Technical Advance

Software Tools for High-Throughput Analysis and Archiving of Immunohistochemistry Staining Data Obtained with Tissue Microarrays

Chih Long Liu,* Wijan Prapong,[†]
Yasodha Natkunam,[‡] Ash Alizadeh,*
Kelli Montgomery,[‡] C. Blake Gilks,[§] and
Matt van de Rijn[‡]

From the Departments of Biochemistry, Genetics,[†] and Pathology,[‡] Stanford University Medical Center, Stanford, California; and the Department of Pathology and Genetic Pathology Evaluation Centre,[§] Vancouver General Hospital, Vancouver, British Columbia*

The creation of tissue microarrays (TMAs) allows for the rapid immunohistochemical analysis of thousands of tissue samples, with numerous different antibodies per sample. This technical development has created a need for tools to aid in the analysis and archival storage of the large amounts of data generated. We have developed a comprehensive system for high-throughput analysis and storage of TMA immunostaining data, using a combination of commercially available systems and novel software applications developed in our laboratory specifically for this purpose. Staining results are recorded directly into an Excel worksheet and are reformatted by a novel program (TMA-Deconvoluter) into a format suitable for hierarchical clustering analysis or other statistical analysis. Hierarchical clustering analysis is a powerful means of assessing relatedness within groups of tumors, based on their immunostaining with a panel of antibodies. Other analyses, such as generation of survival curves, construction of Cox regression models, or assessment of intra- or interobserver variation, can also be done readily on the reformatted data. Finally, the immunoprofile of a specific case can be rapidly retrieved from the archives and reviewed through the use of Stainfinder, a novel web-based program that creates a direct link between the clustered data and a digital image database. An on-line demonstration of this system is available at <http://genome-www.stanford.edu/TMA/explore.shtml>. (*Am J Pathol* 2002, 161:1557-1565)

TMAs allow for the immunohistochemical analysis of large numbers of tissue samples.¹⁻⁴ With as many as 500 different tumors in a single tissue microarray (TMA) paraffin block, thousands of tumor samples can be screened for protein expression using a few array sections. Commercial digital-imaging systems can rapidly store thousands of images, but tools for quick retrieval of these images and for analysis of staining results are lacking. In particular, traditional means of recording and analyzing immunohistochemical staining are based on a small number of cases with a few antibodies and are not well-suited for the huge amounts of data that are rapidly generated with TMA experiments. Because it is possible to do hundreds of different immunostains per case using TMAs, there will be a cumulative accrual of data linked to each of thousands of individual cases, leading to an urgent need for tools to handle this data efficiently and effectively.

Gene expression studies using cDNA microarrays have necessitated the development and introduction of new analytical and data management tools into molecular biology laboratories.⁵ We have drawn on our experience with gene expression microarray experiments⁶⁻⁸ to develop a robust system for the comprehensive management of data generated through immunohistochemical staining of TMAs. This includes using pre-existing analytical tools (Cluster and TreeView software), that were specifically created for analysis of complex microarray expression data,⁵ but are here applied to data derived from immunohistochemical staining of TMAs. Development of this system required integration of these software applications with commercially available hardware and software. Additional novel software (TMA-Deconvoluter and Stainfinder, described herein) were designed to link these components to create a fully functional system. To date we have accumulated more than 3000 cases and

Supported by the National Institutes of Health-National Cancer Institute (grant 1U01 CA 85129).

Accepted for publication July 18, 2002.

Address reprint requests to Matt van de Rijn, M.D., Department of Pathology, Stanford University Medical Center, 300 Pasteur Dr., Stanford, CA 94305. E-mail: mrjijn@stanford.edu.

Figure 1. Flow diagram of TMA lab data acquisition, analysis, and archival storage system. **Gray boxes** indicate components directly relevant to the TMA data management system introduced here.

more than 65,000 different immunostaining results in our archive. Our system allows us to easily add cases or immunostaining results to this data bank, with rapid retrieval of both data sets for further analysis, and of archived digital images of immunostained tissues for review.

Organization of the Stanford TMA Data System

A flow diagram showing the overall organization of the system we have designed is described in Figure 1. A description of the components of the system follows,

using a lymphoma TMA with 265 cases as an example,⁹ with a more detailed explanation available at the accompanying website: <http://genome-www.stanford.edu/TMA>.

TMA Construction

A case selected for TMA construction is first assigned a unique, anonymous, identifying number that is maintained in a separate Excel workbook (Lookup table, Figure 2). This is preferable because the data are easily stripped of patient identifiers to protect patient confidentiality, permitting the data to be shared with collaborators. This identifier will be permanently associated with the case and will provide the key for associating information from scoring worksheets (see next section) to a master database containing the relevant patient/tumor diagnosis information, follow-up data, and any previous immunohistochemical or fluorescence *in situ* hybridization data on that case. As a result, this identifier can be used in any subsequent TMA that contains the corresponding case. For example, a block is taken from a case of large cell lymphoma used in a TMA to study immunohistochemical markers of possible prognostic significance in lymphoma. This same case may next be used in a TMA to test the utility of a marker for differentiating between germ cell tumors and lymphomas occurring in the mediastinum, or as a negative control in a TMA designed to test new epithelial markers. It should be noted that the unique

Microsoft Excel - par_blocks_array_03_15_02.xls										
File Edit View Insert Format Tools Data Window Help										
G543										
1	A	B	C	D	E	F	G	H	I	J
no	organ	class	diagnosis	modifier 1	modifier 2	modifier 3				
527	526	uterus	malignant	carcinoma	endometrial					
528	527	ovary	malignant	carcinoma	papillary serous		grade 2			
529	528	lymph node	malignant	lymphoma	DLCL		grade 3			
530	529	lymph node	normal	normal						
531	530	colon	normal	normal		tonsil				
532	531	lymph node	malignant	lymphoma	DLCL					
533	532	lymph node	malignant	lymphoma	DLCL	Transformed from low grade				
534	533	skin	malignant	lymphoma	DLCL					
535	534	placenta	normal	normal						
536	535	lymph node	malignant	lymphoma	DLCL					
537	536	ent	malignant	lymphoma	DLCL					
538	537	lymph node	malignant	lymphoma	DLCL					
539	538	lymph node	malignant	lymphoma	DLCL	Transformed from low grade				
540	539	liver	benign	cirrhosis						
541	540	lymph node	malignant	lymphoma	DLCL					
542	541	lymph node	malignant	lymphoma	DLCL					
543	542	lymph node	malignant	lymphoma	DLCL					
544	543	uterus	benign	leiomyoma						
545	544	small bowel	malignant	lymphoma	DLCL					
546	545	liver	malignant	lymphoma	DLCL					
547	546	lymph node	malignant	lymphoma	DLCL					
548	547	soft tissue	malignant	lymphoma	DLCL					
549	548	kidney	normal	normal						
550	549	breast	benign	pregnant						
551	550	lymph node	malignant	lymphoma	DLCL					
552	551	lymph node	malignant	lymphoma	DLCL					
553	552	breast	malignant	lymphoma	DLCL					
554	553	lymph node	malignant	lymphoma	F & DLC					
555	554	lymph node	malignant	lymphoma	FLCL					
556	555	prostate	normal	normal						
557	556	spleen	malignant	lymphoma	F & DLC					
558	557	lymph node	malignant	lymphoma	FLCL					
559	558	colon	normal	normal						

Figure 2. A user-supplied lookup table of diagnosis information for a given core sample. The unique TMA case numbers are listed in **column A**.

number in the Lookup table correlates not only with a specific paraffin block but also with a particular diagnosis within that block. The area of interest corresponding to that diagnosis is circled on the accompanying glass slide, and it is from this area that tissue cores are taken. To avoid confusion, we do not retrieve different diagnostic tissues from the same block (ie, *in situ* ductal carcinoma and invasive ductal carcinoma), but instead use two different blocks from the same case for this situation. We have used more than 3000 cases for TMA construction in our laboratory. The first six columns in the Lookup table after the unique identifier can be used to describe a variety of other features than those shown in Figure 2, such as age, gender, clinical outcome, treatment, and so forth. The data in these six columns will appear in the TreeView output file (see below).

The TMAs are constructed using a tissue arrayer (Beecher Instruments, Silver Spring, MD), as described previously.⁹ We routinely use 0.6-mm diameter cores. Sections are cut from the array, and each level is numbered sequentially. The number of levels per block is variable, depending on the thickness of the tissue blocks from which the cores are obtained, but we routinely obtain more than 100 sections per TMA. The number of cores and their site within the tumor (ie, central *versus* peripheral) are functions of the experimental design and hypothesis to be tested, and accordingly, vary from array to array. Each TMA block is subdivided into sectors that measure maximally 11 columns wide, because we have found that larger sectors are harder to score because of the increased likelihood of losing one's place on the grid. In building the array, a space is left between adjacent sectors. The grid pattern is not completely symmetrical, and irrelevant tissues (eg, mouse kidney) are included on each TMA to facilitate orientation. Obviously, great care must be taken in creation of the grid and recording of data to ensure that immunostaining results correspond to the appropriate case.

Recording Immunohistochemical Staining Data

Each individual array has its own Excel scoring workbook, based on the three-dimensional layout of the array, and consists of multiple worksheets. The first worksheet, or master, contains the two-dimensional layout of the TMA, with each field containing the anonymized numerical code from the Lookup table for a particular case (Figure 3A). Subsequent worksheets, copies of the first but with blank cells corresponding to each core, are then used to record the staining interpretations for individual antibodies. Hence, the data worksheets, collectively representing the third dimension of the TMA, each correspond to an interpretation of a specific slide taken from the same TMA paraffin block (Figure 3B). We record data directly onto printed versions of these worksheets, while scoring arrays under the microscope. The use of a worksheet that structurally parallels the layout of the TMA is critical, in our experience, to ensure accurate recording of the data. It is possible to record data by hand in a

linear or other format and later transcribe it into a spreadsheet-type database, but this is labor intensive, and the extra step increases the likelihood of introducing recording errors. Additional data recorded on each worksheet include the section number taken from the array, the staining date, the antibody used, the antigen retrieval method, and the name of the interpreter of the stains, together with the date the stain was interpreted. As such, a workbook can contain separate worksheets containing the staining interpretations of multiple pathologists for a large number of antibodies.

Reformatting Data with TMA-Deconvoluter Software

The TMA-Deconvoluter is a novel program designed to solve the logistical problem of translating large amounts of the three-dimensional TMA data from these raw workbooks into a two-dimensional spreadsheet-type tabulated table format in Excel. This deconvolution process outputs reformatted data that also incorporates the diagnostic and/or clinical information provided by up to six columns in the user's lookup table. The reformatted data are now amenable to further analysis, including hierarchical clustering, generation of survival curves, construction of Cox regression models, and assessment of intra- or interobserver variation.

Operation of the TMA-Deconvoluter is a straightforward process. First, the user specifies the scoring workbook files and the file containing the lookup table of the individual core descriptor information. When the worksheet data are converted into the single text, tab-delimited table format used for hierarchical clustering (Figure 4), each data worksheet is converted into a column of data corresponding to the staining results for the antibody used in that sheet. The descriptor information (for example "bax" in column E) is placed into the appropriate column, and the UID column contains the filename and URL information that is passed on by TreeView into the Stainfinder program (see below). The EWEIGHT and GWEIGHT columns are used by the Cluster software for hierarchical cluster analysis (see below). The TMA-Deconvoluter software is freely available at <http://genome-www.stanford.edu/TMA>.

The TMA-Deconvoluter has a variety of other helpful features. For example, one scoring system for grading immunohistochemical results can be rapidly transformed into a different scoring system. Our usual scoring system designates 0 = negative, 1 = uninterpretable, 2 = weak positive, and 3 = strong positive. The "score conversion utility" enables rapid reformatting of the data such that scores are symmetrical around 0, ie, negative is scored as -2, weak positive as 1, strong positive as 2, and uninterpretable data are left blank. This conversion is necessary because of the design of TreeView for gene microarray data—normally, negative values represent down-regulated genes, whereas positive values represent up-regulated genes. For TMA data, a negative stain should therefore be converted to a negative value, and positive stains be kept as a positive value. The end result

A

B

Figure 3. Worksheets in TMA scoring workbook. **A:** Master worksheet, containing the unique case numbers from the lookup table. Sectors are padded by one row and one column of blank space. The location of a given sector is defined by the cell address at the **top left corner** of the worksheet, which indicates the location of the **top left** data cell in the sector. Each data cell within a sector defines a unique case. **B:** Worksheet containing immunostaining scoring results. Layout parallels master worksheet layout.

is proper visualization in TreeView. In addition, this utility permits results from different labs using different scoring systems to be reformatted to allow direct comparisons to be made.

Hierarchical Cluster Analysis

Hierarchical cluster analysis of our TMA data are performed using software tools that were originally designed for analyzing cDNA microarray data.⁵ Hierarchical cluster analysis has been successfully applied to gene expression data, based on the expression of thousands of genes. It has also been remarkably successful in its ability to group tumors according to their primary site, as

well as subgrouping tumors that are not reproducibly subclassifiable based on conventional morphology.^{6-8,10} Hierarchical clustering can work in two dimensions. For TMA data, the information is used to group tumors together based on the relatedness of their immunostaining with a number of different antibodies, whereas in the second dimension the antibodies are grouped together based on their degree of relatedness in immunoreactivity across the TMA samples examined. Hierarchical clustering can be done using the Cluster program, and the clustered output can be viewed using TreeView, which graphically displays relatedness in both dimensions as dendrograms. Cluster and TreeView are freely available programs that can be obtained at <http://rana.lbl.gov/>

1	UID	NAME	GWEIGHT	alk1	bax	bcl1	bcl2	bcl6	bclb	bcls	cd3	cd5	cd10	cd
2	EWEIGHT			1	1	1	1	1	1	1	1	1	1	
3	1_4_1_1_12_1013_1.jpg	164 alk1 bax bcl1 164	muscle normal normal				-2	-2						-2
4	1_4_2_1_12_1013_2.jpg	169 alk1 bax bcl1 169	muscle normal normal											
5	1_4_3_1_12_1013_3.jpg	115 alk1 bax bcl1 115	lymph node normal normal	1	-2	1	-2	1	1	1			1	1
6	1_4_4_1_12_1013_4.jpg	116 alk1 bax bcl1 116	lymph node normal normal	1	-2	1	-2	2	2	-2	1	2	2	1
7	1_4_5_1_12_1013_5.jpg	168 alk1 bax bcl1 168	lymph node normal normal tonsil	1	-2	1	-2	2	1	1	1	-2	1	1
8	1_4_6_1_12_1013_6.jpg	126 alk1 bax bcl1 126	lymph node normal normal tonsil	1	-2	1	-2	2	2	1	1	-2	1	2
9	1_4_7_1_12_1013_7.jpg	197 alk1 bax bcl1 197	spleen normal normal	1	-2	1	-2	1	1	1	2	-2	1	1
10	1_4_8_1_12_1013_8.jpg	258 alk1 bax bcl1 258	skin malignant melanoma	1	1	1	-2	2	-2	1	2	-2	-2	2
11	1_4_9_1_12_1013_9.jpg	260 alk1 bax bcl1 260	skin malignant melanoma	1	1	2	-2	1	-2	2	2	-2	-2	1
12	1_4_10_1_12_1013_10.jpg	108 alk1 bax bcl1 108	breast malignant carcinoma ductal invasive	1	-2							-2		
13	1_4_11_1_12_1013_11.jpg	95 alk1 bax bcl1 95	breast malignant carcinoma ductal invasive	1	-2		-2	1	-2	-2	1	-2		-2
14	1_4_12_1_12_1013_12.jpg	552 alk1 bax bcl1 552	breast malignant lymphoma DLCL	1	-2	2	-2	2	2	1	1	-2	1	2
15	1_4_2_2_12_1013_21.jpg	541 alk1 bax bcl1 541	lymph node malignant lymphoma DLCL	1	-2	1	-2	2	1	1	1	-2	1	1
16	1_4_3_2_12_1013_20.jpg	537 alk1 bax bcl1 537	lymph node malignant lymphoma DLCL	1	-2	-2	-2	2	2	1	2	-2	1	1
17	1_4_4_2_12_1013_19.jpg	242 alk1 bax bcl1 242	spleen malignant lymphoma DLCL	1	-2	1	-2	2	1	-2	1	-2	-2	1
18	1_4_5_2_12_1013_18.jpg	974 alk1 bax bcl1 974	lymph node malignant lymphoma DLCL	1	-2	1	-2	1	2	1	2	-2	-2	2
19	1_4_6_2_12_1013_17.jpg	224 alk1 bax bcl1 224	lymph node malignant lymphoma DLCL	1	-2	2	-2	2	1	1	1	-2	-2	1
20	1_4_7_2_12_1013_16.jpg	620 alk1 bax bcl1 620	lymph node malignant lymphoma DLCL	1	-2	2	-2	1	2	-2	2	-2	2	1
21	1_4_8_2_12_1013_15.jpg	540 alk1 bax bcl1 540	lymph node malignant lymphoma DLCL	1	1	1					1	-2	1	
22	1_4_9_2_12_1013_14.jpg	466 alk1 bax bcl1 466	lymph node malignant lymphoma DLCL	1	-2	1	-2	1	1	-2	2	-2	-2	1
23	1_4_10_2_12_1013_13.jpg	979 alk1 bax bcl1 979	ent malignant lymphoma DLCL	1	-2	2	-2	2	2	1	1		-2	2
24	1_4_11_2_12_1013_12.jpg	535 alk1 bax bcl1 535	lymph node malignant lymphoma DLCL	1	-2	1				-2	-2	-2	-2	
25	1_4_13_1_12_1013_23.jpg	546 alk1 bax bcl1 546	lymph node malignant lymphoma DLCL	1	-2	2	-2	2	2	1	1	-2	-2	1
26	1_4_2_3_12_1013_24.jpg	551 alk1 bax bcl1 551	lymph node malignant lymphoma DLCL	1	-2	1	-2	2	1	1	1	-2	1	1
27	1_4_3_3_12_1013_25.jpg	545 alk1 bax bcl1 545	liver malignant lymphoma DLCL	1	-2	1	-2	-2	1	-2	1	-2	-2	2
28	1_4_4_3_12_1013_26.jpg	544 alk1 bax bcl1 544	small bowel malignant lymphoma DLCL	1	-2	2	-2	1	2	-2	2	-2	-2	2
29	1_4_5_3_12_1013_27.jpg	547 alk1 bax bcl1 547	soft tissue malignant lymphoma DLCL	1	-2	2	-2	2	1	1	2	-2	1	1
30	1_4_6_3_12_1013_28.jpg	542 alk1 bax bcl1 542	lymph node malignant lymphoma DLCL	1	-2	1	-2	2	-2	-2	1	2	-2	
31	1_4_7_3_12_1013_29.jpg	320 alk1 bax bcl1 320	thyroid malignant lymphoma DLCL	1	-2	1	-2	1	2	1	2	-2	-2	-2
32	1_4_8_3_12_1013_30.jpg	334 alk1 bax bcl1 334	breast malignant lymphoma DLCL	1	-2	1	-2	2	1	1	2	-2	-2	1
33	1_4_9_3_12_1013_31.jpg	330 alk1 bax bcl1 330	lymph node malignant lymphoma DLCL	1	-2	-2				-2	2	-2	-2	
34	1_4_10_3_12_1013_32.jpg	327 alk1 bax bcl1 327	lymph node malignant lymphoma DLCL	1	-2	1						-2	2	
35	1_4_11_3_12_1013_33.jpg	313 alk1 bax bcl1 313	lymph node malignant lymphoma DLCL	1	-2	1	-2	1	-2	-2	2	-2	1	1

Figure 4. Deconvoluted output file. Typical output file from TMA-Deconvoluted, ready for analysis by Cluster software.

EisenSoftware.htm. The EWEIGHT and GWEIGHT columns in the deconvoluted data, as described previously (Figure 4), can be modified manually, if it is desired that a given antibody or patient sample should be given a different weight in the clustering process (see also additional information on Cluster at <http://rana.lbl.gov/EisenSoftware.html>).

The data deconvoluted by TMA-Deconvoluted is in a format suitable for immediate analysis by Cluster, and the resulting clustered file can then be opened in TreeView (Figure 5A). In this example, a strong positive score is represented as a bright red block and a weak positive score a dark red block, and a negative score appears as a green block. Stains that could not be interpreted because of technical reasons (eg, loss of tissue from the slide) are represented by gray blocks. Dendrograms showing the two dimensions of clustering are seen at the top and left-hand sides of Figure 5.

Other Data Analyses

TMA-Deconvoluted can also be used to generate files with a layout usable for statistical analysis programs (eg, SPSS or WinSTAT) to generate Kaplan-Meier survival curves, multivariate analysis, and so forth. Reanalysis can be readily performed as new data are accrued and added to the database.

Capture and Retrieval of Digital Images Using Stainfinder Software

We use the Bliss system from Bacus Laboratories (Lombard IL; <http://bacuslabs.com>) for digital image collection. The images were collected using Tracer v0.87h software supplied with the Bliss system. Each tissue core section is digitally recorded in six separate images that are automatically tiled together by the Bliss system. Each tiled image is stored as a JPG compressed image of ~400 kb. TMAs are scanned one sector at a time, and the images from each sector are stored in a separate folder using a standardized nomenclature. For a detailed description of the standardized nomenclature, see <http://genome-www.stanford.edu/TMA>. To date, we have archived digital images on more than 65,000 immunostained tissue cores.

Stainfinder is a novel web-based program that links a selected row in the TreeView graphical output file (corresponding to all immunostaining data for one arrayed core) to the digital images recorded from that core. For TreeView figures showing gene expression microarray data, each row corresponds to a gene or clone, and the link provides access to additional information about this particular gene or clone. In the TMA laboratory, Stainfinder serves a similar but distinct function, and clicking on a row from a TreeView document retrieves a web page

Figure 5. A: Screenshot of Tree View. TreeView output file, example of deconvoluted, clustered data in TreeView on lymphoma array.⁹ **B:** Stainfinder screenshot, with thumbnail representation of selected images.

Figure 6. Example of clustering result of lymphoma TMA containing 265 lymphomas and stained with 27 different antibodies. Clustering of the data results in partial grouping together of tumors with the same or similar diagnoses.

that lists all stains scanned in for this core and stored on the server (see screenshot on accompanying website). We currently use a 400GB RAID server to store digital images. Clicking on the display images button rapidly retrieves digital images in a thumbnail format (Figure 5B) that can be enlarged by clicking on the name of the thumbnail. Used in this manner, one can rapidly verify the score given to a particular antibody-stained core sample, by comparing the color block representing the score in the TreeView document to the corresponding digital image. The ability to do this is important and valuable, because the interpretation and determination of scores is subjective and often requires review. This is also valuable for allowing one to compare multiple stains done on the same core, allowing verification of scoring data taken from the same sample. In addition, variables such as nuclear versus membrane staining, the presence of stromal cell reactivity, and so forth, are features that can best be evaluated by re-examining the digital images. In comparison, reanalysis of multiple stains on one core by using the TMA slide sections would be extremely laborious. When immunostains for TMAs are scored by more than one pathologist, or if levels of the same array are stained with the same antibody by more than one laboratory, it is possible to rapidly identify interobserver or interlaboratory differences in results. The Stainfinder software is freely available at <http://genome-www.stanford.edu/TMA>.

An example of an application of the software is shown in Figure 6, where a visualization of data obtained from a TMA with 265 lymphomas, stained with 27 antibodies is shown. Immunostaining results were recorded in an Excel workbook and reformatted with TMA-Deconvoluter.

Clustered data were visualized in TreeView. The visual representation of the entire dataset is shown on the left panel with magnification of specific areas shown on the right. The dendrogram reveals that two different interpretations (in this case by the same pathologist, YN) of the same section stained with MUM-1 clustered tightly, indicating their high degree of correlation. The TreeView linkage to the image files allows for a rapid inspection and final scoring of the stains in which interpretations differed. Clustering of tumors, based on their immunoreactivity, exhibits a partial grouping together of lesions with the same diagnosis. For example, all 14 T-lineage lymphomas on this TMA clustered together while 5 of 8 Mantle cell lymphomas, and 4 of 8 CLL/SLL were located on unique terminal branches of the dendrogram (Figure 6).

Discussion

TMAs are a significant advance over previous attempts to put multiple samples in a single paraffin block, and there are many potential benefits of using TMAs. One is the ability to screen large numbers of cases in a single staining run, thereby minimizing run-to-run variability in immunohistochemical staining. They dramatically decrease costs of conducting immunohistochemical studies, and increase the numbers of studies that can be done on small pieces of tissue, by using small cores of tissue rather than cutting sections of every block for each study.

Initial concerns about technical difficulties in creating and staining of array slides were rapidly put to rest; in our experience, the TMA paraffin blocks can be handled like

regular blocks for cutting and staining. A more significant concern centered on how representative small cores (typically 0.6-mm cores in our laboratory) of tissue would be for either morphological or immunohistochemical assessment. The available data indicate that the use of small cores is, perhaps surprisingly, highly representative of the entire tumor, and this has served as an impetus for rapidly increasing the introduction of TMA technology. For example, Nocito and colleagues,³ in a study of more than 2000 cases of bladder cancer, found that four 0.6-mm cores yielded highly concordant information on tumor grade and proliferative activity, when compared to whole sections of tumor. Camp and colleagues² found that analysis of two 0.6-mm cores yielded comparable information to whole section immunohistochemical analysis for estrogen receptor in more than 95% of cases of breast cancer studied. Most significantly, Torhorst and colleagues⁴ found that in a series of 553 cases of breast cancer, a single core of tumor was sufficient to demonstrate the prognostic significance of estrogen receptor, progesterone receptor, and p53 immunostaining, and that immunostaining of a single core was equivalent or superior to staining of whole sections, in demonstration of the prognostic significance of these markers. Thus, intratumoral heterogeneity, although an important theoretical consideration in the use of TMAs, has not proven to be an insurmountable problem in the application of TMA technology.

There are numerous potential applications of TMAs beyond testing of prognostic markers. TMAs are a superior way of testing new markers of potential diagnostic utility against a large panel of different tumors. We anticipate that the phenomenon of newly described antibodies being maximally specific in the first 6 months after they are described, with a progressive decline in specificity as more comprehensive studies are undertaken and more data accumulates, will end with the introduction of TMAs. For example, we were able to test MUM1/IRF4, a marker of plasmacytic differentiation, against 1335 different human tumors and normal tissues, showing that although it is a sensitive marker of plasmacytic differentiation, it lacks specificity, because it also stained other hematolymphoid and melanocytic tumors.⁹ Use of TMAs should become routine in this important application. We have also used TMAs for routine quality assurance; the use of a 351-case TMA block allows us to rapidly determine whether there has been change throughout time in the staining pattern of diagnostic antibodies, and to confirm that new lots of antibody have maintained their sensitivity and specificity.¹¹ In a test of interlaboratory variability in estrogen receptor staining, we were able to rapidly identify both highly concordant staining and interpretation for four of five participating laboratories, as well as a significant trend to weaker staining and resulting discrepant results from one laboratory.¹² TMA experiments are a logical follow-up to gene expression microarray experiments that show expression of groups of genes of prognostic or diagnostic significance, and we have undertaken such experiments in follow-up to our earlier gene expression studies of lymphoma,⁶ breast carcinoma,⁷ and sarcoma.¹³ This may be the most fruitful avenue of study as it allows rapid testing of novel potentially

diagnostically relevant immunohistochemical markers identified by genome-wide gene expression studies.

Given the utility of TMAs, it is essential that the tools for data manipulation and storage be sufficiently evolved to meet these emergent needs. As we started to do TMA studies we were immediately struck by the inadequacy of existing approaches for recording immunohistochemical data. TMAs allow a large number of results to be generated, but equally importantly, dramatically increase the number of results that can routinely be generated on a single case. This increased complexity required that we rethink how we manage our data, and move away from project-by-project data management to a larger system in which all data for the TMA laboratory is centrally stored. TMA-Deconvoluter assists us in this goal by allowing rapid conversion of worksheet data to tabular form. This enables a more conventional assessment of the data (eg, Kaplan-Meier survival curves, multivariate analysis) to be done, and also allows hierarchical cluster analysis to be applied to the immunohistochemical staining data. It is not clear yet if hierarchical cluster analysis will be as powerful a tool in analysis of TMA experiments as it has been in gene expression microarray experiments, in which it has been critically important. However, the complexity of data from TMA experiments parallels that of gene expression microarrays and suggests that it will be applicable. Hierarchical cluster analysis may prove useful not just as a research tool, but also in diagnosis. A significant problem in the interpretation of immunohistochemical staining results occurs when there are apparent inconsistencies in the staining profile based on a panel of immunostains (for example a pleural tumor that is calretinin-, CEA-, and B72.3-positive, and negative for WT-1, Ber-EP4, and CK5/6). Although a completely typical immunophenotypic profile is readily interpretable, having extended panels of antibodies applied to difficult cases increases the likelihood that there will be apparent inconsistencies in the staining profiles generated, with one or more antibodies giving aberrant staining results. Currently no systematic approach exists to the interpretation of an extended data set generated by staining a case with a large panel of antibodies. However, application of the system described in this study may be a first step toward more complex analyses of immunostaining results. In a separate experiment, we have shown that with a set of 351 cases stained with 22 antibodies, hierarchical cluster analysis allows a significant degree of clustering of tumors according to tissue of origin.¹¹ In the current study we show that lymphomas of one type tend to cluster together. There are impediments to the diagnostic application of hierarchical cluster analysis, especially the lack of sufficiently tissue-specific markers (although newer markers such as myogenin and thyroid transcription factor are a significant advance). As well, the dynamic range of immunostaining scores is significantly narrower than that obtained for mRNA levels in gene microarray experiments and, as a result, the clustering of tumors based on immunoreactivity can be expected to be less well defined than is seen in gene expression studies. However, new technologies that can quantify the intensity of immunoreactivity may improve the dynamic range of immunostain data.

The main cost of the system we describe is the Bliss system for capture and storage of digital images. Although such a system is desirable, it is possible to set up a data management system for a TMA lab with nothing more than a PC with Excel and a statistical analysis package, as the other software components we describe are freely available. To realize the full potential of TMA technology, however, a digital image collection will allow for revisiting and reanalyzing immunostain interpretations.

The software system described in this article has been used successfully by groups outside our institution. By using the same software, it has made it possible to share the same arrays and array data, facilitating collaborations around areas of unique expertise and further maximizing the important resource that TMAs represent.

Acknowledgments

We thank F. Hsu for comments and suggestions on the use of TMA-Deconvoluter and Drs. Pat Brown and David Botstein for their support. Supported by grant NIH-CA-85129.

References

1. Kononen J, Bubendorf L, Kallioniemi A, Barlund M, Schraml P, Leighton S, Torhorst J, Mihatsch MJ, Sauter G, Kallioniemi OP: Tissue microarray for high-throughput molecular profiling of tumor specimens. *Nat Med* 1998, 4:844–847
2. Camp RL, Charette LA, Rimm DL: Validation of tissue microarray technology in breast carcinoma. *Lab Invest* 2000, 80:1943–1949
3. Nocito A, Bubendorf NA, Maria Tinner E, Suess K, Wagner U, Forster T, Kononen J, Fijan A, Bruderer J, Schmid U, Ackermann D, Maurer R, Alund G, Knonagel H, Rist M, Anabitar M, Hering F, Hardmeier T, Schoenberger AJ, Flury R, Jager P, Luc Fehr J, Schraml P, Moch H, Mihatsch MJ, Gasser T, Sauter G: Microarrays of bladder cancer tissue are highly representative of proliferative index and histological grade. *J Pathol* 2001, 194:349–357
4. Torhorst J, Bucher C, Kononen J, Haas P, Zuber M, Kochli OR, Mross F, Dietrich H, Moch H, Mihatsch M, Kallioniemi OP, Sauter G: Tissue microarrays for rapid linking of molecular changes to clinical end-points. *Am J Pathol* 2001, 159:2249–2256
5. Eisen MB, Spellman PT, Brown PO, Botstein D: Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998, 95:14863–14868
6. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J, Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, Staudt LM: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000, 403:503–511
7. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lønning PE, Borresen-Dale AL, Brown PO, Botstein D: Molecular portraits of human breast tumours. *Nature* 2000, 406:747–752
8. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lønning P, Borresen-Dale AL: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 2001, 98:10869–10874
9. Natkunam Y, Warnke R, Montgomery K, Falini B, van de Rijn M: Analysis of MUM1-IRF4 protein expression using tissue microarrays and immunohistochemistry. *Mod Pathol* 2001, 14:686–694
10. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999, 286:531–537
11. Hsu FD, Nielsen TO, Alkushi A, Dupuis B, Huntsman D, Liu CL, van de Rijn M, Gilks CB: Multiple tumor tissue microarrays are an effective quality assurance control for diagnostic immunohistochemistry. *Mod Pathol* 2002, in press
12. Parker RL, Huntsman DG, Lesack DW, Cupples JB, Grant DR, Akbari M, Gilks CB: Assessment of inter-laboratory variation in the immunohistochemical determination of estrogen receptor using a breast cancer tissue microarray. *Am J Clin Pathol* 2002, 117:723–728
13. Nielsen TO, West RB, Linn SC, Alter O, Knowling MA, O'Connell JXO, Zhu S, Fero M, Sherlock G, Pollack JR, Brown PO, Botstein D, van de Rijn M: Molecular characterisation of soft tissue tumors: a gene expression study. *Lancet* 2002, 359:1301–1307