

# Table of Contents

1. [General Information](#)
2. [TIGR Software Overview](#)
3. [Spotfinder Introduction](#)
4. [Procedure Description](#)
5. [Quality Controls](#)
6. [Program settings](#)
7. [Database Operation](#)
8. [Interface Description](#)
  - [8.1 MenuBar](#)
    - [8.1.1 Image Menu](#)
    - [8.1.2 Grid Menu](#)
    - [8.1.3 Data menu](#)
    - [8.1.4 Settings Menu](#)
    - [8.1.5 Help Menu](#)
  - [8.2 Status Bar](#)
  - [8.3 Left Pane](#)
    - [8.3.1 Thumbnail View](#)
    - [8.3.2 Gridding and Processing Dialog](#)
    - [8.3.3 Post-Processing Dialog](#)
    - [8.3.4 Cell Editor Dialog](#)
  - [8.4 General Page](#)
  - [8.5 Overlay Page](#)
  - [8.6 Analysis Page](#)
  - [8.7 RI plot Page](#)
  - [8.8 Data Page](#)
  - [8.9 QC view Page](#)
9. [MEV file](#)
10. [Annotation file](#)
11. [References](#)

# 1. General Information

## 1.1. Obtaining TIGR Spotfinder

[www.tigr.org/software/tm4/](http://www.tigr.org/software/tm4/)

### Maintainer / Contact Information

TIGR Spotfinder Team - [spotfinder@tigr.org](mailto:spotfinder@tigr.org)

### Platforms:

- TIGR Spotfinder was tested on MS Windows 98, NT, 2000, XP. While we did not have a chance to test it on Windows 95 and ME it should work on those OS also.
- Linux
- Mac OSX

### System Requirements:

Recommended RAM 512 MB

## 1.2. Referencing TIGR Spotfinder

Users of this program should cite:

Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques*. 2003 Feb;34(2):374-8.  
[http://www.tigr.org/software/tm4/menu/TM4\\_Biotechniques\\_2003.pdf](http://www.tigr.org/software/tm4/menu/TM4_Biotechniques_2003.pdf)

## 1.3. A note on non-Windows operating systems

MacOSX users can emulate the mouse 'right-click' by using Ctrl- click.

## 2. TIGR Software Overview

TIGR Spotfinder is one member of a suite of microarray data management and analysis applications developed at The Institute for Genomic Research (TIGR). Within the suite, known as TM4, there are four programs: MADAM, Spotfinder, MIDAS and MeV. Together, they provide functions for managing microarray experimental conditions and data, converting scanned slide images into numerical data, normalizing the data and finally analyzing that normalized data. These tools are all OSI certified (see section 12) open-source and are freely available through the TIGR website, [www.tigr.org/software/tm4](http://www.tigr.org/software/tm4). The Microarray Data Manager (MADAM) is a data management tool used to upload, download, and display a plethora of microarray data to and in a database management system (MySQL). An interface to MySQL, Madam allows scientists and researchers to electronically record, capture, and administrate annotated gene expression and experiment data to be shared with and ultimately used by others within the scientific community. TIGR Spotfinder is image-processing software created for analysis of the image files generated in microarray expression studies. TIGR Spotfinder uses a fast and reproducible algorithm to identify the spots in the array and provide quantification of expression levels. TIGR Microarray Data Analysis System (MIDAS) is an application that allows the user to perform normalization and data analysis by applying statistical means and trim the raw experimental data, and create output for MeV. TIGR MultiExperiment Viewer (MeV) is an application that allows the viewing of processed microarray slide representations and the identification of genes and expression patterns of interest. Slides can be viewed one at a time in detail or in groups for comparison purposes. A variety of normalization algorithms and clustering analyses allow the user flexibility in creating meaningful views of the expression data.

# Introduction

Spotfinder is an image processing program created at The Institute for Genomic Research (TIGR) for analysis of the image files generated in microarray expression studies. Spotfinder uses a fast and reproducible algorithm to identify the spots on array and provide quantitation of expression levels.

The Spotfinder 3.0 has multichannel functionality, i.e. it can analyze microarray slides labeled by more than two dyes. The hardcoded maximum channel number is four but can be changed by request. As input Spotfinder expects up to four separate 16 bit or 8 bit gray scale TIFF image files, generated by a confocal laser scanner. Each file is scanned for one fluorescence dye labeling the query or reference samples used in most microarray experiments. Typically, the spots on array are divided into contiguous sections or subarrays or grids. Spotfinder builds set of grids for whole slide in user assisted mode and then process whole slide in one pass. The user is required to enter a number of parameters to build the grid: the spot spacing, number of printing pins in vertical and horizontal dimensions, distance between them, number of columns and rows in each grid. User needs to modify and align the left most top grid in array manually and then propagate alignment to the rest of the grids automatically. During the process pass the program automatically searches each grid element to identify the spots, measures the signal and local background. Output data can be exported to tab-delimited MEV file for whole slide. The MEV file can be passed to other TM4 programs such as TIGR MultiExperiment Viewer (MeV), or MIDAS.

Program limitations.

The use of Spotfinder is limited to the microarray image types and problems we identified and targeted during development. The software requires multiple images generated by scanning a single slide in both the reference(s) and query(s) channels using identical definitions of the scan area. The sensitivity and gain settings can be different for the all channels. Spotfinder uses data from all channels (all loaded images) to define spot boundaries. Non-overlapping images or those of different sizes will cause the program to fail.

# Procedure description

## Quick Start

This section describes in brief the main steps of microarray image processing using Spotfinder.

1. Check program settings. On **Menu Bar** go to **Settings->General Settings**. Make sure that the **Channel Number** is 2 and **Scale Factor** is 2. For the first run uncheck the check boxes **Use RI plot** and **Use Full Table for MEV data**: we are not sure you computer is advanced enough to digest the full data set. [More details...](#)
  2. Load Images. Select both images at once by holding keyboard Ctrl- key to click on one file then on another one.
  3. Load grid from file. Alternatively compose grid from scratch. [More details...](#)
  4. You should be switched to **Analysis Page**. If not go to it. Move manually whole set of grids to overlay with the image. Place the top left grid in correct position using [All Grids Menu](#), ignore at this moment other grids. Place top left cell in correct position over top left spot on first (top left) grid. Expand or shrink ALL grids in both dimensions. If needed, change appropriately the cell size for ALL grids. [More details ...](#)
  5. Align grid automatically or manually. To make automatic alignment enter mask size on **Gridding and Processing Dialog** and run **Adjust All** .[More details...](#)  
To align grid manually: on **Analysis Page** use **Move All** from [All Grids Menu](#) at first to move all grid set in proper location.  
If it is needed use **Move** for every grid separately. [More details...](#)  
Store aligned grid template in SFG file.
  6. Set Processing parameters in [Gridding and Processing](#) dialog: set **Segmentation Algorithm** to *Otsu*. Leave default settings for **Minimum** and **Maximum** spot size on this dialog. Click **Process All** button. [More details...](#)
  7. Check the alignment on [QC View Page](#). Compare the the size of green area on **QC View**. The grid with green area larger than that of the rest grids on array indicates a possible grid misalignment. Go to **Analysis Page** to check this and realign if it is necessary.
  8. Create MEV file. On **Menu Bar** go to **Data**, select **Save Data to MEV file**.
- 

## Procedure steps in details

To adjust grid size:

1. Move the whole grid set to align top left spot with top left cell of first grid. To do it click right mouse button when mouse pointer does not overlay with any grid. **All Grids** menu will be activated. Choose **Move All** from the **All Grids** menu and move mouse slightly or use keyboard arrows keys to move all grids together to necessary position. When it is done terminate **Move** mode by pressing **End** keys on keyboard or click left mouse button.
2. If number of rows and columns is set correctly but grids do not fit with the image spots you might need expand or shrink grid. Invoke **All Grids** menu and choose **Expand All** command. By using keyboard arrows keys user can expand or shrink all grids simultaneously in any dimension. Both expansion and shrinking are performed while keeping left and top edges of each grid fixed. Only the right and bottom edges of each grid are moved.
3. If it is necessary adjust cell size using command **Cell Size All** from [All Grids Menu](#). Use keyboard arrow keys to increase or decrease cell size in vertical and horizontal dimension. One arrow key click corresponds to cell size increase/ decrease by one pixel. Terminate **Cell Size All** mode when finish by clicking **End** on keyboard or left mouse button. When increase cell size try not to touch the spots in neighboring cells by the adjusting cell box. Unless the adjusting cell box does not reach the adjacent cells it is safe and desirable to increase the area around spot for local background calculation.
4. User can rotate grids if needed. Activate this command mode **Rotate All** from [All Grids Menu](#) and use Up and Down arrows keys on keyboard to rotate all grids together. Rotation of each grid is performed around its top left corner. Again, at this stage use only top leftmost grid as indicator of alignment. The rest of the grids are expected to have the same angular offset due to the parallel orientation of pins during the slide printing.
5. User can repeat steps 1-4 in any sequence any times to reach better grid size adjustment. Make sure at this point to use only first grid (top left) as indicator of proper grid size .

To adjust grid locations on image:

1. By this point first grid (top left) has to be aligned perfectly in size, orientation and location. The rest of the grids should have proper size and orientation, but probably not location. They may have some misalignment in vertical and horizontal position because the pins in reality are not exactly parallel to each other in the robot printing head. When grid size and orientation are set user can start to adjust location for all grids. To make automatic alignment enter mask size on [Gridding and Processing](#) dialog and push **Adjust All** button or choose automatic command **Adjust All** from **All Grids** menu. The mask size should be larger than the typical spot size on array but smaller than the cell size or spacing. It applies adjustment algorithm to each grid one by one and find the best location for each grid on image. This procedure can be applied as many times as it needed.
2. User also can apply this automatic procedure individually to each grid using **Grid Menu** command **Adjust**. To do it user needs to move mouse pointer on top of the selected grid, this grid becomes colored in magenta. Then pressing right mouse button user invokes **Grid Menu** and selects **Adjust** from the menu list. This procedure can be repeated as many times as needed.
3. If automatic **Adjust** procedure does not provide satisfactory alignment user can manually move

selected grid. Select a grid by placing mouse pointer on top of it and activate **Move** mode from **Grid Menu** by right clicking mouse. By moving mouse or applying the keyboard arrow keys user can move selected grid to proper location. When done terminate **Move** mode by pressing keyboard **End** or left mouse button.

To process spots in cells:

1. Before to start of grid processing user must enter experimentally defined parameters: estimates of **Spot Size**, which is the typical spot diameter in pixels. Depending on selected algorithm for spotfinding user must set one or two values for **Spot Size** . Currently three spotfinding methods are implemented in the program.

*Histogram thresholding* method was historically the first implemented in Spotfinder. It requires only one value to enter: the estimate of **Spot Size** in pixels. This value can be estimated by zooming image in **Analysis page** and placing mouse pointer on left and right edges of typical spot (alternately, top and bottom edges of spot). By reading the X (or Y) coordinates in image pixels on **Status Bar** and calculating difference the spot size estimate can be done. It is safer to use slightly overestimated then underestimated value for **Spot Size**. For this reason if it exists a noticable difference in spot sizes across the image it is better to use largest value for **Spot size**. *Otsu thresholding* method is default and requires user to set two values for spot size: minimum and maximum. These values specify the range, which program uses to search for spot boundaries. Default settings for *Otsu* algorithm cover almost all possible spot sizes observable in real microarray images. However, we recommend to narrow it to the expected real spot size range. For *Histogram thresholding* set **Spot Size** in one entry boxes on [Gridding and Processing](#) dialog. The rule of thumb: the entry should be slightly larger than the biggest spots on array.

*Fixed Circle* segmentaion method is the most simple of all. It takes as a spot the area inside the circle centered in the cell. The only parameter user must set is a circle diameter in pixels.

2. Press the **Process All** button, or go to **Process All** in **All Grids Menu**. Spotfinder runs through all grids one by one. Each cell in the grid is analyzed to detect spot boundaries using the specified **Spot Size** and data from both channels A and B. After removing the saturated pixels from analysis, intensities for signal and background are calculated. Spotfinder holds internally the all raw data after processing. The next step is post processing data maipulation, which is performed automatically in a row with processing. In post-processing steps background subtraction is performed and background corrected integrated inetnsities or medians are reported by default. Also the default QC filter procedure is applied and spots flags are set. [More about QC filter and quality controls...](#) User can reset default setting for post-processing any time and update reported data set on **RI Plot Page** and **Data Page**. For more details of post-processing data manipulation go to [Post-Processing Dialog](#)
3. It is possible to process each grid individually by starting **Process** from **Grid Menu** of selected grid.
4. Check box **Contour** to visualize the spots boundaries colored in accordance with flags set by **QC Filter**. Good spots are colored in magenta. Spot's contour is drawn in green if QC filter defined spot as bad at least in one channel.





# Quality controls

Spotfinder uses three types of quality control parameters to evaluate the quality of spot in each channel and in total goodness. The output MEV file reports flags for each channel, QC score total and for every channels, and P values for each channel.

## Spotfinder Flags

There are two types of flags. The flags A, B, and C are assigned to the good spots and differentiate good spots by there size.

- A - spot area is larger than 50 pixels;
- B - spot area is in the range between 30 and 50 pixels;
- C - spot area is smaller than 30 pixels;

The flags S, X, Y, and Z are introduced to the rest of the spots. The next is there meaning:

- S - spot is fully or partially saturated. This is a warning flag , user should evaluate spot and make a decision about it.
- U - spot is set by user for not to be processed, i.e. ignored.
- X - bad spot; this flag is assigned by QC filter. QC filter dependent flag, i.e. when QC filter is off there are no spots with X flag.
- Y - bad spot. Spot for which background is higher then signal. QC filter independent flag. This flag presents even when QC filter is off.
- Z -spot is not detected. QC filter independent flag.

## Spotfinder QC score

QC score is calculated depending on signal-to-noise ratio for every channel in spot and shape quality for whole spot. Spot shape QC score is defined as ratio of spot area to spot perimeter scaled into the range between 0 and 1. It is the same for both channels in the spot, because the shape of spot is identical in all channels for given spot. Signal-to-noise QC score is calculated as percentage of pixels in a spot with values higher than 2\*median(local BKG). The QC score in each channel is calculated as geometric mean of shape and Sgnal-to-noise QC scores for this channel. The total spot QC score is taken as a mean of QC scores in all channels.

## P values

P-values are calculated for each channel as result of t-test comparing the spot pixel set and surrounding

background pixel set. We apply simple procedure based on mean and standard deviation analysis for both spot and background pixel sets. The lower P-values the better spot. One can say that spot with p-value larger than 0.3-0.35 is not good.

# Program Settings

The program settings are stored on local computer when the program is terminated correctly, i.e. without crash. The saved settings are read and set during the next program launch. Currently the settings parameters are:

- Scale Factor for Image Display
- Channel Number
- MEV file preferred interactive RI plot
- Usage of full or reduced table for MEV data on DATA page

**Scale Factor** changes only visualization of tiff images on the screen. For faster screen drawing the typical microarray 16 bit image is converted to 8 bit grey scale image. In addition the image can be compressed in each dimension by factor 2 or 4. This trick significantly decreases the size of the image and speed up drawing. This transformation does not change the original input 16 bit image used for analysis.. Default value for **Scale Factor** is 2.

**Channel number** equals to the number of dyes used in the microarray experiment. Default is 2.

**MEV file preferred fields** can be switched between *integrated intensities* and *medians* for columns 2 and 3. The alternative fields are still reported in MEV files in the columns located far to the right.

User may choose to use or not the interactive **RI plot** that allows to view and select individual spots on RI plot. This option normally takes a significant amount of RAM which is proportional to the number of spots on array.

Another way to speed up operation and optimize memory usage is to set reduced table view in *Data page*. *Data page* is a spreadsheet representation of MEV file. Full size MEV data table has 30 columns for two dye samples and takes a lot of RAM resources. The reduced size table shows user only first eleven columns. It takes less memory and also speeds up the data exchange between internal data structure and the table. This settings has no effect on output MEV file.

Go to [Settings dialog](#) to change them.

## Database Operation

**Spotfinder** can upload the results of image analysis directly to the TIGR compliant microarray database. However, **user must submit experiment info using MADAM all the way through *Scan Page* before to start any data submission from Spotfinder directly or from MEV files.** This is necessary because **Spotfinder** uses the loaded tiff image file names for database submission. The loaded files have to have the same names as files submitted on *Scan Page* of **MADAM**.

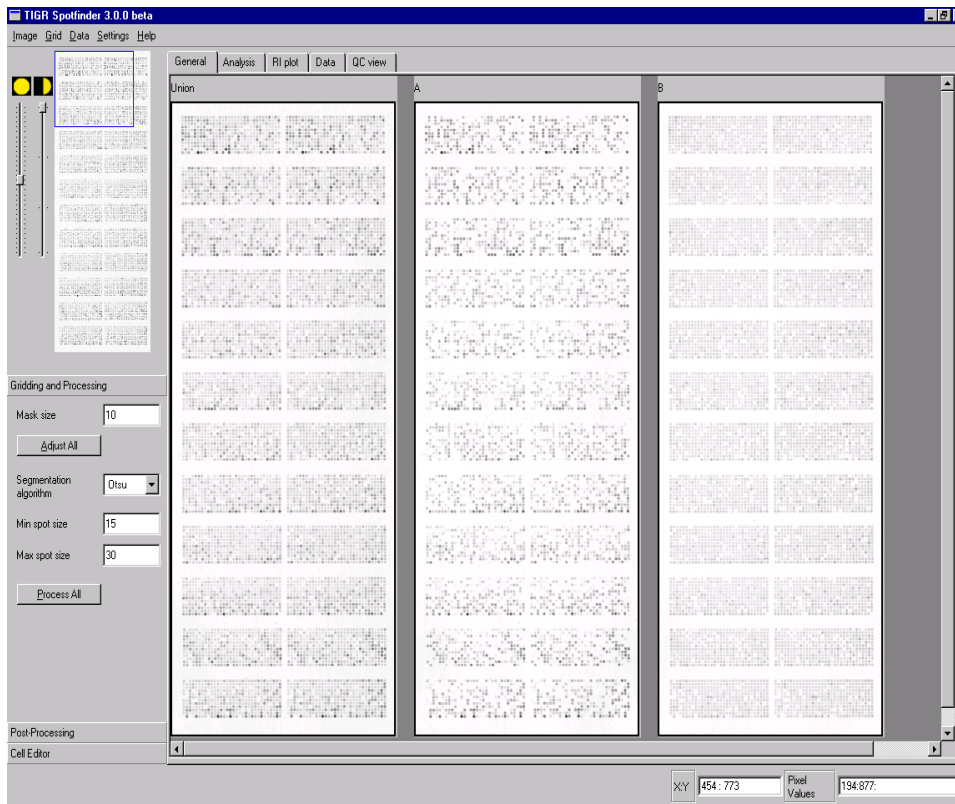
Actual data upload from **Spotfinder** to database is performed with the help of a program called a **SpotfinderDBclient**. **SpotfinderDBclient** is written in java and can be launched locally or from any location on network, where it is accessible from local machine. By default **SpotfinderDBclient** is installed in the same directory as **Spotfinder** itself. **SpotfinderDBclient** uses *jdbc* to connect to database and requires java runtime environment (JRE) to be installed. Distribution comes with two jdbc driver files: one for SYBASE, and another one for MYSQL database. Before to start data submission to microarray database user should launch **Database Settings Dialog** from **Settings** menu to make appropriate settings. User would need to select database type (SYBASE or MYSQL), server name, database name, user name, and password. Go to [Database Settings Dialog](#) for details.

To start database submission user should go to **Data** on menu bar and select command **Upload results to DB**. This function creates temporary mev file in the same folder where **SpotfinderDBclient** is located. The name of temporary mev file is based on the tiff file name currently loaded in channel A. Besides this the function creates DBlog.txt file, if it does not exist yet, or update it, if it is located in the same folder as **SpotfinderDBclient**. DBlog file is tab delimited text file and is used by the **SpotfinderDbClient**. The DBlog file keeps records of all submitted jobs: active, pending, and completed. User may rename or move DBlog file to another location when it becomes too big in size. DBlog file has five columns: analysis ID, temp mev file name, tiff filename in channel A, tiff filename in channel B, and comments. When user selects **Upload results to DB** one line is added to DBlog file that sets analysis ID to zero and stores all needed file names. The comment line default string is set to "not processed". After that **Spotfinder** launches **SpotfinderDBclient**, which in its turn reads DBlog file and processes data line by line. If during the data submission to database user send another data set to DB from Spotfinder, new line of records is added to DBlog file. The new record will be processed after the previous one is completed.

When submission is completed the active record is changed. Actual analysis ID is set in the first column and the information about the number of rows submitted to the expression table on selected database is stored in comments section. If for some reason the data submission to database failed the analysis ID remains set to zero. That means that next time when user starts DB submission **SpotfinderDBclient** will read unprocessed line from DBlog file and try to submit it to database. In this way DBlog file can work as a dispatcher for batch database submission.

# Interface Description

Spotfinder has intuitive graphical user interface (GUI), which is implemented as a single window maximized at the program start up. If user needs to change screen resolution we recommend to restart program to let it fit the full screen.



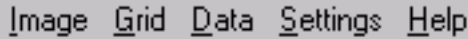
[Click to enlarge](#)

The main program window is a container holding all controls, tab pages, and windows that the user needs for microarray image analysis. By default the program window is maximized to full screen at the start up time. The top and bottom of program interface are the **Menu Bar** and **Status Bar** correspondingly. The leftmost part of the screen is a vertical Pane called **Left Pane**. The rest of the screen area is occupied by the **Tab pages**, which are the containers for the major interactive views. The image analysis is handled through the set of **Tab pages** that are described in details in the further sections.

The main program window is set to fill the full screen area for better image appearance. However in some cases not all controls are visible especially at the bottom of the program interface. In situation like this user can start **Spotfinder** from command line with two parameters indicating program window width and height, for instance: `Spotfinder 1000 800`, to start Spotfinder with 1000 pixels wide and 800 pixels high window.

User might create batch file (BAT) on Windows and script file on Linux/MacOSX with this only command and use it in future for quick program start up.

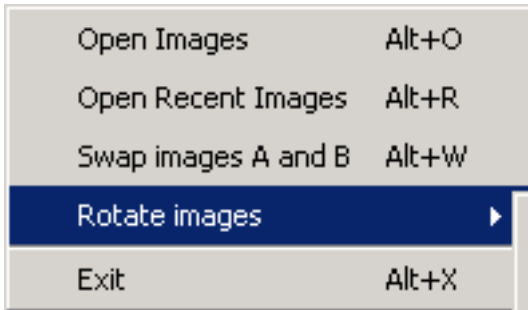
# Menu Bar

A screenshot of a menu bar with five items: 'Image', 'Grid', 'Data', 'Settings', and 'Help'. Each item has a small underlined letter as a mnemonic: 'I' for Image, 'G' for Grid, 'D' for Data, 'S' for Settings, and 'H' for Help.

The Menu Bar is a home for a number of Menus that help user to load in Spotfinder and save images, grids, and data. The settings and help menus are hosted on the Menu bar also.

---

## Image Menu



The Image Menu can be used only for the loading images in the program. Menu command **Load Images** launches the input image files section dialog. Depending on the number of channels selected in settings user needs to select an appropriate number of image files for input in the File Open Dialog. Use CTRL key to make this selection from the list of files. The order in which images are placed in channels A, B and so on depends not on the selection order but on file names. Image name filter sorts the input files alphabetically according to theirs filenames. This mechanism guaranties that the images having suffices Cy3 and Cy5 in filenames are loaded in channels A and B in the same order. This also is true for suffices 532 and 635 used by **Axon** GenePix scanning software to name files.

Currently **Spotfinder** supports multi-tiff images files; therefore the multi-tiff file created by **Axon** scanner can be loaded with just one click. The order of files in **Axon** multi-tiff image file is defined at the moment of file creation and can be changed by applying **Swap Images** procedure (see below). In addition to the regular tiff files Spotfinder can load gel files since gel type is a kind of tiff image format.

Menu command **Open Recent Images** loads the most recent set of images analyzed by Spotfinder on user desktop.

Menu command **Swap Images A and B** allows to swap images in channels A and B if default image order is not appropriate.

Menu command **Rotate Images** is useful when original image was scanned not in traditional way from slide top to the bottom. We consider the traditional image orientation where the image of size the standard microscope slide is oriented vertically such that the image width is smaller than the image height. This command rotates all loaded images by 90 degrees in clockwise or counterclockwise, depending on user choice. Remember that the image rotation results in generation of output MEV file with orientation different from originally loaded image . It may require the remapping of original annotation if such MEV file is used in further analysis.

---

## Grid Menu

<u>C</u> ompose Grid	Alt+C
<u>L</u> oad Grid from file	Alt+L
<u>S</u> ave Grid in file	Alt+S
<u>D</u> elete Grid	Alt+D
I:/TSF3_beta_win/Data/grid5alignedA.gr	
I:/TSF3_beta_win/Data/grid5alignedA.gr	
I:/TSF3_beta_win/Data/grid5alignedA.gr	
I:/TSF3_beta_win/Data/grid5alignedA.gr	
I:/TSF3_beta_win/Data/grid5alignedA.gr	

The Grid Menu provides a set of commands to create, delete, save, and load grids used in program for analysis.

The **Compose Grid** menu command launches a dialog for creating grid from scratch.

See [ComposeGrid](#) page for more details.

The **Load Grid from file** menu command brings up file dialog to select platform independent grid file created and saved previously by **Spotfinder**.

Starting from version 3.1.1 user may load grid from standard GAL file. Note that only GAL file header is used by Spotfinder to define the grid spatial parameters.

The **Save Grid to file** command allows user to save the currently active grid in platform independent grid file.

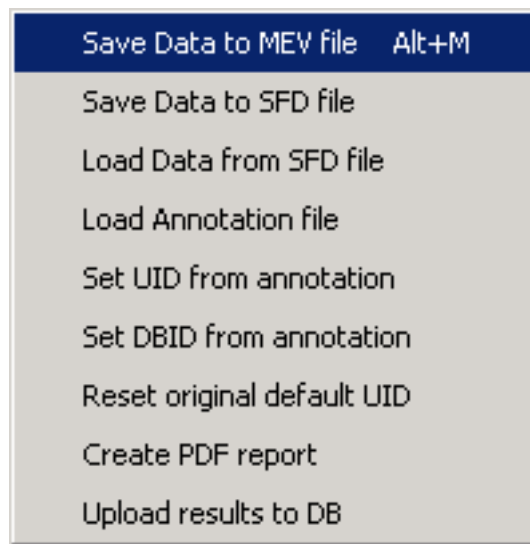
The **Delete grid** command deletes the active grid.

Since introduction of Undo/Redo support for grids in version 3.1.0 the active grid should be understood as currently visible on **Analysis** page.

The list of five recent grid files is shown at the bottom of the **Grid menu**. User can pick any of file in this list for quick loading. These file names are stored in program settings when program is closed and can be accessible in the next program session.

---

## Data Menu



The **Save Data to MEV file** command saves the current data set to MEV file. See [MEV file](#) for details.

The **Save Data to SFD file** command saves data in platform independent binary file. The SFD file holds not only information available in MEV file but all raw data and the information about spot contours.

The **Load Annotation file** command loads annotation data from TIGR type annotation file (\*.ANN) or **Axon** annotation file (\*.GAL). Remember, if the image was rotated by using operation from image menu the annotation file has to be remapped in accordance with the image geometrical transformation.

The **Set UID from annotation** command replaces default UIDs in first column of the MEV data table on **Data Page** with UIDs from annotation file.

The **Set DBID from annotation** command adds DBID column at the end of MEV data table on the **Data Page** and sets DBIDs from annotation file.

The **Reset original default UIDs** command restores original IDs on the first columns of the MEV Data table. The MEV file created by Spotfinder is the WYSIWYG type object for MEV table in **Data Page**.



The **Create PDF report** command writes PDF report for processed images. PDF report is three pages document designed currently for two colors array slides.

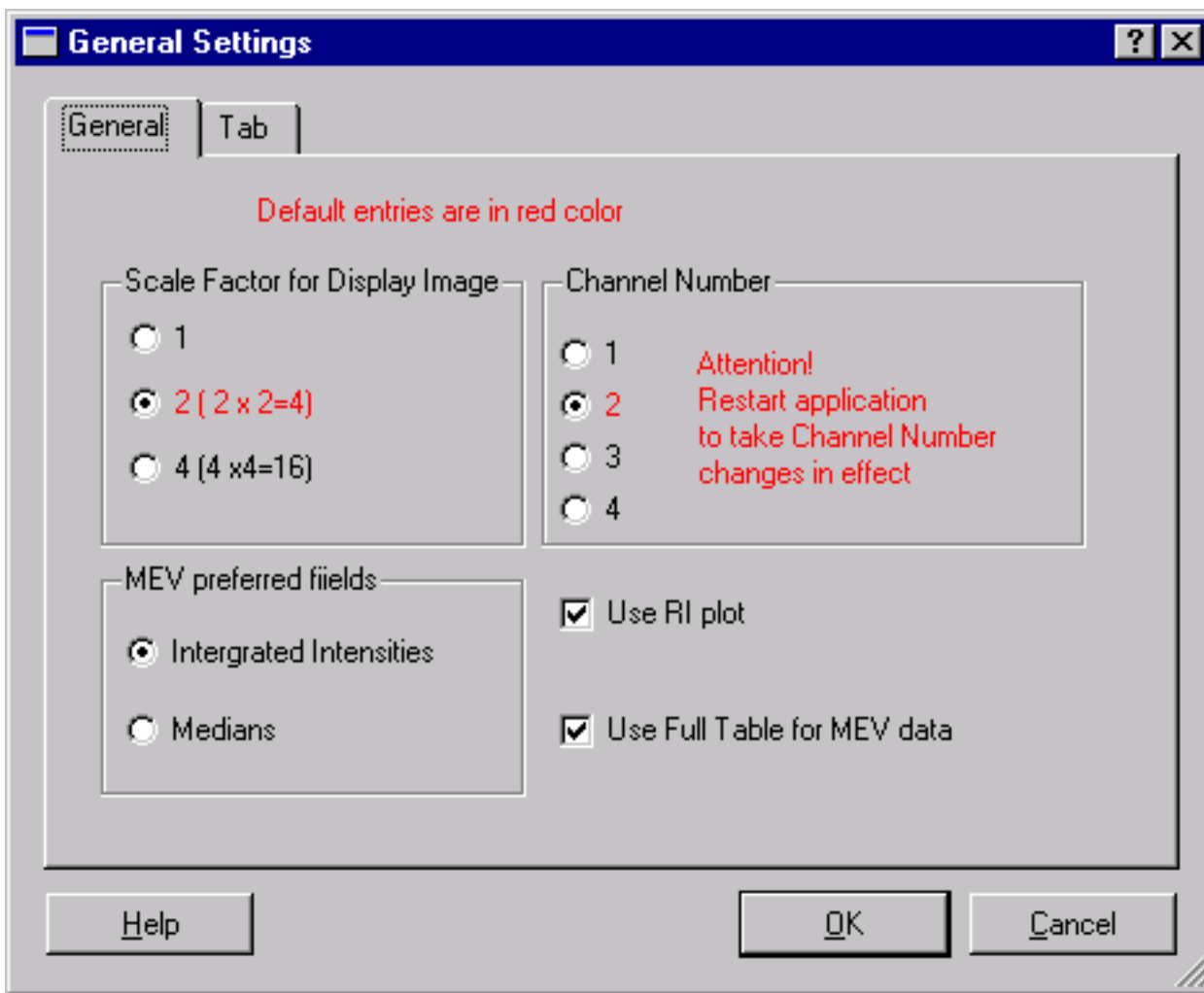
The **Upload data to DB** command creates temp MEV file for current data set in program buffer and starts **SpotfinderDBclient**, which in its turn launches the database upload session. The **SpotfinderDBclient** is written in java and can be installed not only locally but on any machine that can be reached through the local network from host machine. See [Database Operation](#) for details.

---

## Settings Menu



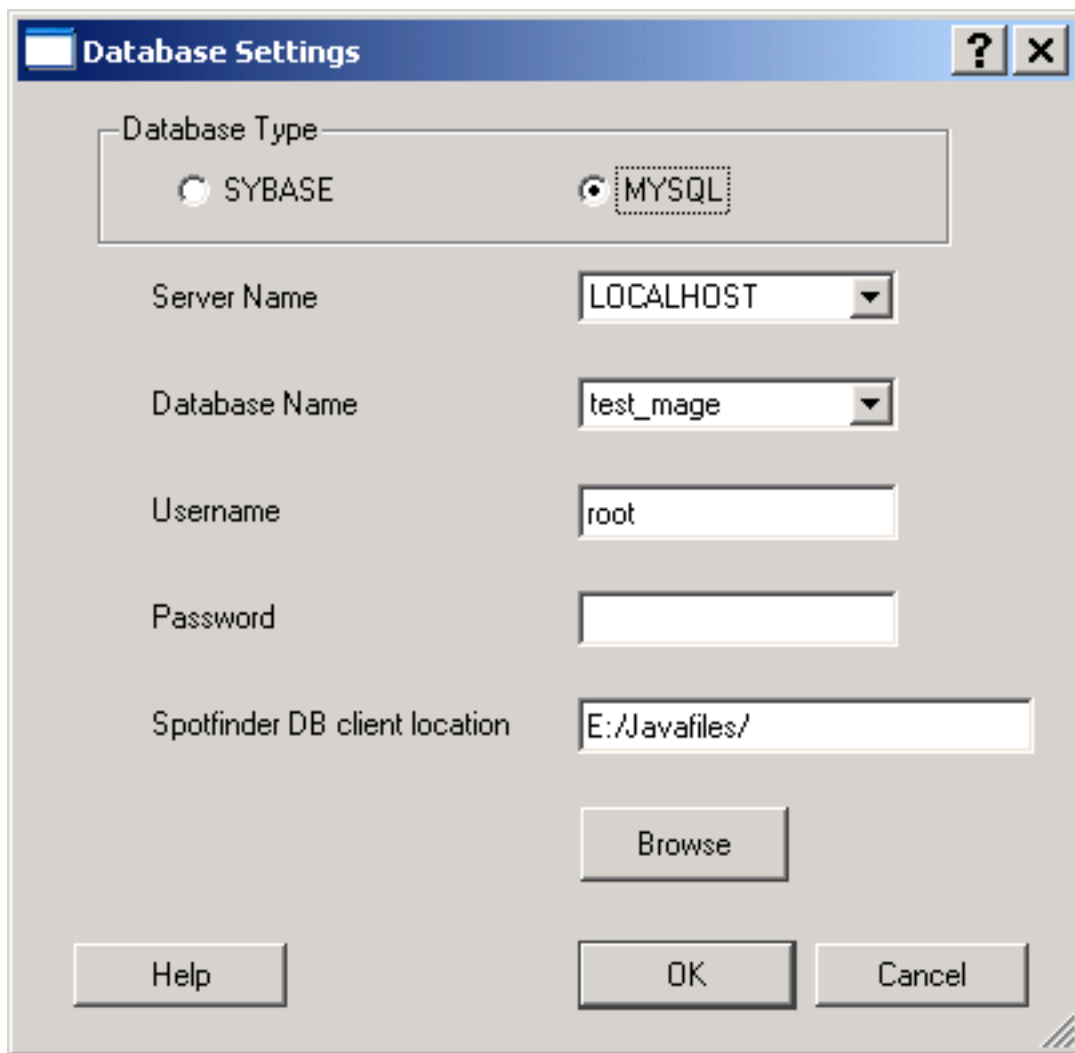
Currently **Settings Menu** has only two commands: **General Settings** and **Database Settings**. It can be extended in future.



The **General Settings** dialog was designed for making changes in program settings. Programs accepts all **Settings** from this dialog without restart except **Channels number**. The change in **Channel Number** requires user to close program and restart it. All other settings will be taken into effect during subsequent program operation. The program settings are stored locally when program is normally terminated. The settings are not saved if program is crashed, therefore it is better to close program and restart to save changes in settings.

---

## Database Settings Menu



The **Database Settings** dialog is used to set and change the parameters needed for upload of the data generated by **Spotfinder** to Database. See [Database operation](#) for details. All database settings besides password are stored locally when program is normally terminated. The settings are not saved if program is crashed. Password is stored in program buffer for use in the current **Spotfinder** session only.

Depending on the database type user should select SYBASE or MYSQL.

## SYBASE

TIGR users should use SYBASE since this is a TIGR home database. The server should be set to SYBTIGR, which hosts all TIGR production microarray databases. TIGR users may set SYBIL for work with one of the test databases in TIGR. As username and password TIRG database userID and password should be used. Make sure that *Spotfinder DB client location* is set correctly and point to the current location of **SpotfinderDBclient**. The default location is the same as **Spotfinder** executable file. However user might choose to move the DB client ( SpotfinderDBclient.jar file) along with DB drivers to another location on site network. We recommend to leave default setting for *Spotfinder DB client location*.

## MYSQL

MYSQL microarray database was designed for users outside TIGR and available for download from TM4 project web site [www.tm4.org](http://www.tm4.org). If MYSQL default installation is performed , which is local, the server name should be LOCALHOST and the default demo database name is *test\_mage*. The default user name and password are *root*. Make sure that *Spotfinder DB client location* is set correctly and point to the current location of DB client. The default location is the same as **Spotfinder** executable file. However user might choose to move the DB client ( SpotfinderDBclient.jar) along with DB drivers to another location including another machine accessible from user machine through the network.

---

## Help Menu



There are three commands in **Help Menu**.

**About** brings up a dialog with **Spotfinder** license information.

**Help** opens this interactive html help document (the one user reads now) in separate window. The keyboard key F1 does the same.

**What's This** command changes mouse cursor to a question mark. When user point this cursor to a some control or dialog on program GUI and click left mouse button the small window with the short explanation for that control appears. Some of these explanatory texts have additional hyperlink, which redirects user to the specific place in Spotfinder html help system.

## Status Bar

Images loaded	Grid Indices	4 : 1	X:Y	226 : 1574	Pixel Values	1:1:
---------------	--------------	-------	-----	------------	--------------	------

The **Status bar** at the very bottom of the main window serves for displaying textual information. The leftmost part of the **Status Bar** reports information about current process status, annotation data for selected spot on **Analysis** page, or loaded file names on **General** page.

The next is a list of possible usage of left segment of the **Status Bar**:

- Display filenames during the loading image (\*.tif ) and grid (\*.sfg ) files in the programs. See [Image Menu](#) for details.
- Display loaded filename when click on the corresponding image of the [General Page](#).
- Display annotation string when user select cell/spot on **Analysis Page** or **Data Page** if annotaion file is loaded in Spotfinder.

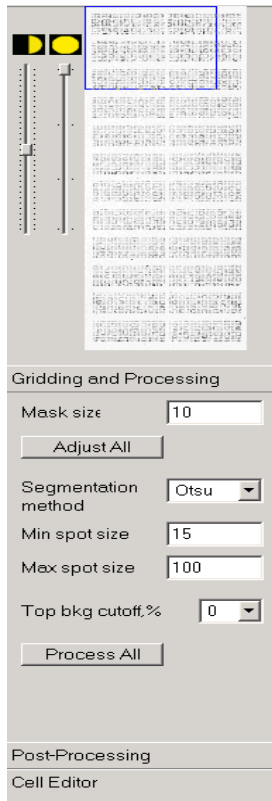
The rightmost part of the **Status Bar** has three text boxes, which display the current grid indices, the pixel coordinates and values for current mouse pointer position on image when **Analysis Page** is active. The coordinate box always reports two values: X an Y values separated by colons. The **Pixel Values** box reports the pixel value at (X,Y) position in accordance with the current channel number settings (number of dyes) in the program. The pixel position is shown in **X:Y** box. When **RI plot Page** is active the **X:Y** indicator displays the coordinates of RI plot in  $\log_2$  scale. The RI plot limits are X[0,32] , Y[-16,16].

The **Grid Indices** control displays the index of the grid, which is pointed to by the mouse cursor on the **Analysis** page.

---

# Left Pane

[click to enlarge](#)



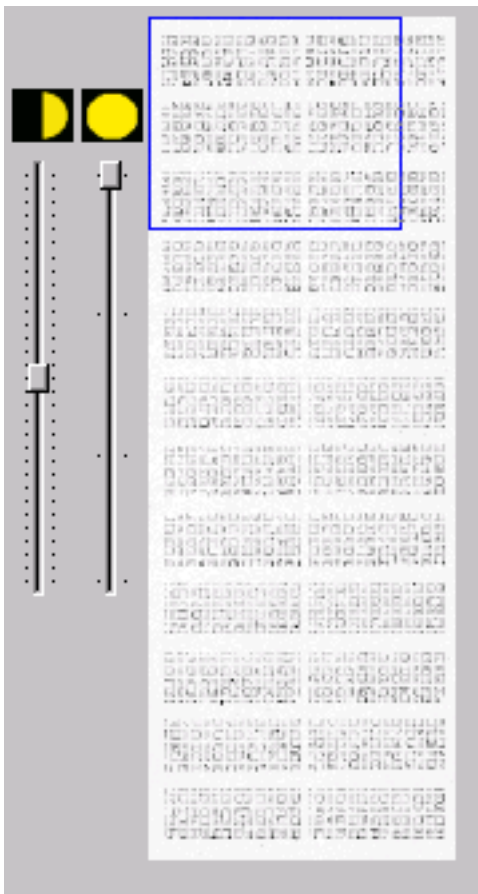
The **Left Pane** is a home for the dialogs and controls that are permanently visible to user.

The top part of **Left Pane** is occupied by [Thumbnail View](#) and the sliders for settings [Brightness](#) and [Contrast](#). The bottom part is a toolbox control that is a container for three dialogs: [Gridding and Processing Dialog](#), [Post-Processing Dialog](#), and [Cell Editor Dialog](#). Only one of these dialogs can be shown at once. User can switch from one dialog to another by clicking the horizontal tab with the name of dialog.

---

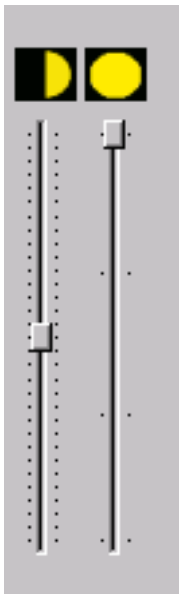
## Thumbnail View and Image Control Sliders



### Thumbnail View



The **Thumbnail View** is a small image view showing the **Union** image. In addition the blue overlapping rectangle shows user the location of current visible image area on [Analysis Page](#) and [QC View Page](#). The position and size of the blue rectangle are changed when user scrolls though the image or zoom it in or out on those pages.

## Brightness and Contrast slide controls



 Contrast and  Brightness slider bars help to adjust the appearance of the grayscale image. The initial settings are: contrast is in the middle of its range while brightness is set to zero (top position). These controls change only the Spotfinder display image views but do not effect the original image data.

## Tool Box Control

Tool Box control is a container for many dialogs. Only one of these dialogs can be shown at the time. User can switch from one dialog to another by clicking the horizontal tab with the name of selected dialog.

## Gridding and Processing Dialog

This dialog helps user to set parameters for automatic grid alignment, spot finding, and background cutoff.

### Auto Grid Adjustment Settings

**Mask Size** is the only parameter needed to be set for auto grid adjustment. The automatic adjustment procedure uses the mask to set grid in place. The mask value has to be set equal or slightly higher than the expected spot size. This parameter is measured in pixels. Default value is 10. The button **Adjust All** starts auto alignment for all grids (blocks).

The screenshot shows a software dialog box titled "Gridding and Processing". It features a tabbed interface with three visible tabs: "Gridding and Processing" (the active tab), "Post-Processing", and "Cell Editor". Under the active tab, there are several input fields and buttons. The "Mask size" field is set to 10. Below it is an "Adjust All" button. The "Segmentation method" is set to "Otsu" with a dropdown arrow. Below that, "Min spot size" is set to 15 and "Max spot size" is set to 100. The "Top bkg cutoff,%" field is set to 0 with a dropdown arrow. At the bottom of this section is a "Process All" button.

### Segmentation Method Settings

In image processing the identification of regions within an image is called segmentation. Hence, the segmentation in microarray image analysis simply means the spot detection: the search for the line (boundary) between spot and surrounding background. **Spotfinder** currently uses three segmentation methods. One is **Histogram thresholding** method, which requires one input parameter: an estimated spot size measured in pixels. Another method is **Otsu thresholding** method, which needs two parameters from user to entry: maximum and minimum spot size. The default maximum value is limited by to the cell size. The default minimum spot size equals to 3 pixels across the spot. The last segmentation method is **Fixed Circle** method. It takes as a spot the area inside a circle centered in the cell. The only parameter user can set for this method is the circle diameter in pixels. The button **Process all** launches the processing loop; segmentation is a key task of this processing.

### Top Background Cutoff Settings

Starting from version 3.1.1 user has an option to set a portion of top background pixels that are discarded from background set before to the local background estimate. It helps to prevent background



overestimation by avoiding high intensity pixels. Those can be bright pixels next to the spot contour. In addition this feature can be helpful in situation where there is a local spikes in background. The default settings is 5%. Any changes can be taken into effect only in processing step hence reprocessing is needed after changing **Top Background Cutoff** parameter.. Zero parameter value is equivalent to the functionality of Spotfinder versions prior 3.1.1.

## Post-Processing Dialog

This dialog is a container for the controls that can be used to set parameters for QC filter. These settings change the parameters that does not require reprocessing and for this reason is referred as post-processing dialog.

Group of radio buttons **QC Filter** can be used to turn on and off QC filter. Default value is on.

Radio button group **Keep Flagged Values** allows user to set the reported intensities of flagged bad spots to zero (default). As option user can keep bad spot values in output data set.

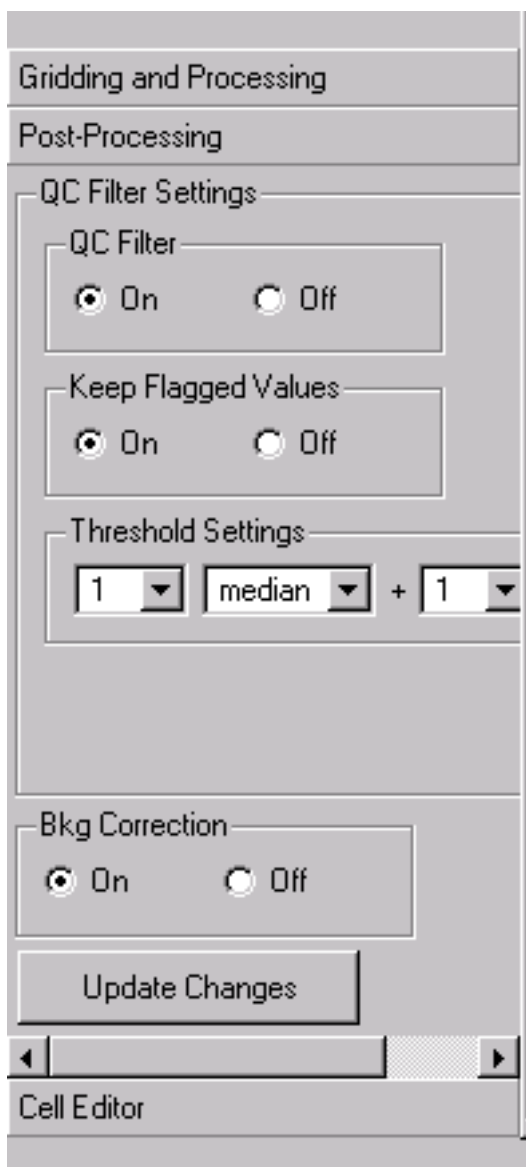
When **QC Filter** is enabled the threshold settings can be selected from dialog. The threshold is defined in the form of  $n * (\text{background median}) + m * (\text{background standard deviation})$ . Default values are  $n=1$ , and  $m=1$ . User may adjust this threshold cutoff level by changing parameters  $n$  and  $m$ .

Also user can set background correction on and off by using button group **Bkg Correction**.

Non of these settings require data reprocessing, however user have to click button **Update Changes** to take all changes into effect.

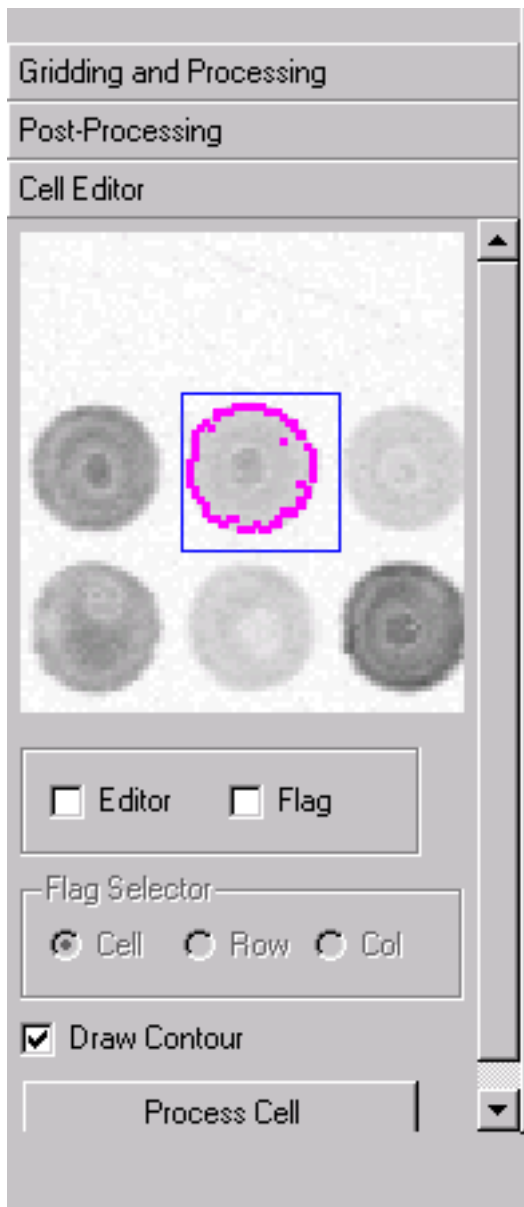
Note that non of those settings change raw data in the program internal data set.

**Post-Processing dialog** can be shown with vertical or/and horizontal scroll bar depending on the current computer screen resolution. Adding scrolling functionality gives user a possibility to view all dialog controls without distortion.



## Cell Editor Dialog

This dialog can be used to view, align, and process selected spot. The channel of the selected spot is controlled by the **Image Selector** combo box on **Analysis Page**. The check box **Editor** must be checked to make **Cell Editor** active. When active **Cell Editor** displays the spot clicked by user on the **Analysis Page** the selected spot cell. In addition the spot can be selected from **RI plot** or from **MEV** table on **Data Page**. To display the spot contour in **Cell Editor** view user should check **Draw Contour** check box. When **Flag** check box is checked user may set processing flag for spot, row or column by mouse click on the selected object on **Analysis Page** or **Cell Editor** itself (spot only). The processing flag sets to certain spot prevent this spot be processed at the processing step. The **Flag Selector** defines what object flagging apply to: spot, row or column. Default **Flag Selector** setting is **Spot**. The color of the cell or cells in row/column is changes from blue to green (Windows and Linux only) or set cell crossed (MacOSX) for processor flagged cells. Flagged cells are not processed during processing. Flagging can be turned back by repeated clicking on cell, row or column. Starting from version 3.1.1 user can set flags to whole group of genes/spots based on there annotation information. This can be done interactively using [Annotation table](#) on **Data Page**.



The button **Process Cell** initiates processing for selected cell only.

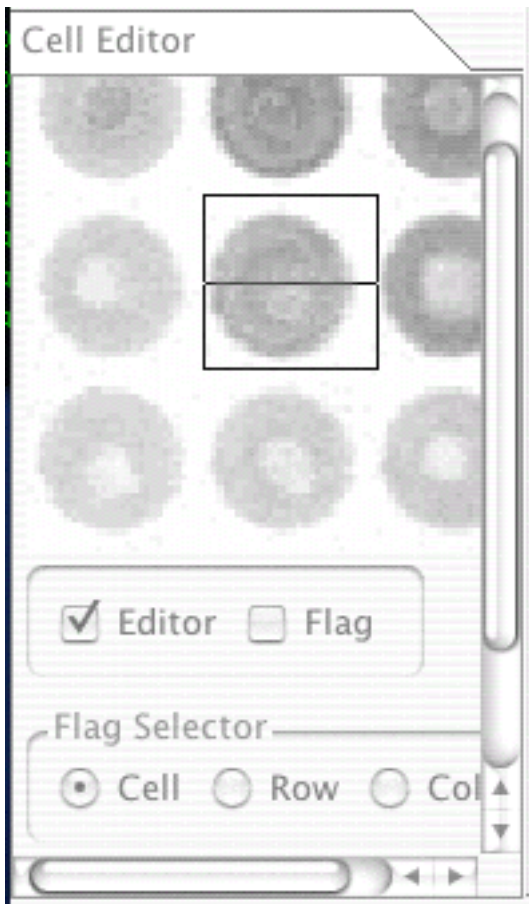
The selected spot cell can be edited, i.e. moved and resized. To move cell user should point the mouse pointer inside the cell on **Cell Editor**, make left mouse click and hold it while moving mouse to a new location. Release the mouse button to drop a cell at new position.

To change the cell rectangle size user should right click mouse when mouse pointer at the right bottom corner of the cell and hold at the bottom right corner of the cell, then drag the corner to a new position. Move cell after that if needed.

**Cell Editor Dialog** can be scrollable depending on the current computer screen resolution. Adding scrolling functionality gives user a possibility to view all controls without distortion. The

shown dialog has vertical scroll bar.

Note: the small image view on this dialog is implemented in self-scaling mode. It won't show selected cell in full view at the first time after program loaded. Select any spot/cell few times to get full view scaling.



This is how **Cell Editor Dialog** appears on Mac platform. Because coloring of movable objects is not supported on MacOSX we use horizontal crossing line to label flagged cells.

---

# General Page

[Click to enlarge](#)



---

**General page** gives a wide angle view for all images loaded in Spotfinder. Every image is shown in full view. The number of the image panels shown depends on the number of channels selected by the user. See [Program settings](#) for more details.

User may need to scroll panes with images horizontally to see all of them.

The first pane displays so called *Union Image*. The *Union Image* is an image internally created by **Spotfinder** from original images for grid alignment and spotfinding. When *Union Image* is generated each pixel in *Union Image* is set to the maximum of the pixels from all original channels (A and B for two color array) at the same location. As result of this definition *Union Image* keeps all strongest pixels and can be used as target for grid alignment and spotfinding.

The rest of the panes on this dialog display the loaded images in all channels labeled starting from letter **A**. User can see the full path to the file loaded in the selected channel by clicking on the corresponding

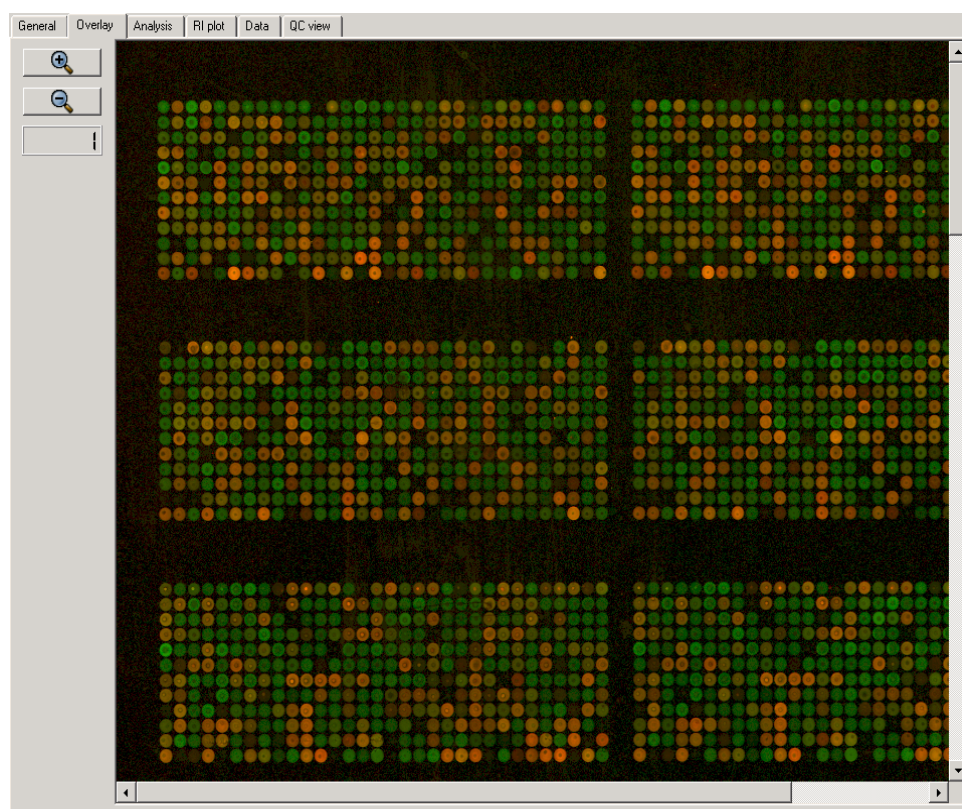
pane of this dialog, The full filepath appears on the left segment of the [Status Bar](#).



# Overlay Page

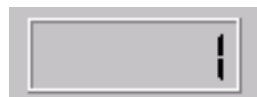
**Overlay page** is the helper view representing as false color view of the first two channels A and B of loaded slide. The next view depicts **Overlay Page** appearance in **Spotfinder**.

[Click to enlarge](#)



The default colors are shades of red for channels **A** image and shades of green for channel **B** image.

 **Zoom In** and  **Zoom Out** buttons can be used for zooming image in and out.

 **Zoom factor** indicator shows the zooming factor relative to the original image scale.

Original image scale is the image view where one screen pixel corresponds to one pixel of displayed image. For instance, the displayed image is twice as less as original 16 bit image in each dimension for [Scale factor](#) =2, which is a default settings. All zooming and scrolling actions on this **Overlay Page** are synchronized for image view on page **QC View Page** and **Analysis Page**.

# Analysis Page

**Analysis page** is the central GUI control object of **Spotfinder**. All steps of the image analysis process are managed from **Analysis Page**. When interested cell is selected it is shown in the center of the red (black on Mac OSX) circle drawn on top of the selected grid. The next view shows **Analysis Page** appearance in **Spotfinder**.

[Windows and Linux View \(Click to enlarge\)](#)

[Mac view \(Click to enlarge\)](#)



**Zoom In** and

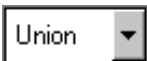


**Zoom Out** buttons can be used for zooming image in and out.



**Zoom factor** indicator shows the zooming factor relative to the original image scale.

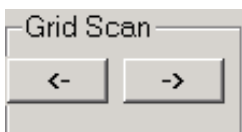
Original image scale is the image view where one screen pixel corresponds to one pixel of displayed image. For instance, the displayed image is twice as less as original 16 bit image in each dimension for [Scale factor](#) =2, which is a default. All zooming and scrolling actions on this **Analysis Page** are synchronized with **Overlay Page** and **QC View Page**.



**Image Selector** combo box is intended to be used for switching between different views in **Analysis page**. The default view is **Union**.



**Contour Selector** check box is designed for setting in and out the detected spot contours on the image view. The contours can be visible only after processing. Default value is off: no contour display. If image is already processed and **Contour** is checked every detected spot has color contour drawn around it. The magenta color is used for displaying good spots. If **QC Filter** is enabled the green color is used for spots flagged as "bad" by **QC filter**. The spot contour is colored in green if spot is "bad" at least in one channel. See [Post Processing Dialog](#) for details.



**Scan grid** button group can be used for a sequential scrolling across the grids. The default starting point is a top left grid indexed by **1:1**. If some grid is selected previously in **Grid Button Control** it becomes the starting point. The **Right** arrow labeled button "->" scrolls grids from left to the right first, then top to the bottom. The **Left** arrow labeled button scrolls grids in reverse order. The top left corner of the selected grid is guaranteed to be at the top left corner of the **Analysis Page** view. Appropriate choice of zooming can be used independently to display selected grid in full.

1:1	1:2
2:1	2:2
3:1	3:2
4:1	4:2
5:1	5:2
6:1	6:2
7:1	7:2
8:1	8:2
9:1	9:2
10:1	10:2
11:1	11:2
12:1	12:2

**Grid Button Control** is a set of the buttons arranged exactly in the same way as the grids on the image. The buttons are labeled with two indices as the drawn grids on **Analysis Page** view. If selected button is pressed the corresponding grid is guaranteed to be at the top left corner of the **Analysis Page** view. Appropriate choice of zooming can be used independently to show selected grid in full.

After loading a grid from file or making it from scratch using the grid creation dialog from [Grid Menu](#) the grid alignment has to be performed. Alignment in general is viewed as process of grid superimposing on image; it requires grid movement, expansion or shrinking, rotation, and fine adjustment on image. There are two types of menu in **Analysis Page** for grid manipulation and processing. **Grid Menu** provides access to commands to control the selected grid. The **ALL Grid Menu** provides access to commands to control ALL grids. There are two types of commands or functions in each menu. The top four commands are interactive and require a user input to perform operations. All interactive commands are performed on step-by-step basis in small increment. They need to be terminated when user decides that selected command completes the task. Interactive command termination is made by pressing END on keyboard or left mouse button. The current list interactive commands is: **Move**, **Expand**, **Cell Size**, and **Rotate**. All four arrows keys are active for **Move**, **Expand**, **Cell Size** commands and only keys **Up Arrow** and **Down Arrow** can be used for **Rotation**. The mouse movement can be used for **Move** command.

The rest of commands in both menu are automatic. Being launched by user they enter iteration loop and when it's done terminate themselves. After starting interactive command user has to use keyboard arrow keys to enter changes in desirable direction.





User can bring up **Grid Menu** by clicking right mouse button when mouse pointer overlays the selected cell. The selected grid is painted in magenta when mouse is on top of it. Unselected grids are painted in blue color.

Interactive commands:

- **Move** command moves selected grid up, down, left, and right by mouse or by using keyboard arrow keys.
- **Expand** command resizes grid while keeps the top and left edges in there positions:
  - expands grid from left to the right by pressing right arrow key;
  - shrink grid from right to the left by pressing left arrow key;
  - expands grid from top to the bottom by pressing down arrow key;
  - shrink grid from bottom to the top by pressing up arrow key;
- **Cell Size** resize all cells in grid simulteneously:
  - up arrow key moves top cell edge up and bottom cell edge down by one pixel each; cells become bigger in height;
  - down arrow key moves top cell edge down and bottom cell edge up by one pixel each; cells become smaller in height;
  - right arrow key moves every cell right edge to the right and left edge to the left by one pixel; cells become wider;
  - left arrow key moves right cell edge to the left and left cell edge to the right by one pixel; cells become smaller in width;
- **Rotate** command rotates grid around its top left corner:
  - up arrow key rotates the grid counterclockwise;
  - down arrow key rotates the grid clockwise;

### Accelerator Key

User may increase step for **Move**, **Expand**, **Cell Size**, and **Rotate** commands by pressing and holding **CTRL** when using selected keyboard arrow key. This key combination increases the step of current action by factor of 10 compare to the keyboard key action without **CTRL** pressed.

Undo/Redo functions:

- **Undo** command restores previous grid.
- **Redo** command reverse **Undo** command action.

Currently the depth of undo/redo buffer are implemented in **Spotfinder** equals to 3.

Automatic grid commands:

- **Adjust** command begins the grid adjustment procedure. The **Mask Size** has to be set in [Gridding Processing Dialog](#) before this command action.
- **Process** command begins the grid processing. User should set all needed [Algorithm Settings](#) prior to this command.
- **Visible** command toggles on and off the grid drawing on **Analysis Page**. This makes selected grid [visible](#) or [invisible](#). Only the grid enclosing rectangle is shown on the **Analysis Page** when grid is set invisible.



User can pop up **All Grids Menu** by right clicking mouse when mouse pointer is outside of any grid.

Interactive commands:

- **Move All** moves all grids up, down , left, and right by moving mouse or by keyboard arrow keys.
- **Expand All** resizes all grids while keeps the top and left edges without changes:
  - expands all grids from left to the right by pressing right arrow key;
  - shrink all grids from right to the left by pressing left arrow key;
  - expands all grids from top to the bottom by pressing down arrow key;
  - shrink all grids from bottom to the top by pressing up arrow key;
- **Cell Size All** resize all cells in all grids simultaneously:
  - up arrow key moves top cell edge up and bottom cell edge down by one pixel each; cells become bigger in height;
  - down arrow key moves top cell edge down and bottom cell edge up by one pixel each; cells become smaller in height;
  - right arrow key moves right edge to the right and left edge to the left by one pixel; cells become wider;
  - left arrow key moves right cell edge to the left and left cell edge to the right by one pixel; cells become smaller in width;
- **Rotate All** command rotates all grids around there top left corner:
- top arrow key rotates all grids counterclockwise;
- down arrow key rotates all grids clockwise;

### Accelerator Key

User may increase step for **Move**, **Expand**, **Cell Size**, and **Rotate** commands by pressing and holding **CTRL** when using interactive command by applying the keyboard arrow keys. This key sequence increases the step by factor of 10 compare to the keyboard key action without **CTRL** key pressed.

All Grids Undo/Redo functions:

- **Undo All** command restores ALL previous grids.
- **Redo All** command restores the **Undo All** command action.

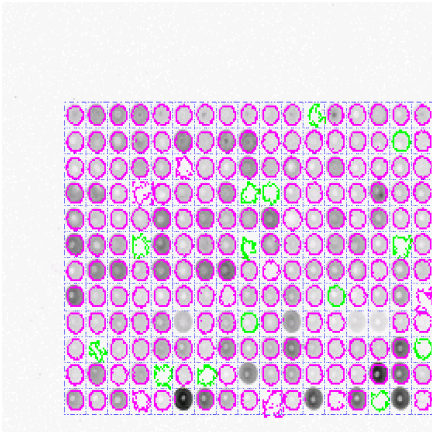
Currently the depth of implemented in **Spotfinder** undo/redo buffer equals to three.

Automatic grid commands:

- **Adjust All** command begins all grids adjustment procedure. The **Mask Size** has to be set in [Gridding Processing Dialog](#) before the use of this command.
- **Process All** command begins all grids processing. User should set all needed [Algorithm Settings](#) prior to this command.
- **Visible All** command toggles on and off all grids drawing on **Analysis Page**. Only the ALL grids enclosing rectangles are shown on the **Analysis Page** when ALL grids are set *invisible*. This makes all grids [visible](#) or [invisible](#).

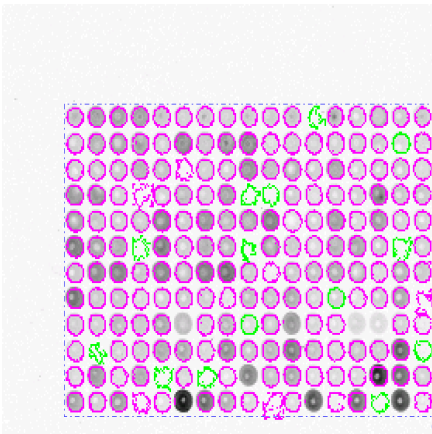
---

Grid is Visible



[click to enlarge](#)

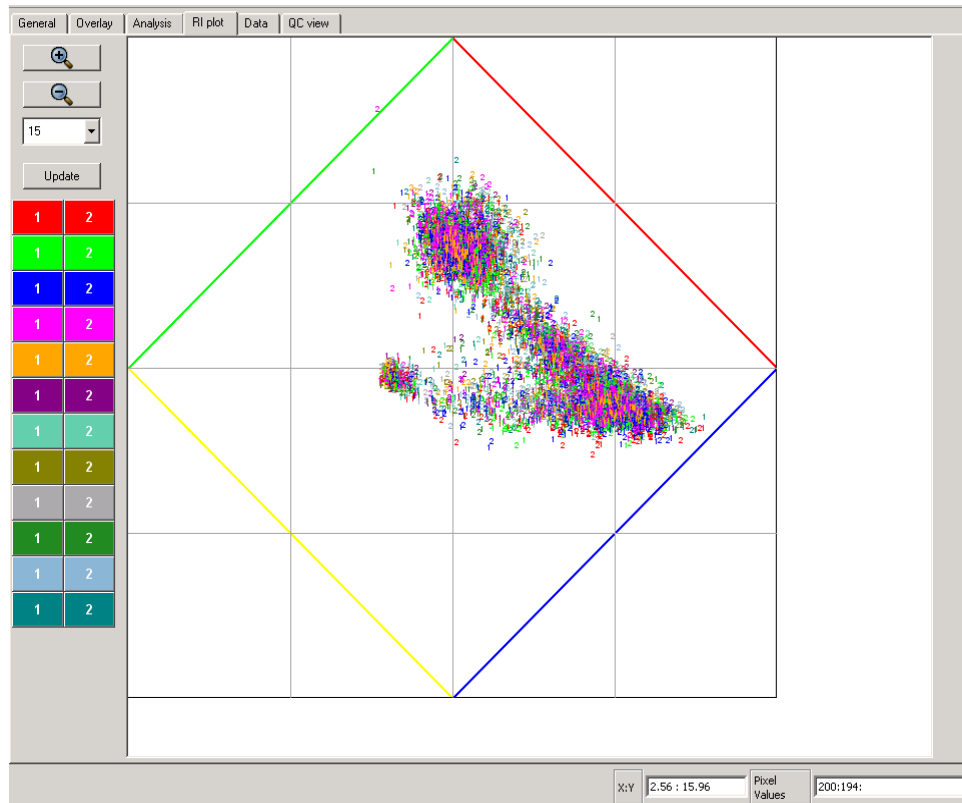
Grid is Invisible



[click to enlarge](#)

# RI Plot Page

[Click to enlarge](#)



**RI Plot Page** is a home of an interactive **RI plot view** of plot  $\text{Log}_2(A/B)$  vs.  $\text{Log}_2(A*B)$ , where A and B are the background corrected medians for channels A and B respectively. Every spot/gene is presented by a colored number on **RI plot view**. Color represents metarow, and number represents metacolumn. User can zoom **RI plot view** in and out, as well as scroll it, and change symbol size and visibility for every grid individually. Interactivity is implemented as ability for user to click on selected spot/gene and automatically switch to the **Data page** where the same selected spot/gene highlighted in the MEV table.

**Note:** Because of the interactive functionality **RI plot view** can be very slow when redrawing and updating itself with large data sets (high density slides). User can disable **RI plot view** from [Program Settings](#) to improve program performance and responsiveness in interactive mode.

The yellow, red, blue, and green lines on **RI plot view** are the limits of log-ratios described in [4]. The meaning of these limit lines is that the data points corresponding to the real spots cannot be shown outside of the diamond. Any data points shown outside the diamond lines is result of some errors in the procedure. It is most likely to be the incorrect background correction as result of the wrong spot boundaries detection. For this reason the **RI plot view** can be used as quick visual quality control of spotfinding procedure.

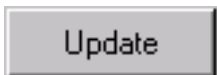


**Zoom In** and



**Zoom Out** buttons can be used for zooming image in and out.

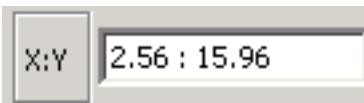
This combo box is a **Symbol Size Selector**. It allows user to set the symbol size (digits) showing the spots/genes on **RI plot view**. Default value is 2, which is good for detailed view at high zoom values.



The push button **Update** forces to redraw the whole **RI plot view**. It should be used to update the changes made in symbol size or/and **Grid Selector**.



The **Grid Selector** is a set of the colored buttons arranged in the same way as the grids on a slide. Every button corresponds to one grid of slide. Geometry of **Grid Selector** is defined at the time of grid loading from file or grid creation from dialog. See [Grid Menu](#) for details of this. The color buttons are labeled by numbers rising from left to the right. The numbers are the indices of metacolumns. The buttons in one row are colored in the same color and denotes one metarow. User can turn on and off the drawing of the spots/genes on **RI plot view** for any grid by pushing the corresponding button on **Grid Selector**. The pushed (shaded) button stands for grid set no to be painted on **RI plot view**, while normal button stands for painted grid. For instance, red symbol **1** is used for drawing on **RI plot view** all genes/spot of the grid in first metacolumn and first metarow. A number of symbols **1** will be visible on **RI plot view** when red button labeled by number one is at normal. Similarly, when the same button is pushed down non of red ones ( **1** ) is shown on the **RI plot view**.

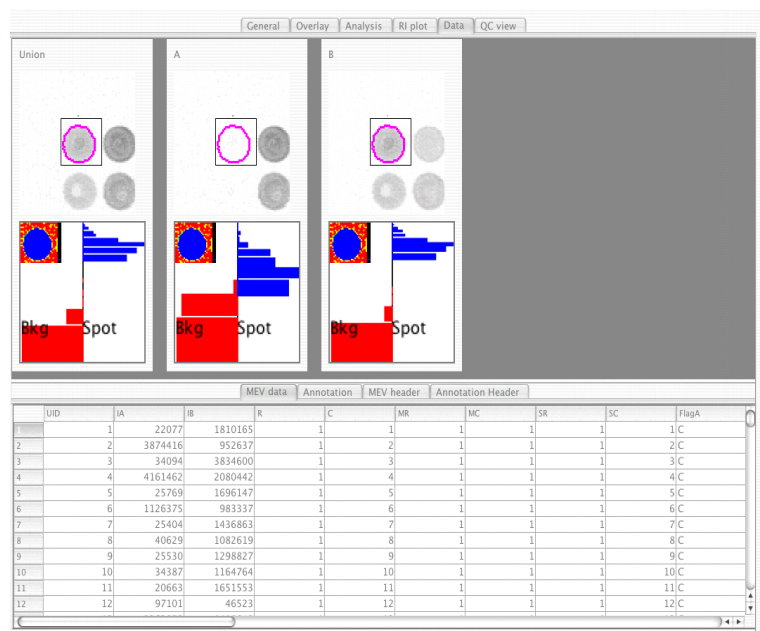
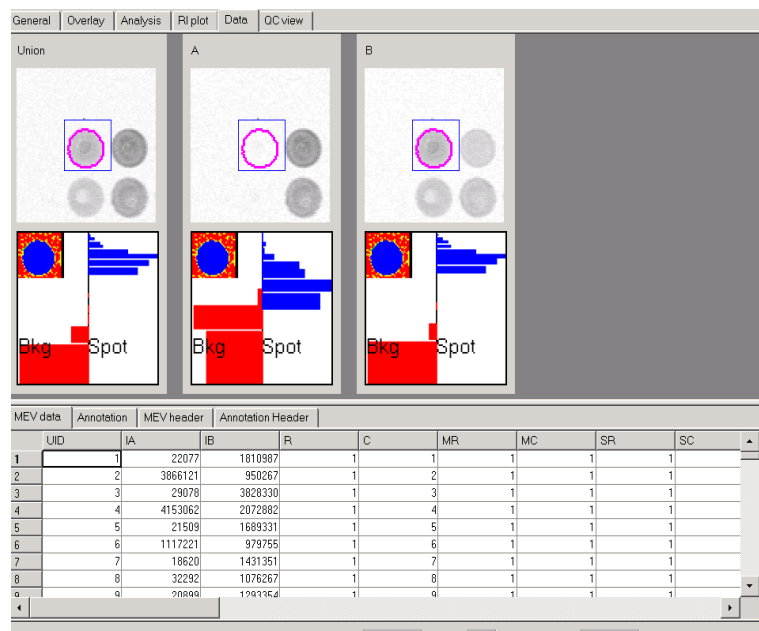


The **X:Y** indicator on status bar shows the RI pot coordinates when RI plot is visible. The cordinates are shown in  $\log_2$  scale, as result the whole plot area has limits from 0 to 32 in X coordinate and from -16 to 16 in Y coordinate.

# Data Page

[Windows and Linux View. Click to enlarge](#)

[MacOSX View. Click to enlarge](#)



**Data Page** was designed to display spots data after processing. The displayed data include spatial information about detected spot shape and its numerical parameters shown below on spreadsheet like control. **Data Page** is separated in two views. The vertical location of the separator can be adjusted by user. Its upper part consists of a number of panels with selected spot view and spot/background histogram in all channels. The lower part is a **Table View**.

## Table View

**Table View** is a set of the interactive tables for viewing processed data in MEV file format shown as if it is opened in spreadsheet. The **Data Page** depicts MEV file header and body separately in two different tables: **MEV header** and **MEV data**. These two tables give user complete view of the future mev file, which program creates when user stores data in mev file. Before to store data to MEV file user may add comments to MEV header. The last line of **MEV header** can be used for this. If annotation file is available it can be loaded and displayed in tables **Annotation** and **Annotation header**. All tables are *read only* (above mentioned comment line in MEV header is the only exception) and interactive. For more details on [MEV file](#) or [Annotation file](#) click appropriate link.

The **Annotation** table has a simple interactive functionality that allows to turn on and off the flags for all spots with selected annotation field. Flagged spots are not processed by Spotfinder and result in zero values in **MEV data**. In downstream analysis they can be easily extracted from the future data analysis. To flag all spots with the chosen field in selected column in **Annotation** table right click on this field and select the command **Flag/Unflag spots with selected fields**. The repeated action unflags cells. The flags are stored in grid file when grid is saved to file. When loaded this file brings the grid with the same cells flagged across the slide. The cell flagging on **Data Page** does not require checking **Flag** check box on the **Left Pane Cell Editor**. The screenshot below shows the field "null" selected for flagging in column "Guess" in **Annotation** table.

MEV data		Annotation	MEV header	Annotation Header	
		GB#	TC#	Guess	
1		AF159801	null	null	
2	UI	AI836658	TC966777	TC: Ubiquinone biosy	Flag/unflag spots with selected fields biosynthesis protein 7
3	UI	AI839716	TC1085623	TC: mKIAA0184 protein {Mus musculus}	
4	UI	AI843348	TC1011100	TC: Similar to Npw38-binding protein NpwBP {Mus musculus}	
5	UI	AI846314	TC1016244	TC: ancient conserved domain protein 1 {Mus musculus}	
6	UI	AI847690	TC1038488	TC: ribosomal protein S4 cytosolic [validated] - rat	
7	UI	AI848929	TC986232	TC: WNT-4 protein precursor. [Mouse] {Mus musculus}	
8		AI853382	TC1030885	TC: RNA binding motif protein 7 {Mus musculus}	

**Note:** Because of the interactive functionality **Data Page** consumes a lot of memory resources and can be very slow when deals with large data sets ( high density arrays ). User may choose to display the reduced size of MEV data file on **Data Page** to improve program performance and optimize memory use. In this mode only the first ten columns will be shown in **MEV data** view. This visualization mode does not limit content and the size of output MEV file. The settings can be changed from [Program Settings](#) dialog.

Interaction assumes that user can click on any row in the table and select a spot/cell for viewing:

- in all channels in **Spot Histogram View** on this page;
- in **Left Pane** [Cell Editor](#);
- in **RI Plot Page** the selected spot drawing symbol is displayed in black rectangle;

The interested spot selection can be done from **Analysis Page** by clicking on cell of interest and from **RI Plot Page** by clicking on the selected symbol on RI Plot.

## Spot and Histogram Panel

### Spot View

The selected spot is displayed in the vicinity of its close eight neighboring spots. The selected spot cell is editable in the same way as on [Cell Editor](#).

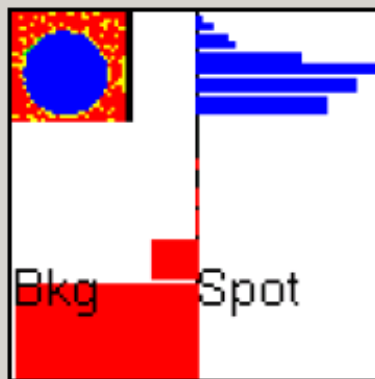
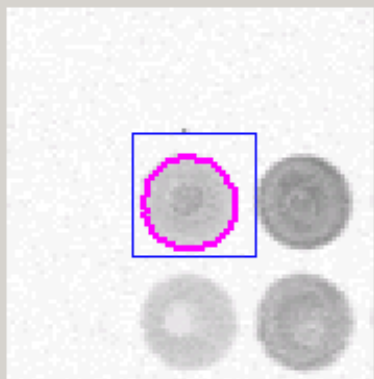
Selected spot cell can be edited, i.e. moved and resized. Spot cell flagging is implemented in this control in the same way as in **Cell Editor Dialog**. To move cell just make left mouse click and hold it while moving mouse to the new location. To change the cell rectangle size right click mouse and hold at the bottom right corner of the cell, than drag the corner to a new position. Move cell after that if needed. Spot cell flagging is implemented in this control in the same way as on [Cell Editor](#).

### Histogram Plot

The **Histogram Plot** helps user to visualize the spot/background separation performed by the segmentation procedure. The spot and background pixels histograms are displayed on the same panel for every channel. The histograms for both background and spot pixels are plotted vertically, with minimum pixel at the bottom and maximum pixel at the top of the panel. Vertical middle line divides pane in two halves: left for background pixels set, right one for spot pixels set. On the right side of the panel the blue color displays the spot pixels histogram in log scale. On the left side red color displays the background pixels histogram in log scale. The spot pixels histogram peak is at the extreme right side, while background pixel histogram maximum is at the



Union



extreme left side. Location of the blue histogram on the right above the red one on the left means a good separation between a spot and surrounding background. In this case one can tell about real spot detection with a perfect separation between spot and background pixels. When both red and blue histogram peaks are at the same vertical level or close to each other there is no difference between spot and background: the spot cannot be considered as real.

### Colored cell mask

Starting from version 3.1.1 a colored mask is drawn in the top left corner of the **Histogram Plot**. This mask is introduced to visualize the result of advanced cell pixels clustering available in 3.1.1. Four sorts of pixels are colored in four different colors. **Blue** pixels belongs to the detected spot, those are signal pixels. The spot is quantitized using this set of pixels. Saturated pixels are shown in blue also. **Green** pixels if shown mark the high value pixels outside spot area. **Red** pixels form the background area. Background calculation uses this set of data. **Yellow** pixels mark the subset of background pixels ( T, %) that are not used in background calculation. The T- **Top Background Cutoff** in % is set on [Gridding and Processing Dialog](#). By setting higher or lower **Top Background Cutoff** user may increase or decrease the red background area. Marked in yellow color pixels can originate from local background spike pixels and from the area closed to the detected spot contour. By setting higher **Top Background Cutoff** user can be assured that background is not effected by the pixels outside spot area.

## QC View Page

**QC View Page** is a set of the quality control views. Currently there are implemented two types of the views: View A and ViewB.

View A is color display that shows the spots painted in different colors depending on there expression ratio relatively to two preset threshold levels and flag value. Using two thresholds subdivides all spots in three categories: below low threshold, above high threshold, and between these two. After applying total slide normalization the spots with  $\log_2$  expression ratio above high threshold (1 is default) shown in yellow, below low threshold (-1 is default) - in blue, and between these two threshold painted in gray color. Additionally bad and non-detected spots forms the forth category and are shown in green color. View A has two flavors . One of them (1) shows all spots in grid with expression ratios in the same category as one colored rectangle for every grid. The rectangle area is proportional to the number of spots in the category. As result the grid rectangle looks like the superposition of four big color rectangles. Another flavor (2) of View A displays spots colored by using the same schema but located in there original positions. View A has a number of controls described below.

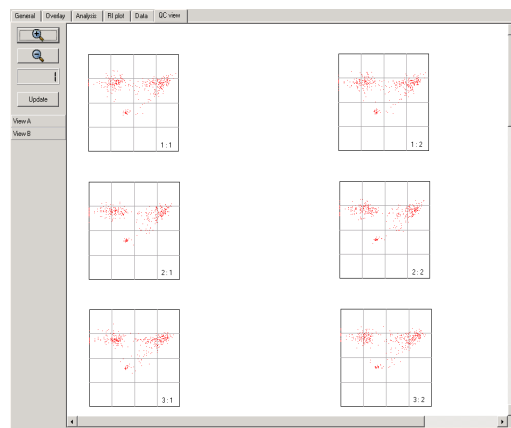
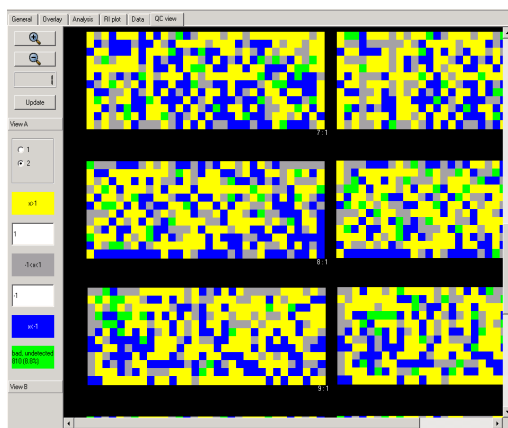
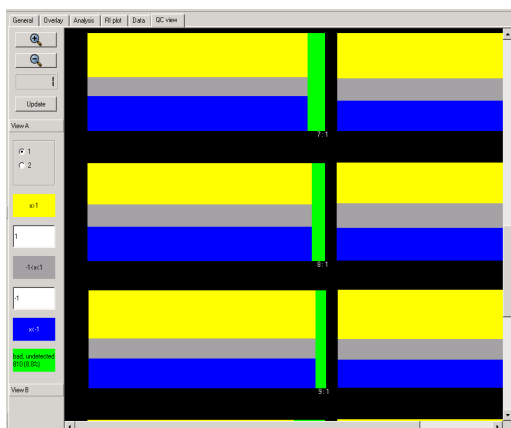
The quality control function of **QC Page** can be demonstrated when use it for visual comparative analysis of the relative number of bad/undetected spots in individual grids on array. Under assumption that all grids on array have approximately the same number of the bad spots we can quickly locate the grid with abnormally high or low number of bad/undetected spots. These grids have to be chacked for the proper alignment at first.

View B shows the Intensity B vs. Intensity A for every grid or block. In this view grid are shown in there original location and labeled by indices in the same way as on View A. View B (3) has no controls currently.

( 1 ) [Click to enlarge](#)

( 2 ) [Click to enlarge](#)

( 3 ) [Click to enlarge](#)



Zoom In and



Zoom Out buttons can be used for zooming image in and out.

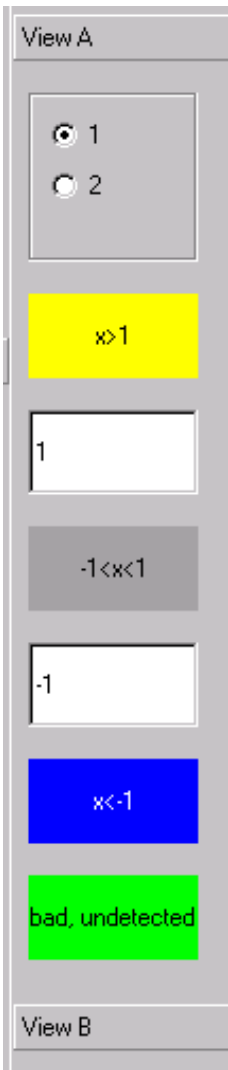


**Zoom factor** indicator shows the zooming factor relatively to the original image scale. All zooming and scrolling actions on **QC View Page** are synchronized with **Analysis Page** and **Overlay Page**.

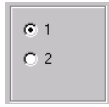


The push button **Update** forces to redraw the whole **QC View Page**. It should be used to update changes made by using any control in **Control Panel** at this page.

## Control Panel



The **Control Panel** currently is implemented as a holder for only one dialog with controls that allow to set two thresholds. It might be extended in future to manage more QC views.



Radio button selector swiches between **View 1** and **View 2**



Yellow legend label with number in it shows the upper expression ratio threshold. Spots with log ratio above this value are shown in yellow.



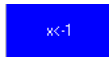
Text entry box can be used for setting the upper threshold.



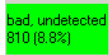
Gray legend label holds the expression with the range between upper and low threshold. Spots in this range of expression ratio are painted in gray color.



Text entry box can be used for entering the low expression threshold.



Blue legend label with number in it shows the low threshold of expression ratio. Spots with log ratios below this threshold are shown in blue color.



Green legend label with text reminding that the bad and undetected spots are shown in green. The numbers show the total number of bad/ undetected spots and there percentage relatively to the total number of spots on the slide. The numbers give correct information only after the full slide is processed since always uses the total number of spots as reference. The number of bad/undetected spots is reported in **MEVheader** of [Data Page](#) next to the *input\_row\_count* parameter which is the total number of spots in array. This number is reported in ouput MEV file.

# MEV file description

This is short .mev File Format Description. Please look TM4 suite package for full document.

Production Revision 4.0 (July 16, 2004)

A MultiExperimentViewer or .mev file is a tab-delimited text file that contains coordinate and expression data for a single microarray experiment. A single header row is required to precede the expression data in order to identify the columns below. With the exception of optional comment lines, each remaining row of the file stores data for a particular spot/feature on the array.

MeV and other TM4 software tools will consider comment lines non-computational. A comment line must start with the pound symbol '#', and can be included anywhere in the file. If the pound symbol is the first character on a line, the entire line (up to the newline character '\n') will be ignored by the software tool.

The mev files created at TIGR will typically contain at least one comment at the top of the file with the following information. The format and fields contained within these comments are subject to change. See Appendix 2 for details.

version	Version number based on revisions of expression data
format_version	The version of the .mev file format document
date	Date of file creation or update
analyst	Owner or the person responsible for creating the file
analysis_id	id from the analysis table that corresponds to this set of expression values
slide_type	slide_type from the slide_type table that this array is based on
input_row_count	Number of rows of expression (eg. non-header) data in input files
output_row_count	Number of rows of expression (eg. non-header) data in this file
created_by	Software tool used to create the file
description	Common name or other details about the experiment

An example of the leading comments:

```
# version: V1.0
# format_version: V4.0
# date: 10/06/2004
# analyst: aisaeed
# analysis_id: 10579
```

```
# slide_type: IASCAG1
# input_row_count: 32448
# output_row_count: 32448
# created_by: TIGR Spotfinder 2.2.3
# TIFF files processed: gpc30025a_532_nm.tif, gpc30025a_635_nm.tif
# description: Tumor type comparison
# This is the 4th experiment in a series of 20 to identify tissue-specific genes.
```

The header row consists of the field names for each subsequent row in this file (with the exception of comment lines). A minimum of seven columns must be present, and these must use a set of specifically named headers. Any number of additional columns may be included. The seven required column headers are:

UID Unique identifier for this spot

IA Intensity value in channel A

IB Intensity value in channel B

R Row (slide row)

C Column (slide column)

MR Meta-row (block row)

MC Meta-column (block column)

As of version 4.0 of this file format the IA and IB columns can be substituted with MedA and MedB. The new requirement is that at least one integrated intensity (IA, IB, etc.) or one median (MedA, MedB, etc.) value be reported for each channel in the microarray. For example, a two channel microarray .mev file would require either IA and IB or MedA and MedB.

MedA Median intensity in channel A

MedB Median intensity in channel B

The mev files created at TIGR may use one of the two formats for the header row, depending on the origin of the mev file. The non-required columns (i.e. anything after the 7th column) may be rearranged and their names are subject to change at this time.

1) Spotfinder created mev file:

```
UID \t IA \t IB \t R \t C \t MR \t MC \t SR \t SC \t FlagA \t FlagB \t SAA \t SAB \t SFA \t SFB \t QCS \t
QCA \t QCB \t BkgA \t BkgB \t SDA \t SDB \t SDBkgA \t SDBkgB \t MedA \t MedB \t AID
```

UID Unique identifier for this spot

IA	Intensity value in channel A
IB	Intensity value in channel B
R	Row (slide row)
C	Column (slide column)
MR	Meta-row (block row)
MC	Meta-column (block column)
SR	Sub-row
SC	Sub-column
FlagA	TIGR Spotfinder flag value in channel A
FlagB	TIGR Spotfinder flag value in channel B
SA	Actual spot area (in pixels)
SF	Saturation factor
QC	Cumulative quality control score
QCA	Quality control score in channel A
QCB	Quality control score in channel B
BkgA	Background value in channel A
BkgB	Background value in channel B
SDA	Standard deviation for spot pixels in channel A
SDB	Standard deviation for spot pixels in channel B
SDBkgA	Standard deviation of the background value in channel A
SDBkgB	Standard deviation of the background value in channel B
MedA	Median intensity value in channel A
MedB	Median intensity value in channel B
MNA	Mean intensity value in channel A
MNB	Mean intensity value in channel B
X	X coordinate of the spot cell rectangle
Y	Y coordinate of the spot cell rectangle
PValueA	P-value in channel A
PValueB	P-value in channel B
DBID	Data Base ID (used when UID is substituted)

The first seven fields (UID, IA, IB, R, C, MR and MC) are required as specified above.

This flexible format allows users to track slide-specific data of interest, such as background, spot size and alternate intensities without requiring them of all users or adopting a limited 'vocabulary' of field names. This header row serves to identify the required and additional data columns. UID must be the

left-most column in the mev file. Other columns do not need to be present in a fixed order.

For mev files generated at TIGR, the UUIDs may be of the form: database\_name:spot\_id (eg. cage:20238). For any given microarray database, the id field in the spot table will be unique. The combination of database and spot\_id will therefore uniquely identify any spot on any array created at TIGR. It is important to note that this is not enough information to distinguish between spots in the same location on two slides of the same slide\_type, as this would typically require an analysis\_id. Since annotation data is based on slide\_type, it is not necessary to make this distinction, as all slides of a given type will use the same annotation file.

The AID column will usually contain an incremental sequence of numbers starting at 1. These can be used to return the file to the original sorted order and can function as a unique row identifier if necessary.

Applications that generate files of expression data (commonly in tav format) by retrieving records from the database access the spot table. TIGR Spotfinder, Midas and Madam are all capable of generating UUIDs of the form described above in addition to the typical coordinate and intensity data.

mev files are required to end with the extension '.mev'. At this time there are no further naming conventions for mev files.



# Annotation file

## Description of Annotation Files

An annotation file is a tab-delimited text file containing annotation data for a specific `slide_type`. `mev` files can be associated with an annotation file only if both types of files are based on the same `slide_type`. The keys to this association are the unique ids in both files. Rows of `mev` and annotation files can be associated with each other if the unique ids are identical. A single header row is required to precede the annotation data in order to identify the columns below. Each remaining row of the file stores annotation data for a particular spot/feature on the array.

Annotation files may contain any number of non-computational comment lines. These lines, starting with '#', will be treated identically to comment lines in `mev` files, and should precede the header row.

Annotation files created at TIGR will use UIDs that match the format used in the `mev` files, most likely `database_name:spot_id`. The structure of each annotation file is detailed below. The header row consists of headers that identify each column of data. Each subsequent row of the file stores data for a particular spot/feature on the array. The annotation files created at TIGR will typically contain at least one comment at the top of the file with the following information:

version	Version number based on revisions of annotation data
format_version/td>	The version of the .mev file format document
date/td>	Date of file creation or update
analyst/td>	Owner or the person responsible for creating the file
created_by/td>	Software tool used to create the document
gi_version/td>	Version of the Gene Indices (or db?) that produced this annotation data
slide_type/td>	type from the slide_type table that this array is based on
output_row_count/td>	Number of rows of annotation (eg. non-header) data
description/td>	Common name or other details about the experiment

An example of the leading comments:

```
# version: V3.0
# format_version: V4.0
# date: 04/20/2004
# analyst: jwhite
# created_by: Database script
```

```
# gi_version: 3.0
# slide_type: IASCAG1
# output_row_count: 32448
# description: Standard annotation file
```

The header row consists of the field names for each subsequent row in this file. Only the UID field is required. It must be the first field present and it must be named 'UID'. Any number of additional fields may be included. Annotation files created at TIGR will always contain the following columns:

UID unique identifier for this line of annotation

R row (slide row)

C column (slide column)

The remaining fields may vary, and a standard set has yet to be determined. Such a list will be published on a future date. R and C have been included to allow for manual alignment of the mev and corresponding annotation files in the event that the mev files were not generated in a traditional manner (ie. using Madam, etc.).

Some varieties of annotation files follow. The format may vary depending on the purpose of the file:

```
UID \t R \t C \t FeatN \t GBNum \t TCNum \t ComN \t ...
```

```
UID \t R \t C \t GeneN \t Rxn \t PathwayN \t ...
```

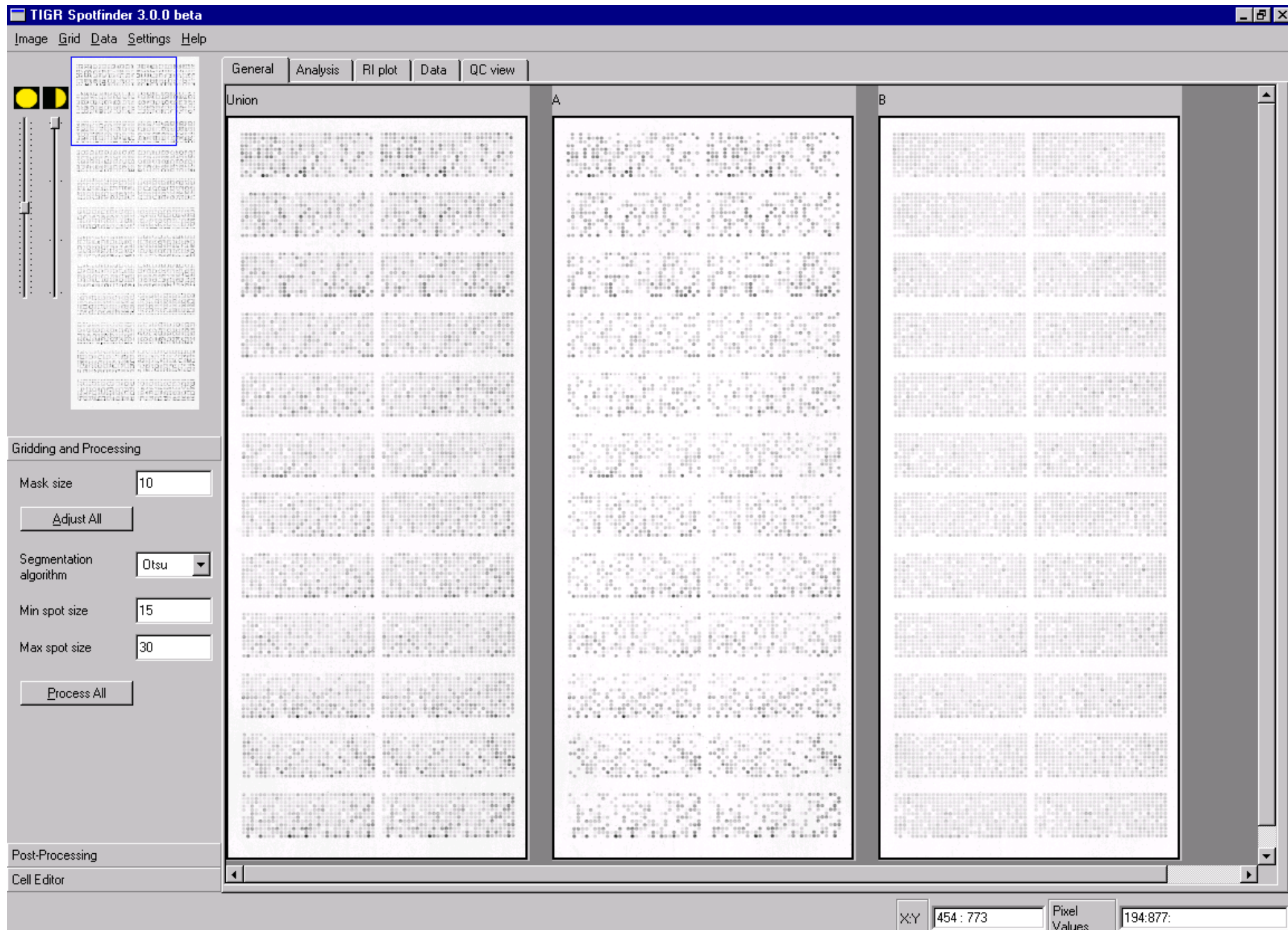
```
UID \t R \t C \t FeatN \t End5 \t End3 \t ChrNum \t ...
```

Of course, it would be possible to combine the fields of these files, or add fields that have not been mentioned here. The goal is to keep the annotation flexible and the processing seamless.

There are not any naming conventions for annotation files at this time. If such a standard is introduced in the future, it will be detailed here.

## References

1. A.I.Saeed, V.Sharov, J. White, J. Li, W.Liang, N. Bhagabati, J. Braisted, M. Klapa, T. Currier, M. Thiagarajan, A. Sturn, M. Snuffin, A. Rezantsev, D. Popov, A. Ryltsov, E. Kostukovich, I. Borisovsky, Z. Liu, A. Vintsevich, V. Trush, and J. Quackenbush 2003.  
TM4: A free, open-source system for microarray data management and analysis.  
BioTechniques 34(2): 374-378.
2. This is an implementation of Otsu algorithm used in **Spotfinder**:  
P. Liao, T. Chen, and P. Chung 2001. A fast algorithm for multilevel thresholding.  
Journal of Information Science and Engineering 17: 713-727.
3. This is a implementation of algorithm for cluster labeling in percolation theory:  
J.Hoshen and R.Kopelman  
Physical Review B, vol.14, number 8, 1976, pp.3438-3445.  
Percolation and cluster distribution. I. Cluster multiple labeling technique and critical concentration algorithm.
4. This is a description of **RI Plot** limits:  
Vasily Sharov , Ka Yin Kwong , Bryan Frank , Emily Chen , Jeremy Hasseman , Renee Gaspard , Yan Yu , Ivana Yang and John Quackenbush  
[The limits of log-ratios](#), BMC Biotechnology 2004, 4:3 doi:10.1186/1472-6750-4-3  
(<http://www.biomedcentral.com/qc/1472-6750/4/3>)

[Back](#)



# Compose Grid dialog

Grid Design			
<div>Standard   Custom</div>			
MetaRow	12	MetaColumn	2
Pin X, pix	900	Pin Y, pix	450
Row Number	12	Column Number	32
X Spacing, pix	28	Y Spacing, pix	28
<div>Help   OK   Cancel</div>			

## Fields description

User needs to enter all field in **Grid Design** dialog. Only **Standard Grid Design** currently implemented. This design assumes that slide was printed by using rectangular pin (pen) settings. Every distinctive grid (block) on slide was printed by one pin. The pins are arranged in rectangular pattern: in metarows and metacolumns.

In our dialog **MetaRow** is the number of pins in vertical dimension, **MetaColumn** is the pin number in horizontal dimension.

**PinX** is a distance between pins in horizontal dimension measured in image pixels. **PinY** is a distance between pins in vertical dimension measured in image pixels.

**Row Number** is the number of rows in every grid printed by the same pin.

**Column Number** is the number of columns in every grid printed by the same pin. For **Standard Grid**

**Design** all grids are the same.

**X Spacing** is distance between adjacent spots' centers in horizontal dimension measured in image pixels.  
**Y Spacing** is distance between adjacent spots' centers in vertical dimension measured in image pixels. In most cases they are the same.

## How to get values for these entries

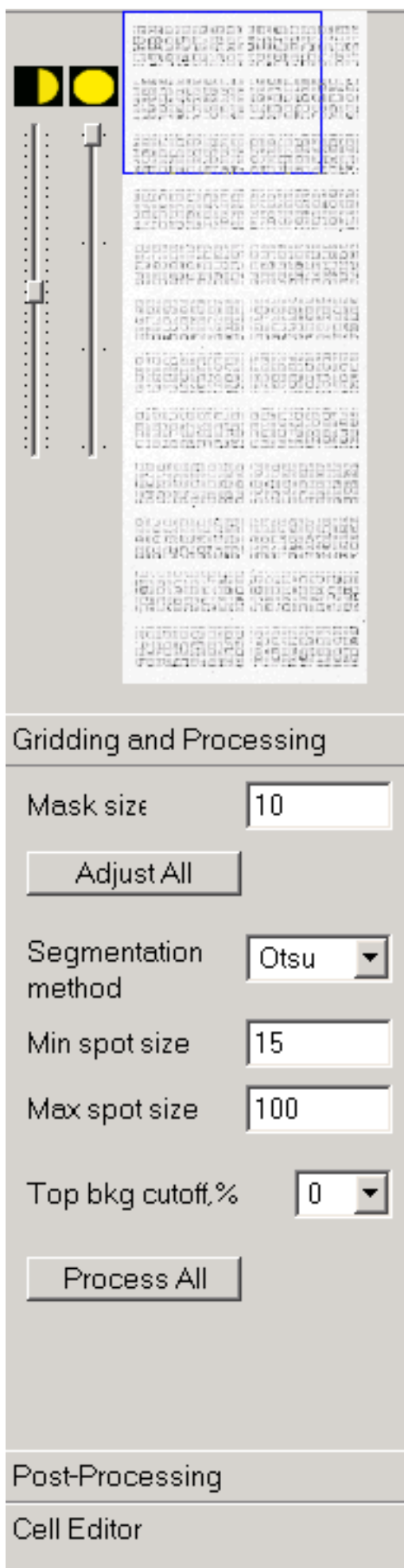
User can get information about these fields from:

- Printing robot settings
- By measuring the parameters on slide. User can use **Analysis Page** to make these measurements. Use mouse pointer and read the X and Y coordinates on the controls of **Status Bar** at the bottom right corner. [More details...](#)

## How to get the image pixel size

It is defined by the scanner resolution used when the slide was scanned. Typically, nowadays the microarray slides are scanned at resolution 10 mk per pixel, which prived us with the image file in the range 20 - 30 MB. Some researcher scan at higher resolution 5 mk per pixel ( *Axon Instruments* scanner can do it). The output file size can be around 100 MB in this case. Hence, rule of thumb is to look at the image file size.

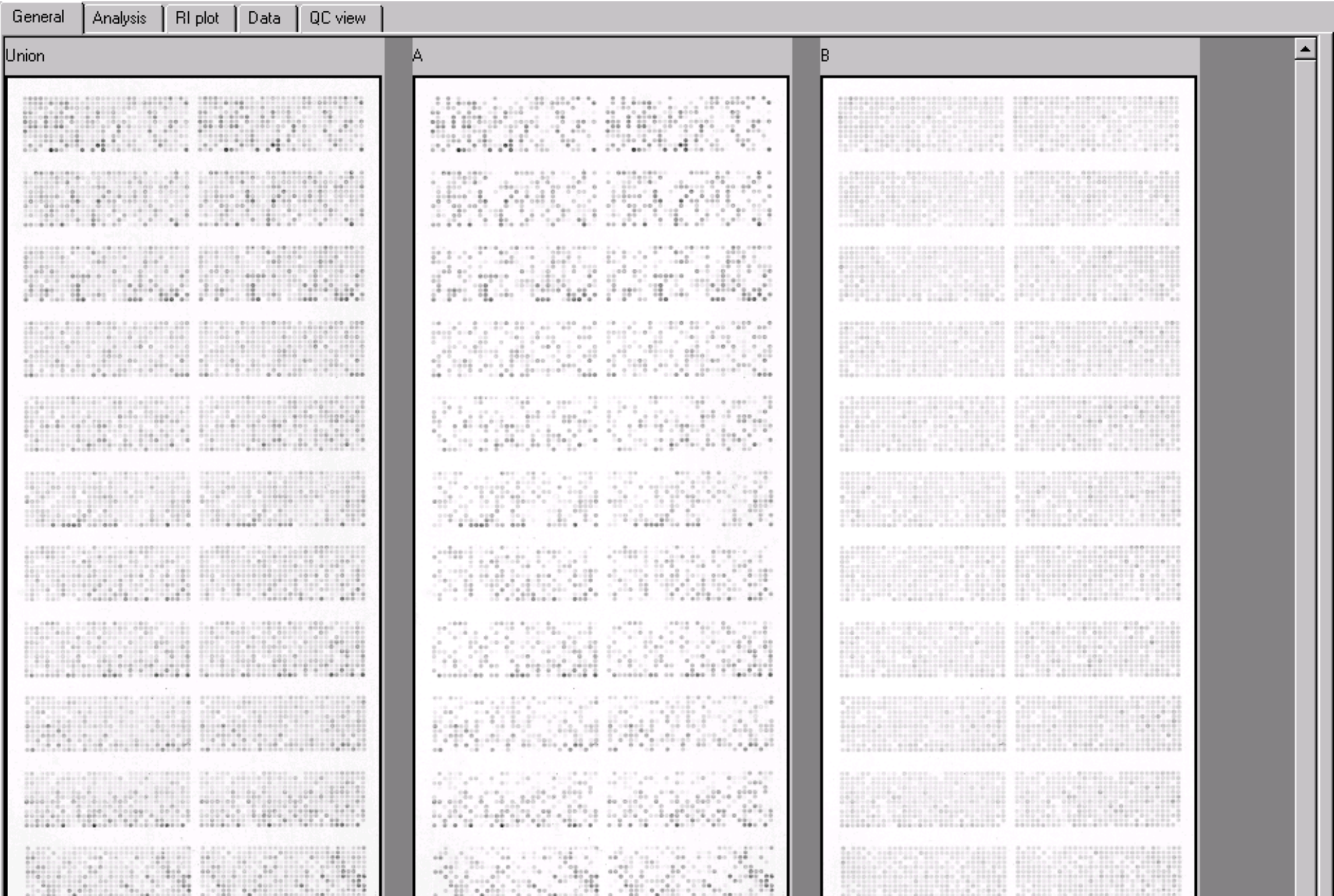


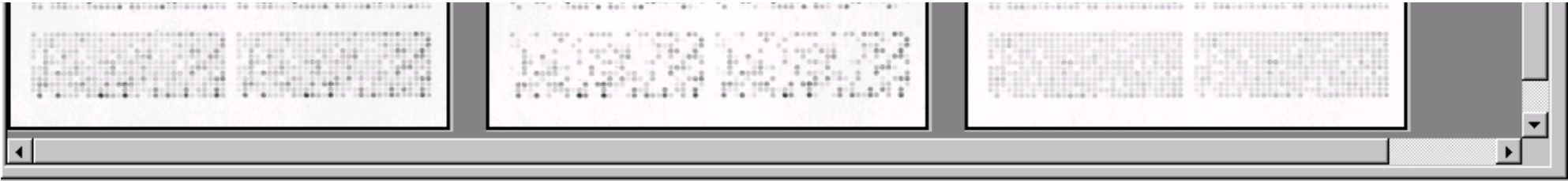




# General page view

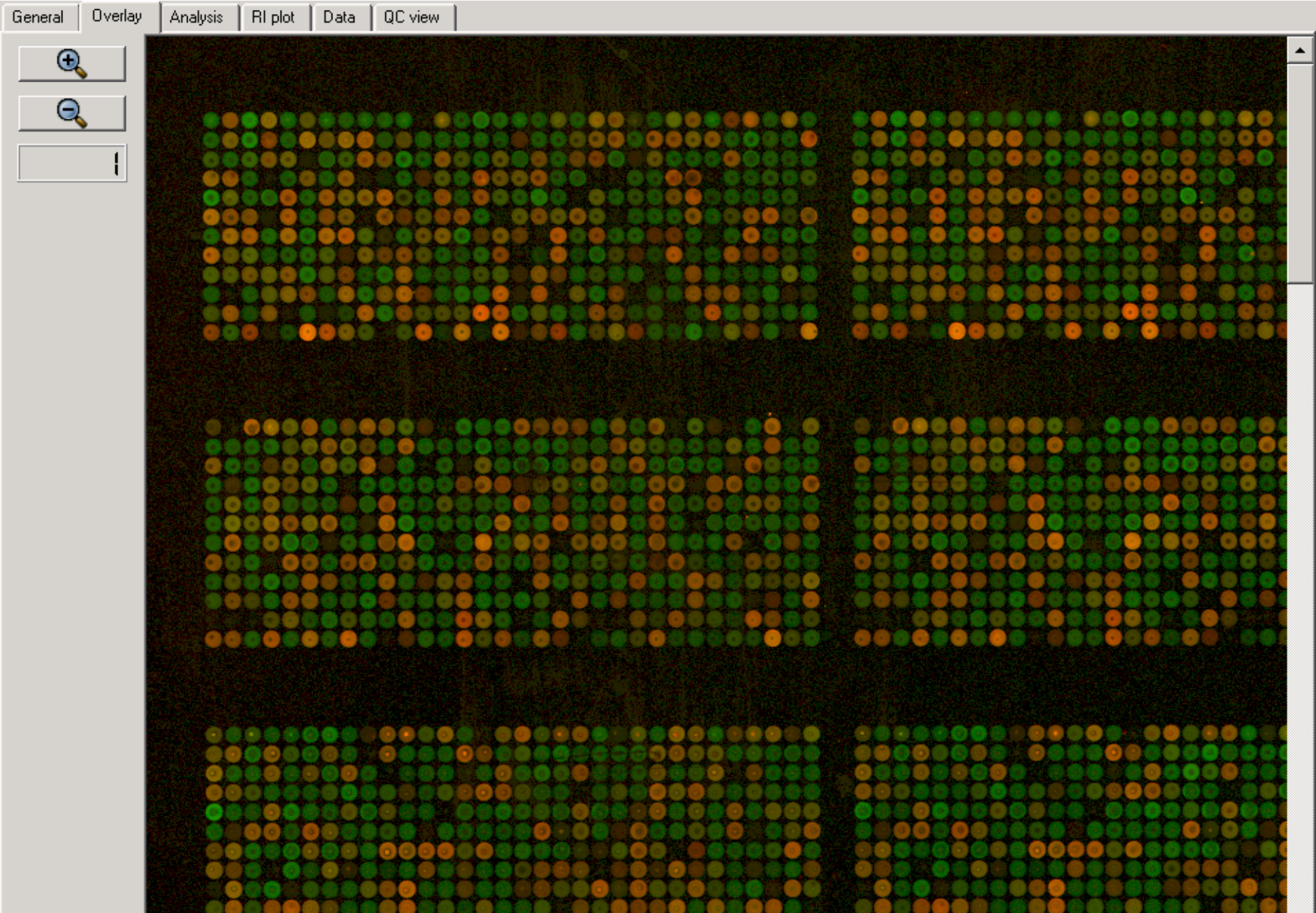
[Back](#)





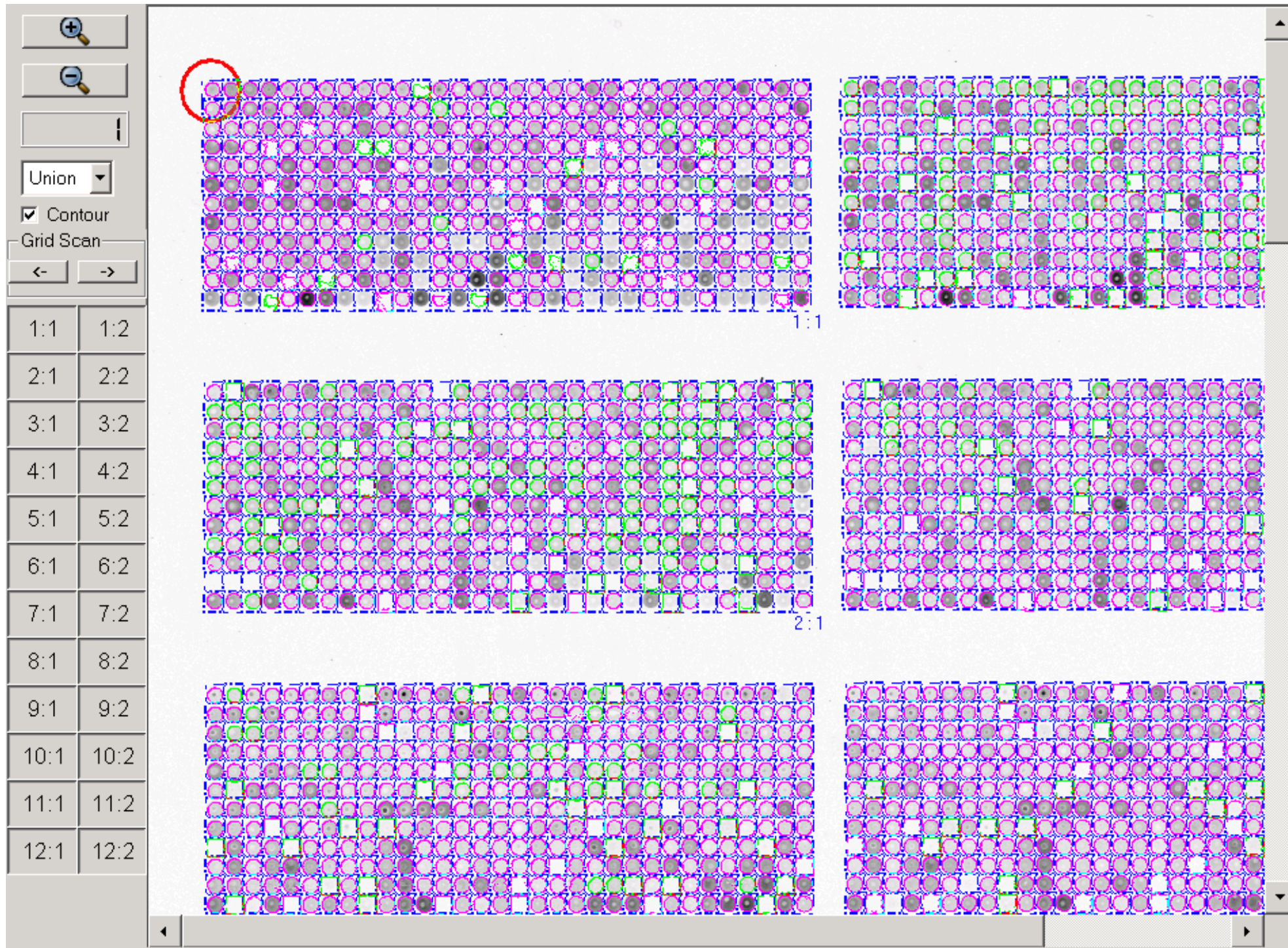


# Overlay Page View



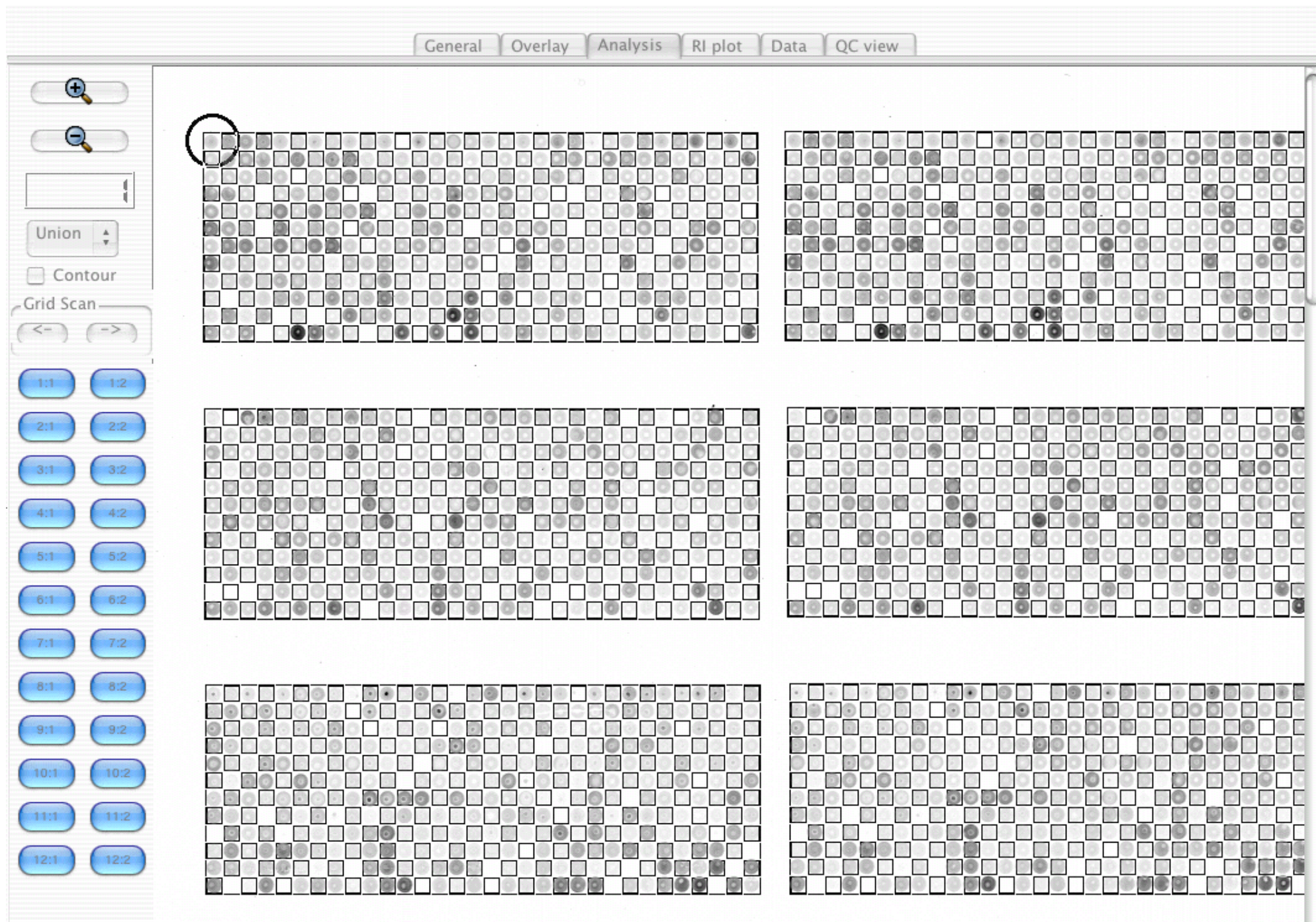


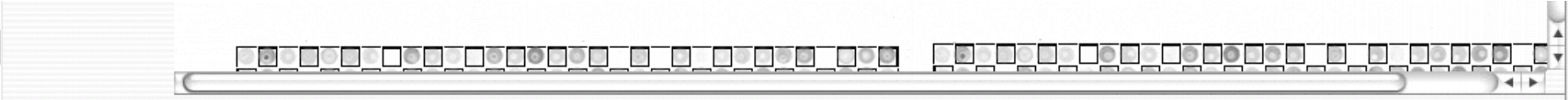






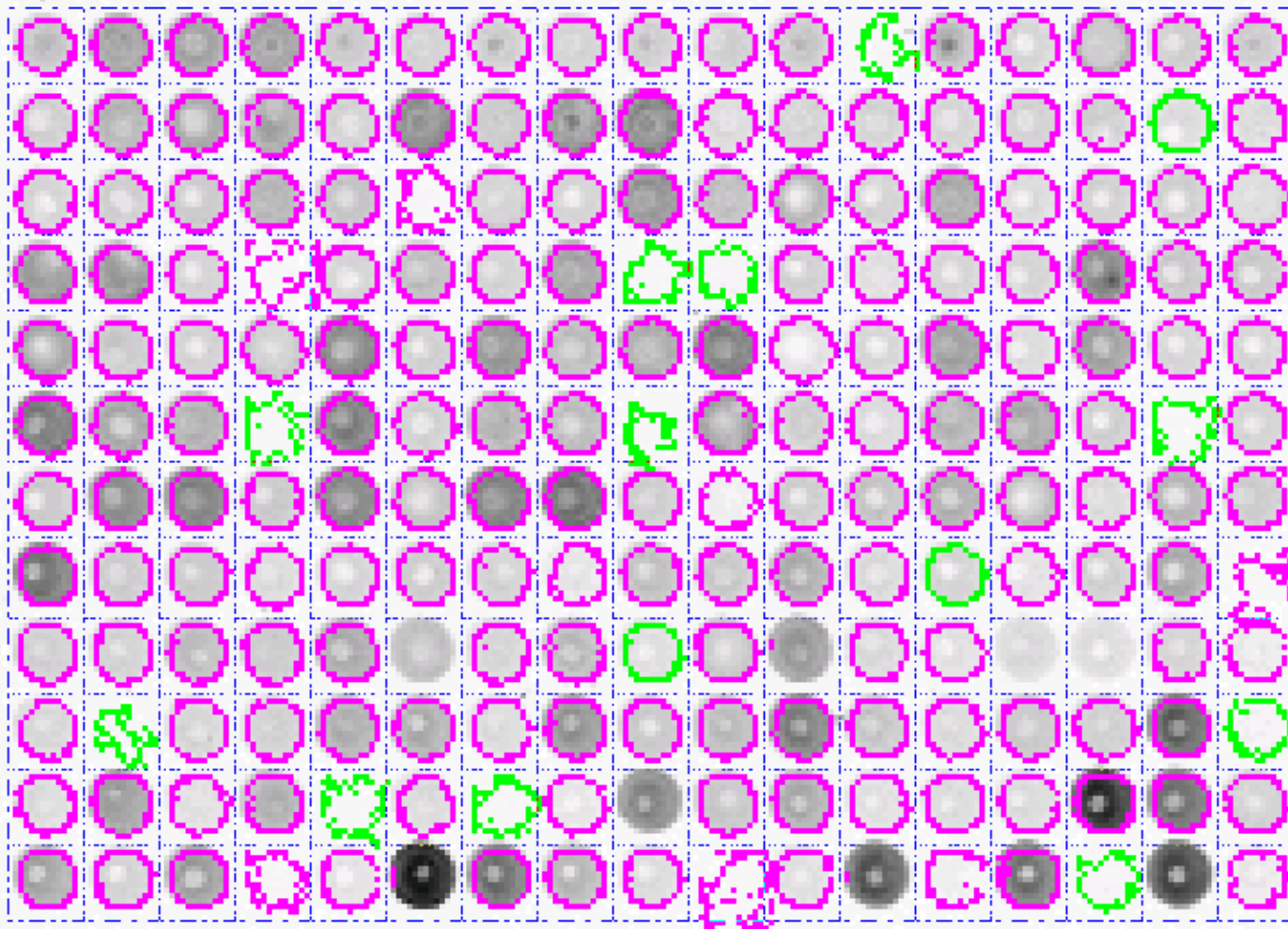
## Analysis Page Mac view





## Grid is visible

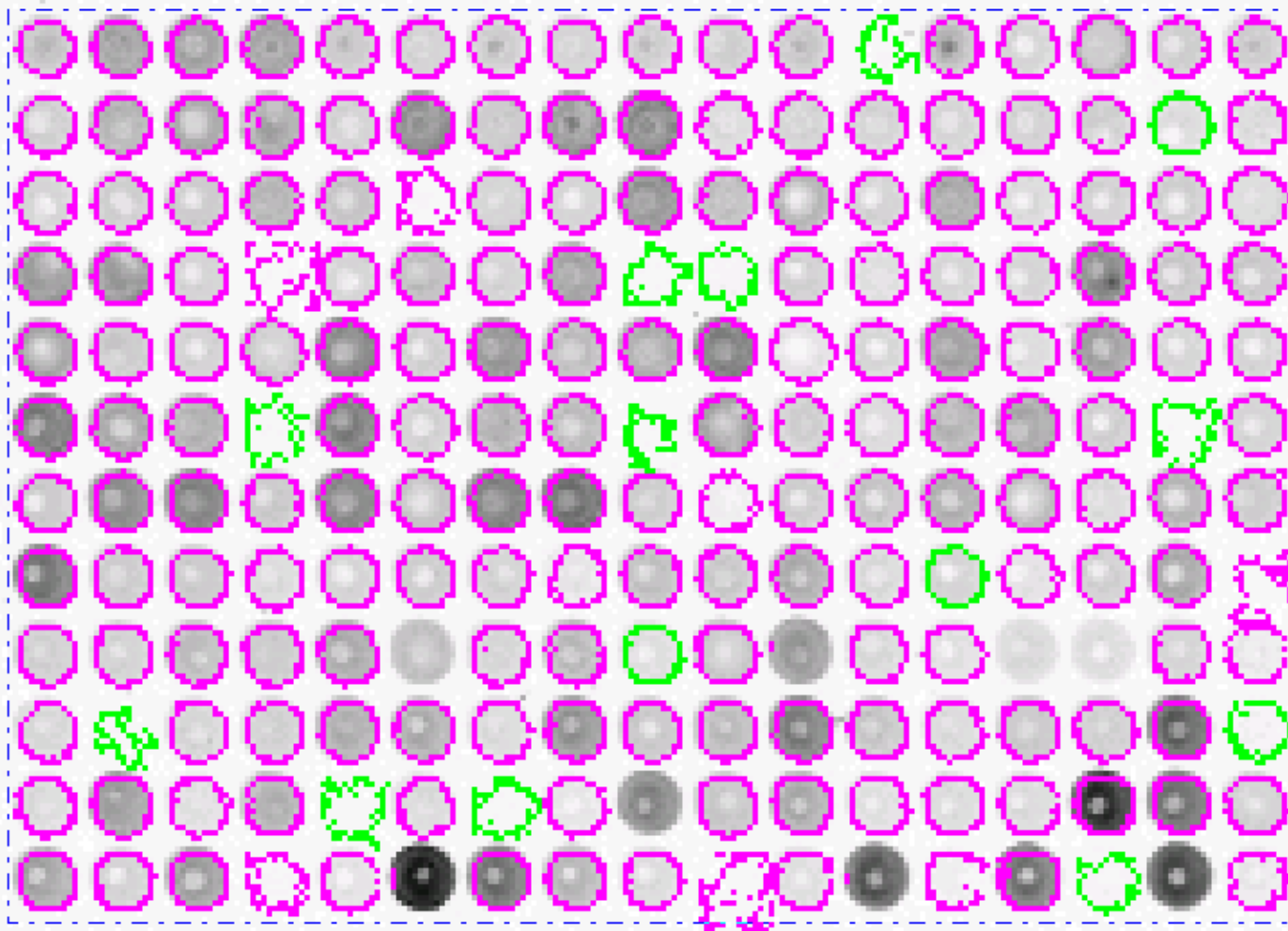
[Back](#)

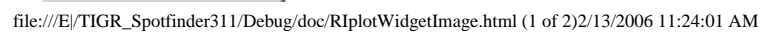




## Grid is invisible

[Back](#)

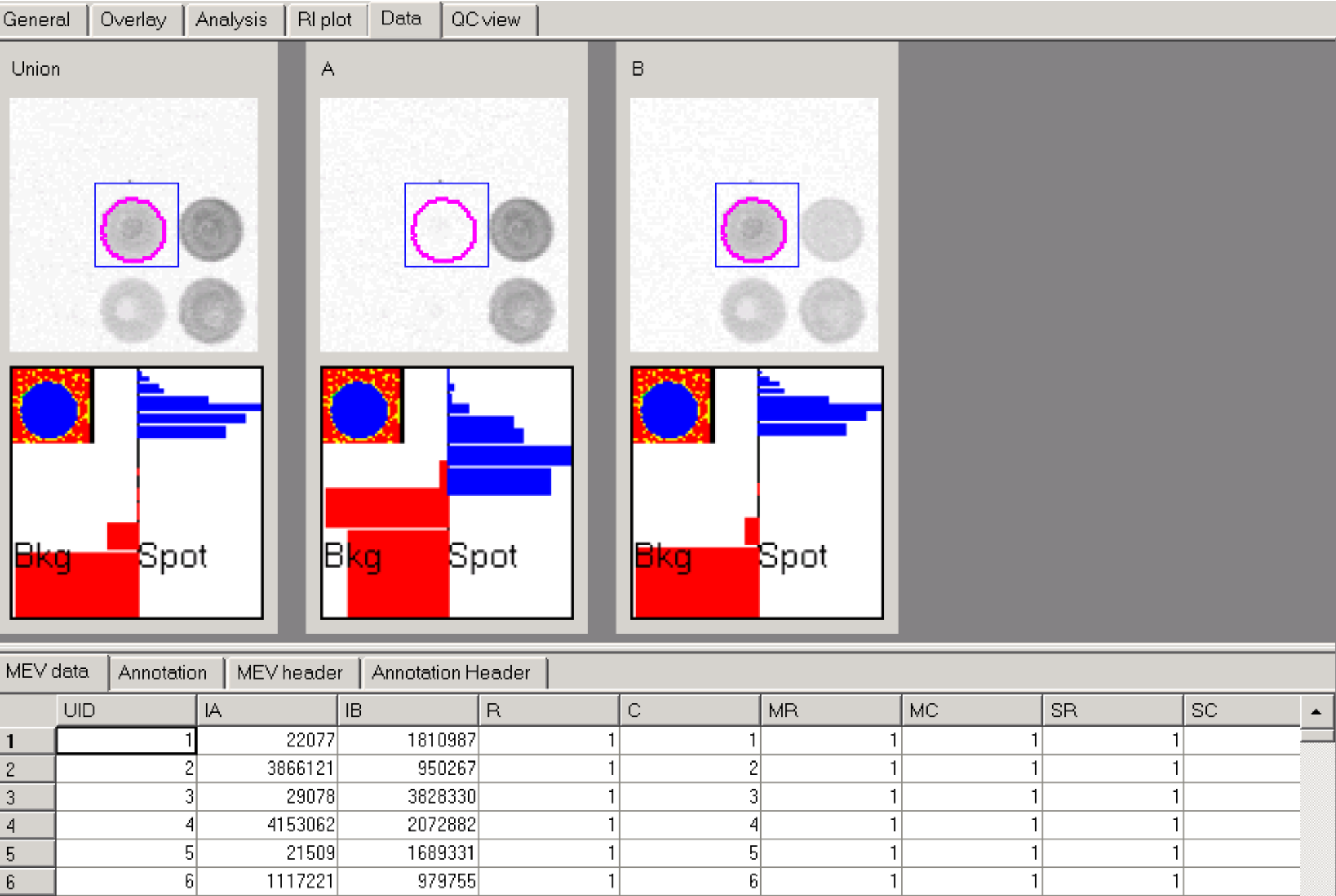






Data page view

[Back](#)



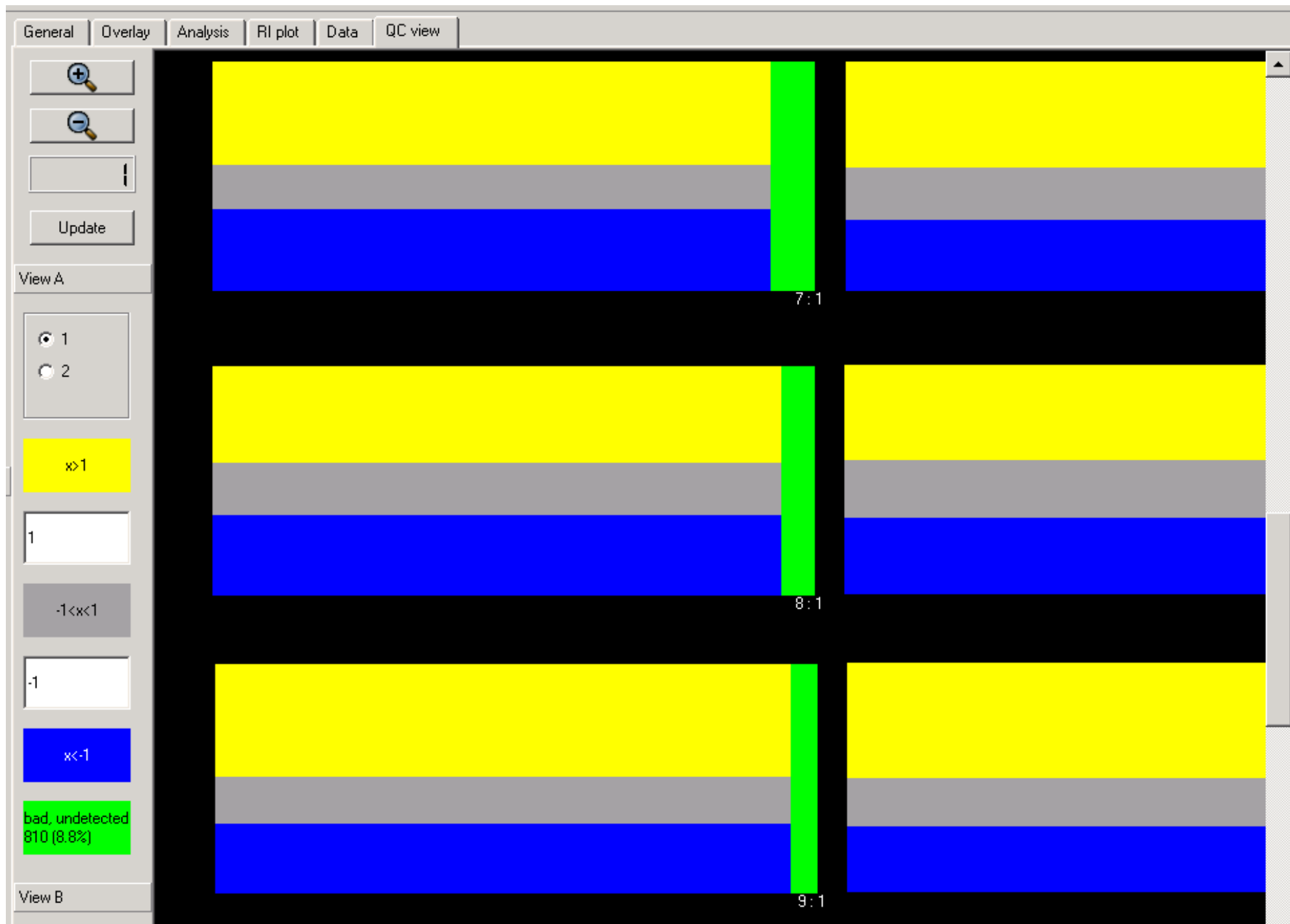


7	7	18620	1431351	1	7	1	1	1		
8	8	32292	1076267	1	8	1	1	1		
9	9	20899	1293354	1	9	1	1	1		
◀										▶

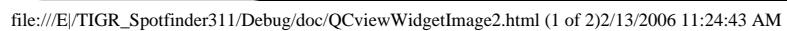
## Data Page on MacOSX



12	12	97101	46523	1	12	1	1	1	1	12	C	

[Back](#)







# QC View Plot

