistic and associated  $P$  value using the function coxph in the R package survival.

To assess the robustness of the logistic regression coefficients, we performed a bootstrap analysis on the NKI data set, implemented with the "boot" package in R (39). On the basis of this analysis, for each of the 6642 features, we obtained a 95% CI for the feature's coefficient estimate. To assess the performance of the model on the VGH data set, we trained the prognostic model on the full NKI data set and tested the model on the VGH data set, excluding the 42 VGH cases that had been used for training the epithelial stromal classifier.

Additional description of methods used in the analysis is provided in the Supplementary Materials and Methods.

## SUPPLEMENTARY MATERIAL

[www.sciencetranslationalmedicine.org/cgi/content/full/3/108/108ra113/DC1](http://www.sciencetranslationalmedicine.org/cgi/content/full/3/108/108ra113/DC1)

Materials and Methods

Fig. S1. Histologic grade and overall survival.

Fig. S2. Model predictions when epithelial-stromal classifier training was limited to 107 images from NKI data set.

Fig. S3. Reproducibility analysis of 5YS model performance on breast cancers with replicate cores in the VGH data set.

Table S1. Univariate survival analysis.

Table S2. Top features in 5YS model from bootstrap analysis.

Table S3. Full 5YS model and full list of image features.

Table S4. Epithelial-stromal classifier.

Table S5. Data table to generate epithelial-stromal classifier.

Table S6. Full image feature data set (values averaged per patient and scaled).

Table S7. Full raw image feature data set (raw values for each individual image).

Table S8. Definiens rule set to extract features from each superpixel.

Table S9. Definiens rule set to apply epi-stroma classifier and generate feature set.

## REFERENCES AND NOTES

1. J. Müller, C. West, *On the Nature and Structural Characteristics of Cancer, and of Those Morbid Growths Which May Be Confounded with It* (Sherwood, Gilbert, and Piper, London, 1840).
2. D. H. Patey, R. W. Scarff, The position of histology in the prognosis of carcinoma of the breast. *Lancet* **211**, 801–804 (1928).
3. H. J. Bloom, W. W. Richardson, Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. *Br. J. Cancer* **11**, 359–377 (1957).

4. V. Le Doussal, M. Tubiana-Hulin, S. Friedman, K. Hacene, F. Spyrtos, M. Brunet, Prognostic value of histologic grade nuclear components of Scarff-Bloom-Richardson (SBR). An improved score modification based on a multivariate analysis of 1262 invasive ductal breast carcinomas. *Cancer* **64**, 1914–1921 (1989).
5. C. W. Elston, I. O. Ellis, Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: Experience from a large study with long-term follow-up. *Histopathology* **19**, 403–410 (1991).
6. C. Sotiriou, M. J. Piccart, Taking gene-expression profiling to the clinic: When will molecular signatures become relevant to patient care? *Nat. Rev. Cancer* **7**, 545–553 (2007).
7. J. S. Reis-Filho, B. Weigelt, D. Fumagalli, C. Sotiriou, Molecular profiling: Moving away from tumor philately. *Sci. Transl. Med.* **2**, 47ps43 (2010).
8. M. J. Bissell, D. Radisky, Putting tumours in context. *Nat. Rev. Cancer* **1**, 46–54 (2001).
9. B. S. Wiseman, Z. Werb, Stromal effects on mammary gland development and breast cancer. *Science* **296**, 1046–1049 (2002).
10. R. B. West, D. S. Nuyten, S. Subramanian, T. O. Nielsen, C. L. Corless, B. P. Rubin, K. Montgomery, S. Zhu, R. Patel, T. Hernandez-Boussard, J. R. Goldblum, P. O. Brown, M. van de Vijver, M. van de Rijn, Determination of stromal signatures in breast carcinoma. *PLoS Biol.* **3**, e187 (2005).
11. A. H. Beck, I. Espinosa, C. B. Gilks, M. van de Rijn, R. B. West, The fibromatosis signature defines a robust stromal response in breast carcinoma. *Lab. Invest.* **88**, 591–601 (2008).
12. A. Bergamaschi, E. Tagliabue, T. Sørlie, B. Naume, T. Triulzi, R. Orlandi, H. G. Russnes, J. M. Nesland, R. Tammi, P. Auvinen, V. M. Kosma, S. Ménard, A. L. Borresen-Dale, Extracellular matrix signature identifies breast cancer subgroups with different clinical outcome. *J. Pathol.* **214**, 357–367 (2008).
13. A. E. Karnoub, A. B. Dash, A. P. Vo, A. Sullivan, M. W. Brooks, G. W. Bell, A. L. Richardson, K. Polyak, R. Tubo, R. A. Weinberg, Mesenchymal stem cells within tumour stroma promote breast cancer metastasis. *Nature* **449**, 557–563 (2007).
14. G. Finak, N. Bertos, F. Pepin, S. Sadekova, M. Souleimanova, H. Zhao, H. Chen, G. Omeroglu, S. Meterissian, A. Omeroglu, M. Hallett, M. Park, Stromal gene expression predicts clinical outcome in breast cancer. *Nat. Med.* **14**, 518–527 (2008).
15. G. Bianchini, Y. Qi, R. H. Alvarez, T. Iwamoto, C. Coutant, N. K. Ibrahim, V. Valero, M. Cristofanilli, M. C. Green, L. Radvanyi, C. Hatzis, G. N. Hortobagyi, F. Andre, L. Gianni, W. F. Symmans, L. Pusztai, Molecular anatomy of breast cancer stroma and its prognostic value in estrogen receptor-positive and -negative cancers. *J. Clin. Oncol.* **28**, 4316–4323 (2010).
16. G. G. Van den Eynden, C. G. Colpaert, A. Couvelard, F. Pezzella, L. Y. Dirix, P. B. Vermeulen, E. A. Van Marck, T. Hasebe, A fibrotic focus is a prognostic factor and a surrogate marker for hypoxia and (lymph)angiogenesis in breast cancer: Review of the literature and proposal on the criteria of evaluation. *Histopathology* **51**, 440–451 (2007).
17. T. R. Fanshawe, A. G. Lynch, I. O. Ellis, A. R. Green, R. Hanka, Assessing agreement between multiple raters with missing rating information, applied to breast cancer tumour grading. *PLoS One* **3**, e2925 (2008).
18. J. P. Baak, P. H. Kurver, S. de Graaf, M. E. Boon, Morphometry for prognosis prediction in breast cancer. *Lancet* **2**, 315 (1981).
19. C. Cordon-Cardo, A. Kotsianti, D. A. Verbel, M. Teverovskiy, P. Capodiceci, S. Hamann, Y. Jeffers, M. Clayton, F. Elkhettabi, F. M. Khan, M. Sapir, V. Bayer-Zubek, Y. Vengrenyuk, S. Fogarsi, O. Saidi, V. E. Reuter, H. I. Scher, M. W. Kattan, F. J. Bianco, T. M. Wheeler, G. E. Ayala, P. T. Scardino, M. J. Donovan, Improved prediction of prostate cancer recurrence through systems pathology. *J. Clin. Invest.* **117**, 1876–1883 (2007).
20. M. J. Donovan, S. Hamann, M. Clayton, F. M. Khan, M. Sapir, V. Bayer-Zubek, G. Fernandez, R. Mesa-Tejada, M. Teverovskiy, V. E. Reuter, P. T. Scardino, C. Cordon-Cardo, Systems pathology approach for the prediction of prostate cancer progression after radical prostatectomy. *J. Clin. Oncol.* **26**, 3923–3929 (2008).
21. A. H. Beck, M. B. Resnick, K. C. Drumea, E. Sabo, *Quantitative Image Analysis for the Classification of Epithelial Neoplasia in Image Analysis in Medical Microscopy and Pathology*, H. S. Wu, A. Einstein, Eds. (Research Signpost, Kerala, India, 2007).
22. J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
23. R. J. Tibshirani, B. Efron, Pre-validation and inference in microarrays. *Stat. Appl. Genet. Mol. Biol.* **1**, Article1 (2002).
24. M. J. van de Vijver, Y. D. He, L. J. van't Veer, H. Dai, A. A. Hart, D. W. Voskuil, G. J. Schreiber, J. L. Peterse, C. Roberts, M. J. Marton, M. Parrish, D. Atsma, A. Witteveen, A. Glas, L. Delahaye, T. van der Velde, H. Bartelink, S. Rodenhuis, E. T. Rutgers, S. H. Friend, R. Bernards, A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347**, 1999–2009 (2002).
25. C. Sotiriou, P. Wirapati, S. Loi, A. Harris, S. Fox, J. Smeds, H. Nordgren, P. Farmer, V. Praz, B. Haibe-Kains, C. Desmedt, D. Larsimont, F. Cardoso, H. Peterse, D. Nuyten, M. Buyse, M. J. Van de Vijver, J. Bergh, M. Piccart, M. Delorenzi, Gene expression profiling in breast cancer: Understanding the molecular basis of histologic grade to improve prognosis. *J. Natl. Cancer Inst.* **98**, 262–272 (2006).

26. R. Liu, X. Wang, G. Y. Chen, P. Dalerba, A. Gurney, T. Hoey, G. Sherlock, J. Lewicki, K. Shedden, M. F. Clarke, The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N. Engl. J. Med.* **356**, 217–226 (2007).
27. J. T. Chi, Z. Wang, D. S. Nuyten, E. H. Rodriguez, M. E. Schaner, A. Salim, Y. Wang, G. B. Kristensen, A. Helland, A. L. Borresen-Dale, A. Giaccia, M. T. Longaker, T. Hastie, G. P. Yang, M. J. van de Vijver, P. O. Brown, Gene expression programs in response to hypoxia: Cell type specificity and prognostic significance in human cancers. *PLoS Med.* **3**, e47 (2006).
28. T. Sorlie, C. M. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, T. Thorsen, H. Quist, J. C. Matese, P. O. Brown, D. Botstein, P. Eystein Lønning, A. L. Borresen-Dale, Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10869–10874 (2001).
29. D. S. Nuyten, T. Hastie, J. T. Chi, H. Y. Chang, M. J. van de Vijver, Combining biological gene expression signatures in predicting outcome in breast cancer: An alternative to supervised classification. *Eur. J. Cancer* **44**, 2319–2329 (2008).
30. D. Hanahan, R. A. Weinberg, The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
31. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).
32. W. Tan, W. Zhang, A. Strasner, S. Grivennikov, J. Q. Cheng, R. M. Hoffman, M. Karin, Tumour-infiltrating regulatory T cells stimulate mammary cancer metastasis through RANKL-RANK signalling. *Nature* **470**, 548–553 (2011).
33. S. P. Shah, R. D. Morin, J. Khattri, L. Prentice, T. Pugh, A. Burleigh, A. Delaney, K. Gelmon, R. Guliany, J. Senz, C. Steidl, R. A. Holt, S. Jones, M. Sun, G. Leung, R. Moore, T. Severson, G. A. Taylor, A. E. Teschendorff, K. Tse, G. Turashvili, R. Varhol, R. L. Warren, P. Watson, Y. Zhao, C. Caldas, D. Huntsman, M. Hirst, M. A. Marra, S. Aparicio, Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461**, 809–813 (2009).
34. L. Ding, M. J. Ellis, S. Li, D. E. Larson, K. Chen, J. W. Wallis, C. C. Harris, M. D. McLellan, R. S. Fulton, L. L. Fulton, R. M. Abbott, J. Hoog, D. J. Dooling, D. C. Koboldt, H. Schmidt, J. Kalicki, Q. Zhang, L. Chen, L. Lin, M. C. Wendl, J. F. McMichael, V. J. Magrini, L. Cook, S. D. McGrath, T. L. Vickery, E. Appelbaum, K. Deschryver, S. Davies, T. Guintoli, L. Lin, R. Crowder, Y. Tao, J. E. Snider, S. M. Smith, A. F. Dukes, G. E. Sanderson, C. S. Pohl, K. D. Delehaunty, C. C. Fronick, K. A. Pape, J. S. Reed, J. S. Robinson, J. S. Hodges, W. Schierding, N. D. Dees, D. Shen, D. P. Locke, M. E. Wiechert, J. M. Eldred, J. B. Peck, B. J. Oberkfell, J. T. Lofie, F. Du, A. E. Hawkins, M. D. O’Laughlin, K. E. Bernard, M. Cunningham, G. Elliott, M. D. Mason, D. M. Thompson Jr., J. L. Ivanovich, P. J. Goodfellow, C. M. Perou, G. M. Weinstock, R. Aft, M. Watson, T. J. Ley, R. K. Wilson, E. R. Mardis, Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* **464**, 999–1005 (2010).
35. A. Marusyk, K. Polyak, Tumor heterogeneity: Causes and consequences. *Biochim. Biophys. Acta* **1805**, 105–117 (2010).
36. L. Mulrane, E. Rexhepaj, S. Penney, J. J. Callanan, W. M. Gallagher, Automated image analysis in histopathology: A valuable tool in medical diagnostics. *Expert Rev. Mol. Diagn.* **8**, 707–725 (2008).
37. J. P. Baak, The framework of pathology: Good laboratory practice by quantitative and molecular methods. *J. Pathol.* **198**, 277–283 (2002).
38. C. L. Hitchcock, The future of telepathology for the developing world. *Arch. Pathol. Lab. Med.* **135**, 211–214 (2011).
39. A. C. Davison, D. V. Hinkley, *Bootstrap Methods and Their Applications* (Cambridge Univ. Press, Cambridge, UK, 1997).
40. **Funding:** A.H.B. is an Advanced Residency Training at Stanford fellow and is the recipient of a Residents Research Grant from the College of American Pathologists. This project was funded by the Stanford University Department of Pathology and by the Stanford University Bio-X Interdisciplinary Initiatives Program. **Author contributions:** A.H.B., D.K., R.B.W., and M.v.d.R. conceived the study. A.H.B. and D.K. designed the methods and performed the analysis. A.R.S. contributed histopathological analysis of images. S.L., R.J.M., T.O.N., and M.J.v.d.V. contributed images and image annotation data. All authors contributed to preparation of the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data availability:** Image files used in this analysis are available at [tma.stanford.edu/tma\\_portal/C-path/](http://tma.stanford.edu/tma_portal/C-path/).

Submitted 25 April 2011

Accepted 21 October 2011

Published 9 November 2011

10.1126/scitranslmed.3002564

**Citation:** A. H. Beck, A. R. Sangoi, S. Leung, R. J. Marinelli, T. O. Nielsen, M. J. van de Vijver, R. B. West, M. van de Rijn, D. Koller, Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Sci. Transl. Med.* **3**, 108ra113 (2011).

## Editor's Summary

### An Automated Pathologist Reads Cancer Biopsies

How is a camera different from the human eye? Only the eye's images undergo extensive secondary processing as they are interpreted by the human brain. But what if we could program a computer to do the secondary processing? A pathologist reading a cancer biopsy slide matches his or her brain's memory of certain cancer-related features (tubules, atypical nuclei, and mitosis) against the tissue. This decades-old scoring system is still standard in most places for prognosis and treatment of cancer, despite its variability and often unreliability. Now, Beck *et al.* have created an automated pathologist by replacing the human brain with sophisticated image processing software and instructing it to find quantitative aspects of breast cancer tissue that predict prognosis. The software located a set of features that strongly predicted breast cancer outcome in both training and validation samples.

With an image analysis protocol they termed C-Path, the authors set their program loose on a set of samples from patients in the Netherlands. From more than 6000 features, the software found a set that were associated with samples from patients who had died sooner. The key aspect of this analysis was that these features were not predefined by a pathologist as being relevant to cancer; instead, the software itself found the cancer-related features among the very large set of measurements of the image. Classifying the tissue as epithelial or stromal, an important part of cancer diagnosis, took a bit of extra work: The authors needed to provide the software with some hand-marked samples so it could learn the difference. The C-Path score yielded information above and beyond that from many other measures of cancer severity including pathology grade, estrogen receptor status, tumor size, and lymph node status. In another, completely independent group of women from Vancouver, the C-Path score was also associated with overall survival.

An unexpected finding was that the features that were the best predictors of patient survival were not from the cancer itself but were from the adjacent stromal tissue. Women with worse outcomes tended to have thin cords of epithelial cells infiltrating the stroma, which resulted in high-risk stromal matrix variability scores. These patients also tended to have more inflammatory cells in the stroma (picked up as dark areas by the software). Replacing the human brain with an unbiased image processing system can extract more information from microscopy images and discover new biological aspects of cancer tissue.

**A complete electronic version of this article** and other services, including high-resolution figures, can be found at:

<http://stm.sciencemag.org/content/3/108/108ra113.full.html>

**Supplementary Material** can be found in the online version of this article at:

<http://stm.sciencemag.org/content/suppl/2011/11/07/3.108.108ra113.DC1.html>

**Related Resources for this article** can be found online at:

<http://stm.sciencemag.org/content/scitransmed/4/157/157ra143.full.html>

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>