Software Tools for Managing and Analyzing TMA Data

Contains documentation for the following:

The TMA-Deconvoluter
Version 1.06
Usage Walkthrough

Stainfinder
Version 1.00
Usage Walkthrough

Compiled by Chih Long Liu
7-1-2002

Table of Contents

Terms and Conditions for Use

Software is being provided free of charge, provided that:

1. User is affiliated with an academic or non-profit organization

2. User may not distribute any modified versions of the software or any of its accompanying documentation without prior consultation with the authors

3. User provides proper citation or acknowledgment if any of the software programs documented in this file are used in a paper that has been accepted for publication

Because the software is being provided free of charge, the authors do not claim responsibility for any damages, monetary or otherwise, stemming from misuse of this software.

Commercial interests must contact the authors for proper licensing arrangements prior to use.

Please refer to the TMA website, http://genome-www.stanford.edu/TMA, for correspondence information of the authors.

Overview

The TMA-Deconvoluter is a deconvolution program designed for tissue microarrays (TMAs) scanned with the Bliss microscope system. It facilitates conversion of raw scoring worksheets, formatted in a manner similar to the three-dimensional layout of the tissue microarrays, into the two-dimensional text tab-delimited format which has been optimized for hierarchical clustering with Mike Eisen's Cluster program. Output from this program permits analytical techniques such as clustering to be performed on tissue microarrays. Additionally, this program produces output that is compatible with a web-based image retrieval system based on the BLISS system and the Stainfinder program.

In this document, you will find instructions on how to set up a system for managing your TMA data in this fashion. You will find descriptions on producing or acquiring the necessary components of this system in this document. Afterwards, you will find instructions on the actual operation of the TMA-Deconvoluter and Stainfinder, along with the requisite components of this system.

For further details on how you can adapt the TMA-Deconvoluter and Stainfinder to manage your TMA needs, please refer to the corresponding paper:

Liu et al., "Software Tools for High-throughput Analysis and Archiving of Immunohistochemistry Staining Data Obtained with Tissue Microarrays," (2002), *in review*.

The instructions and documentation you will find here will largely overlap the contents of the TMA website, http://genome-www.stanford.edu/TMA. The only differences will involve features not documented in the website that relate to the finer points of operating the TMA-Deconvoluter.

Please note that important feature additions, changes, bug fixes, and any other changes affecting this TMA data management system will be updated on the website more frequently than in this documentation.

System Requirements

This is a quick summary of the items you will need to set up your data management system for your TMAs. You will be needing the following:

- Excel-format raw scoring workbooks representing data in the layout of your TMA. **More** on the format of these workbooks, in the Appendix..

- A lookup file, in text tabulated table format, containing your case descriptor information. **More** on the format of these workbooks, in the Appendix..

- The TreeView program. You will use this to rapidly visualize the scoring results of large sections of your TMA data. This program can be used in conjunction with Stainfinder. You can obtain TreeView at http://rana.lbl.gov/EisenSoftware.htm.

- The Cluster program. You will use this to perform hierarchical clustering analysis on your data, if desired. The files output by this program are designed to work with TreeView. You can obtain Cluster at http://rana.lbl.gov/EisenSoftware.htm.

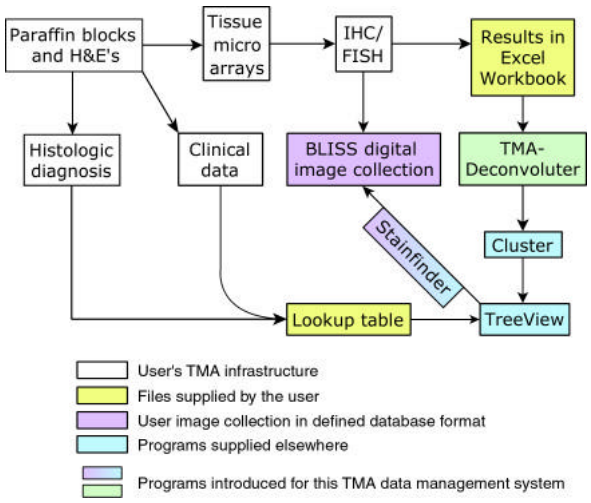
- An on-line database of stored images. In the present configuration, the TMA-Deconvoluter and Stainfinder have been configured to work with image filenames defined by the nomenclature of the Bliss microscope system (Bacus Laboratories Inc.). Your database can run on the platform of your choice.

- A web server that is capable of running CGI programs written in PERL. The Stainfinder walkthrough will include instructions on setting up a web server. ***Note: it is highly recommended that a system administrator or some other person experienced in networking and internet technologies set up the web server, since an improperly configured server can be vulnerable to attack by malicious users.***

- A Windows PC running Microsoft Excel 97 or later. This is necessary to run the TMA-Deconvoluter program. **Note: the Mac OS is NOT supported**, because the TMA-Deconvoluter makes the use of Active X controls, which are not available in Macintosh versions of Microsoft Office. See the **FAQ** section for details.

Once you have all of these components assembled, proceed to the **walkthrough** section.

Walkthrough



The diagram above provides an overview of the TMA data management system and where the various system components lie in relation to each other.

Once you have produced the necessary files (as outlined in yellow in the diagram above), proceed to the **walkthrough** for using the **TMA-Deconvoluter** to generate files for data analysis, and performing clustering analysis on your TMA data.

Afterwards, proceed to the **walkthrough** for setting up the **Stainfinder** to link datasets viewed under TreeView with the on-line Bliss digital image collection.

TMA-Deconvoluter Walkthrough

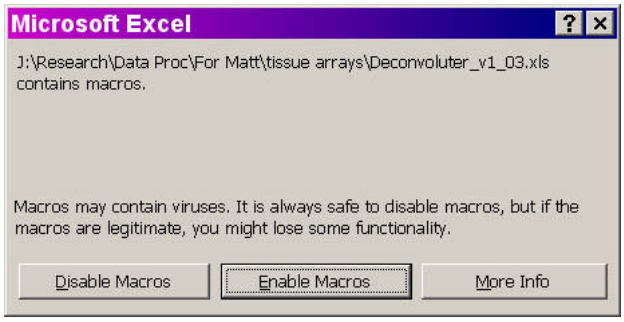
Note: if this is your first time using this TMA data management system, it is recommended that you proceed to the Downloads section of the TMA website, http://genome-www.stanford.edu/TMA/download.shtml and download the Demo Suite files. The examples presented in this walkthrough employ those files, and you should obtain the same results in using these files and with the other necessary programs available elsewhere.

What you will need for this walkthrough:

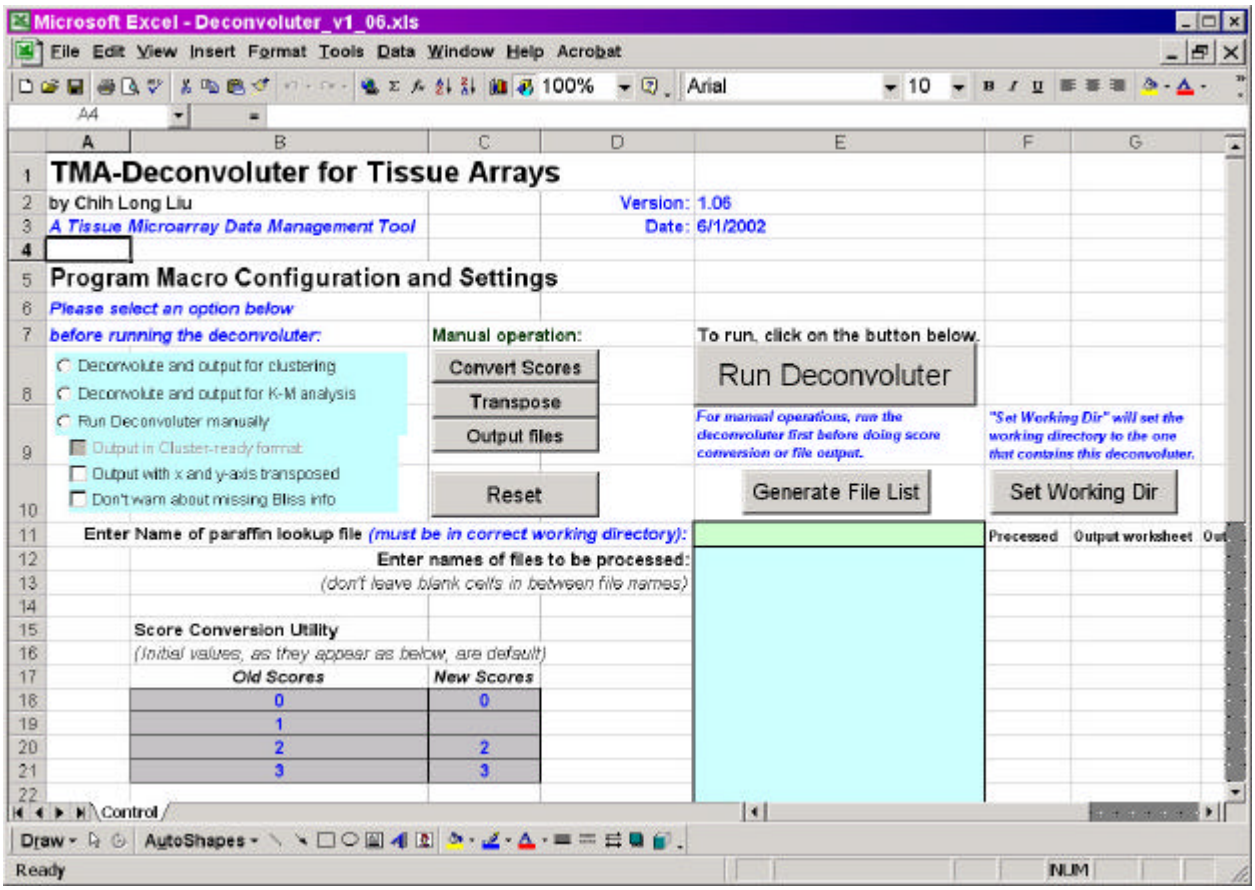
- Raw scoring worksheets containing your TMA data
- A lookup file containing information about each case
- The TMA-Deconvoluter
- The Cluster program
- The TreeView program
- Microsoft Excel for Windows with Visual Basic support installed
- A Windows PC

Begin by storing a copy of the TMA-Deconvoluter, your raw scoring worksheets, and the lookup file in the same folder. (If you decide to put the files elsewhere, you will need to know how to specify the appropriate path information in the entry fields for those files).

You may now open the TMA-Deconvoluter. A dialog box might pop up before the TMA-Deconvoluter opens, asking you if you want to enable macros (as shown below). Click "Enable Macros". If you leave them disabled, the TMA-Deconvoluter will not run.



After opening the TMA-Deconvoluter, you should be greeted with the screen below:

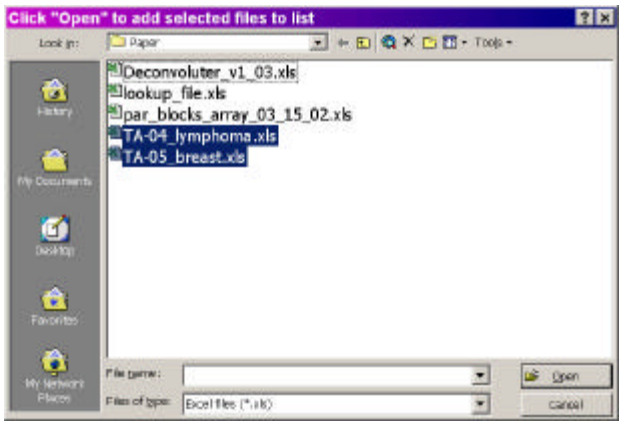


Excel normally sets the current working directory based on your default settings, which may not necessarily be the directory containing the TMA-Deconvoluter and the other files that you placed in that folder. **Click on "Set Working Dir"** (right portion of screenshot) to set the working directory to be the container for your TMA files. You may see a dialog box indicating the current working directory and the new working directory, and it will request whether this is correct. Click "Yes" or "No" to continue.

Specifying your input files

You now need to specify the name of the "paraffin" lookup file. Enter the full exact filename, including the extension, in the pale green worksheet cell. (The version available for download comes prepared for operation with the demo files.) Alternatively, you can click on the "Generate File List" button and select your file from a menu (see below).

Specify the names of your raw scoring workbooks in the light blue worksheet cells. The TMA-Deconvoluter is capable of batch operation, so you can process multiple files at a time. A helpful feature is file list generation, where instead of having to type in all the file names of the list of files you wish to process, you can graphically select the desired files in your folder. To do this, click on "Generate File List". An Open file dialog box will appear below:



Select files by clicking and dragging. To select multiple files that are not listed next to each other, hold down the Ctrl key while selecting the desired files, either with the mouse or by using the arrow keys and the spacebar. **Click "Open"** to add those files to your list.

You will now see an additional dialog box asking whether or not you wish to specify the lookup file. If you click "Yes", you will get another Open file dialog box very similar to the one above, except that you will be allowed to select only one file at a time. Select your lookup file and click "Open" to add the file name in the light green worksheet cell.

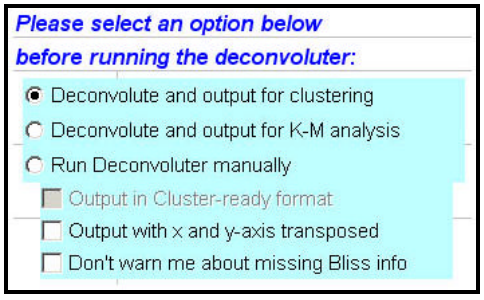
Selecting Output Options

Optional: You may specify the names of the output files, which will be in text tab-delimited format. Specify the names of the output files (without the **.txt** extension) in the appropriate cells (see below).

lookup_file.xls	Processed	Output worksheet	Output filenames
TA-04_lymphoma.xls			4
TA-05_breast.xls			5

As shown in the example above, the output from TA-04_lymphoma.xls will be outputted to the file 4.txt. Output from TA-05_breast.xls will be outputted to the file 5.txt. A later section of the walkthrough will further describe the "Processed" and "Output worksheet" cells.

Now, choose an output option (such as that shown below).



For clustering, the default is shown above. The next step in this walkthrough will also show what the output will look like if you had chosen the K-M output option.

A note on scores: the TMA-Deconvoluter will assume that you are using the following scoring convention:

Score	Description	Treeview score	Appearance under Treeview
	Missing datapoint		
0	Negative	-2	
1	equivocal/uninterpretable		
2	weak	1	
3	strong	2	

Please note that our initial scoring system was as follows: 0 = negative, 1 = uninterpretable/equivocal, 2 = weakly positive, 3 = strongly positive. The score conversion utility allows us to automatically convert these values to values that are symmetric around zero. In this case negative becomes -2, an uninterpretable score (old score= 1) will be left open in the new scores and subsequently will show a gray box in Treeview. Weakly positive (old score = 2) will become 1 and strongly positive (old score = 3) will become 2. Accordingly, the default scoring conversion values have been set in this way:

Score Conversion Utility

(Initial values, as they appear as below, are default)

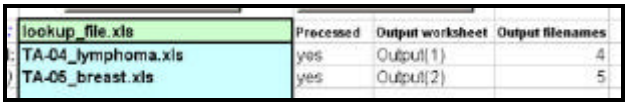
Old Scores	*New Scores*
0	-2
1	
2	1
3	2

Score conversion can be carried out separately under manual operation, but you will need to select the "Run Deconvoluter manually" option and **Run Deconvoluter** first before converting the scores.

You are now ready to run the Deconvoluter. Click on the button as shown above.

After the TMA-Deconvoluter has completed its operation, you will find yourself back at the main screen of the TMA-Deconvoluter (at the "Control") worksheet. You will notice that the

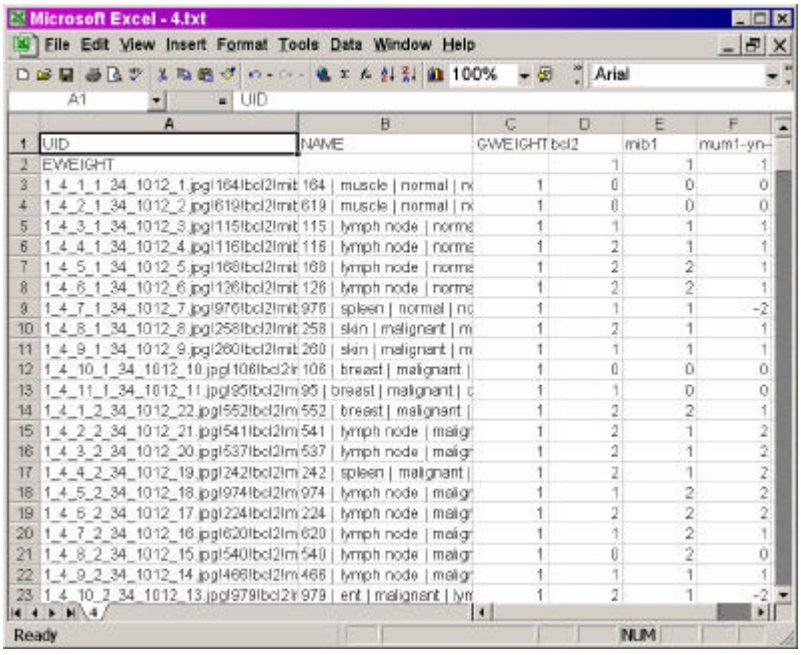
output appears in additional worksheets within the TMA-Deconvoluter, and that a report of the activity of the TMA-Deconvoluter appears, as shown below:



lookup_file.xls	Processed	Output worksheet	Output filenames
TA-04_lymphoma.xls	yes	Output(1)	4
TA-05_breast.xls	yes	Output(2)	5

The "Processed" column indicates whether or not the raw scoring workbook was successfully processed. The "Output worksheet" column indicates the name of the worksheet within the TMA-Deconvoluter that contains the output data, and the "Output filenames" column contains the file names of the output files. If you had specified your own filenames, these would remain unchanged, and the corresponding files (4.txt, 5.txt in this example) would appear in the current working directory. If you had not specified any file names, the file names will be the same as the worksheet names.

Opening one of the output file names, or clicking on the corresponding worksheet tab within the TMA-Deconvoluter, should result in the screen below (if you had chosen to output in the PCL format):



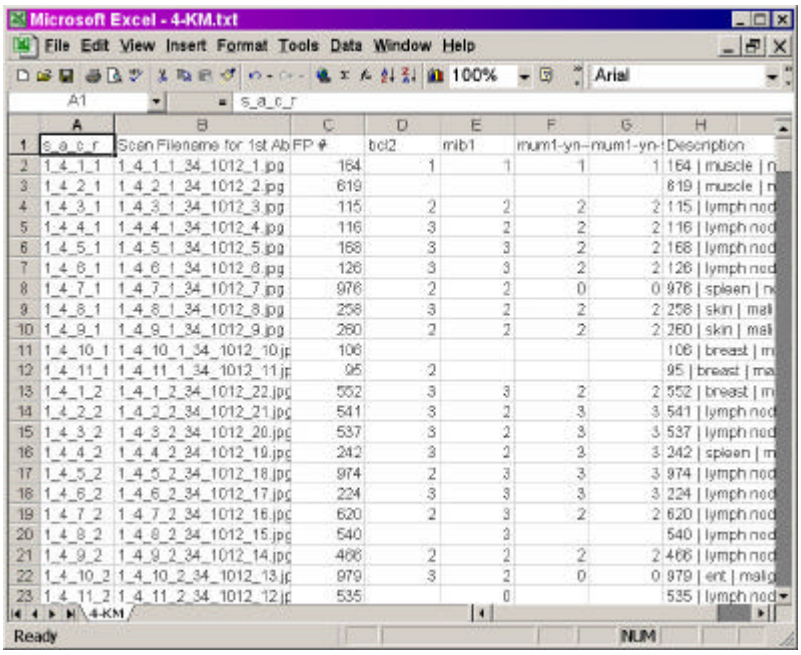
	A	B	C	D	E	F
1	UID	NAME	GWEIGHT	bcl2	mib1	mum1-yn-
2	EWEIGHT			1	1	1
3	1_4_1_1_34_1012_1.jpg\164\bcl2\mit	164 | muscle | normal | n	1	0	0	0
4	1_4_2_1_34_1012_2.jpg\619\bcl2\mit	619 | muscle | normal | n	1	0	0	0
5	1_4_3_1_34_1012_3.jpg\115\bcl2\mit	115 | lymph node | norma	1	1	1	1
6	1_4_4_1_34_1012_4.jpg\116\bcl2\mit	116 | lymph node | norma	1	2	1	1
7	1_4_5_1_34_1012_5.jpg\168\bcl2\mit	168 | lymph node | norma	1	2	2	1
8	1_4_6_1_34_1012_6.jpg\126\bcl2\mit	126 | lymph node | norma	1	2	2	1
9	1_4_7_1_34_1012_7.jpg\976\bcl2\mit	976 | spleen | normal | no	1	1	1	-2
10	1_4_8_1_34_1012_8.jpg\258\bcl2\mit	258 | skin | malignant | m	1	2	1	1
11	1_4_9_1_34_1012_9.jpg\260\bcl2\mit	260 | skin | malignant | m	1	1	1	1
12	1_4_10_1_34_1012_10.jpg\106\bcl2\r	106 | breast | malignant |	1	0	0	0
13	1_4_11_1_34_1012_11.jpg\95\bcl2\m	95 | breast | malignant | c	1	1	0	0
14	1_4_1_2_34_1012_22.jpg\552\bcl2\m	552 | breast | malignant |	1	2	2	1
15	1_4_2_2_34_1012_21.jpg\541\bcl2\m	541 | lymph node | malig	1	2	1	2
16	1_4_3_2_34_1012_20.jpg\537\bcl2\m	537 | lymph node | malig	1	2	1	2
17	1_4_4_2_34_1012_19.jpg\242\bcl2\m	242 | spleen | malignant |	1	2	1	2
18	1_4_5_2_34_1012_18.jpg\974\bcl2\m	974 | lymph node | malig	1	1	2	2
19	1_4_6_2_34_1012_17.jpg\224\bcl2\m	224 | lymph node | malig	1	2	2	2
20	1_4_7_2_34_1012_16.jpg\620\bcl2\m	620 | lymph node | malig	1	1	2	1
21	1_4_8_2_34_1012_15.jpg\540\bcl2\m	540 | lymph node | malig	1	0	2	0
22	1_4_9_2_34_1012_14.jpg\466\bcl2\m	466 | lymph node | malig	1	1	1	1
23	1_4_10_2_34_1012_13.jpg\979\bcl2\r	979 | ent | malignant | lym	1	2	1	-2

For those of you familiar with the PCL (Pre-CLuster) format, you will recognize the formatting of the output.

- In Column A is the UID (Unique IDentifier) column, which contains the information passed on by TreeView to the Stainfinder program (for more information on this, please refer to the Stainfinder walkthrough).

- In Column B is the NAME column, which contains the information obtained from the lookup file. The information present is the same as Columns A-F in the lookup file, except with a pipe ("|") separating the different columns from that file. The unique case identifier, followed by the diagnostic information, appears in the field.
- In Column C is the GWEIGHT column, which defines the absolute weight each case is given in the clustering. The default value for each case is 1, indicating that each cases given equal weight in the hierarchical clustering. You may alter these values to some other number, prior to clustering.
- Columns D and onward: each column represents the scoring data obtained from a slice of your TMA. The name of the antibody, as specified in your original raw scoring workbook, used to stain that slice, appears at the top row of that column.
- In Row 2 is the EWEIGHT column, which defines the absolute weight each slice is given in the clustering. the default value for each slice is 1, indicating that each slice is given equal weight in the hierarchical clustering. You may also alter these values to some other number, prior to clustering.

If you had chosen to output in the K-M format, you will get the following instead:



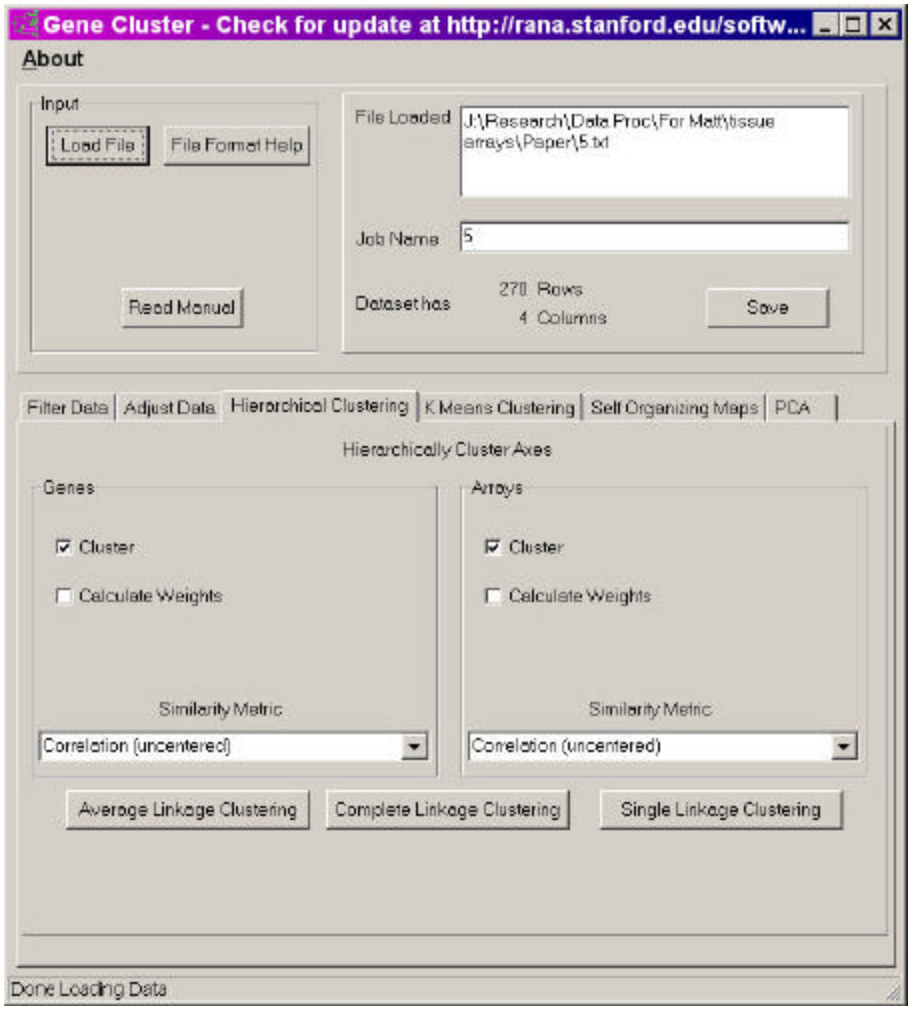
- Column A provides information on the physical location of the spot within the TMA in the following format, s_a_c_r, where:
  s = sector number
  a = array number
  c = column number
  r = row number
- Column B provides information on the corresponding digital image filename for a particular spot. The nomenclature of this filename uses the numeric coding system used is based on the Bliss microscope system Bacus Laboratories Inc., consisting of seven numbers separated by underscores. This is covered in the Stainfinder walkthrough.

- Column C consists of the unique case identifier. It is labeled as "FP#" because the lookup files were generated from a FileMaker Pro database in the van de Rijn laboratory.
- Columns D-G (in this example; TMAs with a larger number of slices will occupy more columns in this range) contain the unmodified score data of the TMA.
- Column H (in this example; in other datasets, it will be the rightmost column in the file) contains the description for that spot and is formatted in the same way as the NAME column in the PCL output file.

You are now ready to proceed with hierarchical clustering or any other sort of statistical analysis you wish to perform on your TMA data.

Clustering deconvoluted data, prior to viewing in Treeview

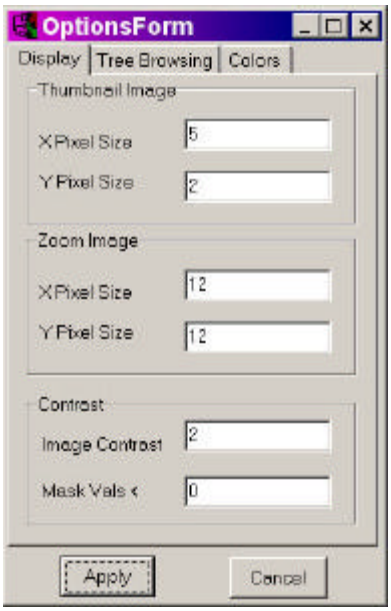
Open the Cluster program. If you don't have it, download it from http://rana.lbl.gov/EisenSoftware.htm. You may pre-filter the data if you wish; otherwise, proceed to clustering by clicking on the "Hierarchical Clustering" tab. You should see the following screen below:



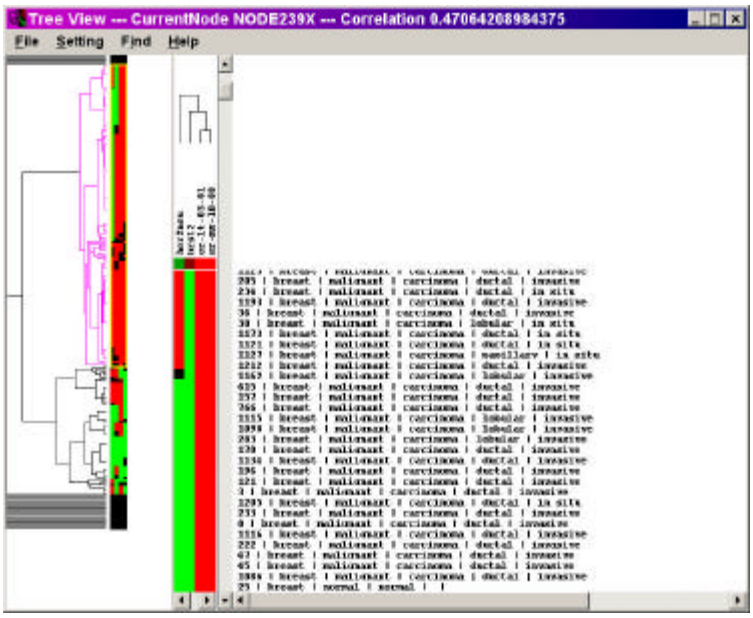
The default parameters for clustering are shown in the screenshot above. If you wish, you may modify these parameters or other parameters available in the Cluster program. However, the demo files provided on this website have been clustered with these default parameters. Click on "Average Linkage Clustering". The resulting CDT, GTR, and ATR files will appear in your working directory and are ready for viewing under TreeView. If you wish to find out more about the various clustering methods and other features of this program, please refer to the corresponding documentation for Cluster.

Open the TreeView program. If you don't have it, download it from http://rana.lbl.gov/EisenSoftware.htm.

Go to **File → Open**, and open your newly clustered file (e.g. 4.cdt, in this example). You may adjust the thumbnail and zoom image pixel size (defined for each individual datapoint, represented by a single block of color) to your preferences. You will also need to adjust the contrast to 2 (representing a dynamic range of 2-fold induction or 2-fold repression in gene microarrays, for which TreeView was originally designed).



If you use the parameters indicated in the screenshot above, and using the demo file 4.cdt, you should get the following:



You may recall that the scoring data is represented in the following manner:

Score	Description	Treeview score	Appearance under Treeview
	Missing datapoint		
0	Negative	-2	
1	equivocal/uninterpretable		
2	weak	1	
3	strong	2	

You may now browse your dataset for additional analysis. To use TreeView with Stainfinder, proceed to the Stainfinder walkthrough, which begins on the next page.

Stainfinder Walkthrough

Note: if this is your first time using this TMA data management system, it is recommended that you proceed to the Downloads section and download the Demo Suite files. The examples presented in this walkthrough employ those files, and you should obtain the same results in using these files and with the other necessary programs available elsewhere. It is also highly recommended that you first go through the TMA-Deconvoluter ***walkthrough*** *if you haven't already done so.*

What you will need for this walkthrough:

- Your digital image collection with files created or named in the Bliss filename nomenclature.
- A computer connected to your intranet or to the Internet that will serve as the web server for your digital image collection.
- A web server program that can support CGI.
- The PERL interpreter appropriate for your web server.
- The Stainfinder program.
- TreeView
- A clustered dataset
- A web browser capable of displaying images

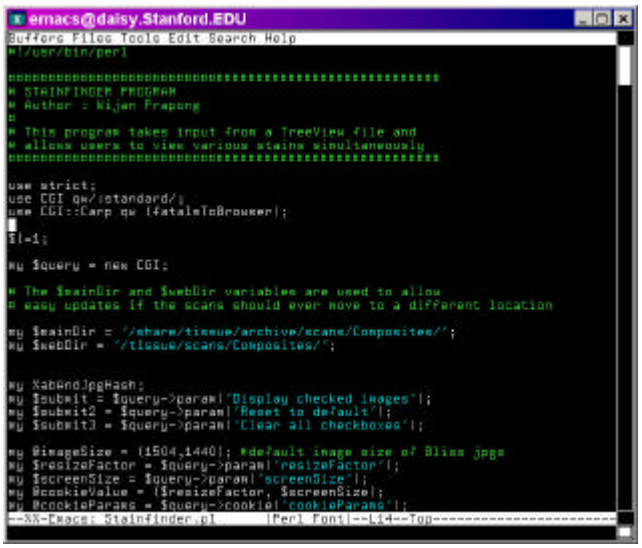
1. Begin by setting up your web server. Proceed to the Appendix for instructions.

2. Deposit your TMA image files in your web server. Please refer to the Bliss file system **nomenclature** in the Appendix for proper file system structure and file names. Be sure that the entire collection resides inside the directory structure recognized by the web server.

3. Deposit your most updated lookup file into your web server, in the "Composites" directory of your TMA image file collection.

4. Configure Stainfinder.

Open the copy of the Stainfinder program stored on your web server with your favorite text editor. If you modify a copy stored elsewhere, be sure to upload it to your web server once you have completed your modifications. An example is shown below.



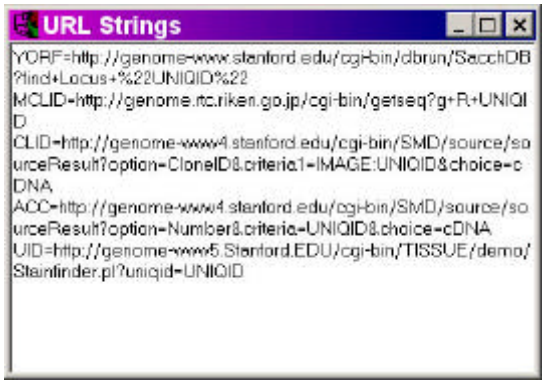
Modify lines 22 and 23 so that they point to the correct location of your files on your web server. The exact location will depend on your server configuration.

Save the changes and close the text editor.

Configuring TreeView

Open the TreeView program and load your clustered dataset. If you have not already done so, proceed to the **walkthrough** for generating a clustered dataset of your TMA data. Alternatively, you may download the demo clustered datasets from the Downloads section and use that instead.

Go to the Setting menu and select “Edit URL strings”. A screen like the one shown below will pop up.



On a new line, enter the URL string corresponding to the address of the Stainfinder program on your web server, plus the parameters. The screenshot above indicates the correct URL string for the demonstration database on the Stanford Genome servers. For example, if the Stainfinder program was stored in the **cgi-bin** directory of your server, and your server address was www.myserver.edu, you would specify the following:

UID=http://www.myserver.edu/cgi-bin/Stainfinder.pl?uniqid=UNIQID

You will notice that TreeView is passing whatever is specified in the UID column to the variable UNIQID. For example, if your UID column contained the following:

1_4_1_1_34_1012_1.jpg!164!bcl2!mib1!mum1!mum1

then it will send the following URL to your default browser (as a single unbroken line):

http://genome-www5.Stanford.EDU/cgi-bin/TISSUE/demo/Stainfinder.pl?uniqid=1_4_1_1_34_1012_1.jpg!164!bcl2!mib1!mum1!mum1

If you have configured Stainfinder and TreeView properly on your computer and on your web server, clicking on a TreeView link should bring up a screen similar to what you will get when you click on the link above.

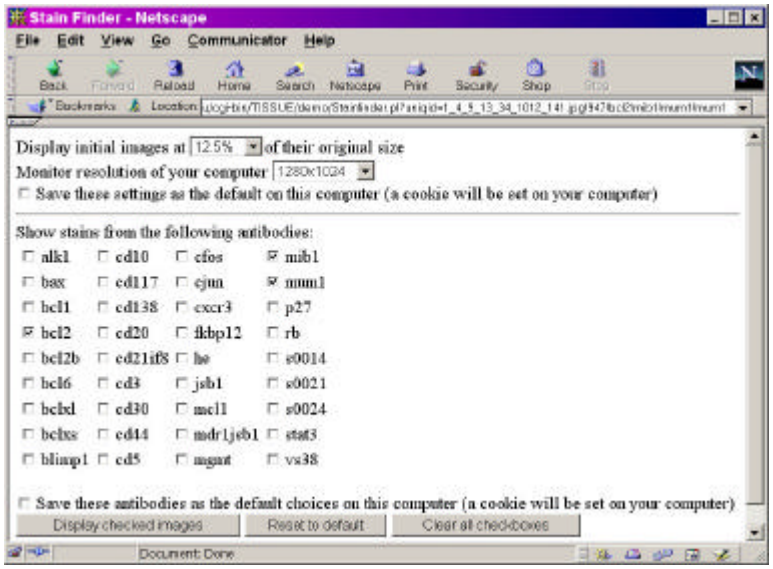
Using Stainfinder with TreeView

If you have correctly configured your web server and Stainfinder, and if you have a clustered data sets prepared, then you are ready to start using Stainfinder with TreeView.

1. Open the TreeView program and load your clustered dataset. If you have not already done so, proceed to the walkthrough for generating a clustered dataset of your TMA data. Alternatively, you may download the demo clustered datasets from the Downloads section and use that instead.

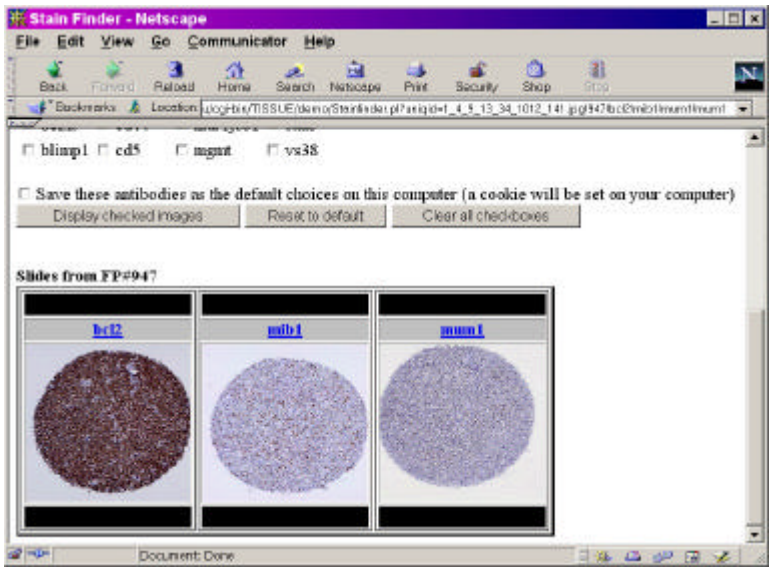
2. You may now click on one of the descriptors (shown below) to view the corresponding spot images from your collection.

3. A new browser window will open, and you should get a screen like the one shown below (depending on how many antibodies stains you have available on your digital image collection for that particular core):



Here, the default settings will be selected. The thumbnail images will be shown in the same browser window, and you can select a percentage value of the original image size for those thumbnails. You should also select the correct maximum resolution for your monitor, since Stainfinder will size the browser window of zoomed images (more on this in the next step). You will notice that Stainfinder will provide a list of all available antibodies for that particular core sample, with the relevant antibodies preselected, based on your data set being viewed in TreeView. You may alter your selections as desired. In addition, if you wish to retain your altered selections, choose your appropriate options, and a cookie will be set on your computer.

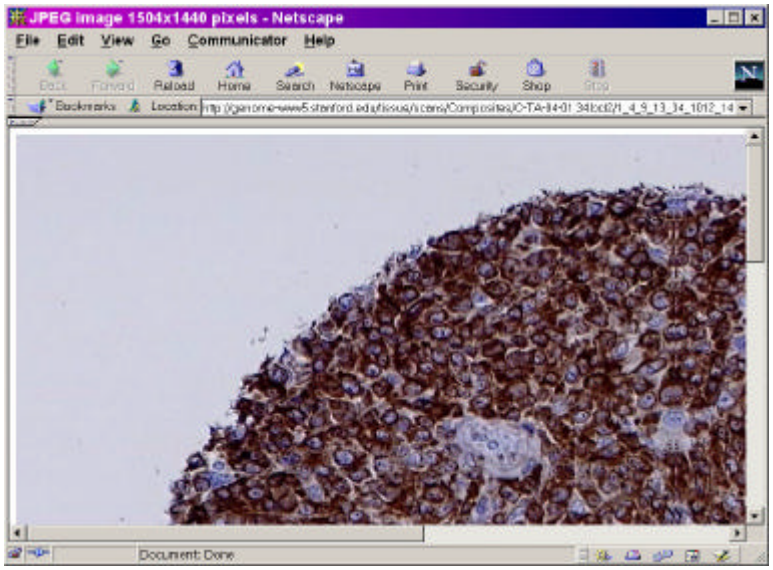
4. Click on “Display checked images”. You should get a screen like the one shown below (depending on your selection of antibodies):



You will notice that the images retrieved correspond to either the antibodies present in your datasets, or your own selections. This is a helpful feature for comparing the scores shown in your dataset under TreeView with the actual stain images.

Note: you may have to scroll down to see your images. In addition, if an antibody appears more than once in your data set (as shown earlier in this TreeView example), it will be displayed only once with Stainfinder.

5. You may now click on the link corresponding to one of the images. This will bring up the browser window containing the full-size image, allowing you to browse the image in full detail:



You are now ready to use Stainfinder with TreeView to verify scores, compare stains, or perform any other function that would assist you in exploration of your digital image collection and corresponding data analysis.

An on-line demonstration of a clustered dataset with Stainfinder is available at http://genome-www.stanford.edu/TMA/explore.shtml.

Additional Features of the TMA-Deconvoluter

Note: manual operation is recommended only for users who are familiar with the TMA-Deconvoluter.

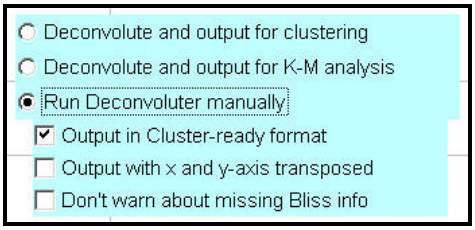
While automatic operation of the TMA Deconvoluter for this system is sufficient for many users, it is possible to run the TMA Deconvoluter manually for increased flexibility. When automatic operation is selected (by choosing one of the two output options and then clicking "Run Deconvoluter", the following is performed:

1. The raw scoring workbooks are deconvoluted and outputted to a new worksheet in the TMA Deconvoluter, based on the output option selected.
2. The scores in each output worksheet is converted to new scores, based on the output option.
3. Each worksheet is output to a text file and saved.

Manual operation entails running each of these steps separately, and the user may choose not to run all of these steps. This added flexibility is helpful – for example, the user may choose to:

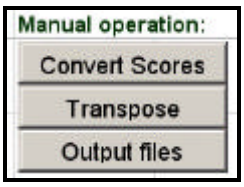
- keep the scores unchanged, skipping step 2
- not output any files, skipping step 3
- perform other data manipulations following deconvolution in step 1 prior to outputting to files.

To perform manual operations, select the "Run deconvoluter manually" option, as shown below:



In this state, the Deconvoluter will, by default, output in a format ready for K-M (or other) analysis. To output in cluster-ready format, check the corresponding option, as shown above. You will notice that this checkbox option is now enabled while the Deconvoluter is in manual mode.

Proceed as you would for normal operation in running the Deconvoluter. You will notice, however, that clicking on "Run Deconvoluter" will run only the Deconvolution step. To execute the other steps, click on one of the manual operation buttons, as shown below:

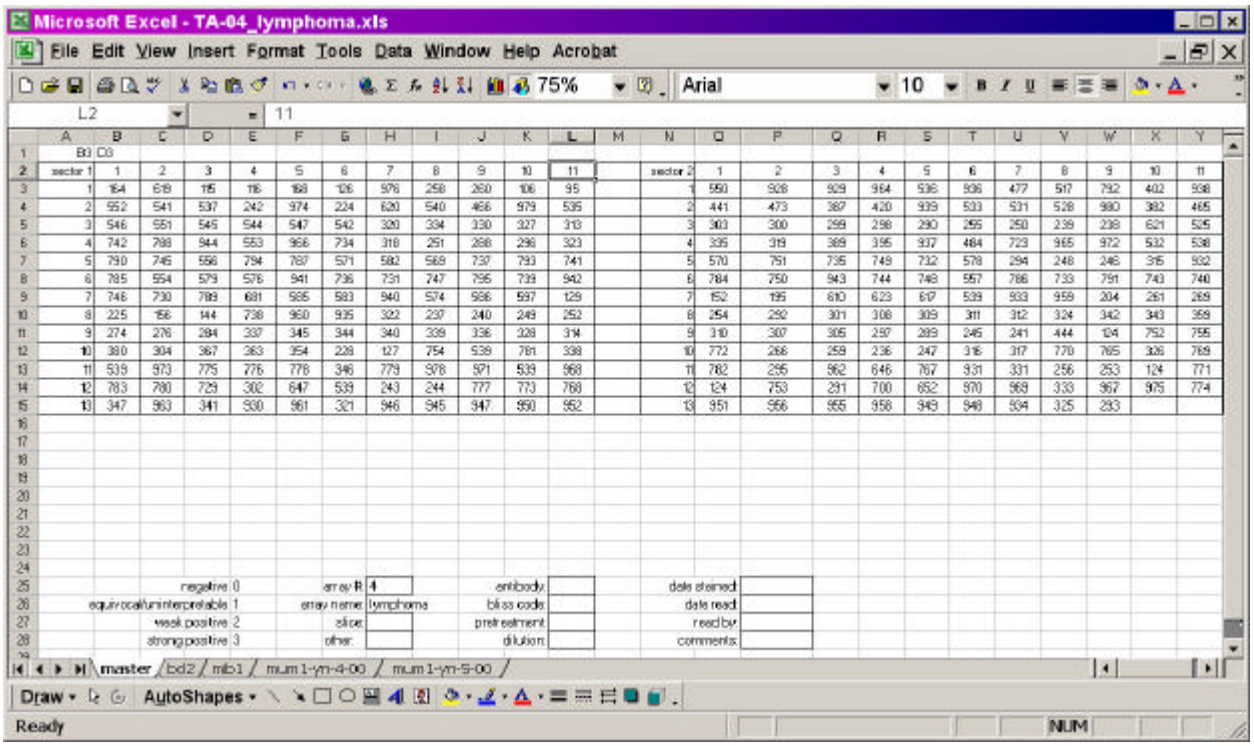


- **Convert Score** – Using the Score Conversion Utility, the scores are converted based on user settings. This process can be performed iteratively – if the user has more than 4 different score types to convert, the use can specify the first four scores, click "Convert Scores", and repeat the process until all scores have been converted. *Note: the TMA-Deconvoluter will not check whether or not the specified scores exist in the dataset prior to performing the conversion and may return an error if such scores do not exist.*
- **Transpose** – transpose the data by swapping the x and y-axis. Note: Because fo Excel's 256-column limit, it is highly inadvisable to perform transposition on a data set that contains more than 250 cases.
- **Output files** – output the data from the output worksheets in the TMA-Deconvoluter to text tab-delimited files.

These operations can be performed in any order (after deconvolution), but the user should know what (s)he is doing to the data.

Appendix

Raw Scoring Worksheet Format



Master worksheet screenshot

- The first worksheet in the workbook should be a master worksheet that defines the layout of the sectors that comprise the tissue microarray. We routinely divide our tissue microarrays in sectors that are separated on the physical microarray by a space. We find it easier to read our tissue arrays if the maximum number of columns on an array sector does not exceed 12-14 columns. In addition we find that scanning images with the Bliss system in four sectors of, say 125, cores is easier than scanning all 500 images on an array in a single run. The drawback of course is that each sector needs to be set up for scanning separately on the Bliss system and that each sector requires its own folder of jpeg files. Should you decide to have all your sectors combined in one, the notation as below should still be maintained but of course you would avoid naming sectors 2, 3 and 4. A screenshot of the master worksheet appears above. The name of this worksheet should be "master" and should be identical in layout to all other worksheets that exist in this workbook.

- Cell "A1" (the top left cell in the worksheet) should contain the cell address, in R1C1 format of sector one, row one, column one. In the vast majority of cases, this will be "B3".

- The addresses for any additional sectors should be similarly defined for all other sectors in the TMA, and they should be specified in the cells directly to the right of cell A1 (i.e.

they should all be located in Row 1 of the worksheet). In addition, the number of total cells containing these addresses will be used to determine the total number of sectors that exist in your tissue microarray.

- Each sector should have a header column and a header row that defines the total number of rows and columns for that sector. The TMA-Deconvoluter explicitly uses the sector headers in the master worksheet in order to determine the sector dimensions. The sector name (e.g. "Sector 1") should be located in the cell where the header column and row meet. This cell should not be empty.

- There must be at least one row and column of separation between sectors. If there is no separation, the dimensions of that sector may be incorrectly detected by the TMA-Deconvoluter. Following this formatting convention, it is possible to have four or more sectors in your TMA.

- For the master worksheet, the data region of the sectors should contain the case identifiers that provide the key to the lookup file containing additional information. The identifiers do not have to be arranged in order, and if there is an empty cell, the corresponding case will not be present in the output file, regardless of whether or not scores exist in the subsequent data worksheets.

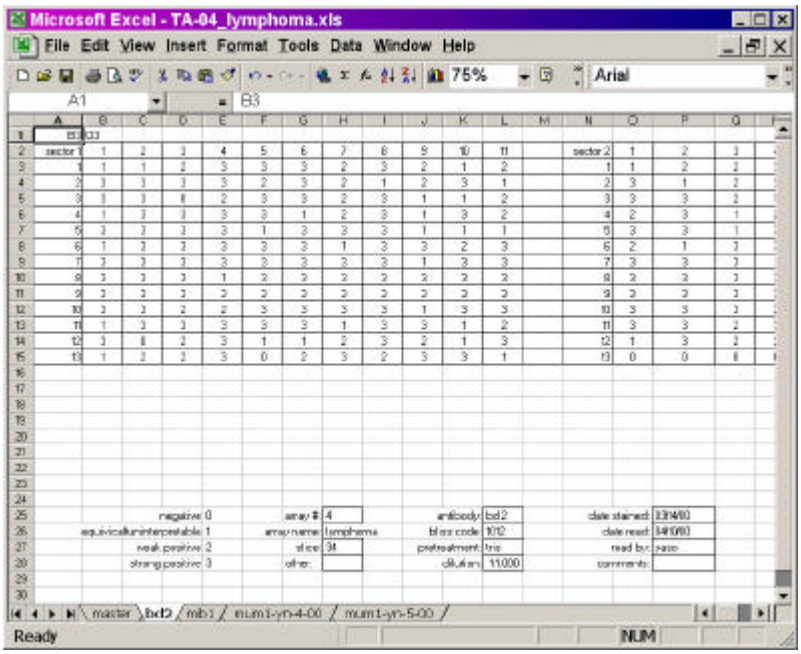
- Below the sector layout, additional information of the TMA can be provided. Please refer to the screenshot for an example. Currently, the following are required for compatibility with Stainfinder; the cells containing the information must be positioned at the indicated addresses below:

Cell Address	Description
H25	array #
H26	array name
H27	slice #
L25	antibody[*,+]
L26	bliss code[+]

* *Note: this is what Stainfinder uses to identify antibodies. It should not have any punctuation characters, and should be identical to name of antibody in the name of the folder in which the images are stored. Take a look at the Bliss filename nomenclature to see how the file system is organized.*

\+ *Note: if this information is missing, TMA-Deconvoluter will warn the user. This information is necessary for proper operation with Stainfinder.*

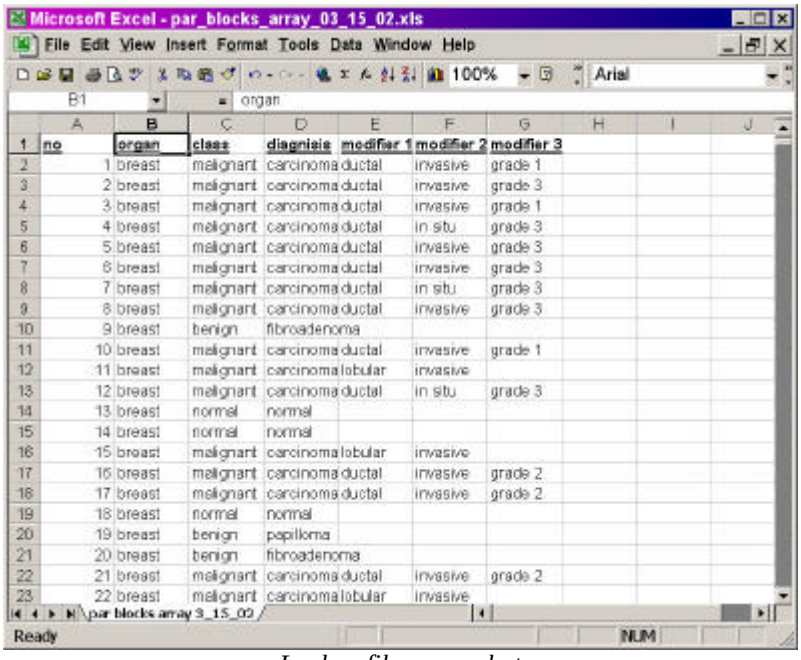
Note that this must also be identical to the rest of the worksheets in the workbook.



Scoring worksheet screenshot

- A scoring worksheet will contain the scoring information of an antibody stain on a slice of the TMA. The name of the worksheet should match the name given in cell L25; the value in cell L25 will be used as the name of the antibody in the output file. Within each sector, scoring data is provided. There is a limit of 253 scoring worksheets per workbook – this corresponds to the 256-column limitation in Microsoft Excel and the three columns required for the format of the pre-cluster file for the Cluster program.

Lookup File Format



	A	B	C	D	E	F	G	H	I	J
1	no	organ	class	diagnosis	modifier 1	modifier 2	modifier 3			
2	1	breast	malignant	carcinoma	ductal	invasive	grade 1			
3	2	breast	malignant	carcinoma	ductal	invasive	grade 3			
4	3	breast	malignant	carcinoma	ductal	invasive	grade 1			
5	4	breast	malignant	carcinoma	ductal	in situ	grade 3			
6	5	breast	malignant	carcinoma	ductal	invasive	grade 3			
7	6	breast	malignant	carcinoma	ductal	invasive	grade 3			
8	7	breast	malignant	carcinoma	ductal	in situ	grade 3			
9	8	breast	malignant	carcinoma	ductal	invasive	grade 3			
10	9	breast	benign	fibroadenoma						
11	10	breast	malignant	carcinoma	ductal	invasive	grade 1			
12	11	breast	malignant	carcinoma	lobular	invasive				
13	12	breast	malignant	carcinoma	ductal	in situ	grade 3			
14	13	breast	normal	normal						
15	14	breast	normal	normal						
16	15	breast	malignant	carcinoma	lobular	invasive				
17	16	breast	malignant	carcinoma	ductal	invasive	grade 2			
18	17	breast	malignant	carcinoma	ductal	invasive	grade 2			
19	18	breast	normal	normal						
20	19	breast	benign	papilloma						
21	20	breast	benign	fibroadenoma						
22	21	breast	malignant	carcinoma	ductal	invasive	grade 2			
23	22	breast	malignant	carcinoma	lobular	invasive				

Lookup file screenshot

- The lookup table should contain a header at Row 1.
- Unique case numbers should be located in column A.
- Columns B-F can contain additional information and will be included in the NAME column of the output file (which will correspond to the descriptors under Treeview). These descriptors can be diagnostic features (tumor grade, etc., as in this example) but can also contain clinical data such as follow-ups, etc.
- A pipe ("|") will appear between the information contained within each column, in the output file.
- Any additional information beyond Columns A-F will not be included in the output file.

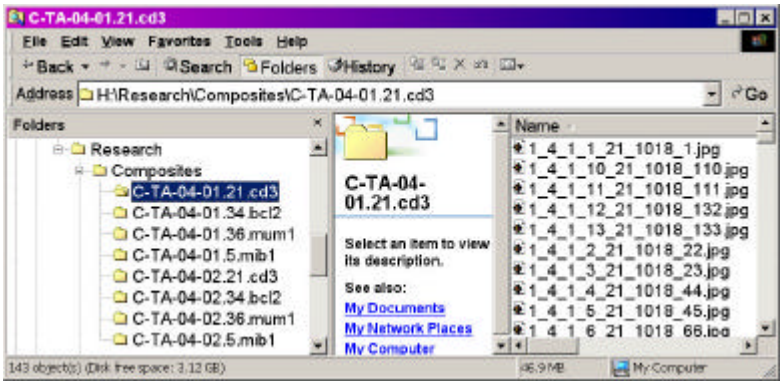
Note: the case identifier must begin with the integer 1, and there should be no skipping in the numbering. Thus, for any given case number n, it should appear on row n + 1.

Please refer to the screenshot for an example of the layout.

Bliss File Name Nomenclature

The TMA data management system described here was designed with the Bliss filename nomenclature in mind. If you created and curate your digital image collection with the Bliss microscope system, your files will already be in this format. If you have not, the following will describe how the nomenclature is structured so that you can adapt your system appropriately.

Images generated by the Bliss microscope system (Bacus Labs Inc.) are stored in the following directory structure:



All TMA directories should be contained in the directory “Composites”, as shown above. Each directory or folder represents a collection of images from a single sector of a given slice of the TMA, as shown above. The total number of folders in the selection will equal the product of the number of arrays, slices, and sectors in your TMA collection. In the example above, which also represents the files stored in the on-line demonstration database, there are 8 folders, representing an array composed of 4 slices and 2 sectors each.

The directory name is composed of the following:

C-TA-0x-0y.z.Ab-Name

An example of this would be the following:

C-TA-04-02.34.bcl2

where:

Variable	Description	Example
x	TMA Array Number	4
y	TMA Sector	2
z	TMA Slice Number	34
Ab-Name	Antibody Name	bcl2

So, in the above example, the folder/directory represents images collected from TMA number 4, sector 2, slice 34, stained with the antibody against bcl2.

Each of the image filenames in these directories consist of the following:

`s_a_c_r_sl_code_d.jpg`

An example of this would be the following:

`1_4_1_1_34_1012_1.jpg`

where:

Variable	Description	Example
s	Sector Number	1
a	Array Number	4
c	Column Number	1
r	Row Number	1
sl	Slice Number*	34
code	Bliss Code Number*	1012
d	Order Number	1

* Note that the raw scoring workbooks used to generate the TMA TreeView files must contain the necessary information in order for Stainfinder to recognize the filenames. This is why the TMA Deconvoluter will warn the user if “Bliss or slice information” is missing from their scoring worksheets.

So, in the above example, the folder/directory represents image collected from Sector 1, Array 4, Column 1, Row 1, Slice 34, Bliss Code designation 1012, and is the first picture scanned in the array.

A note about the Bliss code numbers: The BLISS code number is a numerical code for the name of the antibody used in staining and is required since the name of a JPG image file generated by the Bliss system consists of 7 numbers separated by underscores, but does not allow for letters.

Your digital image collection must follow this nomenclature in order for Stainfinder to work properly. If there are missing or incorrect information, Stainfinder may not be able to retrieve the image file or may retrieve incorrect image files. Please examine the on-line demonstration database and accompanying files from the Downloads section to see how a properly configured system should behave.

Note: The Bliss microscope system employs a serpentine scanning method in generating the order of pictures collected. The Order Numbers reflect this scanning method. An example of a serpentine scan is shown below, where the numbers represent the order in which the images are acquired.

Serpentine Scan

Sector *n*	Column										
Row	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1	1	2	3	4	5	6	7	8	9	10	11
2	22	21	20	19	18	17	16	15	14	13	12
3	23	24	25	26	27	28	29	30	31	32	33

etc....

Scanning beings at Row 1, Column 1 of Sector n and proceeds to the end of the row. It then moves one row down and proceeds to the left, until it reaches the left end of the row. It continues onto the third row and proceeds in like fashion until the end of the sector is reached.

Web Server Setup Instructions

Preface and Disclaimer

Note: Installation packages for web servers are available that run smoothly on the majority of computer configurations. However, setting up a web server is not a trivial task, and should not be handled by a novice without guidance. Please realize that you are exposing a portion of your computer to anybody who has access to the internet. Security is a complicated matter that people make careers out of, and security holes are periodically discovered in even the best web servers. Plan on checking regularly for updates and/or security fixes. With this in mind, these instructions are provided as a quick reference only. Non-technical users are advised to consult a systems administrator for assistance. We are not responsible for any damage, loss, nor compromise of information caused by an improperly configured server.

Instructions are provided here for installing the Apache web server on a Windows-based computer with the basic default settings. Apache is free, open source, and runs on multiple platforms – which helps explain why it is currently the most popular web server on the Internet today. While Apache can be run on Windows 95, 98, or ME, these are operating systems intended for the home user and not a production environment, so these operating systems are not officially supported by Apache. Furthermore, at the time of this writing, there may be a possible bug that affects Windows XP. Hence, it is highly suggested that Apache be run on either Windows NT or Windows 2000. The documentation for Apache (http://httpd.apache.org) is extensive and far more detailed that what you will find here.

Prepare your computer (server) for installation

Go to Windows Update (http://windowsupdate.microsoft.com) and install any security fixes. Security fixes are found in the Critical Updates section.

A server ideally should not be used to run any more applications outside of its main function. Ports likewise should be closed if they are not required for an application. The Apache web server, as most web servers do, is accessed through port 80. If this makes no sense to you, please seek further assistance with your setup.

Download the installation package from the Apache web site

There are currently two versions of Apache in development – Apache 1.3 and 2.0. We currently suggest using version 2.0 as it is better optimized for use on Windows machines. The easiest installation packages for the Windows platform can be found at the Apache Windows Distribution page:

http://www.apache.org/dist/httpd/binaries/win32/

You should read the release notes on this page carefully for any notices or known problems. These packages are provided in the MSI format. For our purposes, the package we will use is

called “apache_2.0.35-win32-x86-no_ssl.msi”. The name of the package will obviously change as new releases appear.

If you are using Windows 2000, XP, or ME, you simply need to open file you just downloaded. If you are using Windows 95/98/NT, try opening the package, but if it does not run, then you need to download the Windows Installer program in order to use the MSI packages (scroll to the bottom of the Apache Windows Distribution page for details and links to the Windows Installer).

Install Apache

Follow the instructions provided by the installer. When asked for the server name and domain name, enter the information as provided by your systems administrator. Alternatively. if you do not have this information, but you have a static IP address, you can enter the IP address instead. Apache is best run as a service, meaning that it runs whenever the computer is started. After installation, there are many useful shortcuts at Start Menu-> Programs->Apache HTTP Server, including links to start and stop the Apache Web Server. You may also see an icon representing the Apache Web Server in the system tray of your computer (lower right hand corner).

If installation was successful, a test page will appear when you type in the address of your computer into your web browser.

Configure Apache

Apache’s options are controlled by one main text file entitled httpd.conf. If you accepted the default settings, this file is located in the “C:/Program Files/Apache Group/Apache2/conf/” directory. There is also a shortcut to this file found at: Start Menu→Programs→Apache HTTP Server→Configure Apache Server→Edit the Apache httpd.conf Configuration File.

For our purposes, we are mainly concerned with the location of your files. Limiting access to these files is beyond the scope of this walkthrough, and the Apache documentation should be closely read if you need to do so.

Open the httpd.conf file in your favorite text editor. Lines that begin with the # sign are informational only. Scroll down through the file until you see a section that reads:

```
# DocumentRoot: The directory out of which you will serve your
# documents. By default, all requests are taken from this directory, but
# symbolic links and aliases may be used to point to other locations.
#
DocumentRoot “C:/Program Files/Apache Group/Apache2/htdocs”
```

The main pages of your web site should be located in the Document Root. You can move the Document Root by simply changing the location specified in the httpd.conf file.

If you continue to scroll down the httpd.conf file, you will see the below section:

```
# ScriptAlias: This controls which directories contain server scripts.
# ScriptAliases are essentially the same as Aliases, except that
# documents in the realname directory are treated as applications and
```

```
# run by the server when requested rather than as documents sent to the client.
# The same rules about trailing "/" apply to ScriptAlias directives as to
# Alias.
#
ScriptAlias /cgi-bin/ "C:/Program Files/Apache Group/Apache2/cgi-bin/"
```

Files in directories specified in this section are executed rather than simply displayed. You can specify multiple directories by simply adding another ScriptAlias line to the httpd.conf file. The Stainfinder Perl script should be placed in one of these directories.

Install Perl

Perl is a programming language commonly used by CGI programs. Because Stainfinder is written in this language, you must install Perl on your server in order to run Stainfinder successfully.

Perl for Windows install packages in MSI format are available from ActiveState (http://www.activestate.com). The current version at the time of this writing is ActivePerl 5.6.1 build 631. Running the installer file will walk you through the Perl installation. The default settings should be sufficient.

Install Stainfinder

Copy Stainfinder (the file Stainfinder.pl) to a directory recognized by your web server. Your web server will have a "root" directory whose location is defined by the configuration of your server, and anything placed outside the "root" directory will normally not be recognized by the server.

Configure Stainfinder to your installation

The sample Stainfinder program, written in Perl, was taken from a Unix system. The opening line of the file tells the program where Perl is located on the system, and must be changed to accommodate differences in the Windows system. On a Windows system, the first line should be changed from:

```
#!/usr/bin/perl
```

to:

```
#!c:/Perl/bin/Perl.exe
```

assuming that you chose the default settings during the installation of Perl. Please be certain to type in both the pound sign and the exclamation mark. Naturally, that line may need to be changed if you installed Perl in a different directory.

At this point, your web server is up and running with Perl. The Stainfinder program can also run, but its configuration has not yet been completed.

If you are a novice user who has been following our instructions up to this point, we strongly urge you to have your configuration checked by a systems administrator to ensure that there are no security issues with your configuration.

Please continue with the rest of the Stainfinder walkthrough to complete configuring Stainfinder with your system.

Revision History

Revision history of Deconvoluter program.

Current version is **1.06** (6-1-02).

*Earliest supported version is **1.06**. If you are encountering problems, ensure that your current version is this version or higher.*

Deconvoluter Version 1.06, 6/1/02

- **Bug Fix:** There was a bug that causes the Deconvoluter to navigate to the incorrect column if the number of columns in the sectors were different than that of the very first sector. Thanks to Brian Ring of **Applied Genomics** for pointing out this bug. This bug has now been fixed.

Deconvoluter Version 1.05a, 4/26/02

- **Bug Fix:** The Deconvoluter was not activating the "Output to cluster format" checkbox option when manual operation mode was selected. This has now been fixed.

Deconvoluter Version 1.05, 4/22/02

- **Bug Fix:** There was a subtle bug in how Deconvoluter handled different sector sizes in the master worksheet of the raw scoring workbooks. The TMA-Deconvoluter had been allocating memory incorrectly for TMAs which had the last sector be smaller than the other sectors, resulting in "Subscript out of range" errors. This has now been fixed.
- **New feature:** Deconvoluter now prompts you if you want to input your lookup file, after you generate your file list. If you select "yes", it will then ask for your lookup file and place the file name in the appropriate field, saving you the hassle of typing it in.

Deconvoluter Version 1.04, 4/7/02

- **New feature:** Deconvoluter now checks your lookup file to determine whether the list of unique case identifiers is correct. This will alert the user to any offset in the case numbering.
- **Modification:** Minor changes made to the "Control" worksheet (main screen) of the Deconvoluter.

Deconvoluter Version 1.03, 3/15/02

- **Bug Fix:** Deconvoluter had always assumed that the dimensions of all sectors would be equal to or less than the dimensions of Sector 1. As a result, cases existing outside the dimensions of Sector 1 in the other sectors were being omitted from the output file. This has now been fixed, and Deconvoluter should support TMA layouts where the dimentions of each sector is independent of any of the other sectors.
- **Modification:** Added an option to turn off warning messages for missing Bliss code/slice information. This is useful for people who haven't yet or don't wish to establish operation with Stainfinder at this time.

Deconvoluter Version 1.02, 3/5/02

- **Modification:** Made the "Set Working Dir" more informative for the user.

Deconvoluter Version 1.01, 2/8/02

- **New feature:** enhanced file list generation. Now, instead of prompting the user whether any of all the XLS files found in the current working directory are to be added to the file list, the user gets an Open File dialog box, whereupon selecting the desired files and clicking "Open" will result in the addition of those files into the file list.
- **New feature:** "Set Working Dir" button. It sets the working directory to be the directory containing the Deconvoluter.

Deconvoluter Version 1.00, 5/28/01

- Initial release version.

Frequently Asked Questions (FAQ)

Q: Why do the case numbers appear offset by one in the output file?

A: Your lookup file may have an offset in your case numbers. The TMA-Deconvoluter expects that for given case number *n*, the row on which it is found is *n* + *1*. For example, if the case identifier is "134", TMA-Deconvoluter expects to find it on row 135. TMA-Deconvoluter also expects the first case identifier to be "1", and versions 1.04 or later will alert you to any errors in your lookup file format.

Q: Why do I get the error: "Runtime error 1004: Cannot rename a sheet to the same name as another sheet, a reference object library, or a workbook referenced by Visual Basic"?

A: You didn't properly reset the TMA-Deconvoluter after your previous use of Deconvoluter. Click "End" to dismiss the dialog box, then go back and reset the Deconvoluter. To do this, click on the "Reset" button in the Control worksheet. The TMA-Deconvoluter needs to be reset after each batch is processed so that it can clean up after itself and prevent errors like these from occurring.

Q: Are there plans for a Macintosh version of the TMA Deconvoluter or any of the other components of the TMA data management system?

A: The short answer: Yes, on a limited basis.

The long answer: The TMA data management system has been optimized for Windows-based systems, primarily because of the availability of Active X controls on the Windows platform and because the **TreeView and Cluster** programs are currently available only for Windows operating systems. Because of ongoing requests, a Macintosh-compatible version of the TMA Deconvoluter is currently under development and is meant to only serve as pre-processing of data for users who will eventually view the data on a Windows-based system. This Macintosh-compatible version will have a more limited user interface and features, and it will not be updated as frequently due to the author's limited accessibility to a Macintosh computer. Java-based programs that possess the capabilities of TreeView and Cluster are currently under development by the Stanford Genome Databases Group.

Update (6-26-02): The Java version of TreeView is now available **here**. Please note that this is a work in progress, so documentation is sparse, and not all features are implemented. You will need to install the Java Runtime Environment in order for the program to work. Please refer to the program website (at the link above) for further details, and please direct questions about this program to the program's author, **Alok Saldanha**.