Proceedings of the American Medical Informatics Association. Vol. 1: 136-139, 2002

Unsupervised Imaging, Registration and Archiving Of Tissue Microarrays

Wenjin Chen BS[1], Michael Reiss MD[2], David J. Foran PhD[1,2]

[1] Center for Biomedical Imaging & Informatics, UMDNJ, Piscataway, NJ
[2] The Cancer Institute of New Jersey, UMDNJ, Piscataway, NJ

ABSTRACT

Tissue microarray (TMA) technology offers several advantages over traditional methods of specimen preparation by maximizing limited tissue resources and providing the means for visualizing molecular targets. Currently, the primary methods used to evaluate the arrays involves the interactive review of TMA samples while they are viewed under a microscope and are subjectively evaluated and scored. The process is slow, tedious and prone to error. In order to facilitate large-scale, multi-institutional studies a more automated and reliable means for processing tissue microarrays is needed. We have developed a web-based prototype which features automated imaging, registration and intelligent archiving of tissue microarrays in multi-user, network environments.

INTRODUCTION

Tissue microarrays are a method of harvesting small cylinders of tissue from a range of standard histological sections and arranging them on a recipient paraffin block in a grid-like manner such that hundreds or thousands can be analyzed simultaneously[1,2]. This technique allows maximization of tissue resources by analysis of small core biopsies of blocks, rather than complete sections. Using this technology, a carefully planned array can be constructed with cases from pathology tissue block archives, such that a 20-year survival analysis can be performed on a cohort of 600 or more patients by use of only a few microliters of antibody.

Another major advantage of the TMA technique is the fact that each specimen is treated in an identical manner. Therefore, reagent concentrations are identical for each case, as are incubation times, temperatures and wash conditions. Using conventional protocols, a study composed of 300 tissue sample would involve processing of 300 hundred slides, which is at least 20 batches of 15 slides. Using tissue microarrays the entire cohort can be processed on a single slide. Currently, the primary methods used to evaluate the arrays involve manual review of TMA samples while they are viewed under a microscope and are subjectively evaluated and scored. An alternate, but less utilized approach is to sequentially digitize specimens for subsequent assessment[3]. Both procedures ultimately involve the interactive evaluation of TMA samples which is a slow, tedious process which is especially error-prone. Often while navigating among the regularly arranged tissue cores it is easy to loose track of the current array location. This is especially true at higher magnifications. To address these issues, a computer assisted navigational approach for assessing TMA samples would be highly desirable and could potentially improve the reliability and reproducibility of results.

Beyond the algorithm and software development that will be required to analyze the arrays, reliable tools will be needed to facilitate large-scale, multi-site collaboration for a broad spectrum of research and clinical activities such as tissue banking, proteomics and outcome studies. Future progress in several key areas will rely upon the capacity of individuals to dynamically acquire, share and assess microarrays and correlated data.

We have developed a web-based prototype for automated imaging, feature extraction, and intelligent archiving of tissue microarrays. The system consists of a robotic microscope interfaced with a JAVA-based micro-controller and imaging workstation. The software is platform- and operating system-independent and with minor configuration changes can interface with any commercially available robotics. The system utilizes a combination of sophisticated image processing and pattern recognition strategies to co-register specimens while the software directs a robotic microscope to systematically image specimens at multiple optical magnifications, delineate array disks, extract spectral and spatial signatures and populate local or distributed relational databases with the resulting data including pointers to imaged arrays. The prototype features stand-alone and network modes. The system also features a visually intuitive interface, which enables local and remote users to manipulate the configuration of digitized

arrays in order to facilitate new experimental designs and data assimilation.

MATERIALS

Four tissue microarrays consisting of 57 primary breast carcinomas and 16 associated normal breast tissue specimens were obtained from women under age 35. Consecutive microarray sections were stained with antibodies for phosphorylated Smad2, and Smad4, and counter stained with hematoxylin. A fourth slide was stained with only hematoxylin.

The prototype TMA analysis system consists of an Olympus AX70 microscope equipped with a Prior 6-way robotic stage and motorized turret. The server workstation was developed on a Pentium II computer, equipped with 256 Mbytes of RAM, and running a Windows 98 operating system. Video images are systematically digitized using an Olympus DC330 720-line, 3-Chip video camera and a Flashpoint 128 high-resolution frame grabber manufactured by Integral Technologies, Inc. A JAVA-based microscope-controller was developed to allow network-based, remote control of the movement of the stage, selection of objectives, adjustment of illumination conditions, shutter speed, and gain of the video camera, and entropy-based, auto-focusing of specimens.

Middleware was developed to allow communication among the robotics, the image processing modules, and an Oracle 8i Database Management System. The database resides on a networked, Pentium III computer with 256 Mbytes of RAM, and running Windows NT 4.0 workstation operating system.

On the client side, a minimum of 64 Mbytes of RAM and a clock speed of 200 MHz is required. The client application has been successfully tested on Windows, Solaris, RedHat Linux and Macintosh OS X.

METHODS

A means for performing unsupervised registration of arrays was designed to recover the grid structure based upon an image map which is generated during a low-resolution pilot scan of the specimen. The software automatically locates, delineates and indexes each disk using proper column and row indices. Technical challenges arose during the course of development of these algorithms since variations in sample preparation and mechanical distortion gave rise to samples which exhibited rows and columns which were often improperly aligned. Slight errors in lens co-focal and co-centering are compensated for using empirically derived

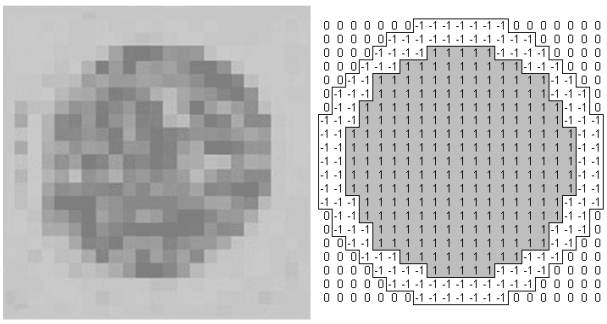

Figure 1. Left: an enlarged view of one tissue disc in the map image. Right: The generated template.

data. An entropy-based, fast auto-focusing algorithm was developed to ensure image quality.

During the course of a low-resolution scan a quilted digital version of the specimen is created by scanning the sample while accurately stitching slightly overlapped frames together. The size of tissue disks, as they appear on the image map, is estimated based upon the approximate core diameter of the physical array as well as system and scan settings. The template in Figure 1 was generated by encoding 1s into an area covering the specimen disks and a rim of –1s two pixels wide at the boundary.

The imaged disks are first convolved with template and then a two-step, top-hat peak detection is used to determine local maximums of the convolution output. Spatial constraints are applied so that there is only a single candidate center point for each disk.

The Hough transformation is used for detecting alignment in the imaged arrays[4]. Generally, to detect straight lines in Cartesian coordinates, every candidate line must be mapped into a point in Hough space, at which point peak detection algorithms are applied to identify local maximum points, which correspond to lines in the original image. In the algorithms that we developed, every pair of candidate disc center points are used to establish a candidate line, and the corresponding two-dimensional Hough space is subsequently analyzed. Due to the grid-like nature of the microarray, when the resulting Hough space is projected onto the θ axis, there can be two peaks detected which are $90°$ from each other corresponding to column and row orientation of the array, i.e., the overall rotation of the array. The intercepts can then be computed by least-square fitting for rows and columns. By doing so, the two-dimensional peak detection process is simplified into two, one-dimensional operations, and the computation time is greatly reduced.

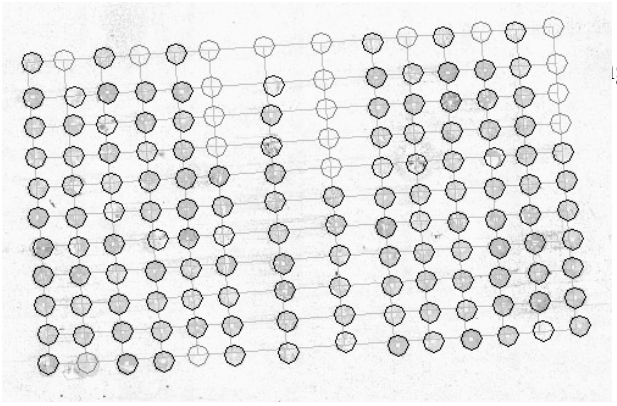

Figure 2. Registered image map with super-imposed grid. Dark circles shows detected tissue discs, light circles show *place holder* grid positions.

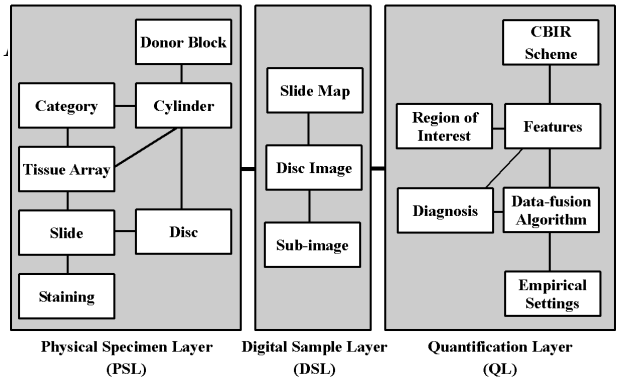

Figure 3. Database Structure

RESULTS

Local/Remote registration. A distributed telemicroscopy (DT) subsystem[5] was integrated with the TMA analysis prototype to enable individuals to operate the systems from remote locations. After user logging into the system and issuing the "Registration" command, the remote microscope automatically begins taking digital images of slightly overlapped frames of the tissue microarray sample in a raster pattern. These frame images are scaled down appropriately before being transmitted to the client so that each frame image is only about 20Kb. It takes only a fraction of second to transmit the images over a fractional T1 Internet connection.

The client application receives the scaled frame images and automatically stitches them together thus producing a map image which then undergoes the unsupervised registration protocol as described in the *Methods* section. Within seconds, the registration results are displayed on the client interface as the recovered grid is superposed on the original map image (see Figure 2). For each grid position for which there is no detected tissue disks a place holder is inserted to compensate for any discs that may have been lost during physical processing, and to provide accurate reporting of those discs exhibiting no expression. Actual disks and place are encircled with different colors to distinguish one from the other. Figure 2 shows some detected discs which deviated from the regular grid positions due to deformation of the paraffin medium during specimen preparation.

To test the prototype system remote registration was performed on four TMA sample slides, which were taken from the same recipient block. In all four cases, the unsupervised TMA registration protocol successfully recovered the grid structures of the samples, despite the fact that staining and rotation of the grids varied significantly among slides.

Distributed archiving. Having determined the physical location of each tissue disc during the course of the registration step, the robotic microscope systematically digitizes and archives each tissue core at a user-defined set of magnifications.

Full size, 640x480, 24-bit colored, non-compressed images are transmitted to the client computer using TCP/IP network protocol. A single image frame includes one entire tissue disk at a 10x magnification. Whenever higher magnification are selected, multiple image frames are digitized and stitched to generate the complete image on the client disk. Correlated array profiles and slide information, such as slice number and staining, are also managed by the database.

Database Design (see figure 3). The physical specimen layer (PSL) of the database relates to the construction and preparation of the actual TMA sample. The specific data which are housed in this layer are referred to as the "array profile". A visually intuitive array profile editor has been developed to facilitate the design, editing and managing of array profiles.

The digital sample layer (DSL) of the database stores archived digital images including the image map as well as images of each tissue disk at multiple resolutions. High-resolution images of tissue disks are broken down into small frames in order to facilitate network access.

Since the TMA technique yielded a standardized set of tissue samples, it provided an ideal data set for developing and evaluating image processing and computer vision protocols, which could be used for performing quantitative immuno-histochemistry. The third layer of our database, the quantification layer (QL), supports automated segmentation and computation of protein expression levels across each disk.

DISCUSSION AND FUTURE DIRECTIONS

Although some DNA microarray readers are capable of reading tissue array slides and some of the software is

quite sophisticated, automatic imaging and evaluation of tissue microarray samples presents a significant set of technical challenges.

First, TMA samples often exhibit morphological irregularities. For example, aside from overall rotation of the grid on slide, some tissue discs are sometimes shifted from regular grid positions as a result of deformation which may occur during preparation. In addition, discs occasionally fall out during preparation[3]. To address these issues it was necessary to develop a robust alignment algorithm which could reliably recover sample grids.

Second, the content of each well on cDNA microarray chips can usually be considered homogeneous, and hence it is sufficient to describe the expression outcome using simple numerical expression levels. This can be accomplished using relatively straightforward image processing protocols. On the contrary, tissue microarrays come in a format of stained tissue discs, which while small are heterogeneous. Depending on the type of tumor or tissue section analyzed, the area of interest may represent nearly the entire disc or only a small percentage thereof. For example, a pancreatic carcinoma or lobular carcinoma of the breast with substantial dysplastic response may show stromal tissue representing a large percentage of the total area of the disk. If the goal of the assay is to determine epithelial cell expression of a given marker, a protocol must be used that evaluates only that region of the disk. The protocol must not only be able to select the region of interest but also normalize it so that the expression level read from any given disk can be compared with that of other disks.

In addition to testing and refining the algorithms, which are used to quantify expression levels, we plan to explore the use of advance computer vision techniques to discriminate among specific staining profiles. For example we have already begun to develop algorithms, which can reliably characterize sub-cellular localization, comparing nuclear and membranous staining patterns.

REFERENCES

1. Kononen J, Bubendorf L, Kallioniemi A, et al. Tissue microarrays for high-throughput molecular profiling of tumor specimens. Nature Med. 4:844,1998.
2. Rimm DL. Camp RL. Charette LA. Costa J. Olsen DA. Reiss M. Tissue microarray: a new technology for amplification of tissue resources. Cancer Journal. 7(1):24-31, 2001.
3. Mucci NR, Akdas G, Manely S, Rubin M. Neuroendocrine expression in metastatic prostate cancer: evaluation of high throughput tissue microarrays to detect heterogeneous protein expression. Human Pathology. 31(4):406-14, 2000.
4. Ross JC. Image processing handbook. 2nd ed. CRC Press; 1995.
5. Foran DJ. Comaniciu D. Meer P. Goodell LA. Computer-assisted discrimination among malignant lymphomas and leukemia using immunophenotyping, intelligent image repositories, and telemicroscopy. IEEE Transactions on Information Technology in Biomedicine. 4(4):265-73, 2000.

ACKNOWLEDGEMENTS

This research was funded in part by a grant from the Whitaker Foundation (grant 98-0202).