Final Progress Report for Research Projects Funded by Health Research Grants

Instructions: Please complete all of the items as instructed. Do not delete instructions. Do not leave any items blank; responses must be provided for all items. If your response to an item is "None", please specify "None" as your response. "Not applicable" is not an acceptable response for any of the items. There is no limit to the length of your response to any question. Responses should be single-spaced, no smaller than 12-point type. The report **must be completed using MS Word**. Submitted reports must be Word documents; they should not be converted to pdf format. Questions? Contact Health Research Program staff at 717-783-2548.

1. **Grantee Institution:** Drexel University

2. **Reporting Period (start and end date of grant award period):** 1/1/2009-12/31/2010

3. **Grant Contact Person (First Name, M.I., Last Name, Degrees):** Anne Martella

4. **Grant Contact Person's Telephone Number:** (215) 895-6471

5. **Grant SAP Number:** 4100047631

6. **Project Number and Title of Research Project:** 11 - Multidimensional Shape/Color Distributions as a Computational Biomarker for Cancer Pathology

7. **Start and End Date of Research Project:** 1/1/2009 – 6/30/2010

8. **Name of Principal Investigator for the Research Project:** David E. Breen, PhD

9. **Research Project Expenses**.

   9(A) Please provide the amount of health research grant funds spent on this project for the entire duration of the grant, including any interest earned that was spent:

   $101,073

   9(B) Provide the last names (include first initial if multiple individuals with the same last name are listed) of **all** persons who worked on this research project and were supported with health research funds. Include position titles (Principal Investigator, Graduate Assistant, Post-doctoral Fellow, etc.), percent of effort on project and total health research funds expended for the position. For multiple year projects, if percent of effort varied from year to year, report in the % of Effort column the effort by year 1, 2, 3, etc. of the project (x% Yr 1; z% Yr 2-3).

Last Name	Position Title	% of Effort on Project	Cost
Breen	Associate Professor	5%	$ 11,040
Reza	Graduate Student	75%	$ 36,599
Garcia	Professor	1.6%	$ 7,562
Milutinovic	Research Engineer	35%	$ 36,638

9(C) Provide the names of **all** persons who worked on this research project, but who *were not* supported with health research funds. Include position titles (Research Assistant, Administrative Assistant, etc.) and percent of effort on project. For multiple year projects, if percent of effort varied from year to year, report in the % of Effort column the effort by year 1, 2, 3, etc. of the project (x% Yr 1; z% Yr 2-3).

Last Name	Position Title	% of Effort on Project
Polikar	Associate Professor, Rowan U.	5%

9(D) Provide a list of **all** scientific equipment purchased as part of this research grant, a short description of the value (benefit) derived by the institution from this equipment, and the cost of the equipment.

Type of Scientific Equipment	Value Derived	Cost
Apple MacPro computer	Performed most of the project's computations. Workstation for grad student.	$6,687

10. Co-funding of Research Project during Health Research Grant Award Period. Did this research project receive funding from any other source during the project period when it was supported by the health research grant?

Yes_____ No____X_____

If yes, please indicate the source and amount of other funds:

11. Leveraging of Additional Funds

11(A) As a result of the health research funds provided for this research project, were you able to apply for and/or obtain funding from other sources to continue or expand the research?

Yes___X_____ No_____

If yes, please list the applications submitted (column A), the funding agency (National Institutes of Health—NIH, or other source in column B), the month and year when the application was submitted (column C), and the amount of funds requested (column D). If you have received a notice that the grant will be funded, please indicate the amount of funds to be awarded (column E). If the grant was not funded, insert "not funded" in column E.

Do not include funding from your own institution or from CURE (tobacco settlement funds). Do not include grants submitted prior to the start date of the grant as shown in Question 2. If you list grants submitted within 1-6 months of the start date of this grant, add a statement below the table indicating how the data/results from this project were used to secure that grant.

A. Title of research project on grant application	B. Funding agency (check those that apply)	C. Month and Year Submitted	D. Amount of funds requested:	E. Amount of funds to be awarded:
Cost-Effective Overall Survival Prediction Based on Image Analysis of Breast Carcinoma Tissue: Correlation with Gene Expression and Lymph Node Status	☐NIH ☒ Other federal (specify: US Army Breast Cancer Research Program) ☐ Nonfederal source (specify: _____)	May 2010	$ 331,617	$ Pending

11(B) Are you planning to apply for additional funding in the future to continue or expand the research?

Yes____X_____ No_____

If yes, please describe your plans:

We plan to submit an R21 proposal to NIH in Fall 2010 in order to continue and extend the research funded by this Health Research Grant (HRG). The HRG-funded research provided the preliminary results needed to make the NIH grant proposal competitive.

12. Future of Research Project. What are the future plans for this research project?

The initial studies were only conducted on a small number of specimens (100). A follow-on study with many more (500) specimens must be contacted. Additionally, a number of other geometric and texture metrics will be implemented and evaluated to determine their prognostic capabilities. We will explore our approach's ability to predict other prognostic factors, such as gene expression and overall survival.

13. New Investigator Training and Development. Did students participate in project supported internships or graduate or post-graduate training for at least one semester or one summer?

Yes____X_____ No_____

If yes, how many students? Please specify in the tables below:

	Undergraduate	Masters	Pre-doc	Post-doc
Male		1		
Female				
Unknown				
Total		**1**		

	Undergraduate	Masters	Pre-doc	Post-doc
Hispanic				
Non-Hispanic		1		
Unknown				
Total		**1**		

	Undergraduate	Masters	Pre-doc	Post-doc
White				
Black				
Asian		1		
Other				
Unknown				
Total		**1**		

14. **Recruitment of Out-of–State Researchers**. Did you bring researchers into Pennsylvania to carry out this research project?

    Yes___X_____ No_____

    If yes, please list the name and degree of each researcher and his/her previous affiliation:

    Aladin Milutinovic, B.S. & M.Eng, Biomedical Engineering, Tufts University

15. **Impact on Research Capacity and Quality**. Did the health research project enhance the quality and/or capacity of research at your institution?

    Yes___X_____ No_____

    If yes, describe how improvements in infrastructure, the addition of new investigators, and other resources have led to more and better research.

    The project allowed us to bring in a graduate-trained biomedical engineer from the Boston area. It also supported the training of a Masters-level computer science student. More importantly the project has allowed us to produce the first verification of an image-based diagnostic technique that we believe will be applicable to many fields of medicine that rely on analyzing and understanding cellular processes and structure, e.g. cancer and aging. Establishing the effectiveness of the technique will therefore have great impact on our

research enterprise as we move to expand its application to other areas of medical investigations.

16. Collaboration, business and community involvement.

16(A) Did the health research funds lead to collaboration with research partners outside of your institution (e.g., entire university, entire hospital system)?

Yes_____ No____X_____

If yes, please describe the collaborations:

16(B) Did the research project result in commercial development of any research products?

Yes_____ No_____X_____

If yes, please describe commercial development activities that resulted from the research project:

16(C) Did the research lead to new involvement with the community?

Yes_____ No____X_____

If yes, please describe involvement with community groups that resulted from the research project:

17. Progress in Achieving Research Goals, Objectives and Aims.

List the project goals, objectives and specific aims (as contained in the grant application's strategic plan). Summarize the progress made in achieving these goals, objectives and aims for the entire grant award period. Indicate whether or not each goal/objective/aim was achieved; if something was not achieved, note the reasons why. Describe the methods used. If changes were made to the research goals/objectives/aims, methods, design or timeline since the original grant application was submitted, please describe the changes. Provide detailed results of the project. Include evidence of the data that was generated and analyzed, and provide tables, graphs, and figures of the data. List published abstracts, poster presentations and scientific meeting presentations at the end of the summary of progress; peer-reviewed publications should be listed under item 20.

This response should be a DETAILED report of the methods and findings. It is not sufficient to state that the work was completed. Insufficient information may result in an unfavorable performance review, which may jeopardize future funding. If research findings are pending publication you must still include enough detail for the expert peer reviewers to evaluate the progress during the course of the project.

Health research grants funded under the Tobacco Settlement Act will be evaluated via a performance review by an expert panel of researchers and clinicians who will assess project work using this Final Progress Report, all project Annual Reports and the project's strategic plan. After the final performance review of each project is complete, approximately 12-16 months after the end of the grant, this Final Progress Report, as well as the Final Performance Review Report containing the comments of the expert review panel, and the grantee's written response to the Final Performance Review Report, will be posted on the CURE Web site.

There is no limit to the length of your response. Responses must be single-spaced below, no smaller than 12-point type. If you cut and paste text from a publication, be sure symbols print properly, e.g., the Greek symbol for alpha (α) and beta (ß) should not print as boxes (□) and include the appropriate citation(s). DO NOT DELETE THESE INSTRUCTIONS.

Multidimensional Shape/Color Distributions as a Computational Biomarker for Cancer Pathology - The purpose of the project is to develop computational techniques for analyzing histology images of breast cancer tumors in order to ascertain the metastasis status of the tumor. The computational techniques will derive shape and color information from the images that will enable automated evaluation of the tumor. Specifically, the techniques will be used to determine if a patient's breast cancer has spread to nearby lymph nodes by examining a primary tumor that has been removed from the patient. This image analysis capability will eliminate the need for exploratory surgical removal of lymph nodes; thus eliminating the associated side effects, e.g. pain, swelling and morbidity, and costs.

Project Overview. We will develop novel computational technologies that may be used to automatically and objectively predict the axillary lymph node status of breast cancer tumors via multiresolution image analysis of the primary tumor. Our hypothesis is that the structure of the nuclear pleomorphisms found in breast cancer tumors can be transformed into high-dimensional shape distributions using geometric measures from stochastic geometry. We will augment this space with information derived from the color/texture variations of the hyperchromatism found in the cancer cell nuclei and the expression of multiple prognostic markers. We hypothesize that, once the structure and color of the cells has been transformed, the resulting distributions of N pathologic staging categories will map into well-defined regions of the high-dimensional distribution space. Mapping an unknown breast cancer sample into this high-D space and determining, via machine learning, to which region it belongs will allow us to automatically predict its axillary lymph node status.

The project will employ a routinely used prognostic immunohistochemistry panel together with H&E staining. The panel includes: estrogen and progesterone receptors, MIB-1 (to calculate proliferative activity), mutated p53 and HER2/neu. Shape distributions represent the structure/shape of 2D/3D objects with a probability distribution produced by repeatedly and randomly applying geometric measures. Stochastic geometry analyzes and quantifies the connections between geometry and probability in order to describe and characterize

aggregate statistical geometric properties. Based on information theoretic distance metrics between the shape distributions, classifiers will be trained to automatically identify lymph node statuses. We will mainly investigate the use of multi-class kernel support vector machines for non-linear classification of all lymph node categories.

Specific Aim 1 – Extend and improve recently developed histology image segmentation and shape distribution generation technologies/algorithms.

Specific Aim 2 – Develop techniques that create color distributions from the pixel regions of the histologic and prognostic marker images that correspond to the segmentation pixels identifying neoplastic cell nuclei.

Specific Aim 3 – Investigate machine-learning technologies for determining which geometric measures and spectral features will best discriminate between lymph node status categories and provide the most reliable/significant lymph node metastasis status classifier.

Specific Aim 4 – Using breast cancer databank specimens, numerous cases will be digitized, processed, and analyzed to validate the approach.

Progress Summary. The project was an outstanding success. Our goal was to develop a method for computationally predicting lymph node metastasis status via inspection of primary tumor histology, and to provide initial verification of its effectiveness. By successfully completing each specific aim of the project, our objective of demonstrating that metastasis to axillary lymph nodes may be determined by image analysis of histology scans of the primary tumor has been achieved.

Methods.

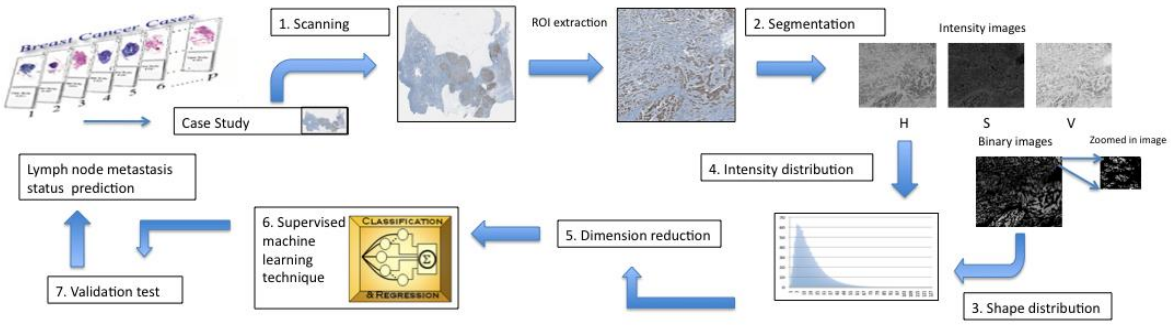


Figure 1. Computational pipeline for automated lymph node metastasis status prediction based on analysis of primary breast tumor histology.

The computational pipeline for automated lymph node metastasis status prediction based on analysis of primary breast tumor histology is presented in Figure 1. The stages of the pipeline

include: 1) selection and scanning of histology slides, as well as gathering the associated de-identified patient data, 2) accurate and robust automated cancer cell and tumor structure segmentation techniques, 3) computing geometric measures from stochastic geometry that transform cancer cell/tumor shapes into shape distributions, 4) creating intensity distributions from the texture variation / nuclear hyperchromasia levels within cancer cells, 5) mapping the high-D shape/color distributions into a lower dimensional feature vector for separation analysis, 6) supervised machine learning technologies that classify test samples into lymph node metastasis status with probabilistic confidence values based on the feature vectors. All of the algorithmic steps needed to produce our computational pipeline were completed on schedule. Technical details describing these steps, as well as the project's results, are given below.

Sample Scanning. Digital slides of cases of invasive mammary carcinoma that have full prognostic panel consisting of six slides were scanned for 55 cases N0 (negative for lymph node metastasis) and 45 cases N1+ (positive for lymph node metastasis) using the ScanScope XT (Aperio, Vista, CA). Each image is scanned using the 20X objective and is on average about 50,000 x 40,000 pixels with a resolution of 0.5 µm/pixel. The full prognostic panel consists of estrogen (ER, clone SP1) and progesterone (PR, clone 1E2) receptors, proliferation antigen MIB-1 (Ki-67, clone 30-9), mutated tumor protein 53 (p53, clone Bp53-11), and Human Epidermal growth factor receptor 2 (HER2/neu, 4B5). Each slide was stained using Benchmark XT (Ventana Medical Systems Inc., Tucson, AZ). These cases were selected in a de-identified fashion by the Honest Broker from the Tissue Procurement Facility from a database of patients that have had consults in the Pathology Diagnostics Laboratory at Drexel University College of Medicine and have a record of positive or negative lymph node metastasis over the past seven years. A 6000 x 6000 pixel region was chosen from the scanned image containing the highest percentage of nuclear staining (as indicated on the prognostic panel report) for each case in order to facilitate segmentation and maximize the number of nuclei available for analysis. The exported region is the one with the highest density of stained cells within a 3mm x 3mm area. A small region of an original image example is shown in the upper right of Figure 2 and in Figure 3a.

Image Processing/Segmentation. The image analysis algorithm has been designed and implemented to isolate the brown-stained nuclei in the scanned images. In the first step, the Red-Green-Blue (RGB) image is converted into the Hue-Saturation-Value (HSV) color space. In the HSV space, to solve to the problem of segmenting the brown areas, all the blue and white regions were filtered out. This can be accomplished by identifying pixels that lie within a specific range (330 → 360, 0 → 90) of the hue channel and the value channel (0 → 0.8). An overview of the process is given in Figure 2. A semi-automated segmentation was developed where a region of the image is pre-segmented. The algorithm then takes the pre-segmented parameters to segment the remainder of the image. A manual saturation threshold is also added to the pre-segmentation parameters to adjust the variability between the staining intensities. The variability in staining can be due to tissue thickness, age, and reagent kit used. After testing several automated methods such as Otsu, Mean of Gaussians, and Background Subtraction, the variability in staining prevented a fully automated algorithm for segmenting the brown nuclei. In order to ensure consistent segmentation of brown nuclei we determined that a semi-automated method was necessary. A morphological opening was

performed to remove noise smaller than 3 pixels in diameter. An example of a final HSV segmented result is shown in Figure 3b.

After the images are segmented based on the HSV color space for brown nuclei, the resulting segmentations still contained clumped regions of cells (Figure 3b). For nuclear analysis to work we needed to ensure that nuclei are not clumped. We processed the segmentations with a watershed algorithm developed in the Cell Profiler system (Broad Institute, Cambridge, MA). The segmentation mask was merged with the saturation channel (Figure 3c). The combination was used in Cell Profiler to identify the nuclear borders in the cases where the nuclei were clumped. The final result of cell nuclei segmentation is shown in Figure 3d.

After the image-processing step, the final segmentations of the brown nuclei are used for multiscale analysis from the nuclear level features (area, perimeter, area v. perimeter, curvature, distance, line sweep, aspect ratio, and color) to large-scale tumor structure. For large-scale tumor structure analysis additional image processing was needed. In order to identify areas of high nuclear staining, the areas were interpolated by a morphological closing. Circular pixel dilation was performed representing one nuclear region (r = 8 pixels) followed by an erosion of the same radius. This action preserves the original boundaries but fills in all the spaces between nuclei that are within one nuclear diameter of each other. Figure 3e shows the results of the large area interpolation.

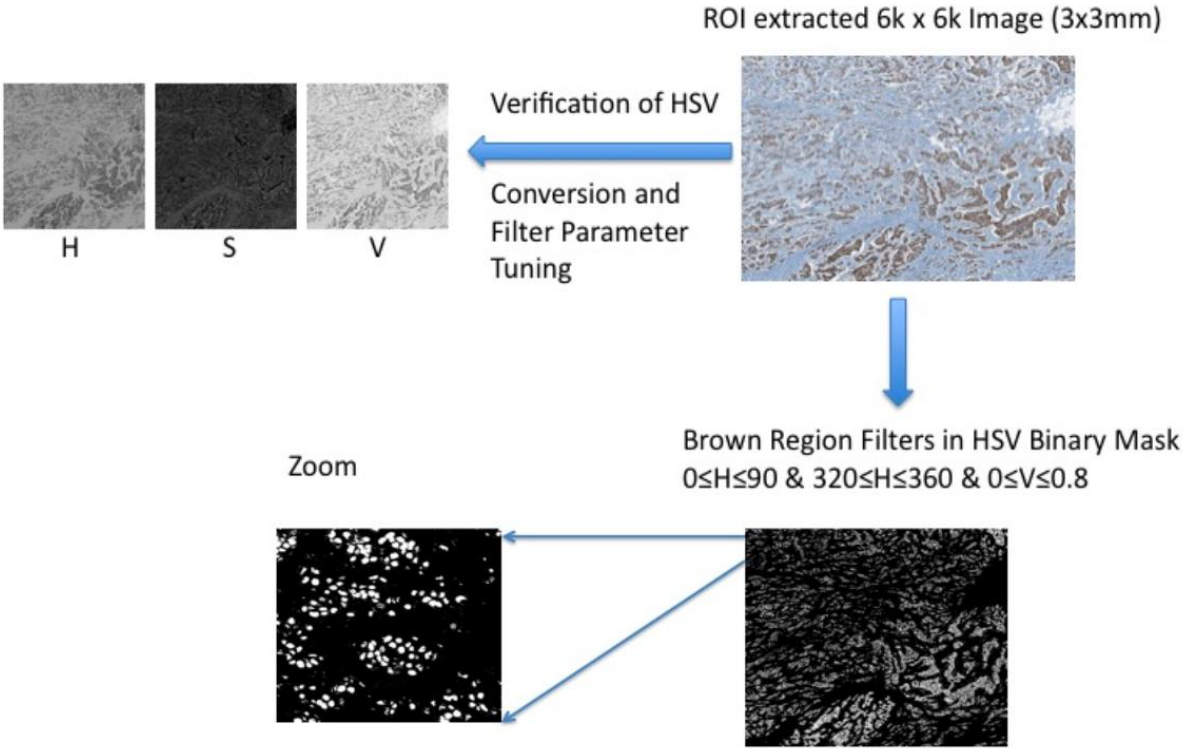
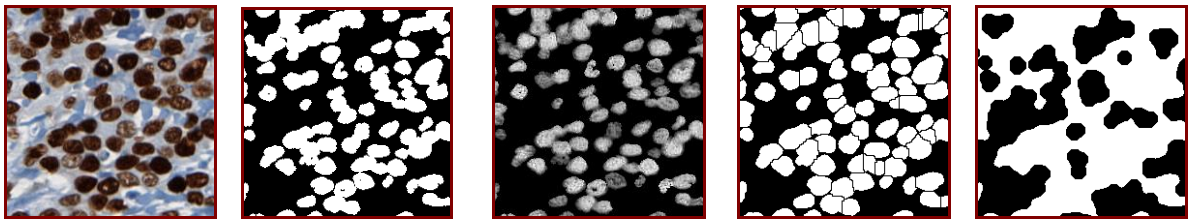


Figure 2. Processing stained histology scans. Separating the image into HSV channels and identifying brown regions.



a. Original Image b. Brown Region c. Mask + Sat. d. Final Segm. e. Large Areas

Figure 3. Results from the histology scan segmentation process. a) Subsection of one of the histology scans. b) Segmentation that identifies the brown regions of the scan. c) Using the segmentation as a mask to identify the brown saturation values in the original image. d) Cell nuclei segmentation produced by Cell Profiler. e) Larger, combined regions produced by applying a morphological closing to the nuclei segmentation image.

Shape Distribution Generation. The C++ programming language was used to write the code that implements the geometric metrics needed to generate shape distributions from the segmented images. Since these segmented images could be huge in size (nearly 1 GigaByte), the existing and new codes needed to be compiled for a 64-bit processor. In that regard, all the libraries (ImageMagick, GNU Scientific Library) that were used in the code were compiled for 64-bit processing, and were linked into a 64-bit executable program.

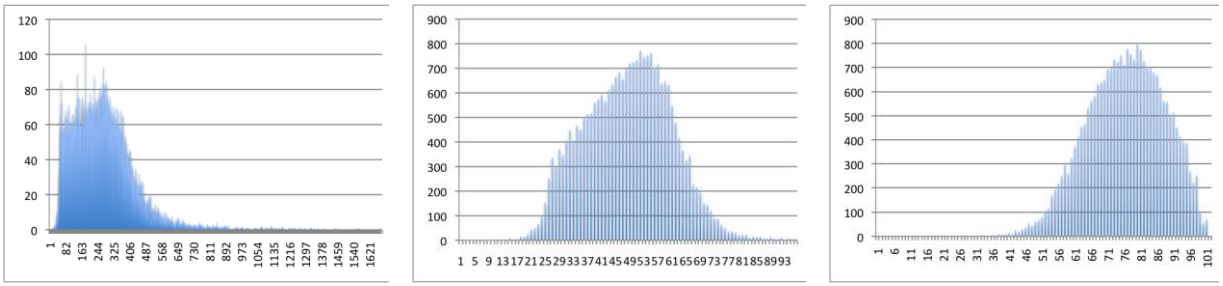


Figure 4. Shape distributions generated from applying the area, area vs. perimeter and aspect ratio metrics to a cell nuclei segmentation image.

Seven geometric metrics have been implemented, debugged and tested, based on programs written in a previous project. They are area, perimeter, area vs. perimeter, aspect ratio, curvature, inside radial contact and line sweep. The area calculates the number of pixels in a single white segmented region. The perimeter metric calculates the number of pixels along the boundary of a white region. The area of a region is divided by its perimeter to produce the area vs. perimeter metric. The aspect ratio metric is defined as the minor axis length of a region divided by its major axis length. The curvature metric is computed as the absolute value of mean curvature at the perimeter pixels. The radial contact measure calculates the shortest distance from a point inside a white foreground region to the black background region. Given a point and a direction within a white region the line contact measure

determines the longest line that can go through the point in the given direction and still fit inside the region.

The shape distributions generated by three of the metrics are presented in Figure 4. Several discrepancies were found while testing and debugging the metric codes. In the area metric the code was processing grayscale values for each pixel of the segmented images. Since the segmented images are binary the area metric code had to be modified to look for binary values (black and white pixel values). The code was fixed to read only binary values and tested against basic geometric shapes (circle, oval etc). In the perimeter metric calculation the same changes were implemented to avoid reading grayscale values. The area vs. perimeter metric calculation follows from the earlier computation of area and perimeter. A slight modification was made to ensure that the codes use the C++ library functions like round, instead of a previously written round function.

The aspect ratio computation uses some library functions from GSL (GNU Scientific Library). While debugging the output from this code certain unexplainable anomalies were found. Later investigation of the code showed that unnecessary rounding operations were being performed on the input being passed to the Eigenvalue computation. To improve the robustness and efficiency of the Eigenvalue computation a new routine was written instead of using the GSL functions. This proved faster and more reliable. Debugging of the curvature metric code on a standard geometric object like an oval showed that the distribution of the curvature values of each boundary pixel was not conforming to the desired histogram. Several errors were found. First. Gaussian blurring was done on the segmented image as a whole. In that case each pixel blur value computation of a segmented nuclei may get some contribution from its neighboring nuclei pixel values. To tackle this problem each nuclei was identified and blurred separately. Further testing helped to find an error in the kernel computation. A problem in computing the kernel was fixed. Testing/fine-tuning was then done by evaluating the results produced by a range of parameter values for the kernel radius and kernel sigma for Gaussian blurring. A test image was blurred various times to investigate the nature/quality of the resulting shape distribution. Finally a radius of 9 pixels and a sigma of value 3 were chosen as optimal for the single blurring that is used during the curvature calculation.

The inside radial contact code was completely rewritten using a flood-fill-like algorithm to calculate the shortest distances needed for the metric. Then it was tested with standard geometric shapes (oval, circle, etc.) to ensure its accuracy and correctness. The line sweep code was not properly computing all the line segments needed to produce the associated shape distribution. One of the corner pixels was overlooked while another corner pixel was considered twice. This problem was fixed. In addition to this, new conditions have been added to detect the following cases: 1) when a line starts in a boundary white pixel and ends in a black pixel, 2) when the line starts in a black pixel then enters a white pixel (that of a segmented nuclei) but does not finish in a black pixel on the image boundary. In both cases the line should be discarded and not contribute to the shape distribution calculation.

Intensity Distribution Generation. Software was developed to generate intensity distributions from each histology scan. This process begins by separating the scans into three color

channels, Hue, Saturation and Value (HSV). An intensity distribution is computed for each channel, producing three distributions per sample. The algorithm for producing the intensity distributions is the following. Given the segmented image of the scan and the three intensity images for Hue, Saturation and Value, iterate through all pixels of the segmented image. White pixels in the segmented image signify stained cancer cell nuclei in the original scans. For each white pixel in the segmented image find the intensity value at the associated pixel in the HSV images. A histogram of intensity values is created for each channel (HSV). The distribution of intensity values within each channel is produced by simply accumulating the values in appropriate bins in the HSV histograms, when processing all white pixels in the segmented image.

Sample Distribution Processing. Using the image processing and segmentation software developed in the early stage of the project, two binary segmented images are generated for each of the 100 histology scans used in our study. One binary image is produced by processing the high resolution scans. The second is produced by applying a morphological closing to the high resolution binary image, and captures the large-scale structure of the invasive cancer cell formations. Our sample set consists of 55 N0 cases (negative for lymph node metastasis) and 45 N1+ cases (positive for lymph node metastasis). These samples were identified and scanned in the early phase of the project. All of the image processing/segmentation, shape distribution and intensity distribution techniques were applied to these sample images. Applying our distribution generation algorithms to these samples produced 11 raw distributions for each case. Six shape distributions (area, perimeter, area/perimeter, aspect ratio, inside radial contact and inside line contact) and three intensity distributions (HSV) are produced from the high resolution images. Two additional shape distributions (inside radial contact and inside line contact) are produced from the closed, large-scale-structure binary images. The curvature metric was not computed because of the excessive computation that would be needed to process all of the histology samples. Future work will investigate techniques for improving the efficiency of this calculation. The range of values present in the inside radial contact and inside line contact distributions were observed to be quite large, varying by many orders of magnitude. Concerned that these great differences in values may adversely affect downstream processing, four additional distributions were produced by transforming the individual bin values for these distributions by the natural log; Thus significantly decreasing the range of values within the inside radial contact and inside line contact distributions.

Dimension Reduction. The distributions, which are represented by histograms, usually have a large and varied number of bins. The number of bins in any one histogram may reach into the 10's of thousands. For statistical analysis each bin may be considered a dimension of a feature space. Having such an enormous feature space makes analysis extremely difficult and problematic. Therefore it is necessary to reduce the dimensionality of the features produced by the shape/color distribution calculations. This is accomplished by calculating characteristic statistics for each distribution, i.e. the distributions are transformed from histograms into a low-dimensional feature vector. For each distribution, the following 10 quantities were computed: mean, median, mode, standard deviation, skew and kurtosis of the values used to make the histograms; and the maximum bin number, maximum value over all bins, mean and standard deviation of the bin values. Additionally, measures that produce distribution histograms of a fixed width (e.g. aspect ratio, hue, saturation and value) can also be averaged down to make a coarser distribution with 10 bins. The dimensionality reduction

step takes histograms that cannot be effectively processed and transforms them into a low dimensional (10) feature vector that is now tractable and analyzable. Once completed this stage produces 17 (13 distributions statistics + 4 averaged down distributions) 10-element feature vectors for each sample. In summary, the 17 feature vectors are: 1. Area, 2. Perimeter, 3. Area divided by Perimeter, 4. Aspect Ratio, 5. Inside Radial Contact, 6. Inside Line Contact, 7. Log of Inside Radial Contact, 8. Log of Inside Line Contact, 9. Log of the Large Scale Structure Inside Radial Contact, 10. Log of the Large Scale Structure Inside Line Contact, 11. Hue, 12. Saturation, 13. Value, 14. Averaged Hue, 15. Averaged Saturation, 16. Averaged Value, and 17. Averaged Aspect Ratio.

Classification via Supervised Machine Learning Techniques. This stage treats the classification process as a data fusion problem, with each one of the 17 feature sets representing a different source of data. Each such source provides some prognostic information, and can be used to train an automated classifier to predict the metastasis status. Based on our preliminary analysis, we concluded that there is complementary information carried by different feature sets, and a strategic combination of such information can significantly improve the prognostic accuracy, compared to any of the feature sets used alone. We used an ensemble of classifiers based approach to data fusion. In this framework different supervised classifiers are trained on data obtained from different sources. In this case, we have 17 such classifiers, specifically 17 Support Vector Machine (SVM) classifiers. Each classifier makes a prediction (yes or no), whether the patient's tumor has metastasized and has a positive lymph node status (N1plus).

This process was further improved, however, by using a "stacked generalization" type of combiner. In stacked generalization, the classifier combiner is an additional "meta-classifier" whose inputs are the outputs of the individual classifiers. In such a framework, the meta-classifier effectively determines how the classifiers should be combined, by analyzing which classifiers are better predictors in which parts of the feature space, or in other words, which data sources / feature sets provide more relevant prognostic information for any given set of values for those features. The meta-classifier can be trained such that their outputs can be interpreted as a lymph node metastasis status probability for each sample. We used Logistic Regression as the "meta-classifier" so that the predictions are not just binary decisions, but also carry a probabilistic measure of the lymph node metastasis status. Logistic Regression fits the data to a logistic sigmoid function with simple probabilistic interpretation of the output, and naturally allows for the creation of an ROC curve, providing effective evaluation of the sensitivity / specificity tradeoff.

In the first stage of the classification process 17 SVMs, one for each feature vector, were used to predict the lymph node metastasis status. This produces a binary classification result, where lymph node positive metastasis status is represented by a "1" and negative status is represented by a "0". The Leave-One-Out (LOO) approach was used to predict the lymph node metastasis status of each sample. 99 samples are used to train the SVM and the remaining sample is then treated as a test sample and classified. This is repeated for each of the 100 samples. Thus for each SVM classifier we compute a 100x1 vector of binary predictions. Once the computation was completed for each of the 17 SVMs, a 100x17 binary matrix is produced, where each row represents the outcomes of the 17 SVMs for each sample. The output of the SVMs for the first 15 samples is given in Figure 5. Both the value of the penalty factor, C, and the selection of kernels were varied to optimize the harmonic

mean of the sensitivity and specificity of the classification results for each SVM. This customized each SVM and ensured that the most effective parameters were used for each feature vector.

DT_log_large	LS_log_large	Area	AR_avg	Asp_ratio	AvP	DT_log	DT	Hue_avg	Hue	LS_log	LS	Perimeter	Sat_avg	Sat	Val_avg	Val	Case
0	0	1	0	1	1	0	0	0	1	1	0	0	1	0	1	0	IG03-015
1	1	1	0	1	0	1	1	0	0	1	1	1	0	0	0	1	IG03-086
1	1	0	1	1	0	0	0	1	1	1	1	1	1	0	1	1	IG03-108
0	1	0	1	1	1	0	0	0	0	0	0	0	0	1	0	1	IG03-159
0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	1	0	IG04-003
0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	IG04-014
0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	IG04-093
1	0	1	0	0	1	1	0	1	0	1	0	0	1	1	0	0	IG04-099
0	0	1	1	1	1	1	1	0	0	1	0	1	0	0	1	1	IG04-242
1	0	1	0	1	1	0	0	0	0	1	0	1	0	1	1	0	IG04-243
1	1	0	0	0	1	0	0	0	1	1	1	1	0	0	0	0	IG04-274
0	1	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	IG05-049
1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	IG05-072
1	0	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0	IG05-088
1	1	0	1	1	1	1	1	0	0	1	1	1	0	0	0	1	IG05-110

Figure 5. SVM classification outputs for the first 15 samples of our study. 1 signifies the sample has been identified as metastatic. 0 signifies the sample has been classified as non-metastatic.

The 100x17 binary matrix was used as input to the "stack-generalization" stage. The Binary Logistic Regression (BLR) classifier found in the SPSS software was used as the "meta-classifier". The outcome of the BLR analysis is the probabilistic confidence measure that a sample has "N1+" status, i.e. that the tumor has metastasized to a lymph node. An outcome of 0.90 for a sample means that the classifier predicts the sample as a N1+ with high confidence. Whereas an outcome of 0.10 means the classifier predicts the sample as a N1+ with low probability, or conversely classifies the sample as N0 with high probability.

Results. The SPSS BLR classifier produced a probabilistic prediction for each sample that it is in the N1+ class. The default threshold value used to classify a sample as "N0" or "N1+" is 0.5. This threshold value produced classifications with a sensitivity of 77.8% and specificity of 85.5% over all 100 samples. Since BLR does not produce a true probability, the threshold may be varied. This generates the Receiver Operating Characteristic (ROC) curve for the overall classification process and allows us to find the optimal threshold value, i.e. the one that produces the best classification results. The optimal threshold value of 0.357 produced a classification sensitivity of 91.0% and specificity of 80.0%, and correctly classified 85 of the 100 samples. The ROC curve is the plot of sensitivity versus (1-specificity) over the full range of threshold values. The ROC curve for the BLR meta-classifier is presented in Figure 6. The area under the curve may be used to characterize the accuracy of the classifier and its ability to separate the two classes. The area under the curve in Figure 6 is 0.90, which qualifies the approach as a good to excellent classifier.

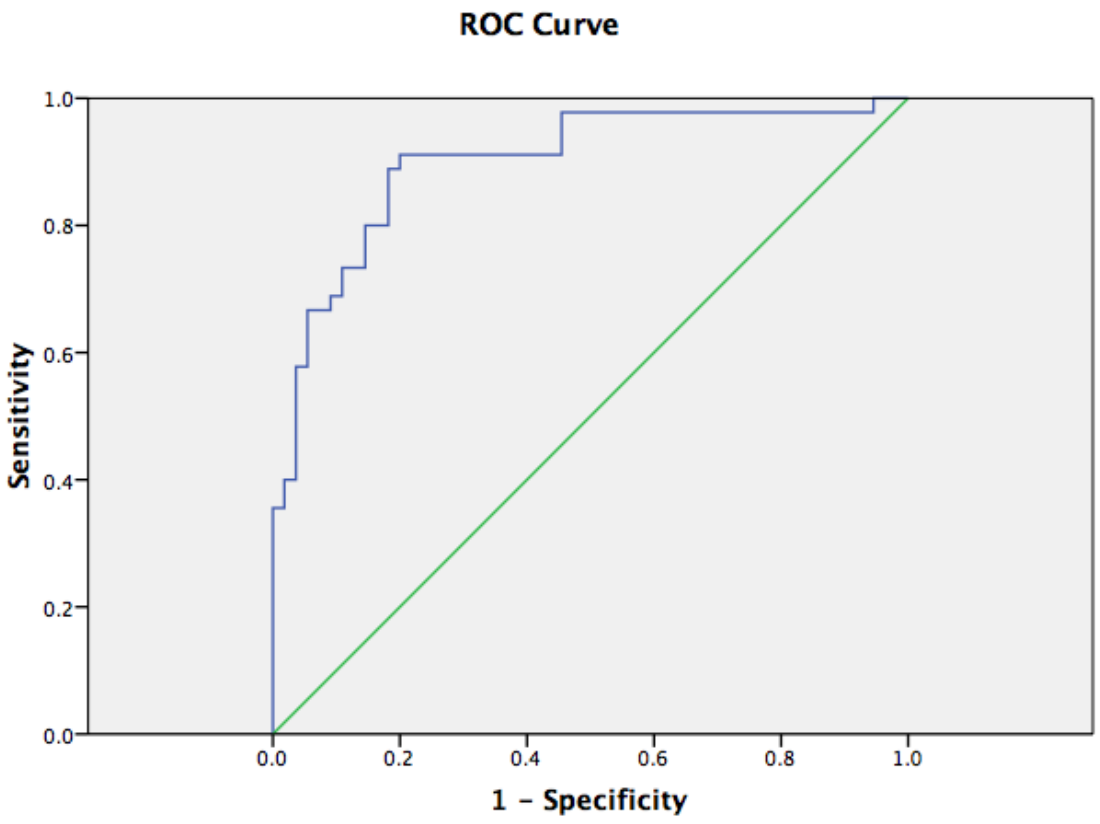


Figure 6. ROC curve for Binary Logistic Regression stacked meta-classifier. Modifying the threshold value that separates N0 and N1+ produces the plotted sensitivity and specificity values.

Presented Posters.

A. Milutinovic, M.A. Reza, D.E. Breen and F.U. Garcia, Multidimensional Shape and Color Distributions as a Computational Biomarker for Cancer Pathology, United States and Canadian Academy of Pathology's 99th Annual Meeting, Washington, DC, March 2010.

M.A. Reza, A. Milutinovic, R. Polikar, F.U. Garcia, D.E. Breen, Multidimensional Shape/Color Distribution for the Prediction of Lymph Node Metastasis Status, Drexel University Research Day, Philadelphia, PA, April 2010.

18. Extent of Clinical Activities Initiated and Completed. Items 18(A) and 18(B) should be completed for all research projects. If the project was restricted to secondary analysis of clinical data or data analysis of clinical research, then responses to 18(A) and 18(B) should be "No."

18(A) Did you initiate a study that involved the testing of treatment, prevention or diagnostic procedures on human subjects?

_____Yes
___X__No

18(B) Did you complete a study that involved the testing of treatment, prevention or diagnostic procedures on human subjects?

_____Yes
___X__No

If "Yes" to either 18(A) or 18(B), items 18(C) – (F) must also be completed. (Do NOT complete 18(C-F) if 18(A) and 18(B) are both "No.")

18(C) How many hospital and health care professionals were involved in the research project?

_____Number of hospital and health care professionals involved in the research project

18(D) How many subjects were included in the study compared to targeted goals?

_____Number of subjects originally targeted to be included in the study
_____Number of subjects enrolled in the study

18(E) How many subjects were enrolled in the study by gender, ethnicity and race?

Gender:
_____Males
_____Females
_____Unknown

Ethnicity:
_____Latinos or Hispanics
_____Not Latinos or Hispanics
_____Unknown

Race:
_____American Indian or Alaska Native
_____Asian
_____Blacks or African American
_____Native Hawaiian or Other Pacific Islander
_____White

_____Other, specify:_____
_____Unknown

18(F) Where was the research study conducted? (List the county where the research study was conducted. If the treatment, prevention and diagnostic tests were offered in more than one county, list all of the counties where the research study was conducted.)

19. Human Embryonic Stem Cell Research. Item 19(A) should be completed for all research projects. If the research project involved human embryonic stem cells, items 19(B) and 19(C) must also be completed.

19(A) Did this project involve, in any capacity, human embryonic stem cells?
_____Yes
___X__ No

19(B) Were these stem cell lines NIH-approved lines that were derived outside of Pennsylvania?
_____Yes
_____ No

19(C) Please describe how this project involved human embryonic stem cells:

20. Articles Submitted to Peer-Reviewed Publications.

20(A) Identify all publications that resulted from the research performed during the funding period and that have been submitted to peer-reviewed publications. Do not list journal abstracts or presentations at professional meetings; abstract and meeting presentations should be listed at the end of item 17. **Include only those publications that acknowledge the Pennsylvania Department of Health as a funding source** (as required in the grant agreement). List the title of the journal article, the authors, the name of the peer-reviewed publication, the month and year when it was submitted, and the status of publication (submitted for publication, accepted for publication or published.). Submit an electronic copy of each publication, listed in the table, in a PDF version 5.0.5 format, 1,200 dpi. Filenames for each publication should include the number of the research project, the last name of the PI, the number of the publication and an abbreviated research project title. For example, if you submit two publications for PI Smith for the "Cognition and MRI in Older Adults" research project (Project 1), and two publications for PI Zhang for the "Lung Cancer" research project (Project 3), the filenames should be:

Project 1 – Smith – Publication 1 – Cognition and MRI
Project 1 – Smith – Publication 2 – Cognition and MRI
Project 3 – Zhang – Publication 1 – Lung Cancer
Project 3 – Zhang – Publication 2 – Lung Cancer

If the publication is not available electronically, provide 5 paper copies of the publication.

Note: The grant agreement requires that recipients acknowledge the Pennsylvania Department of Health funding in all publications. Please ensure that all publications listed acknowledge the Department of Health funding. If a publication does not acknowledge the funding from the Commonwealth, do not list the publication.

Title of Journal Article:	Authors:	Name of Peer-reviewed Publication:	Month and Year Submitted:	Publication Status (check appropriate box below):
1. None				☐Submitted ☐Accepted ☐Published

20(B) Based on this project, are you planning to submit articles to peer-reviewed publications in the future?

Yes____X_____ No_____

If yes, please describe your plans:

A paper describing the project and its results is currently in preparation and will be submitted to the Journal of Pathology Informatics.

21. Changes in Outcome, Impact and Effectiveness Attributable to the Research Project. Describe the outcome, impact, and effectiveness of the research project by summarizing its impact on the incidence of disease, death from disease, stage of disease at time of diagnosis, or other relevant measures of outcome, impact or effectiveness of the research project. If there were no changes, insert "None"; do not use "Not applicable." Responses must be single-spaced below, and no smaller than 12-point type. DO NOT DELETE THESE INSTRUCTIONS. There is no limit to the length of your response.

None, at this time, since the project is still in the laboratory research phase.

22. Major Discoveries, New Drugs, and New Approaches for Prevention Diagnosis and Treatment. Describe major discoveries, new drugs, and new approaches for prevention, diagnosis and treatment that are attributable to the completed research project. If there were no major discoveries, drugs or approaches, insert "None"; do not use "Not applicable." Responses must be single-spaced below, and no smaller than 12-point type. DO NOT DELETE THESE INSTRUCTIONS. There is no limit to the length of your response.

Verification of our image-based approach for determining metastasis status via analysis of the morphology of the primary tumor is a significant technical milestone. We see this achievement as note-worthy because our technology should lead to a low-cost technique for estimating metastasis status of breast cancer patients. More importantly, a considerable

number of patients with low probability of lymph node metastases might avoid axillary dissection and its associated critical side effects (e.g. lymphedema, numbness, pain, infection and compromised immunity) and morbidity. Since the technique is only based on computational analysis of standard histology slides, any health facility that handles and diagnoses breast carcinoma specimens will be able to perform the lymph node metastasis prediction. If the necessary software is locally unavailable, a web-based application can be developed to process the histology slides from underserved areas. Obviating the need for additional surgeries will provide significant cost savings. The immediate impact of our research is that all types of health facilities, e.g. those in rural areas of the United States or even in the Third World, will easily and affordably be able to perform this advanced diagnosis. Image-based diagnostic tools that utilize statistical analysis of histology scans can also potentially be used for cancer at other frequent sites like prostate, lung and colon.

23. Inventions, Patents and Commercial Development Opportunities.

23(A) Were any inventions, which may be patentable or otherwise protectable under Title 35 of the United States Code, conceived or first actually reduced to practice in the performance of work under this health research grant? Yes_____ No___X_____

If "Yes" to 23(A), complete items a – g below for each invention. (Do NOT complete items a - g if 23(A) is "No.")

a. Title of Invention:

b. Name of Inventor(s):

c. Technical Description of Invention (describe nature, purpose, operation and physical, chemical, biological or electrical characteristics of the invention):

d. Was a patent filed for the invention conceived or first actually reduced to practice in the performance of work under this health research grant?
Yes_____ No____

If yes, indicate date patent was filed:

e. Was a patent issued for the invention conceived or first actually reduced to practice in the performance of work under this health research grant?
Yes_____ No____
If yes, indicate number of patent, title and date issued:
Patent number:
Title of patent:
Date issued:

f. Were any licenses granted for the patent obtained as a result of work performed under this health research grant? Yes_____ No____

If yes, how many licenses were granted?_____

g. Were any commercial development activities taken to develop the invention into a commercial product or service for manufacture or sale? Yes___ No____

If yes, describe the commercial development activities:

23(B) Based on the results of this project, are you planning to file for any licenses or patents, or undertake any commercial development opportunities in the future?

Yes_____ No____X_____

If yes, please describe your plans:

24. Key Investigator Qualifications. Briefly describe the education, research interests and experience and professional commitments of the Principal Investigator and all other key investigators. In place of narrative you may insert the NIH biosketch form here; however, please limit each biosketch to 1-2 pages. *For Nonformula grants only – include information for only those key investigators whose biosketches were not included in the original grant application.*

BIOGRAPHICAL SKETCH

Provide the following information for the key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FOUR PAGES.**

NAME	POSITION TITLE
Breen, David Edward	Associate Professor of Computer Science
eRA COMMONS USER NAME BREEN1	

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, and include postdoctoral training.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	YEAR(s)	FIELD OF STUDY
Colgate University, Hamilton, NY	BA	1982	Physics
Rensselaer Polytechnic Institute, Troy, NY	MS	1985	Computer & Systems Engineering
Rensselaer Polytechnic Institute, Troy, NY	PhD	1993	Computer & Systems Engineering

A. Positions and Honors.

Positions and Employment

1982-1985 Graduate Research Assistant, Center for Interactive Computer Graphics
Rensselaer Polytechnic Institute, Troy, NY
1985-1993 Research Engineer, Visual Technologies Program, Design Research Center
Rensselaer Polytechnic Institute, Troy, NY
1987-1988 Visiting Research Engineer, Computer Animation Group
Fraunhofer Institute for Computer Graphics, Darmstadt, Germany
1994-1996 Member of Research Staff, User Interaction & Visualization Group
European Computer-Industry Research Centre, Munich, Germany
1996-2002 Assistant Director, Computer Graphics Laboratory,
Department of Computer Science
California Institute of Technology, Pasadena, CA
2002-2003 Senior Research Scientist, Center for Advanced Computing Research
California Institute of Technology, Pasadena, CA
2003-2009 Assistant Professor, Department of Computer Science
Drexel University, Philadelphia, PA
2009-present Associate Professor, Department of Computer Science
Drexel University, Philadelphia, PA

Professional Memberships

1986-present Member, IEEE Computer Society

1986-present Member, Association for Computing Machinery

2000-present Member, Eurographics Association

Honors

1982 Awarded Garden State Graduate Fellowship
1996 The Literati Club Award for Best Paper of the Year in the International Journal of Clothing Science and Technology (with D. House and R. DeVaul).
2009 NSF CAREER Award
2009 VX Corporation Idea Award (with K. Swanson and K. Brakke) for work related to "a next-generation, unified data representation for digital product design, analysis and manufacture."

B. Selected peer-reviewed publications. *(Selected from 80 peer-reviewed publications)*

Book

1. D.H. House and D.E. Breen (eds.), Cloth Modeling and Animation, AK Peters, Natick, MA, 2000

Book Chapters

2. D. Breen, R. Whitaker, K. Museth and L. Zhukov, "Level Set Segmentation of Biological Volume Datasets," J. Suri (ed.), Handbook of Medical Image Analysis, Volume I: Segmentation Part A, Kluwer, New York, Chapter 8, pp. 415-478, 2005
3. K. Museth, R.T. Whitaker, and D.E. Breen, "Editing Geometric Models," S. Osher and N. Paragios (eds.), Geometric Level Set Methods in Imaging, Vision and Graphics, Springer, New York, Chapter 23, pp. 441-460, 2003

Journal Publications

4. J. Bradshaw, P.L. Gurian, A. Kumar and D. Breen, "Spatial Trends Shed Light on Arsenic Concentrations," OpFlow, Vol. 36, No. 1, pp. 18-21, January 2010
5. M. Eyiyurekli and D. Breen, "Localized Editing of Catmull-Rom Splines," Computer-Aided Design and Applications, Vol. 6, No. 3, 2009, pp. 307-316
6. K.W. Swanson, K.A. Brakke and D. Breen, "Physics-based Surface Modeling using Quasi-Static Liquids," Computer-Aided Design and Applications, Vol. 6, No. 6, pp. 759-768, 2009
7. L. Bai and D. Breen, "Calculating Center of Mass in an Unbounded 2D Environment," Journal of Graphics Tools, Vol. 13, No. 4, pp. 53-60, 2008
8. S. Petushi, J. Marker, J. Zhang, W. Zhu, D. Breen, C. Chen, X. Lin, F. Garcia, "A Visual Analytics System for Breast Tumor Evaluation," Analytical and Quantitative Cytology and Histology, Vol. 30, No. 5, pp. 279-290, October 2008
9. M. Eyiyurekli, P. Manley, P. Lelkes and D. Breen, "A Computational Model of Chemotaxis-based Cell Aggregation," BioSystems, Vol. 93, No. 3, pp. 226-239, September 2008
10. I. Braude, J. Marker, K. Museth, J. Nissanov and D. Breen, "Contour-Based Surface Reconstruction using MPU Implicit Models," Graphical Models, Vol. 69, No. 2, pp. 139-157, March 2007

BIOGRAPHICAL SKETCH

NAME	POSITION TITLE
Fernando U. Garcia, M.D.	Professor in Pathology

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, and include postdoctoral training.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	YEAR(s)	FIELD OF STUDY
Universidad Peruana Cayetano Heredia	MD	1981	Medicine
University of Kansas, Kansas City, Kansas		1982-1986	Resident in Pathology
Vanderbilt University, Nashville, Tennessee		1986-1987	Fellow in Surg Pathology
Vanderbilt University, Nashville, Tennessee		1987-1988	Chief Resident
Vanderbilt University, Nashville, Tennessee		1988-2000	Instructor in Pathology

Professional Experience:

1987-1990 Instructor, Department of Pathology, Vanderbilt University, Nashville, Tennessee.
1988-1990 Consultant, Laboratory Services, Veterans Administration Medical Center, Nashville, Tennessee.
1990-1991 Clinical Instructor, Pathology Department Kansas University, Kansas City, Kansas.
1990-1994 Director, Tissue Culture Core Facility, Pathology Department, Kansas University Medical Center, Kansas City, Kansas.
1991-1994 Medical Director, Cytopathology Laboratory Pathology Department, Kansas University, Kansas City, Kansas.
1991-1994 Assistant Professor, Pathology Department Kansas University, Kansas City, Kansas.
1992-1994 Consultant, Laboratory Services, Veterans Administration Medical Center, Leavenworth, Kansas.
1992-1994 Director of Dermatopathology, Kansas University, Kansas City, Kansas.
1994-1994 Director of Image Analysis, Kansas University, Kansas City, Kansas.
1995-2000 Associate Professor MCP-Hahnemann University Philadelphia, Pennsylvania.
1997-2001 Medical Director, Immunopathology Laboratory, MCP-Hahnemann University, Hahnemann University Hospital.
2001- 2005 Medical Director, Immunopathology, Department of Pathology, Drexel University College of Medicine Hahnemann Hospital, Philadelphia.
1997- Medical Director, Image Analysis Laboratory, Department of Pathology, Drexel University College of Medicine Hahnemann Hospital, Philadelphia.
1997- Medical Director, Tissue Procurement Core Laboratory, Department of Pathology, Drexel College of Medicine, Philadelphia.
1998- Regional Director of Anatomic Pathology, Department of Pathology, Drexel University College of Medicine, Philadelphia.
2000-Present Professor of Pathology, Drexel University College of Medicine Philadelphia, Pennsylvania.
2001-2005 Medical Director & Service Chief, Graduate Hospital
2005- Medical Director & Service Chief, Hahnemann Hospital

Certifications:

1981 Peru Medical License (13087)
1981 Educational Commission for Foreign Medical Graduates Certificate (329-894-0)
1984 Visa Qualifying Examination (VQE)

1985	Federal Licensure Examination (FLEX)
1987	Tennessee Medical License (MD 17629) Not Active
1987	Diplomate of the American Board of Pathology in Anatomic Pathology
1990	Kansas Medical License (4-23066) Not Active
1993	Diplomate of the American Board of Pathology in Cytopathology
1995	Pennsylvania Medical License (MD-054895-L)

Publications (last four years)

1. Wang,JY ,Ho, T., Trojanek,J,, Chintapalli,J, Grabacka,M, Stoklosa,T, **Garcia, FU.** Skorski,T and Reiss,K. Impaired homologous recombination DNA repair and enhanced sensitivity to DNA damage in prostate cancer cells exposed to anchorage-independence. Oncogene (2005) 24, 3748-3758.
2. Popnikolov, NK, Cavone, SM, Schultz, PM, **Garcia, FU**. Diagnostic utility of p75 neurotropin receptor (p75 ^{NTR}) as a marker of breast myoepithelial cells. Modern pathology, 2005, 18:1535-1541.
3. Chen, X, Zhuo, Salim,S and **Garcia, FU**. Intraprostatic spermatozoa: Zonal distribution and association with atrophy. Human Pathology, 2006 37(3):345-351.
4. Petushi, S, **Garcia, FU**, Haber, MM, Katsinis, C , Tozeren, A. Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer. *BMC Medical Imaging* 2006, **6:**14 doi:10.1186/1471-2342-6-14.
5. 15. Raab, SS, Geisinger, KM, Grzybicki, DM, Janosky, JE, **Garcia, FU**, Wojcik, EM, Dahmoush, L, Stone, CH, Meier, FA, Zarbo, RJ. Use of a New Method in Reaching Consensus on the Cause of Cytologic-Histologic Correlation Discrepancy. Am J Clin Pathol 2006;126(6):836-842.
6. Mahmoodi, M, Zhang, S, Salim, S, Hou, JS, **Garcia, FU**. Lipofucsin pigment can be used as a prognostic marker in prostatic adenocarcinoma. Annals in Diagnostic Pathology, 2006; 10(5):257-262.
7. Hu,Y ,Wang, M, Veverka, K, **Garcia, FU** and Stearns, ME. The ABCA5 Protein: A Urine Diagnostic Marker for Prostatic Intraepithelial Neoplasia. Clinical Cancer Research 13, 929-938, February 1, 2007.
8. **Garcia, FU,** Urbanska, K, Koltowski, L, Reiss, K Sell C. IGF-I produced by seminal vesicles: relationship to intraepithelial basal cell hyperplasia in the prostate. *Clinical Cancer Research* 13, 3140-3146, June 1, 2007
9. Vamvakidou, AP, Mondrinos, MJ, Petushi, SP, **Garcia, FU**, Lelkes, PI, Tozeren, A. Heterogenous breast tumoroids: An in vitro assay for investigating cellular heterogeneity and drug delivery. J Biomol Screen 2007 Feb;12(1):13-20.
10. Maraldo,D, **Garcia, FU**, Mutharasan, R. Method for quantification of a prostate cancer biomarker in urine without sample preparation. Anal Chem 2007 Oct 15;79(20):7683-90.
11. **Garcia, FU**, Haber, MM, Chen X. Prostatic basal cells in the peripheral zone and transitional zones: Zonal variation in morphology and in immunophenotype. Prostate 2007 Nov 1;67(15):1686-92.
12. **Garcia, FU**, Sharma, M, Nagle, D, Haber, MM. Increase sensitivity of anal cytology in evaluation of internal compared with external lesions. Acta Cytologica 2007, 51:893-899.
13. Patel, AA, Gilbertson, JR, Showe, LC, London, JW, Ross, E, Ochs, MF, Carver, J, Lazarus, A, Parwani, AV, Dhir, R, Beck, JR, Liebman, M, **Garcia, FU**, Prichard, J, Wilkerson, M, Herberman, RB, Becich, MJ and the PCABC. A novel cross-disciplinary multi-institution approach to translational cancer research: Lessons learned from the Pennsylvania Cancer Alliance Bioinformatics Consortium (PCABC). Cancer Informatics, 2007:3:255-274.
14. Zhang, JZ, Petushi, S, Regli, WC, **Garcia, FU**, Breen, DE, A Study of Shape Distributions for Estimating Histologic Grade. Proc. 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, August 2008, pp. 1200-1205.
15. Petushi, S, Marker, J, Zhang, J, Zhu, W, Breen, DE, Chen, C, Lin, X, **Garcia, FU**, "A Visual Analytics System for Breast Tumor Evaluation," Analytical and Quantitative Cytology and Histology, Vol. 30, No. 5, pp. 279-290, October 2008