A Visual Analytics System for Breast Tumor Evaluation

S. Petushi, M.S.C.E.
Research Engineer
Department of Pathology
245 N. 15th Street, NCB-5138A
Drexel University College of Medicine
Philadelphia, PA 19102
Tel: (215) 762-8657
FAX: (215) 762-3274
email: spetushi@drexelmed.edu

J. Marker, B.S.
Undergraduate Student

J. Zhang, B.S.
Graduate Student

W. Zhu,
Graduate Student

D. Breen, Ph.D.
Assistant Professor

C. Chen, Ph.D.
Associate Professor

X. Lin, Ph.D.
Associate Professor

F. Garcia, M.D.
Professor

The U.S. Army Medical Research Acquisition Activity, 820 Chandler Street, Fort Detrick, MD 21702-5014 is the awarding and administering acquisition office. This investigation was funded under a U.S. Army Medical Research Acquisition Activity; Cooperative Agreement W81XWH 04-1-0419. The content of the information herein does not necessarily reflect the position or the policy of the U.S. Government or the U.S. Army and no official endorsement should be inferred. Additional support was provided by NSF Grant ACI-00832.

Abstract
Objective: To develop a system for the interactive exploration and examination of histology-derived data associated with breast tumors, which may be used to evaluate the histologic grade of the tumor.
Study Design: The system integrates pathologist-generated prognostic data with 2D image analysis data, 2D digital tissue cross-sections and annotations, 3D tumor reconstructions and volumetric analysis, 3D spatial tumor display, and recorded prognostic information from available cases in the DUCOM tumor databank. The system consists of three components: 1) a user interface for applying two-dimensional (2D) image processing, segmentation and annotation to a digitized histology slide, 2) a distance field interpolation method for contour-based three-dimensional (3D) reconstruction of breast tumors, and volumetric model analysis routines, 3) a web-based database management interface for interactive data browsing and searching, and multi-modality visualization.
Results: The system has been implemented and deployed with data from 36 breast cancer cases, 7 of which have been reconstructed in 3D.
Conclusion: Interactive visual analytics technology may be used to create an effective breast tumor evaluation system.

Keywords: visual analytics, image processing, segmentation, 3D reconstruction, image/model analysis, database management, visualization, user interface

Introduction

As part of current standard patient care, pathologist-based evaluation of the specimen and tissue slides of a patient diagnosed with breast cancer is considered the gold standard for tissue neoplasm assessment and the most accurate method for determining histologic grade and pathologic stage. Pathologic findings considered to be of significant prognostic value are the tumor size, and histologic type and grade. To assess each finding pathologists perform specific tasks such as measuring and sectioning the specimen, measuring the tumor size, microscopically examining the H&E stained slides and manually measuring the margins of the specimen. In addition, immunohistochemical profiles of breast cancer specific biomarkers are produced. As a result multiple prognostic information is generated from each task, but the usefulness of the data can be diminished if not visualized and analyzed simultaneously. In addition, a histopathology examination is based on manual observation of two dimensional tissue slides. This examination is strongly dependent on observer variation and, the spatial focus of observation, and does not normally take into account findings from previous breast cancer cases using a systematic quantitatively supported procedure based on the simultaneous viewing and analysis of the prior findings.

This paper presents a visual analytics system for breast tumor evaluation. Visual analytics is the science of analytical reasoning supported by an interactive interface that enables diverse data visualization and visual information correlation. It is clear that an effective breast tumor evaluation system will require the simultaneous consideration of possibly all generated data supported by information recorded for previous breast tumor cases, including the outcomes of those cases.

Related research work on breast cancer diagnosis and prognosis has resulted in computer-assisted systems that mainly focus on data extraction and / or interpretation and / or visualization of individual phases / tasks within the entire diagnosis / prognosis process. These systems do not simultaneously consider and interactively visualize in parallel the diverse information generated for a single patient, in combination with corresponding information from previous cases. Street et al.[1], introduced Xcyt (available for free testing online), a software system that provides diagnosis and prognosis of breast cancer based on FNA (Fine Needle Aspiration). Xcyt is based entirely on information obtained from morphometric analysis of the individual tumor cells, along with the size of the tumor itself. Kayser et al.[2], proposed EAMUS (Electronic Automated Measurement User System), an automated image measurement system for immunohistochemically stained slides including fluorescence images. EAMUS uses an active stratified sampling method to identify and measure objects present in images of immuno-stained slides. In related work Kayser et al.[3] show that texture-based analysis may be used to classify normal histological findings and several tissue-based diagnoses. Görtler et al.[4] contend that tissue-based diagnoses may greatly benefit from grid computing technology. Leong et al.[5], presented a computer-based automated histopathology recognition system to distinguish benign from malignant lesions. A system for combined three-dimensional morphological and molecular analysis of thick tissue samples is proposed by Fernandez-Gonzalez et al.[6]. This system uses more advanced computational techniques and integrates a three-dimensional visualization system with an image analysis system. An extended review of proposed and developed systems related to breast cancer diagnosis is given in [7]. In another medical area, the integrated medical data analysis and visualization system 3D Slicer[8] (open source software) is an advanced tool for

brain-surgery planning and guidance using image fusion and an open MRI (Magnetic Resonance Imaging). 3D Slicer allows the incorporation of multiple data sets into a single display environment and provides capabilities for automatic registration, semiautomatic segmentation, 3D surface model generation, 3D visualization, and quantitative analysis of various medical scans.

The system presented in this paper is composed of three modules and uniquely bundles multiple available breast cancer data with different aspects of analysis and visualization into a single interactive framework. It integrates pathologist-generated prognostic data with 2D image analysis data, 2D digital tissue cross-sections and annotations, contour-based surface reconstruction, 3D tumor volumetric analysis, 3D spatial tumor display, and recorded prognostic information from previous cases. This paper provides an overview of the system by describing the individual modules and illustrating the major steps involved in utilizing it.

Materials and Data

This study employs retrospective breast cancer data available in the Department of Pathology of Drexel University College of Medicine (DUCOM). The original research was approved by the local institutional review board (DUCOM - IRB). Selected data were composed of: a) de-identified prognostic data corresponding to 324 breast cancer cases; b) 44 hematoxylin & eosin (H&E) stained histology slides spanning 7 breast tumors entirely submitted from seven different breast cancer patients that had a lumpectomy as a primary form of treatment.

Prognostic data was comprised of the following information: age, tumor size, tumor stage, tubular formation, mitotic rate, nuclear grade, histologic grade, estrogen receptor value, progesterone receptor value, Ki67 value, p53 value, and Her2neu value.

H&E stained histology slides were created as following: a) in two cases, three cross-sections were cut for each formalin-fixed (paraffin-embedded) block (top, quarter, and middle); b) in the remaining five cases, for each paraffin-fixed block a cross-section slice was cut giving a slide per block, thus each block was represented by only one cross-section cut at the top of the block. See Figure I-E. Note that each tumor case is comprised of multiple paraffin-embedded tissue blocks (Fig.I-B).

As part of the standard protocol in the department, all breast specimens have been submitted entirely using a standardized method. This protocol includes serial sectioning of the specimen at 5 mm intervals with consecutive ordering and maintenance of tissue orientation in 6 planes (Fig. I-A&B). Paraffin embedded 4µm hematoxylin and eosin stained sections of each block have been created and microscopically reviewed to allow classification of the tumor. The database includes all types of breast cancer. For the purpose of this study, only lumpectomy cases were considered for the 2D image processing and 3D reconstructions.

Lumpectomy is the form of breast cancer surgery where the part of the breast containing the tumor (the "lump") and some of the normal tissue surrounding it is removed.
Invasive means that the cancer has "invaded" or spread to the surrounding tissue.
Size is the estimated diameter of the tumor as measured during the tissue preparation.
The next section provides an overview of the system, with each module described and illustrated separately in detail.

System overview

Our visual analytics systems (Figure II) for breast tumor evaluation consists of three major components: tissue image digitization and 2D image processing; 3D tumor reconstruction and volumetric analysis; and web-based database management and data visualization.
In the first component, images of serially-cut tumor histology glass slides, representing the entire tissue, are automatically acquired and converted to digital images. A user interface is utilized to apply 2D image processing, segmentation and annotation in each slide, and to organize and pre-process images for 3D reconstruction. Representative regions of interest are computationally identified and re-scanned at a higher magnification. Image analysis is applied to these regions in order to extract multiple discriminating features. In the second component, a distance field interpolation method for contour-based 3D-reconstruction is applied to the stack of 2D annotated images generated in the first component. This produces a 3D model with separate structures for each of the segmented tumor regions. Volumetric measurements are performed and the 3D tumor reconstructions are interactively visualized. In the third component, a web-based database management interface is developed to allow for interactive data browsing and searching, and multi-modality visualization of the available resources. The following sections provide detailed description and illustrations for each component.

Tissue Image Digitization, 2D-Image Processing/Analysis, and Tumor Annotation

Using the available breast cancer database, de-identified breast cancer data are filtered to create a pool of cases that would satisfy the following criteria:

$$\text{Pool of Cases} \in \{\text{Lumpectomy} \rightarrow \text{Invasive} \rightarrow 1.00\ \text{cm} < \text{Size} < 4.00\ \text{cm}\}$$

A subset of cases is selected from this pool and the associated 2D tumor grossing maps are screened to satisfy the grossing pattern shown in Figure I-B; thus providing consistent input to the 3D tumor reconstruction process. This pattern indicates that the central portion of the specimen containing the tumor is cut only horizontally throughout the vertical axis and the entire tissue of cross-sections cut from each paraffin-fixed block is located in one single slide each (Figure I-B-D). The available H&E histology slides of the selected cases are microscopically reviewed one by one by an expert pathologist to locate the slides that are positive for cancer and have no major cracks or holes caused during slide preparation. In some cases certain slides are re-cut when the tissue cross-section in the slide is not complete or the slide itself is missing from the archive.

The entire tissue area within the H&E glass slides is then automatically digitized, for each stack of serially cut cross-sections, using our motorized histology slide imaging system. This system is comprised of a customized Olympus BX60 microscope equipped with a LUDL's BioPrecision motorized stage, motorized filter wheel, iris and objectives turret, a high-speed acquisition and tiling system (Objective Imaging), which uses a customized PCI card to synchronize camera, software and motorized stage, and a Retiga 2000R digital color camera. The histology slides are initially scanned using a 2X Olympus PlanApo objective to provide an overview of the entire slide and automatically identify the smallest bounding rectangle that surrounds the entire tissue area. The coordinates of this bounding rectangle are used to re-scan

the non-empty tissue area with a 10X Olympus UPlanApo objective, allowing for the visualization of the distribution of cells and tissue structures.

A Dual-Core Intel Xeon 3.0GHz, 6GB Ram, Windows XP 64-bit Dell Precision workstation is used for the image processing and analysis. A prototype user interface is utilized to apply 2D image processing, segmentation and annotation of tumor-tissue structures in each slide, and to organize and pre-process images for 3D reconstruction (Fig. III). As a first step, an automated image processing approach previously described by Petushi et al.[9, 10] is applied to each stack of histology images to perform segmentation. The approach is a hybrid method that bundles within an automated framework: grayscale conversion, segmentation with adaptive thresholding and morphological operations, blob (micro-object) labeling, feature extraction, blob identification, and nuclei classification using supervised learning (Fig. IV). This process identifies the spatial positions of hundreds of thousands of cell nuclei in three morphology groups in a single whole-section image, as well as adipose tissue and extracellular matrix as a background. The result is used to estimate the density distributions of the different cell morphologies identified earlier. The generated density distribution images are then segmented and their morphology and textures are analyzed, using the corresponding original color images and cell-level identification results as a reference, to detect higher order tumor tissue structures/regions such as: invasive tumor, ductal carcinoma in situ (DCIS), cancerization of lobules, fibroadenoma, vessels, necrotic areas, etc. Using the graphical user interface (Fig. III-A) the identified tumor tissue structures are overlaid on the original corresponding areas, enabling manual correction of region boundaries and structure annotation.

For each tumor case, representative/hot regions of interest (ROI) are computationally identified as areas of high cell-concentration and are visually verified. These hot ROIs are re-scanned at a higher magnification and segmented using the method described in[9, 10,]. Image analysis is then applied in order to extract multiple discriminating-capable features[11, 12]. The mean and coefficient of variance for the measurements shown in Table 1 are calculated for each ROI. *Coefficient of Variation* (CoV) is a measure of the dispersion of a probability distribution. It is defined as the ratio of the standard deviation σ to the mean μ: CoV = σ / μ.

Within this module, each stack of H&E stained histology slides is digitized, processed and annotated to create the input data for the second component of our method. A flowchart representing the data flow from the first to the second module is shown in Figure V. In addition cell-level 2D image analysis measurements of representative regions of interest are generated and visualized using the third module of this approach. Discriminating capabilities and the diagnostic/prognostic value of these 2D image analysis measurements are not within the scope of this paper and therefore are not considered here.

3D-Tumor Reconstruction and Volumetric Measurements

The first component of our visual analytics system produces a set of segmented and classified images for each histology scan. See Figure VII (left). The second component of the system contains a computational pipeline that produces smooth surface reconstructions from a set of parallel binary contours[13]. Contours are generated from the boundaries of the classified regions, and are stored as binary images where white pixels represent the contour curves. The

reconstruction process generates an isosurface embedded in a volume dataset by first calculating distance fields in the individual 2D slices. Blending slices are computed between the input contours via spline interpolation of associated pixels in neighboring input slices. The zero isosurface embedded in the resulting volume provides the desired reconstruction. The complete computational process contains several stages. The centers of the contour (white) pixels are interpreted as points in 2D and a Multi-level Partition of Unity (MPU) implicit curve[14] (i.e. a 2D field whose zero level set is the curve) is approximately fit to these points. The narrow band around the MPU curve is swept out by a fast marching method[15] to produce a 2D Euclidean distance field. The medial axis discontinuities inherent in all Euclidean distance fields are smoothed with distance-dependent Gaussian filtering[13]. A volume dataset is produced via monotonicity-constraining spline interpolation[14] of pixels across neighboring distance fields. We employ this type of spline in order to remove undesirable surface artifacts produced by the overshoot normally found in standard splines. A mesh of the isosurface that represents the reconstructed surface may then be extracted from the volume[17].

Once a 3D model has been constructed it is analyzed to produce a set of 3D shape measures. The 3D measures include: surface area, aspect ratio, spherical eccentricity, volume, density, and absolute mean curvature per unit area. **Surface area** over the final reconstruction is calculated by summing the area of the individual triangles in the triangle mesh, which has been extracted from the interpolated volume. The surface area of each triangle is calculated by:

$$\left| \frac{edge_1 \otimes edge_2}{2} \right|.$$

The $\otimes$ operator is the cross product of two vectors defined by two edges *(edge$_1$, edge$_2$)* of the triangle. **Aspect ratio** is obtained by first calculating the eigenvalues of the covariance matrix derived from vertices of the extracted surface,

$$\begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix}.$$

$$D_{\alpha\beta} = \frac{1}{n} \sum_{i=1}^{n} (p_\alpha^i - \overline{p}_\alpha)(p_\beta^i - \overline{p}_\beta),$$

where *n* is the number of vertices in the extracted mesh, $\overline{p}$ is the centroid of all the vertices, and α and β represent the X, Y and Z components of the vertices. Aspect ratio is calculated by dividing the smallest eigenvalue by the largest eigenvalue and indicates if the object is elongated. **Eccentricity** is calculated as the ratio of an object's surface area and the area of a sphere with the same volume,

$$Eccentricity = \frac{SurfaceArea_{Object}}{SurfaceArea_{Sphere}} = \frac{SurfaceArea_{Object}}{4 * \pi * \left(\sqrt[3]{\frac{Volume_{Object}}{(4/3) * \pi}} \right)^2}$$

The volume of a volumetric object is easily approximated by summing the number of voxels marked as "inside" the object of interest. Eccentricity provides a measure of the irregularity of the object's shape. **Density** is a ratio of the object's volume and the volume of the bounding box surrounding the object. The **absolute mean curvature per unit area** is calculated by summing the absolute value of the mean curvature κ over the surface of the object of interest,

$$\frac{1}{Area}\int_S |\kappa| dS$$

The mean curvature calculation can be performed on the derived volumetric dataset at the voxels near the object's surface with the following function[18],

$$\frac{\left(\phi_y^2+\phi_z^2\right)\phi_{xx}+\left(\phi_x^2+\phi_z^2\right)\phi_{yy}+\left(\phi_x^2+\phi_y^2\right)\phi_z'^2-2\phi_x\phi_y\phi_{xy}-2\phi_x\phi_z\phi_{xz}-2\phi_y\phi_z\phi_{yz}}{\left(\phi_x{}^2+\phi_y{}^2+\phi_z{}^2\right)^{\frac{3}{2}}}$$

where ϕ is the implicit function stored in the volume dataset and the subscripts denote specific partial derivatives. This quantity provides a measure of the roughness of the object's surface.

Web-based Interface for Visual Analysis and Tumor Evaluation

The third component of our system provides a Web-based integrated approach to using the datasets and images for visual analysis and tumor evaluation. The integrated interface was implemented as a Java Web application based on open source software (Tomcat Application server and MYSQL database). Described below are its three major components: an image viewing front-end, a back-end data management tool, and a visual analysis tool.

Image Viewing

A main feature of the system allows the user to view four different types of images: raw images, segmented images, 3D moving images (animations), and zoomable high-resolution image (Figure VIII). A great deal of data manipulation is needed to provide these image views. For example, to provide a single high-resolution image that can be enlarged to up to 20x, the system must work with on an average more than 70 Megabytes of image data stored in more than 1000 files. A special viewer is required (Currently, we use Zoomifyer EZ viewer available from http://www.zoomify.com). Similarly, to view a 3D reconstruction model directly would require downloading and processing more than 600 megabytes of data. This is not practical for interactive viewing in the current networked environment. Thus, we have created a QuickTime animation for each case for the 3D displays.

All data manipulation happens on the server. The user only interacts with an easy-to-use Web interface that supports various image browsing and searching functions. For example, the user can view a list of thumbnails of all of the cases at a glance, and select an individual case to view any of the four image types. The user can also search a case by patient's age or by tumor size, histologic grade, etc. (Figure IX).

Data Management

The image browsing front-end interacts with a server-based Java application in the back-end to provide all the data and images. The data are all stored in a relational database. For each case, four types of data are included: metadata, prognostic data, 2D analysis data, and 3D analysis data. Furthermore, a separate Web interface is available for data management for the system administrators. The administrators can insert, edit, or delete data through a Web browser. They can also upload prognostic or analysis data from Excel files to the database. In addition, the administrators also have access to tools for maintaining user accounts for the integrated system.

Visual Analysis

A separate visual analysis tool is provided that allows users to explore relationships between independent and dependent variables for the cancer cases in the database. A network model for the cases is constructed with the dependent variables shown as nodes. These dependent variables represent 9 parameters that are routinely generated and used by pathologist in breast cancer diagnosis (see Table 2). Each case is represented internally as a vector of selected independent variables. The similarity between two cases is calculated with the cosine similarity between the two vectors:

$$Sim(i, j) = \sum_{k=1}^{N} v_{ik} v_{jk} / \sqrt{\sum_{k} v_{ik}^{2} \times \sum_{k} v_{jk}^{2}}$$

Where v_{xy}'s are vector coefficients of the corresponding cases.

The edges between the cases depict the degree of similarity between the cases. The coordinates of these cases in the network are based on the Kamada-Kawai graph layout algorithm[19]. The layout algorithm tends to group similar cases near each other and place dissimilar ones further apart. As a result, a cluster of nodes indicates a group of similar cases in terms of the selected independent variables. It becomes even clearer when we apply colors to nodes. The colors represent the values of any one of the 9 parameters in Table 2. For example, Tubule Formation in the top row has three levels: 1, 2, and 3. The three levels are shown in the visualization with three different node colors, yellow, orange and red. The remaining 8 parameters are used to visualize their relations. In the example shown in Figure X, the three different node colors represent the three values of the Mitotic Rate and show their relation to histological grade. In this visual analysis the 9 prognostic parameters generated by pathologists in 324 breast cancer cases are used. Through these visual displays, users can explore connections between cases by systematically examining a variety of combinations of independent variables and dependent variables. Statistical significance and confidence interval analysis is beyond the focus of this paper and thus is not considered here.

Conclusions

It is extremely challenging to integrate varied, complex data, such as pathologist-generated prognostic data, 2D image analysis data, 2D digital tissue cross-sections and annotations, 3D tumor reconstructions and volumetric analysis, etc., with various images, such as raw images, segmented images, animations, and zoomable high-resolution images, in a practical system. In

this paper, we describe our successful efforts in developing such an integrated system for breast tumor visual analysis and evaluation. We believe that our system will significantly improve the environment for the pathologists and cancer researchers who perform tumor analysis and exploration at a fine and detailed level. The combination of 3-D views and a high-resolution detail view is particularly powerful. We anticipate many potential applications of such a system, including

- computer-based infection detection in soft tissue,
- integrated tumor prediction system,
- research study of the pathogenesis of cancer,
- clinical study to assess sensitivity and specificity of radiologic and imaging techniques effectiveness, and
- models to create better management and treatment algorithms by allowing prediction of growth patterns based on the histologic findings on the biopsy.

Additionally, our proposed system could be a very useful module of a larger internet-based grid-architecture for tissue-based diagnosis; thus benefiting distributed diagnostic support systems in surgical cancer pathology. This kind of system, which incorporates a closed feedback loop by integrating prognostic value reported by pathologists and patient outcome information obtained from a tumor registry database, would provide a self-trained machine-learning system that could be used in surgical cancer pathology to predict tumor behavior, prognosis, and survival rate. Before a specific hypothesis could be supported or implemented, and any predictions made, a statistically significant sample size of tumor cases must be accumulated into the database.

Clearly, more research is needed before the system will be widely adopted and used. Dataflow between the three components needs to be improved and the overall time required to scan a stack of histology glass slides and then produce a 3D reconstructed tumor must also be drastically shortened.

References

1. Street, WN: Xcyt: A system for remote cytological diagnosis and prognosis of breast cancer. In L.C. Jain, editor, Soft Computing Techniques in Breast Cancer Prognosis and Diagnosis, World Scientific Publishing, Singapore, 2000, pp. 297-322

2. Kayser G, Radziszowski D, Bzdyl P, Sommer R, Kayser K: Theory and implementation of an electronic, automated measurement system for images obtained from immunohistochemically stained slides. Anal Quant Cytol 2006; 28(1): 27-38

3. Kayser G, Radziszowski D, Bzdyl P, Sommer R, Kayser K: Towards an automated virtual slide screening: theoretical considerations and practical experiences of automated tissue-based virtual diagnosis to be implemented in the Internet. Diag Path 2006; 1:10

4. Görtler J, Berghoff M, Kayser G, Kayser K: Grid technology in tissue-based diagnosis: fundamentals and potential developments. Diag Path 2006; 1:23

5. Leong FJW: An automated diagnostic system for tubular carcinoma of the breast: a model for computer-based cancer diagnosis. DPhil Thesis, University of Oxford, March 2002

6. Fernandez-Gonzalez R, Jones A, Garcia-Rodriguez E, Chen PY, Idica A, Barcellos-Hoff MH and Ortiz de Solorzano C: A system for combined three-dimensional morphological and molecular analysis of thick tissue samples. Microsc. Res. Tech. 2002; 59(6): 522–530

7. Subramaniam E, Liung TK, Mashor MY, Isa NAM: Breast Cancer Diagnosis Systems: A Review. International Journal of The Computer, the Internet and Management 2006; 14(2): 24 - 35

8. Gering D, Nabavi A, Kikinis R, Hata N, O'Donnell L, Grimson E, Jolesz F, Black P, Wells W: An integrated visualization system for surgical planning and guidance using image fusion and an open MR. J. Magn. Reson. Imaging 2001; 13:967–975

9. Petushi S, Katsinis C, Coward C, Garcia FU, Tozeren A: Automated identification of microstructures on histology slides. IEEE International Symposium on Biomedical Imaging: Macro to Nano 2004; 1:424-427

10. 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

11. C.A. Glasbey, G.W. Horgan, Image Analysis for the Biological Sciences, Wiley, Chichester, 1995

12. C.R. Russ, The Image Processing Handbook, CRC Press, Boca Raton, 1995

13. Marker J, Braude I, Museth K, Breen D: Contour-based Surface Reconstruction Using Implicit Curve Fitting, and Distance Field Filtering and Interpolation. Proc. Int. Workshop on Volume Graphics, July 2006, pp. 95-102

14. Ohtake Y, Belyaev A, Alexa M, Turk G, Seidel H: Multi-level partition of unity implicits. ACM Transactions on Graphics (Proc. SIGGRAPH) 2003; 22(3): 463–470

15. Sethian J: A fast marching level set method for monotonically advancing fronts. Proceedings National Academy of Science 1996; 93(4): 1591–1595

16. Hyman J: Accurate monotonicity preserving cubic interpolation. SIAM Journal on Scientific and Statistical Computing 1983; 4(4): 645–654

17. Lorensen W, Cline H: Marching Cubes: A high resolution 3D surface construction algorithm. Proc. SIGGRAPH 1987, pp. 163–169

18. Osher S, Sethian J: Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations. Journal of Computational Physics 1988; 79(1): 12 – 49

19. Kamada T, Kawai S: An algorithm for drawing general undirected graphs. Information Processing Letters 1989; 31, 7-15.

Figure Legends

Figure I:. A) Model showing the protocol for slicing gross breast lumpectomies: The tissue is oriented based on a surgeon's notation. The tissue after inking (not shown) is serially sectioned to create 5mm slices. These slices are labeled and consecutively submitted for processing, paraffin embedding and slide production; B) Diagram showing 2D map of paraffin-fixed blocks created from the tissue slices in A; C) H&E stained histologic slide; D) Paraffin-fixed block of tissue; E) Cross-section sampling within a paraffin block

Figure II: Web-based breast tumor image and data visualization

Figure III: A) User interface for 2D image processing, segmentation, annotation and pre-processing of histology slides for 3D-reconstruction; B) Annotation of 2D histology images of H&E stained tissue cross-sections

Figure IV: Flowchart of automated cell-level processing of a breast cancer histology image. The stages include segmentation with adaptive thresholding and domain specific morphological operations, blob (micro-object) labeling, feature extraction and selection, and nuclei classification

Figure V: Flowchart of data generated in the first module to serve as input to the second module (3D tumor reconstruction)

Figure VI: Overview of the volumetric reconstruction process. Input is a set of contours represented as binary images. MPU implicit curves are fit to the contours. A Euclidean distance field is generated from the narrow band around the implicit curve. The field is filtered to remove medial axis discontinuities. The filtered fields are interpolated to produce a volume dataset. A mesh of the zero level set is extracted from the volume

Figure VII: A breast cancer tumor model (right) constructed from four histology-based segmentations (left). The model only includes regions with necrotic (purple) and invasive cancer (blue) cells, and the outer specimen membrane

Figure VIII: Four types of image views (clockwise): raw image, segmented image, 3D image, and high-resolution zoomable image

Figure IX: The search interface for the integrated system, PS3. The user can choose any of the criteria listed on the screen to narrow down his/her case selections

Figure X: Network visualization of 345 breast tumor cases based on histologic grade and the three individual component of the Nottingham histologic grading system. Node colors correspond to the 3 levels of mitotic counts

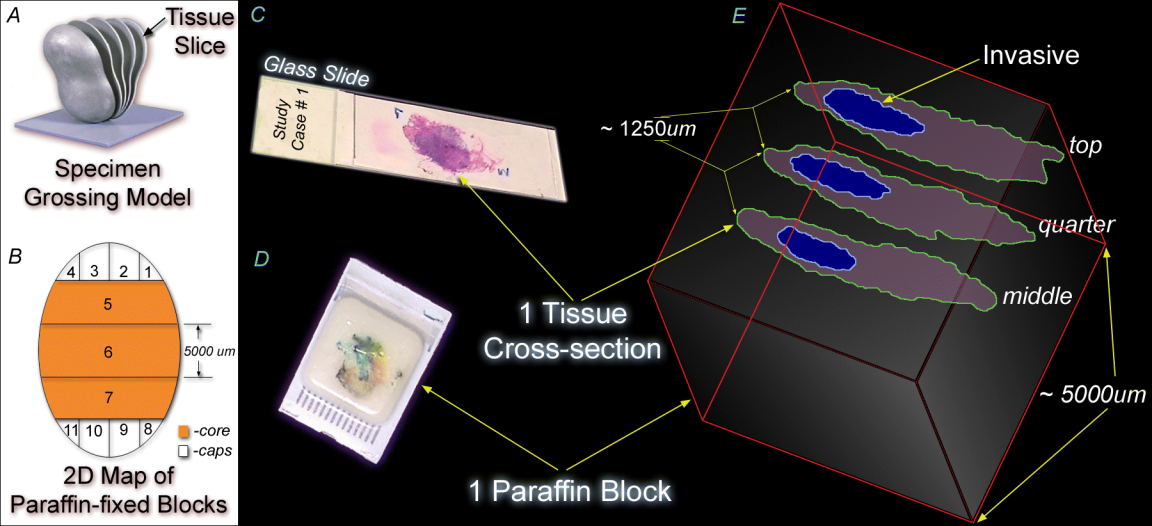
Tables

Table 1: Features extracted in 2D

	Measurement	Description
1	Area	Number of white pixels in a segmented object (blob), after hole filling
2	Clumpiness	Fraction of pixels deviating from the average remaining after a dilation, reflecting texture variations
3	Darkness	Fraction of pixels that deviate more than a certain range (20% default) from the minimum intensity
4	Density Red-Blue	Red and Blue intensities in a segmented object
5	Density Red-Blue Min	Minimum Red and Blue intensities in a segmented object
6	Density Blue Max	Maximum Blue intensity in a segmented object
7	Heterogeneity	Fraction of pixels that deviate more than a certain range (10% default) from the average intensity
8	Number of Holes	Number of black holes within a segmented object, before hole filling
9	Solidity	The ratio between the object area and the convex envelope area

Table 2: Variables for Tumor Cases

Tubule Formation	1	2	3
Nuclear Grade/Pleomorphism	1	2	3
Mitotic Figure/Count	1	2	3
BR Grade/Histologic Grade	1	2	3
ER	Negative	Borderline	Positive
PR	Negative	Borderline	Positive
Ki-67	Low	Borderline	High
p53	Negative	Borderline	Positive
Her2neu	Negative	Borderline	Positive


A
Tissue Slice
Specimen Grossing Model
B
4
3
2
1
5
6
7
11
10
9
8
5000 um
-core
-caps
2D Map of Paraffin-fixed Blocks
C
Glass Slide
Study Case # 1
D
1 Tissue Cross-section
1 Paraffin Block
E
Invasive
~ 1250um
top
quarter
middle
~ 5000um

Click on a thumbnail or case_no to see details

Case_no: 3DB0
Age: 49
Size: 1.3
Br_grade: 3
Er_value: NEG

Prognostic Data
2D Analysis Data
3D Analysis Data

Case_no: 3DB17
Age: 56
Size: 1.6
Br_grade: 2
Er_value: NEG

Prognostic Data
2D Analysis Data
3D Analysis Data

Case_no: 3DB1
Age: 65
Size: 2.0
Br_grade: 1
Er_value: NEG

Prognostic Data
2D Analysis Data
3D Analysis Data

Case_no: 3DB25
Age: 74
Size: 1.5
Br_grade: 3
Er_value: NEG

Prognostic Data
2D Analysis Data
3D Analysis Data

Case_no: 3DB26
Age: 45
Size: 3.0
Br_grade: 3
Er_value: NEG

Prognostic Data
2D Analysis Data
3D Analysis Data

Case_no: 3DB31
Age: 54
Size: 2.0
Br_grade: 3
Er_value: NEG

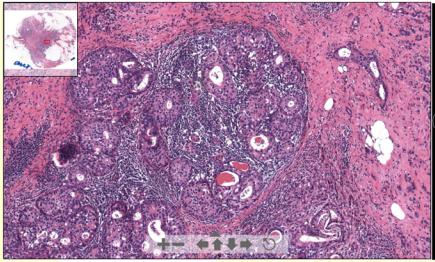
Prognostic Data
2D Analysis Data
3D Analysis Data

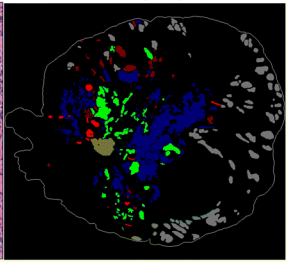
Case_no: 3DB36
Age: 73
Size: 2.0
Br_grade: 3
Er_value: NEG

Prognostic Data
2D Analysis Data
3D Analysis Data

Raw Image | Segmented Image | High Resolution Images | 3D Model

3DB17_2_1 High Resolution Image





Brief Information of Case 3DB17

CASE STUDY #:	3DB17		
AGE:	56	SEX:	F
SPECIMEN:	BR	TUMOR:	BREAST BIOPSY
MRN:	16	TUMOR:	TIC
TUMOR SIZE:	3	NOT GRADED:	2
NOTGIVEN:	1	NON-INVASIVE:	2
FIRSTGIVEN:	0	INVGRADE:	2
ER:	0	ER RATIO:	NEG
PR:	0	PR RATIO:	NEG
HER:	0	HER RATIO:	HIGH
P53:	0	P53 RATIO:	POS
MIB/KI67:	0.8	MIB/KI67 RATIO:	POS

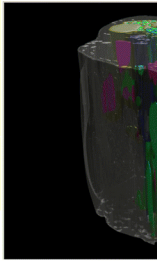
2D Image Analysis of Case 3DB17

Case Number:	3DB17		
Number of NUCLEI:	0	Area: NUCLEI	0
Number of NUCLEI:	0	Area: NUCLEI	0
Number of NUCLEI:	0	Area: NUCLEI	0
Adjacent Tissue:	0	Uniform Area:	0
NUCLEI Distribution:	0	Mid-Density Distribution:	0
Number of NUCLEI High Density:	0.00	Number of Sub-Clusters:	0.00
Intensity Blue Hue:	0.066	Number of Nuclei Holes:	149.361
Nucleus Area:	477.25	Nucleus Area:	0.159
Nucleus Solidity:	0.014	Nucleus Circularity:	0.015
Density Blue Hue:	143.397	Density Blue Hue:	0.063
Density Blue Hue:	0.063	Density Red:	119.43
Density Red:	68.649	Density Red:	0.114
Heterogeneity:	0.132	Darkness:	0.312
Area/Unusual Area:	0.000	Area/Unusual Area:	0.000
Area/Unusual Area:	0.000	Area/Unusual Area:	0.000
Number of Tubular Cross Sections:	0.00	Intensity Blue Hue:	191.611
Nucleus Solidity:	0.37	Nucleus Solidity:	0.876
Nucleus Circularity:	0.184	Density Blue Hue:	167.281
Density Red:	0.089	Density Red:	0.37
Darkness:	0.116		

Show Prognostic Data | Show 2D Analysis Data | Show 3D Analysis Data | Show slide images

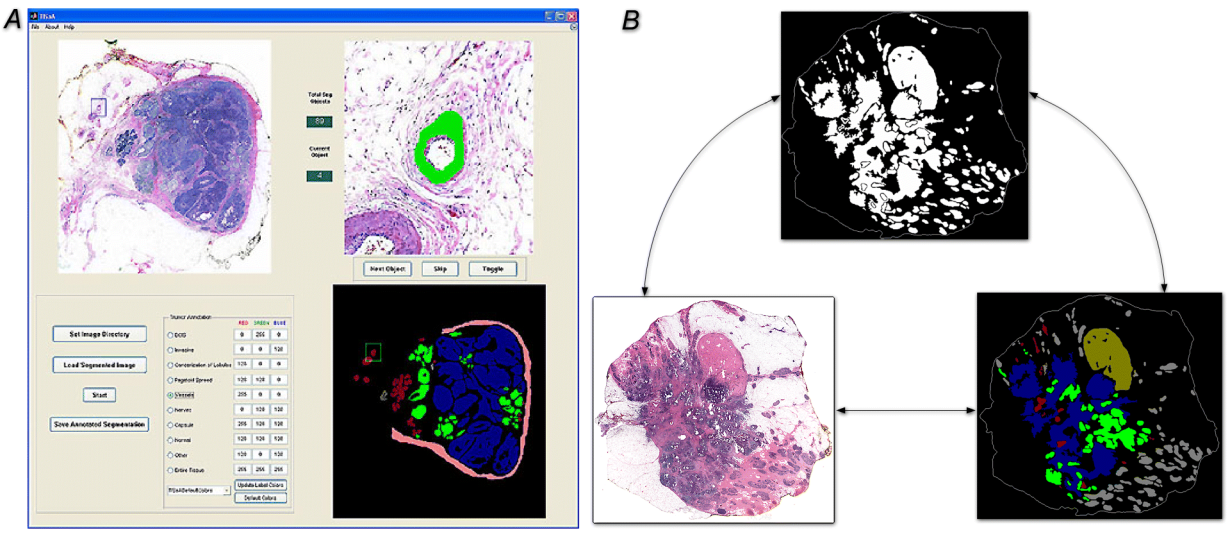
Raw Image | Segmented Image | High Resolution Images | 3D Model

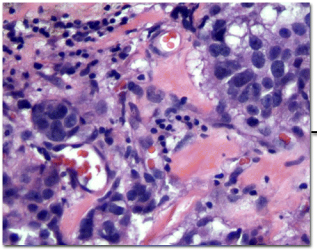
3D Movie for Case 3DB17



Color Scheme

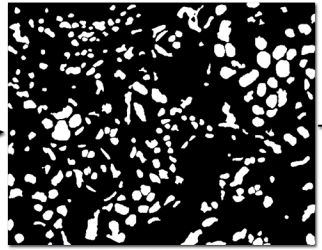
Tissue Structure	Color Name
DCIS:	green
Invasive:	dark blue
Cancenization of Lobules:	dark red
Fibroadenoma:	dark yellow
Vessels:	red
Cysts:	blue
Capsule:	light red
Normal:	gray
Necrotic Area:	purple
Outer Contour:	white
Ductal Hyperplasia:	pink
Thrombosed Vessel:	brown
Fibrosis:	lavender
Sclerosis Adenosis:	plum
Pagetoid Spreed:	khaki
Nerves:	blue
Other:	dark olive green


A
B
Set Image Directory
Load Segmented Image
Start
Save Annotated Segmentation
Next Object
Skip
Toggle
Update Label Colors
Default Colors

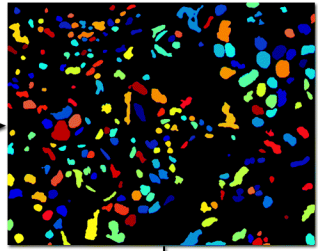


Original Image

Segmentation:
Adaptive
Thresholding
+
Morphological
Operations



Blob Labeling



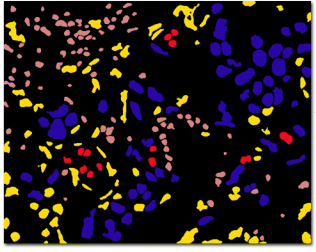
Feature
Extraction

$X_1\ X_2\ X_3 \ldots X_n$

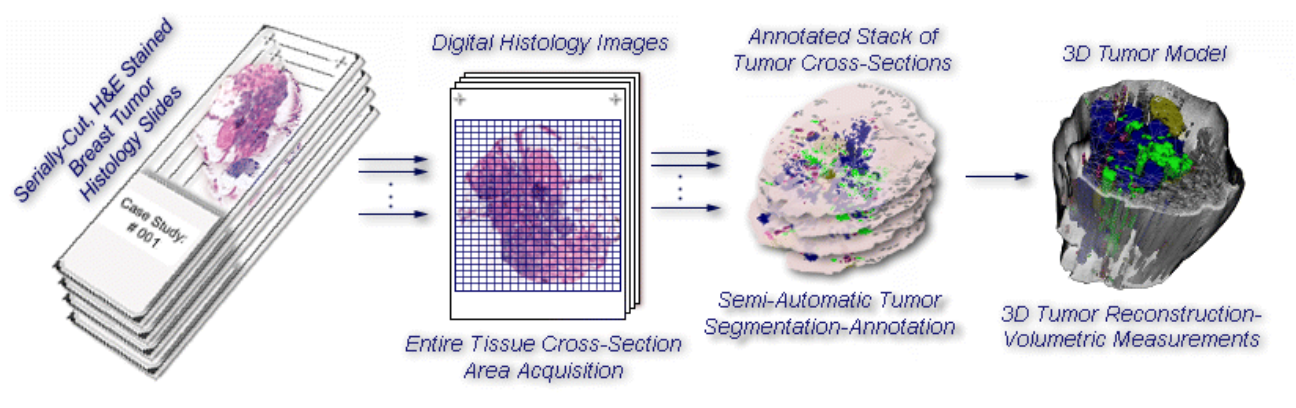
Blob Feature
Vectors

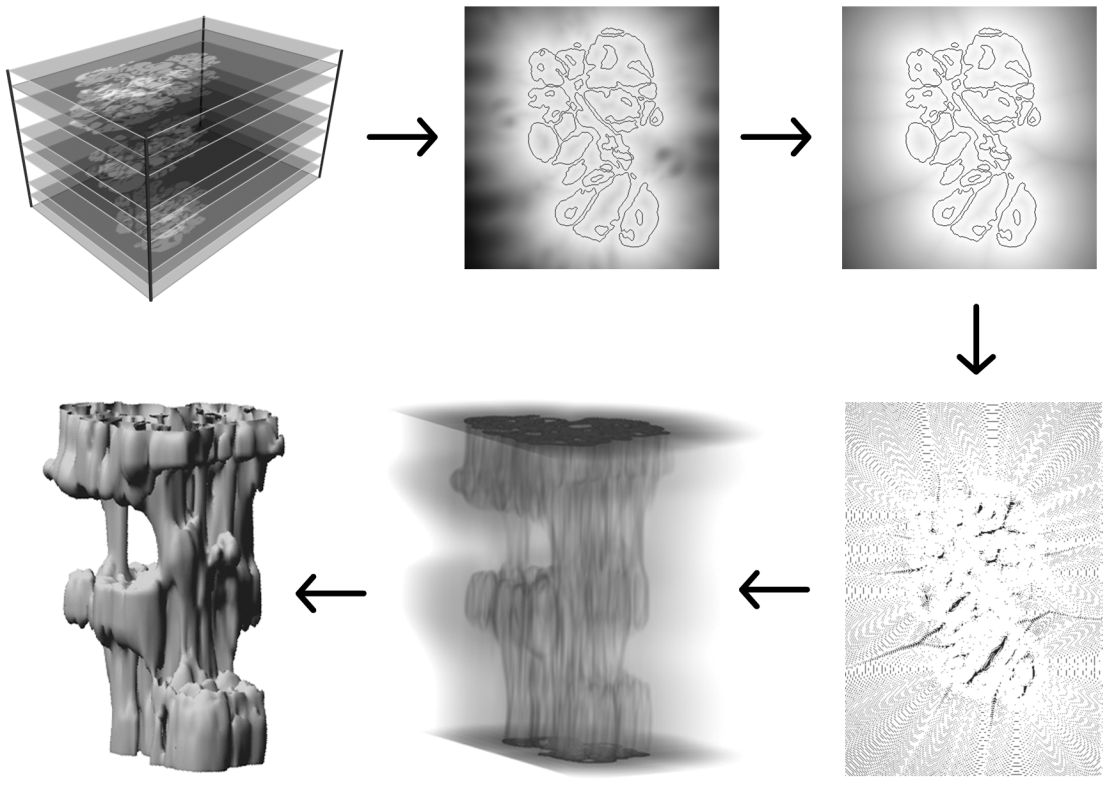
Classifier

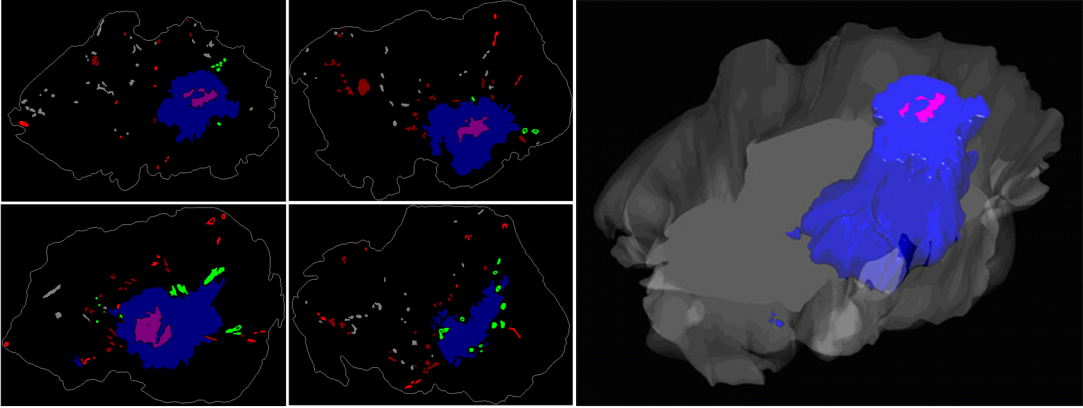
Classified
Cell Nuclei

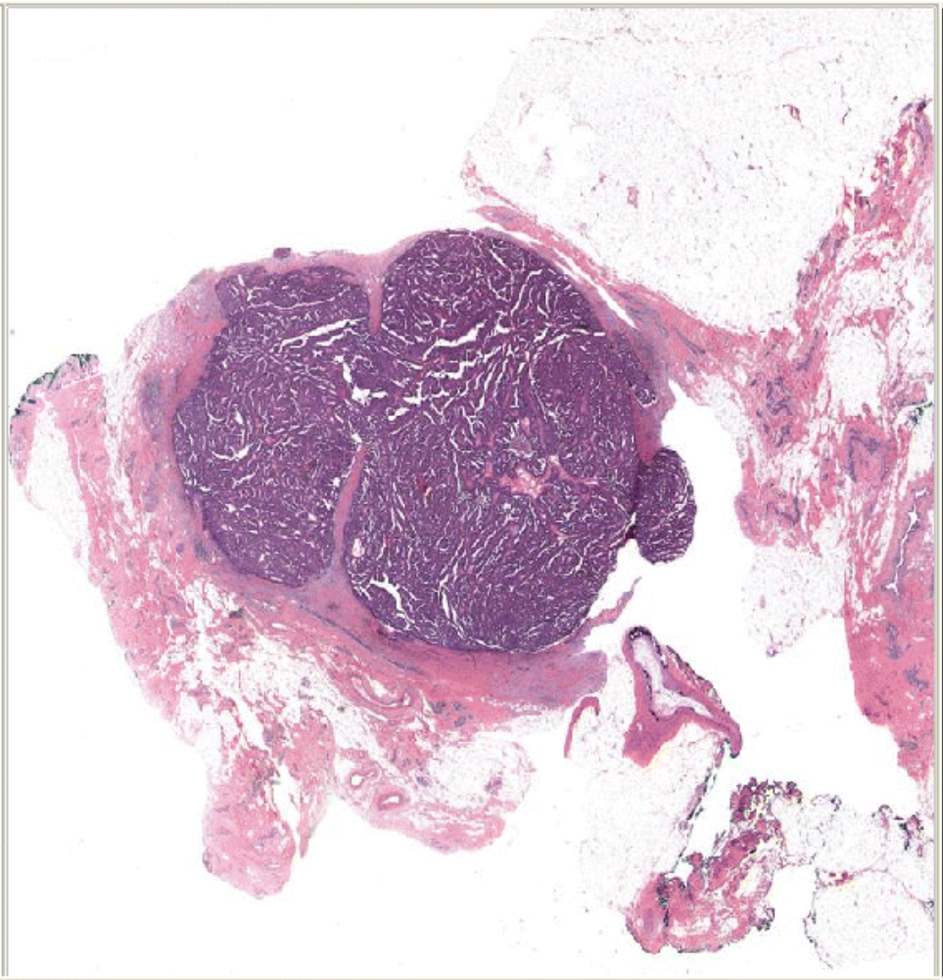


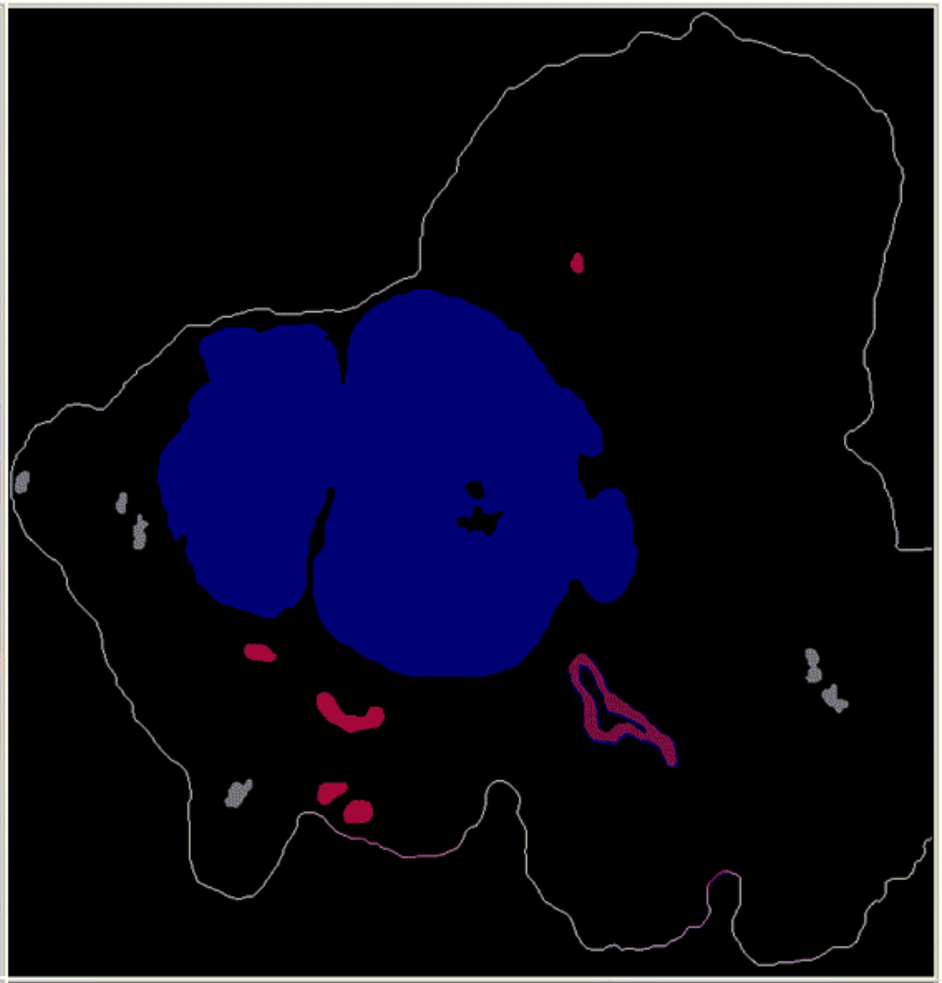
Resulted Image

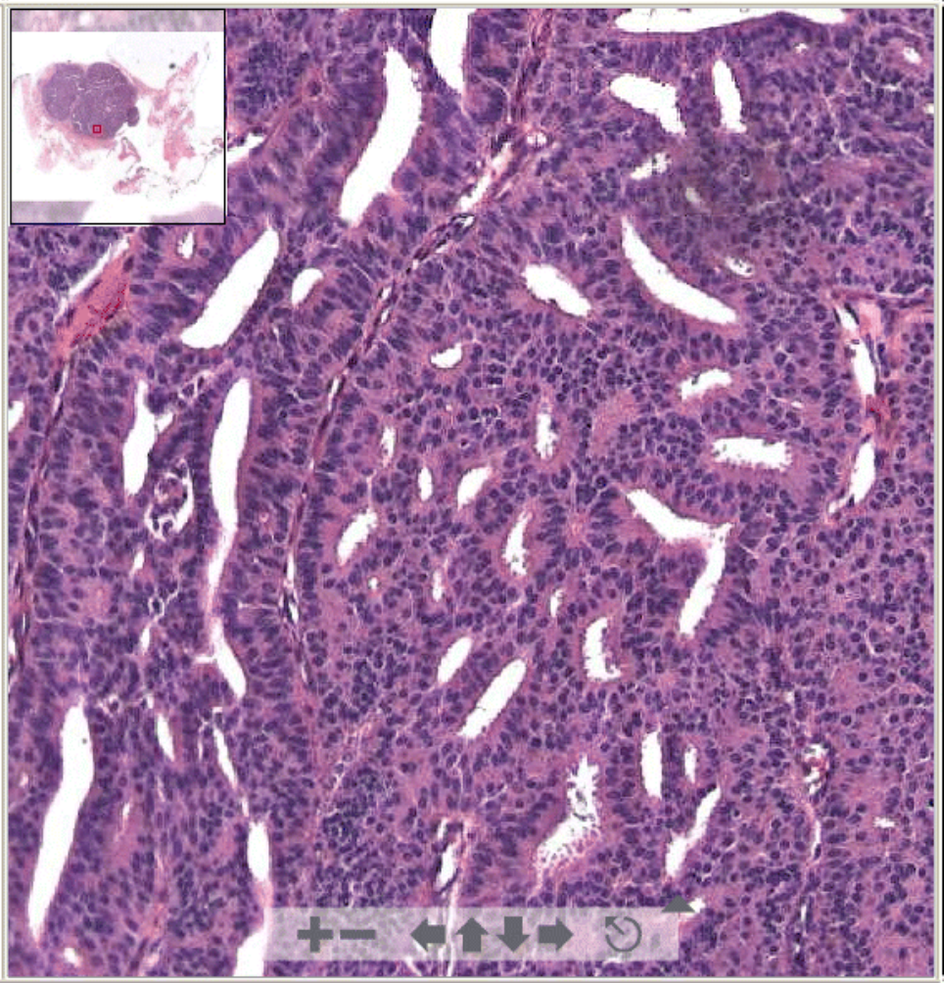

Serially-Cut, H&E Stained Breast Tumor Histology Slides
Case Study: # 001
Digital Histology Images
Entire Tissue Cross-Section Area Acquisition
Annotated Stack of Tumor Cross-Sections
Semi-Automatic Tumor Segmentation-Annotation
3D Tumor Model
3D Tumor Reconstruction- Volumetric Measurements

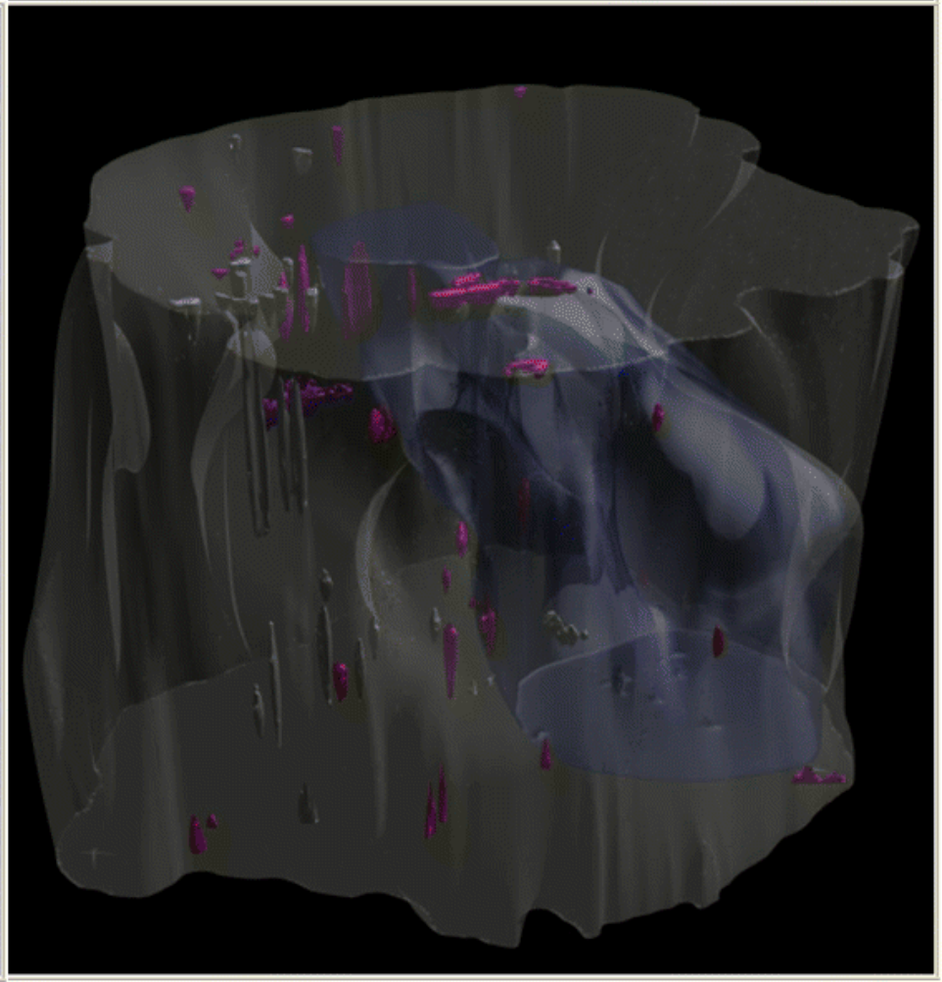












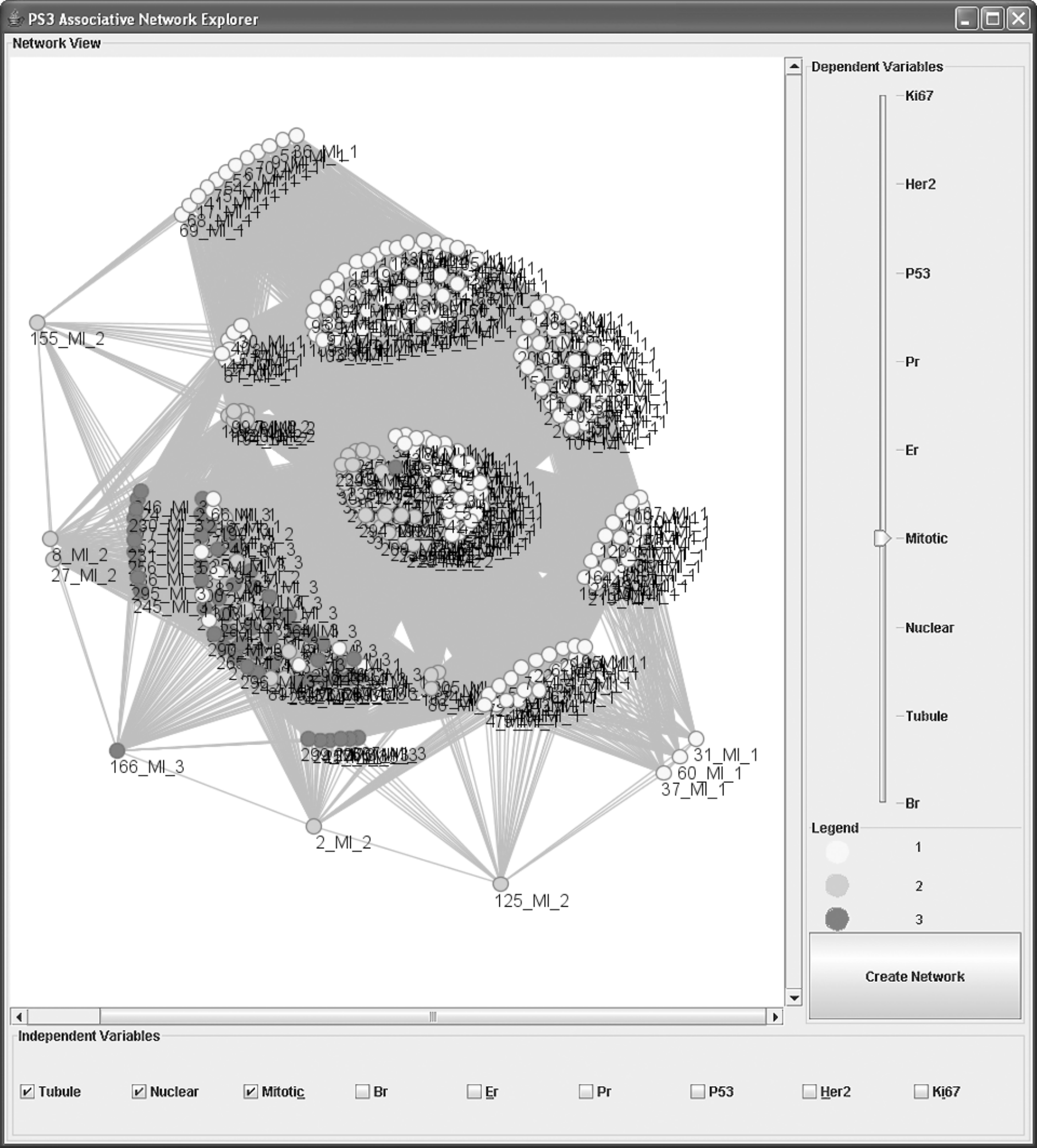
Welcome Sokol

Search by Criteria

AGE: 60-51
BR GRADE: --select BR Grade --
PR VALUE: --select PR Value --
P53 VALUE : --select P53 Value --
Submit

SIZE: --select size range--
ER VALUE: --select ER Value --
KI-67 VALUE: --select KI67 Value--
Reset

CASE STUDY NO	AGE	SIZE	BR GRADE	ER VALUE	PR VALUE	KI-67 VALUE	P53 VALUE
3DB15	57	2.2	2	NEG	NEG	LOW	NEG
3DB17	56	1.6	2	NEG	NEG	HIGH	POS
3DB29	60	4.5	3	POS	POS	HIGH	POS
3DB31	54	2.0	3	NEG	NEG	HIGH	NEG
3DB6	52	1.5	1	POS	POS	BOR	NEG
3DB8	58	1.5	1	POS	POS	LOW	NEG
G3Replace2	58	2.7	3	NEG	NEG	HIGH	POS
G3Replace4	55	1.5	3	POS	BORDERLINE	HIGH	NEG


PS3 Associative Network Explorer
Network View
Dependent Variables
Ki67
Her2
P53
Pr
Er
Mitotic
Nuclear
Tubule
Br
Legend
1
2
3
Create Network
155_MI_2
8_MI_2
27_MI_2
166_MI_3
2_MI_2
125_MI_2
31_MI_1
60_MI_1
37_MI_1
Independent Variables
Tubule
Nuclear
Mitotic
Br
Er
Pr
P53
Her2
Ki67