MAIA - MicroArray Image Analysis
Version 2.5 (4/12/2005)
User Manual

MAIA download page: http://bioinfo.curie.fr/projects/maia/

Author(s): Eugene Novikov (Institut Curie)
E-mail: Eugene.Novikov@curie.fr

System Requirements

MAIA is written in Java (interface) and C++ (algorithms). It runs on Windows platforms 95/98/Me/NT/2000/XP (may be used under Unix after recompiling C++ code) and needs the Java Runtime Environment (JRE) to be installed:
J2SE v 1.4.2_05 JRE (http://java.sun.com/j2se/1.4.2/download.html; http://www.java.com/en/download/)

Software has been tested on the following systems:
(I) Pentium® 4 CPU 3.00GHz and 1 GB of RAM
(II) Intel® Mobile Celeron® CPU 2.00GHz and 256 MB of RAM
No reasons why it should not work with the other configurations.

Time of processing:

Size of an image pair (Cy3/Cy5)	~4MB (~7300 spots)	~40MB (~10000 spots)
System I	~12 sec	~40 sec
System II	~25 sec	For the images of that size, 512MB of RAM is recommended. With 256 RAM, they still can be processed, but it goes slowly due to intensive swapping with hard drive.

Installation

MAIA can be downloaded from the MAIA download page http://bioinfo.curie.fr/projects/maia/

Click MAIA Setup 2.5.exe to start the MAIA 2.5 installer and follow the instructions*.

MAIA 2.5 installation creates a “Curie/MAIA” folder in the list of Programs of the Windows Start menu. This new folder contains the following entries:

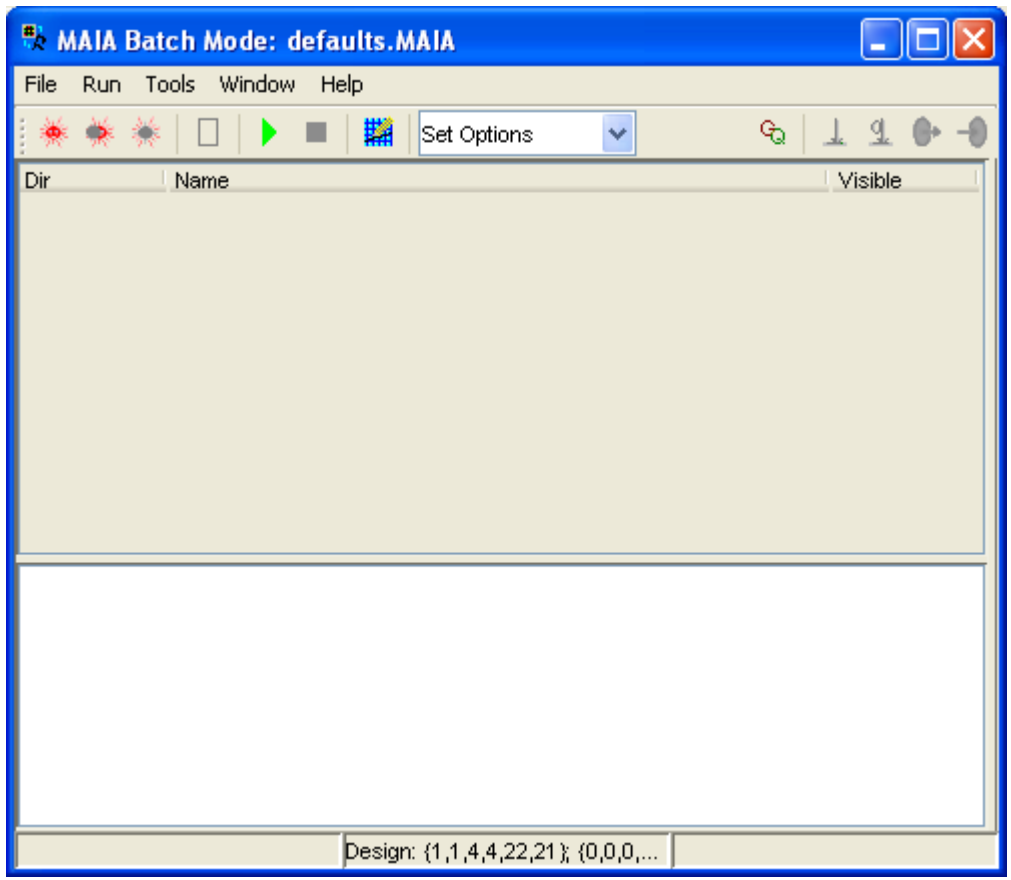
- MAIA starts Microarray image analysis software;
- User Manual is a user manual pdf file;
- Uninstall MAIA will remove MAIA from your computer.

Installation procedure may also create a “MAIA” icon on your Desktop.

*) Installation procedure asks about the default size of the JVM (Java Virtual Machine) memory allocation pool. It is recommended to set it as large as possible, but not larger than the amount of available RAM.

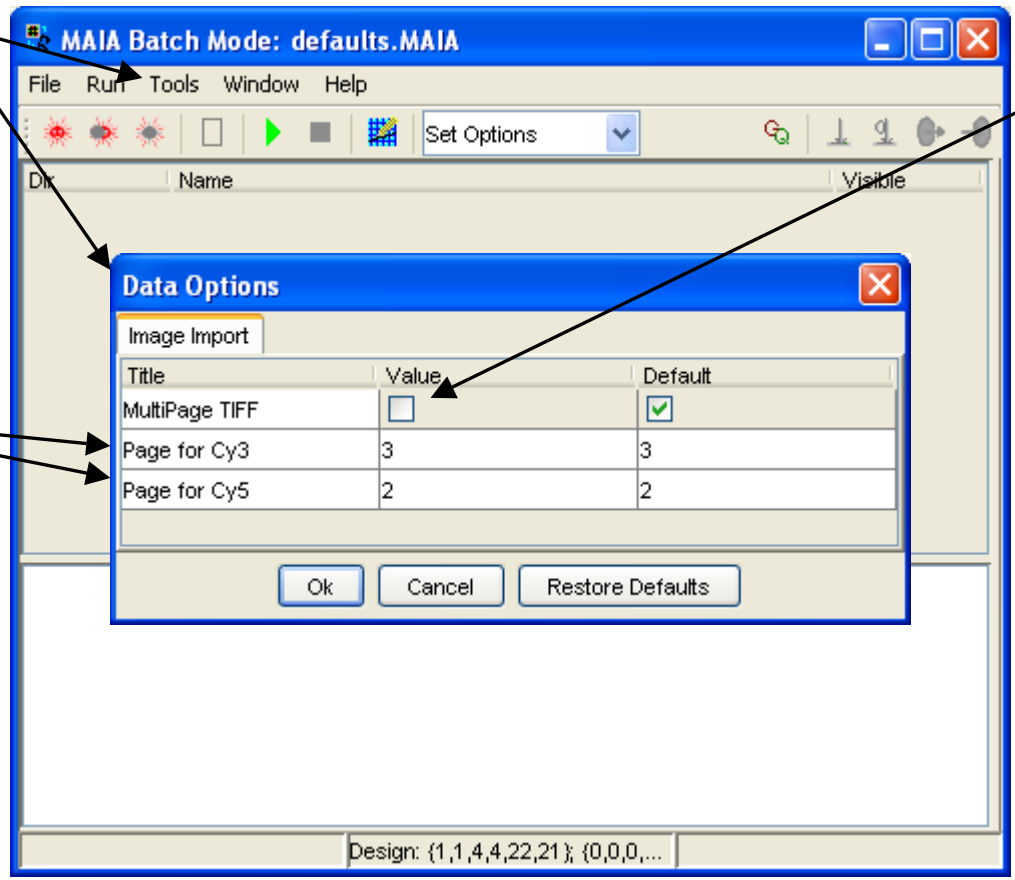
Batch Processing Window

Successful start will bring on the screen the following window:



Data Import Settings

To define the format of the microarray image files select the Menu Item “Tools|Data Options” (Ctrl+D).

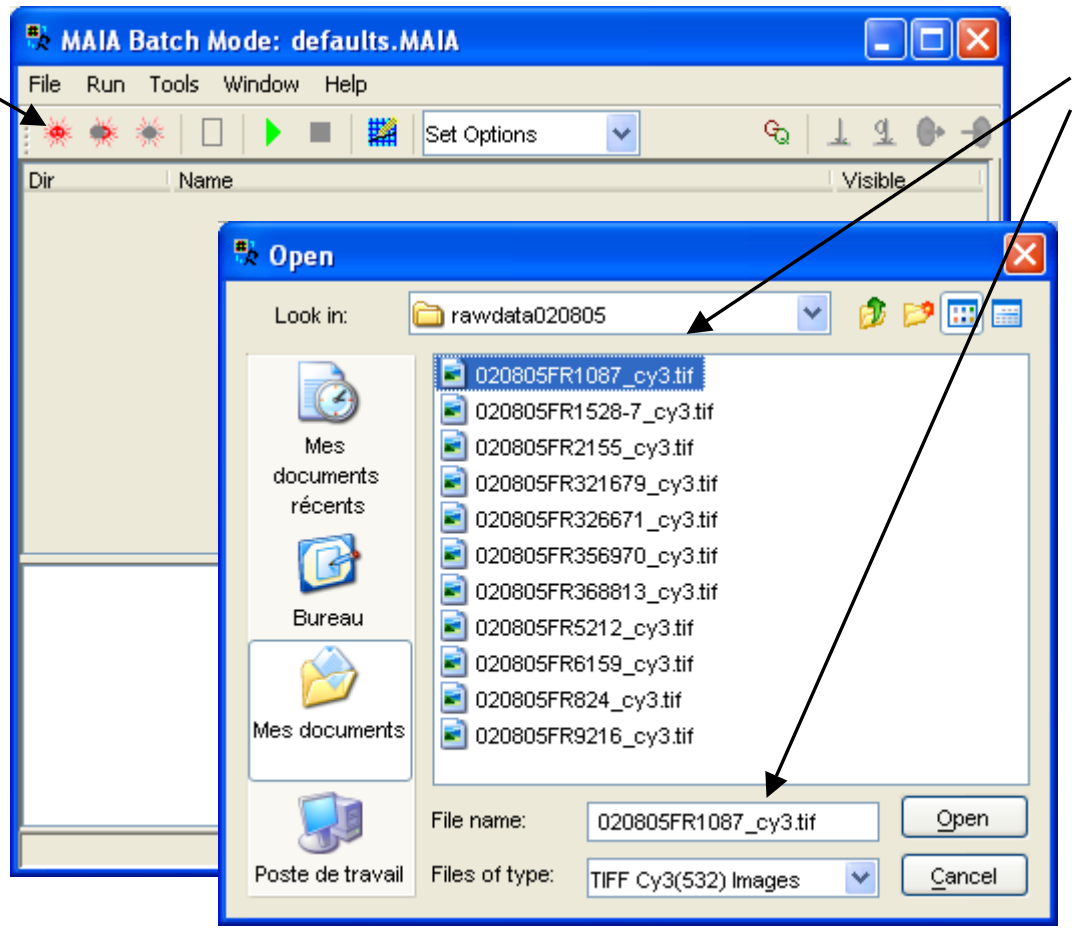


Two options are available:
(i) Cy3 and Cy5 TIFF images are packed into one multi-page TIFF file (checked);
(ii) Cy3 and Cy5 TIFF images are stored in separate files (unchecked).

For multi-page TIFF, specify the pages for the Cy3 and Cy5 images to be analyzed.

File Name Selection

To select microarray images use the Toolbar button “New Experiment” or the Menu Item “File|New Experiment” (Insert).



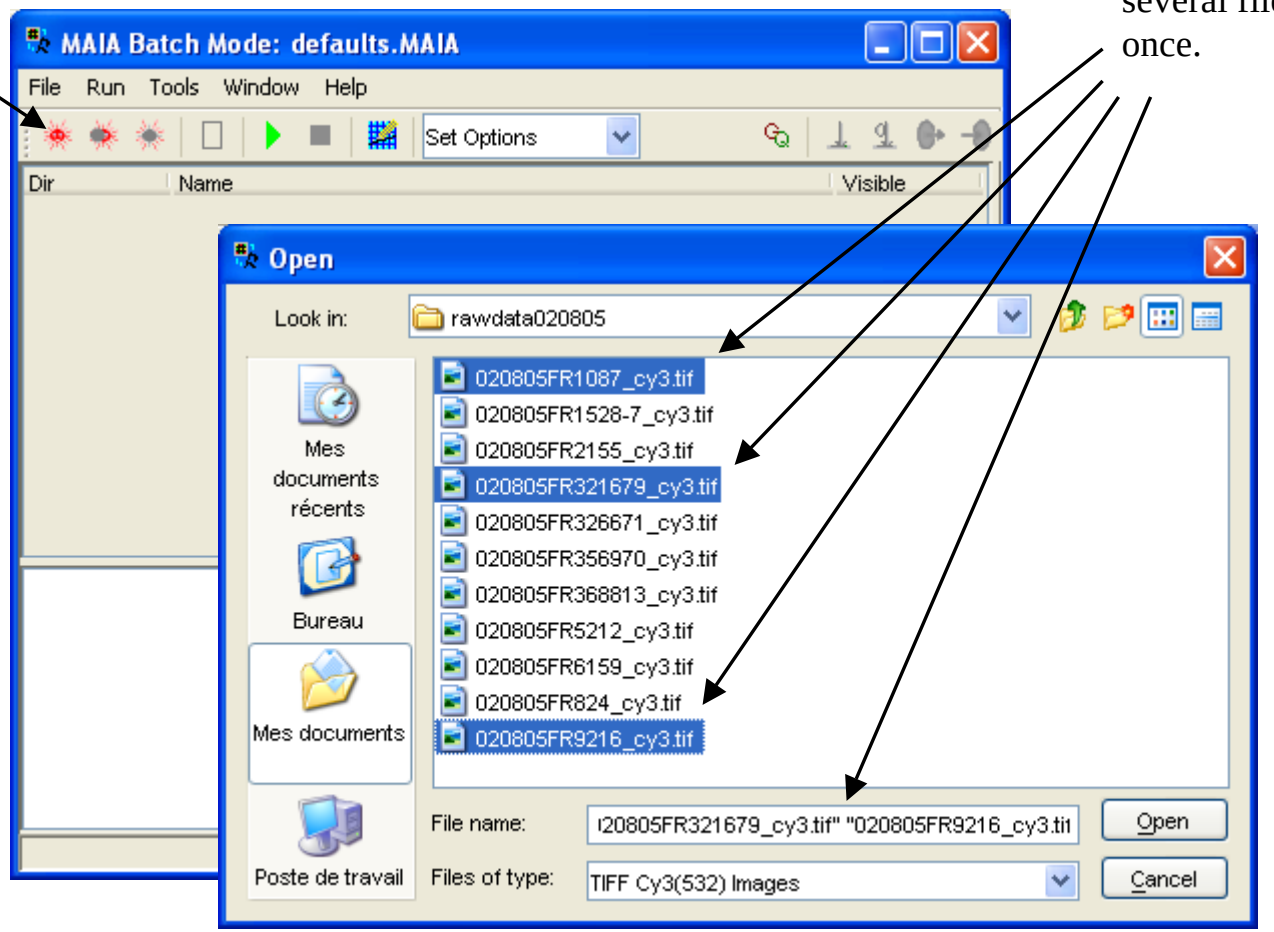
When single-page TIFF files are used, File Browser shows up only Cy3 file names. The correspondent Cy5 file name will be downloaded automatically.
In this case filenames for the pair of Cy3 and Cy5 images must differ only by the suffix: “cy3” or “532” for Cy3 images, and “cy5” or “635” for Cy5 images.

For multi-page TIFF, filenames can be arbitrary.

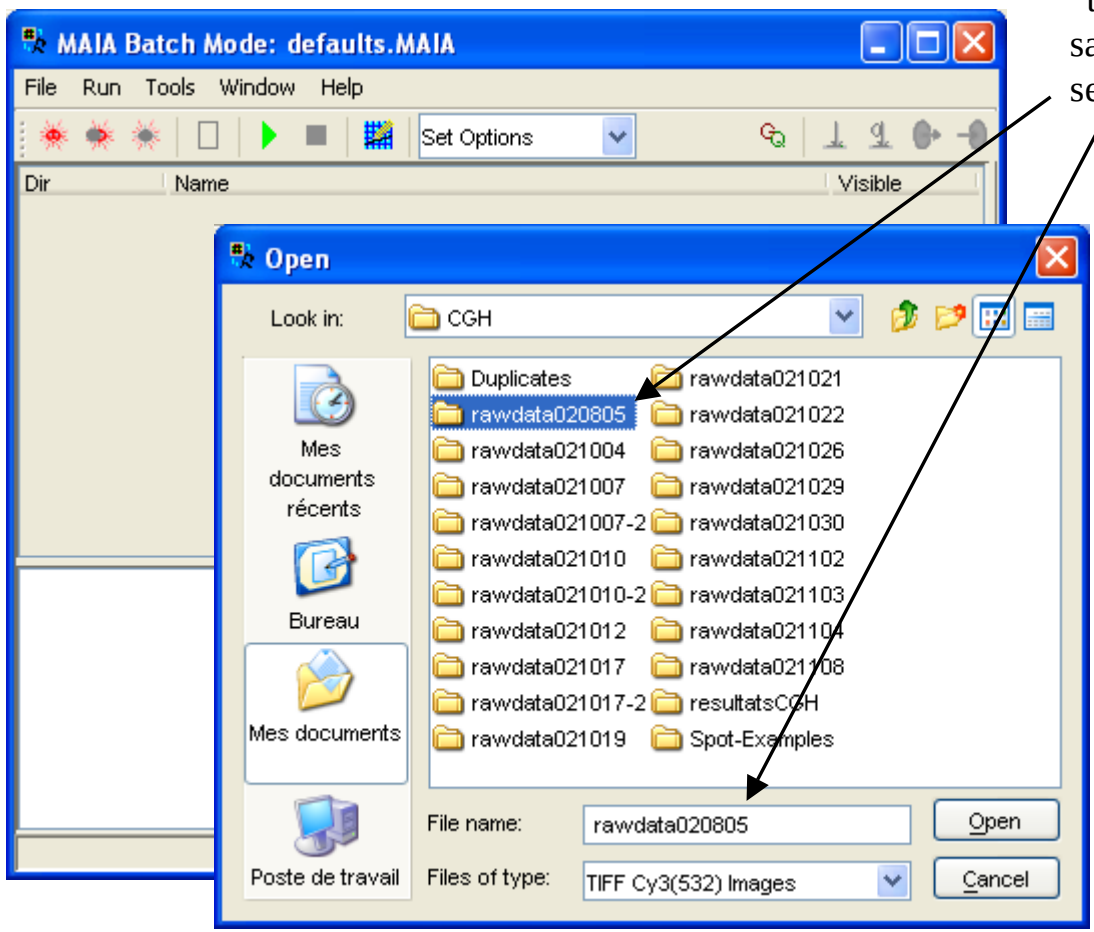
Multiple File Name Selection

Using the Toolbar button “New Experiment” or the Menu Item “File|New Experiment” (Insert) more files can be added into the table.

One can select several filenames at once.



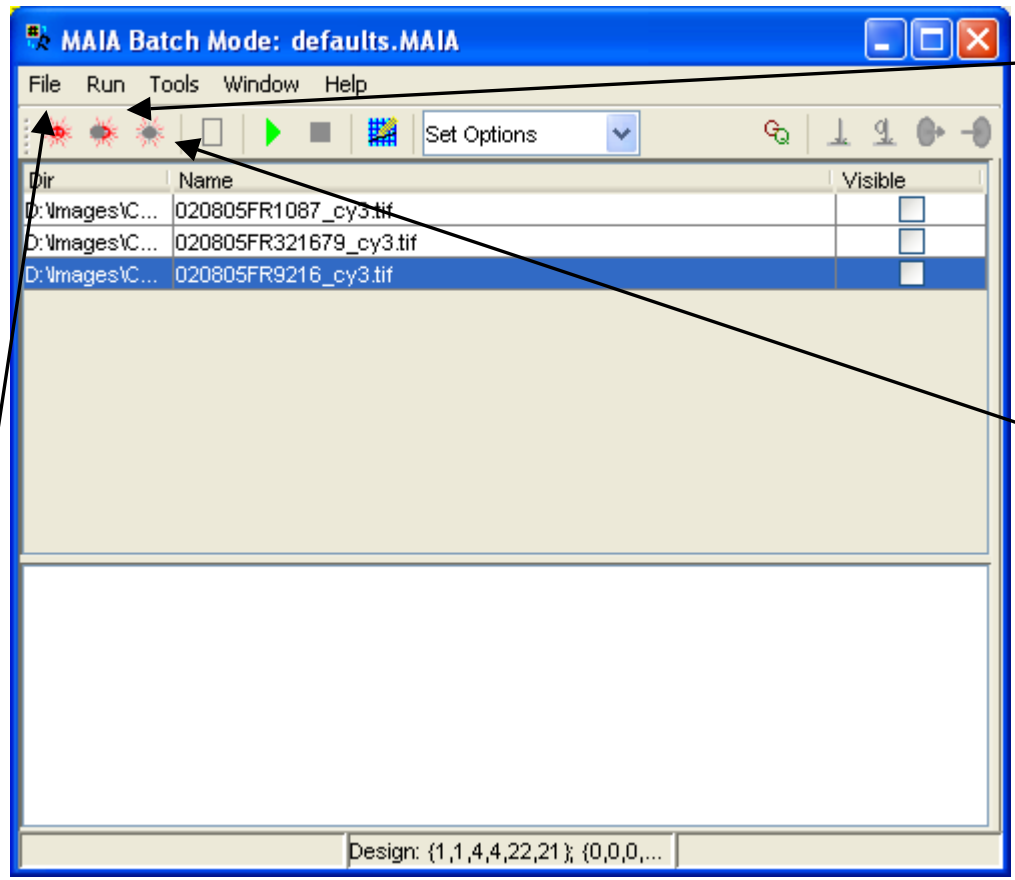
Directory Selection



All TIFF files (extension "tif"/"tiff") from the same directory can be selected.

Batch of File Names

The selected filenames appear in the table.

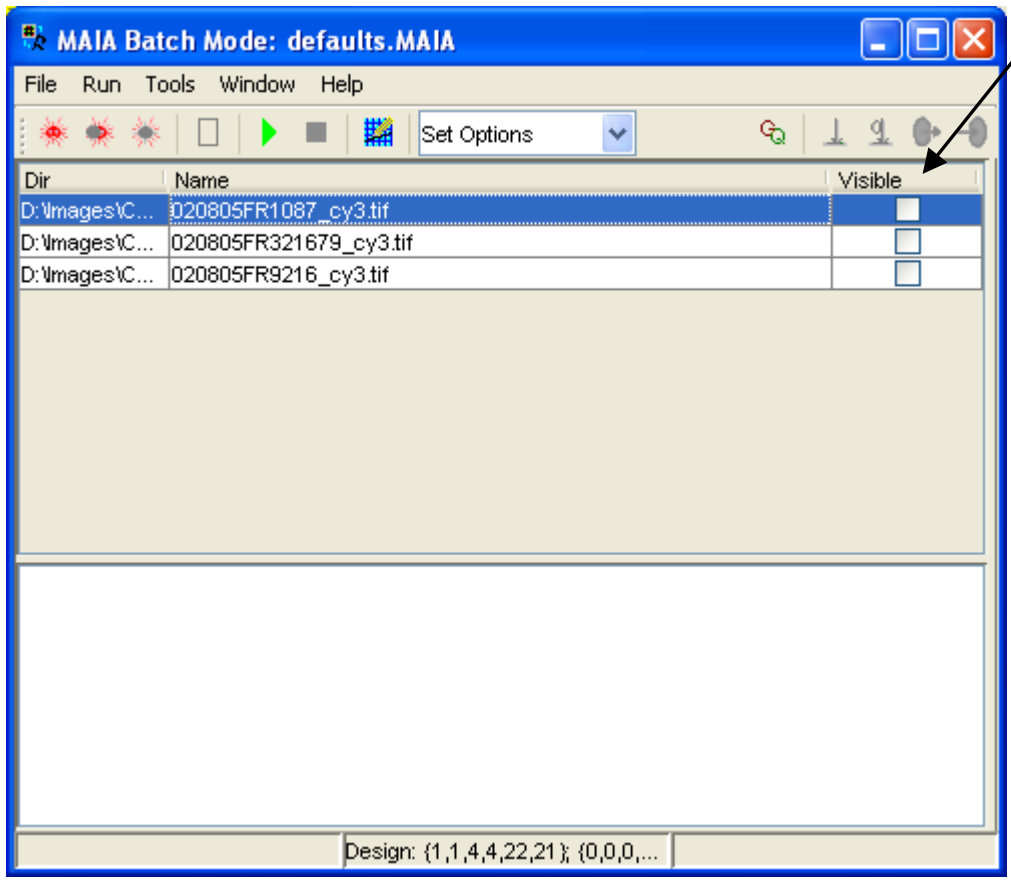


To remove filenames from the batch one may use the Toolbar button "Remove Experiment" or the Menu Item "File|Remove Experiment" (Delete).

The whole batch (a list of files and accompanying options) can be saved on the disk (using the Menu Item "File|Save Group …" (Ctrl+S)) to be able to restore it (using the Menu Item "File|Load Group …" (Ctrl+O)) in the future to reanalyze the batch.

The toolbar button "Remove All Experiments" or the Menu Item "File|Remove All Experiments" will remove all filenames from the batch (Ctrl+Delete).

Ready for Analysis



To open (download) an image check the "Visible" field of the table.

Main Processing Window

Three tabs are created: Ratio Image, Cy3 and Cy5 channel images.

Another pair of images (Cy3/Cy5) can be downloaded using the "Load Data …" button from the Toolbar or the Menu Item "File|Load|Data …" (Ctrl+O).

For the new images, image file format (i.e. multi-page TIFF versus single-page TIFF) can be changed using the Menu Item "Tools|Data Options" (Ctrl+D).

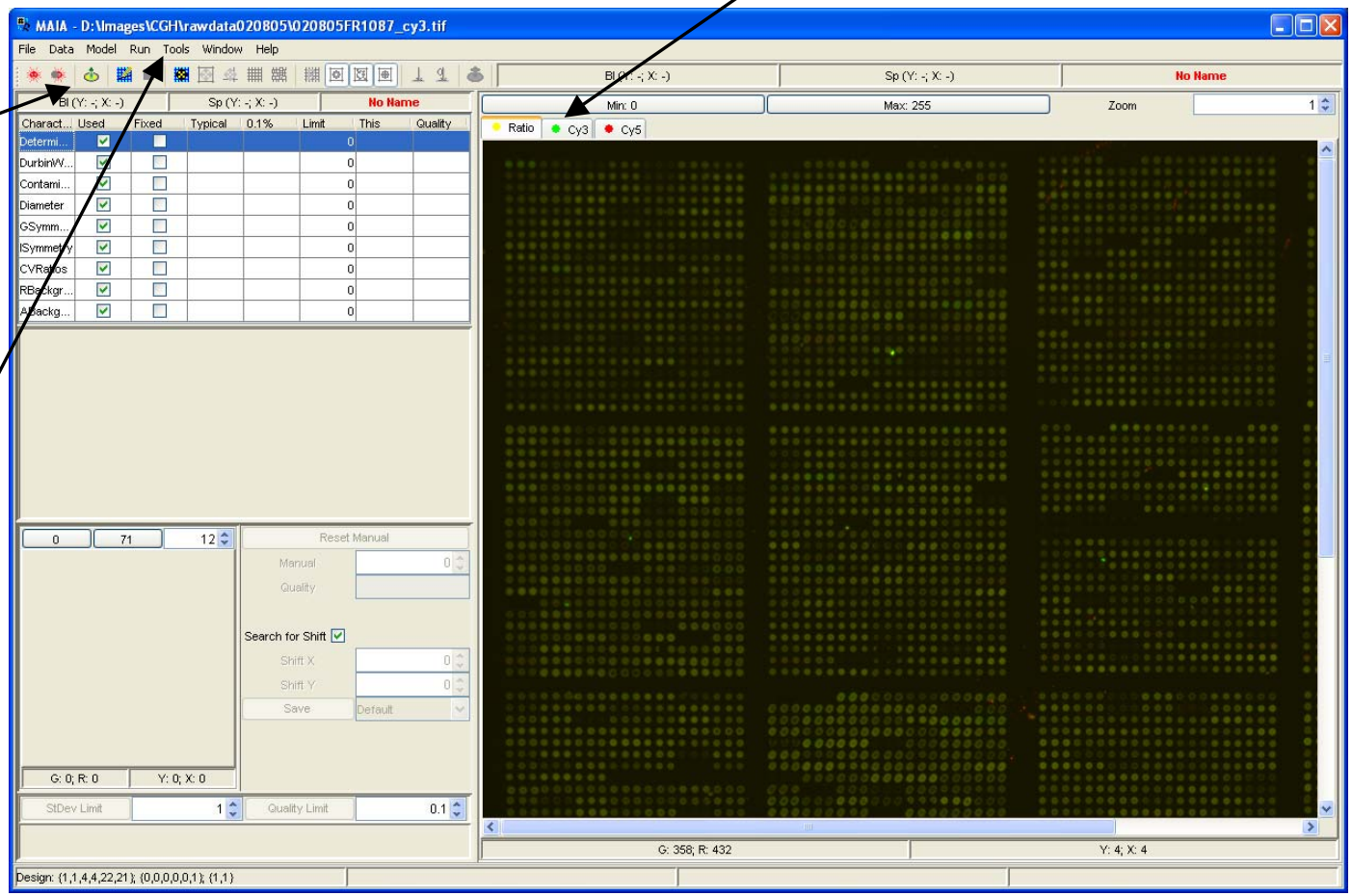
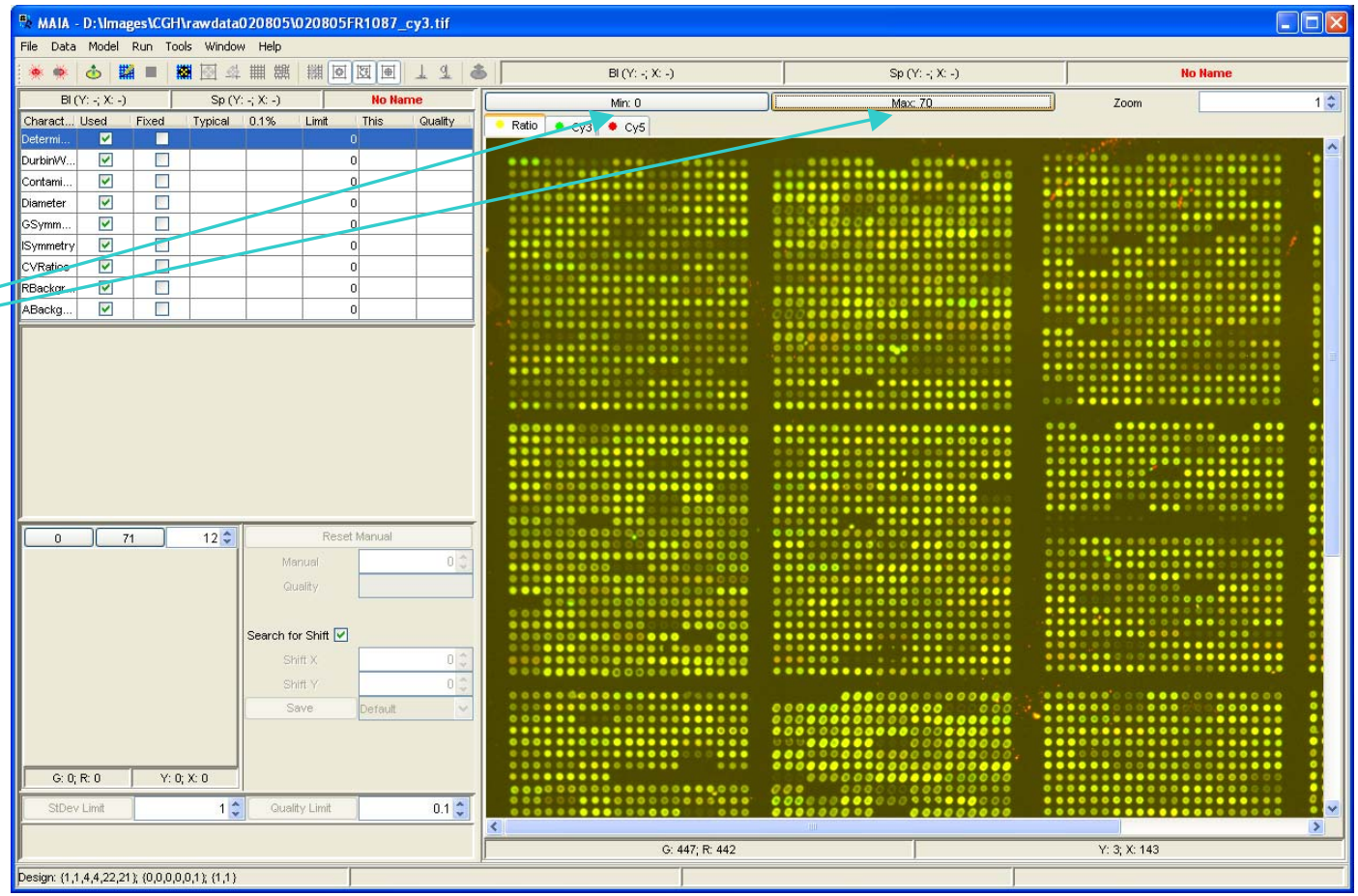


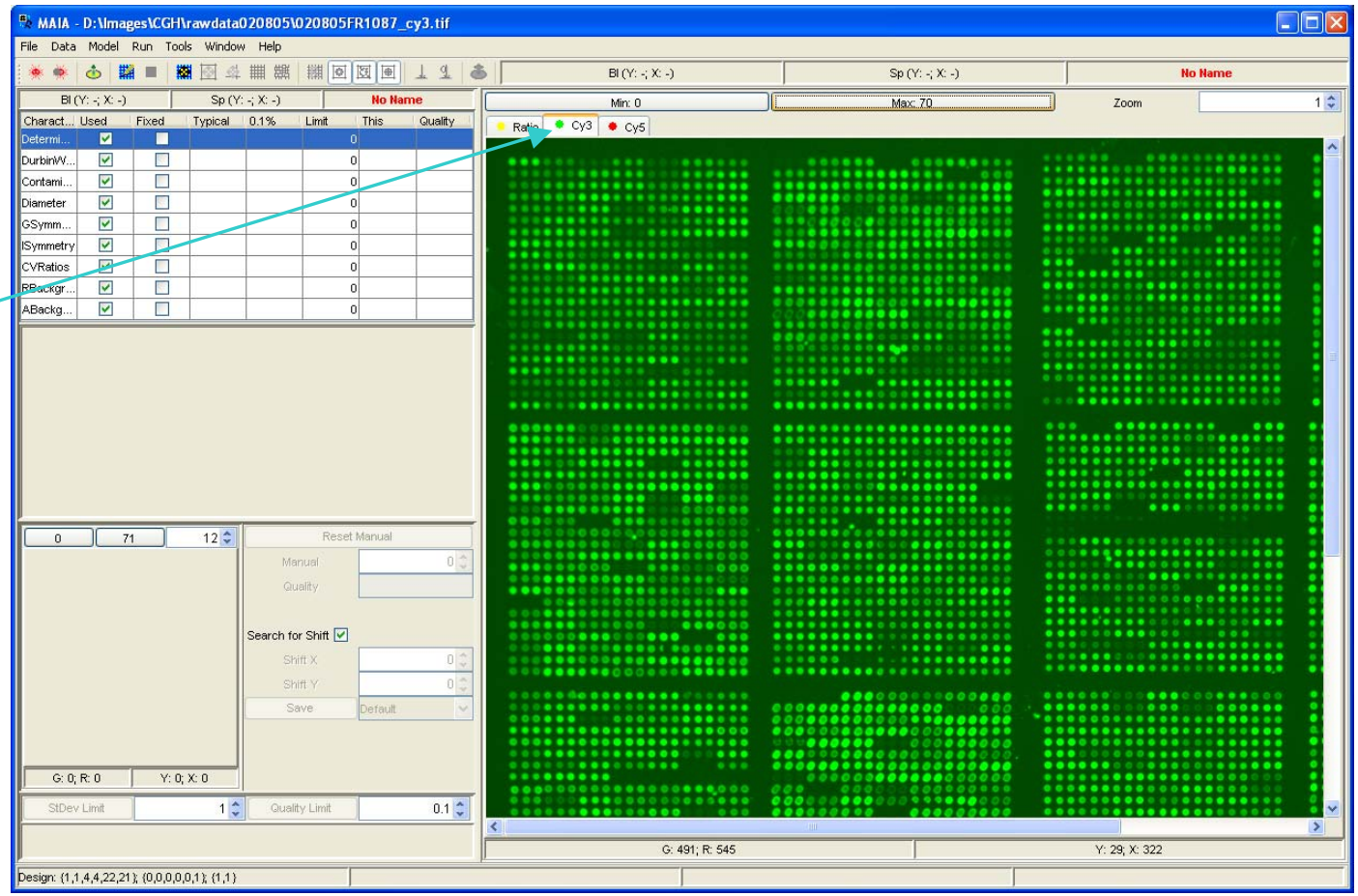
Image Visualization Settings

“Min” and “Max” controls can be used to adjust brightness and contrast of the images.



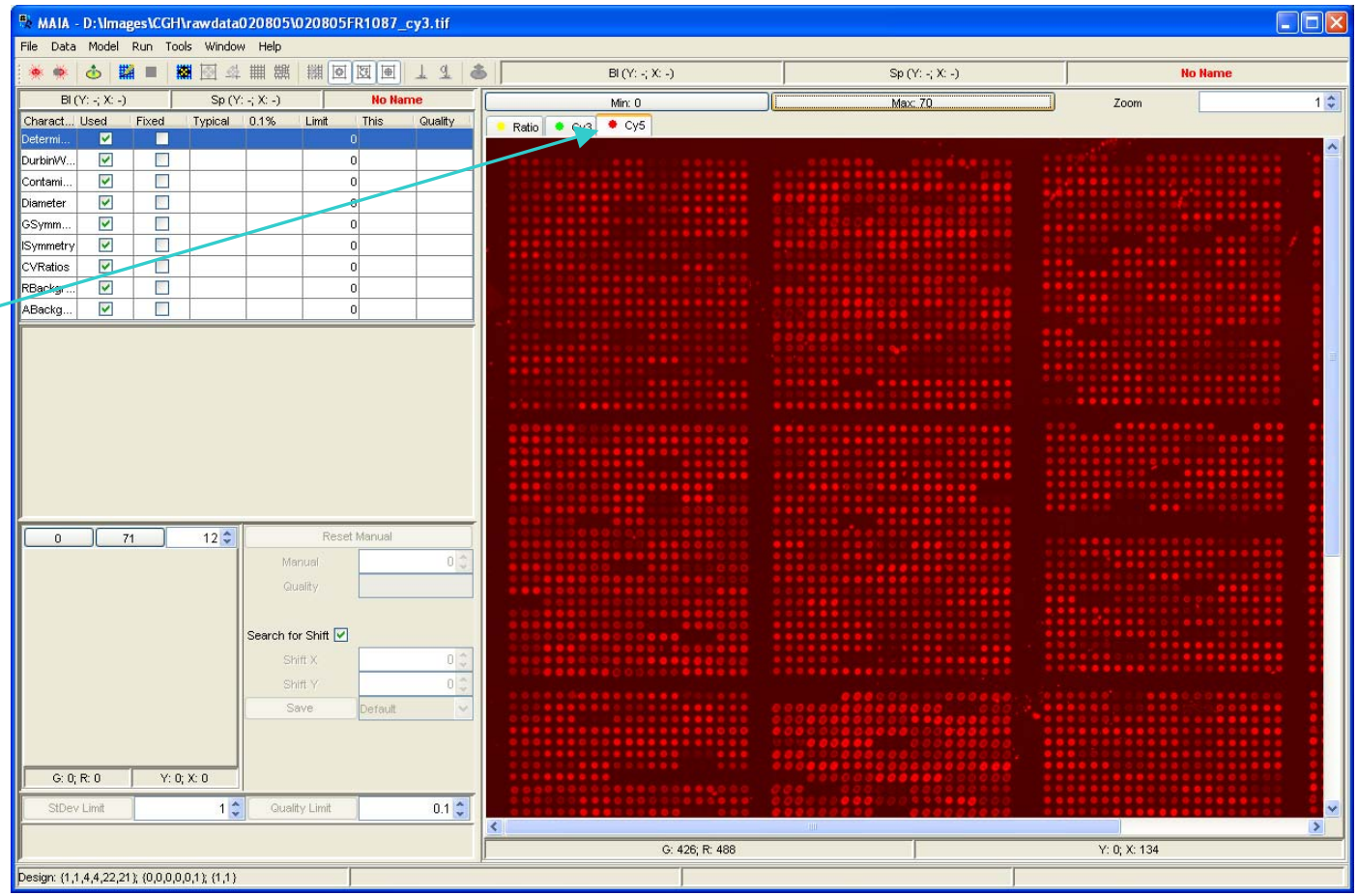
Green Channel

Select the green-dot (Cy3) tab to visualize the image colored in green.



Red Channel

Select the red-dot (Cy5) tab to visualize the image colored in red.



Color Swap

By default, green color is used for the Cy3 image and red color – for the Cy5 image. This assignment can be inverted by the Menu Item “File|Swap Colors”.

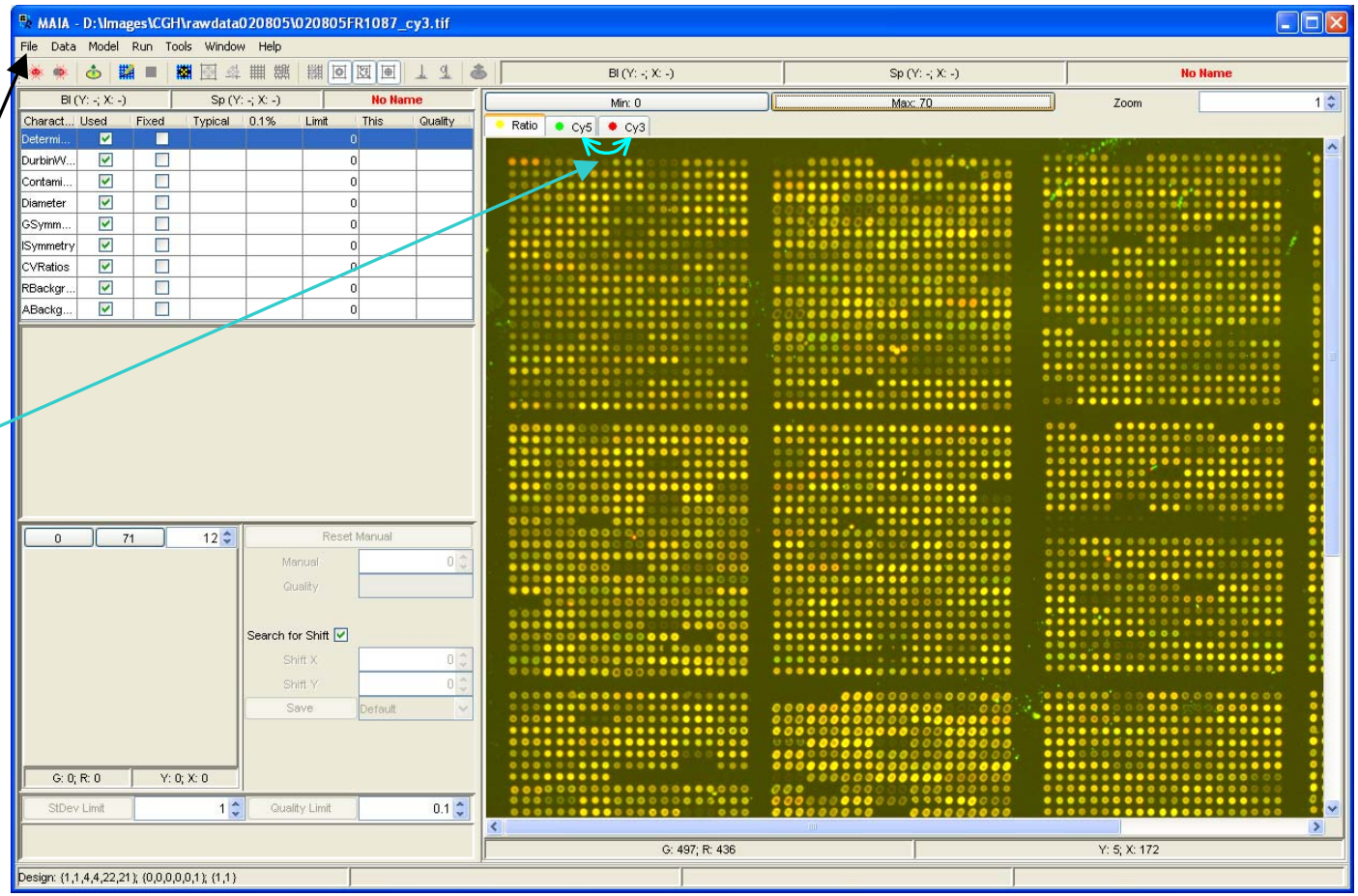
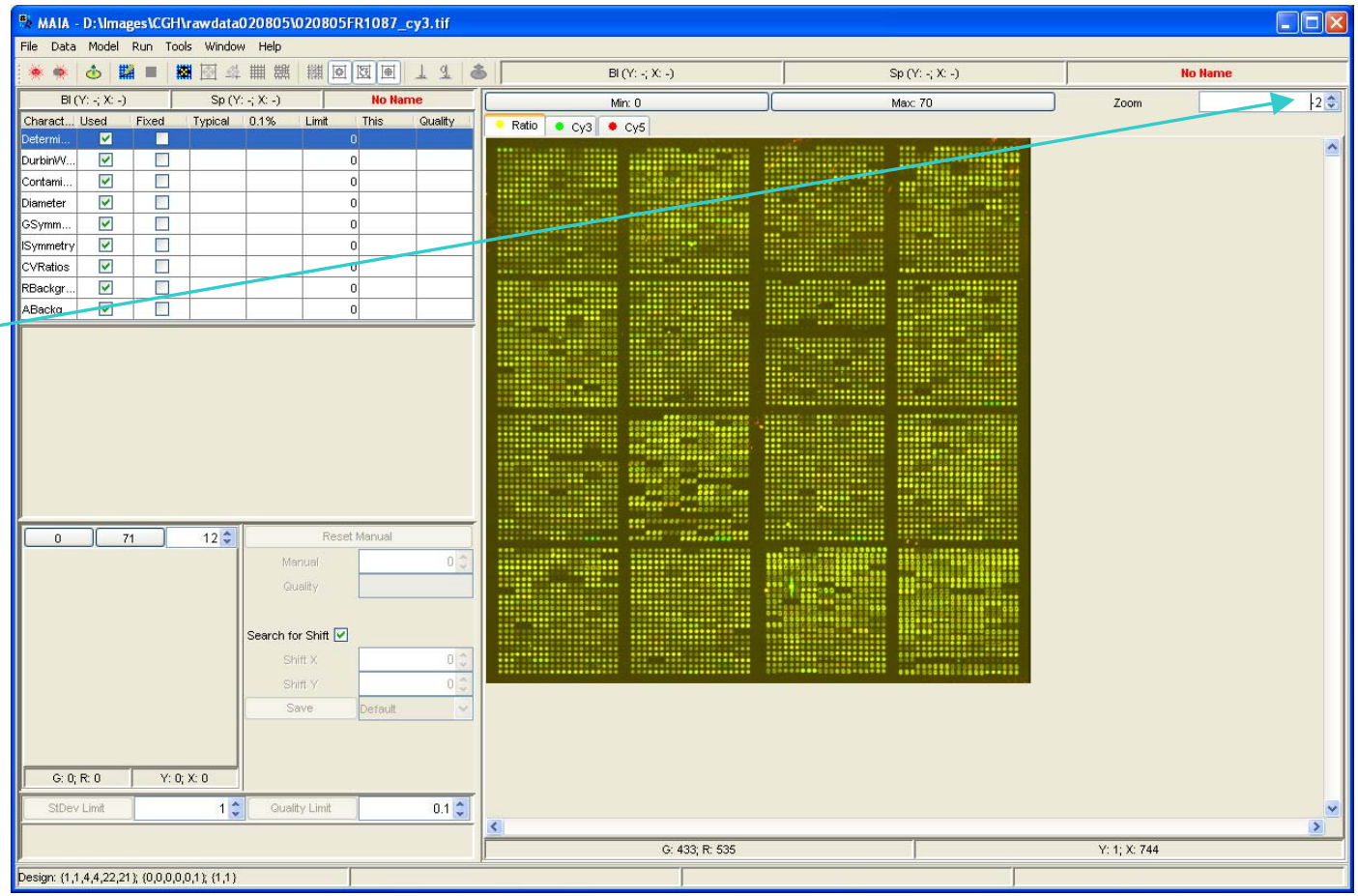


Image Zoom

Image zoom can be changed using either the "Zoom" spinner box or the mouse wheel.

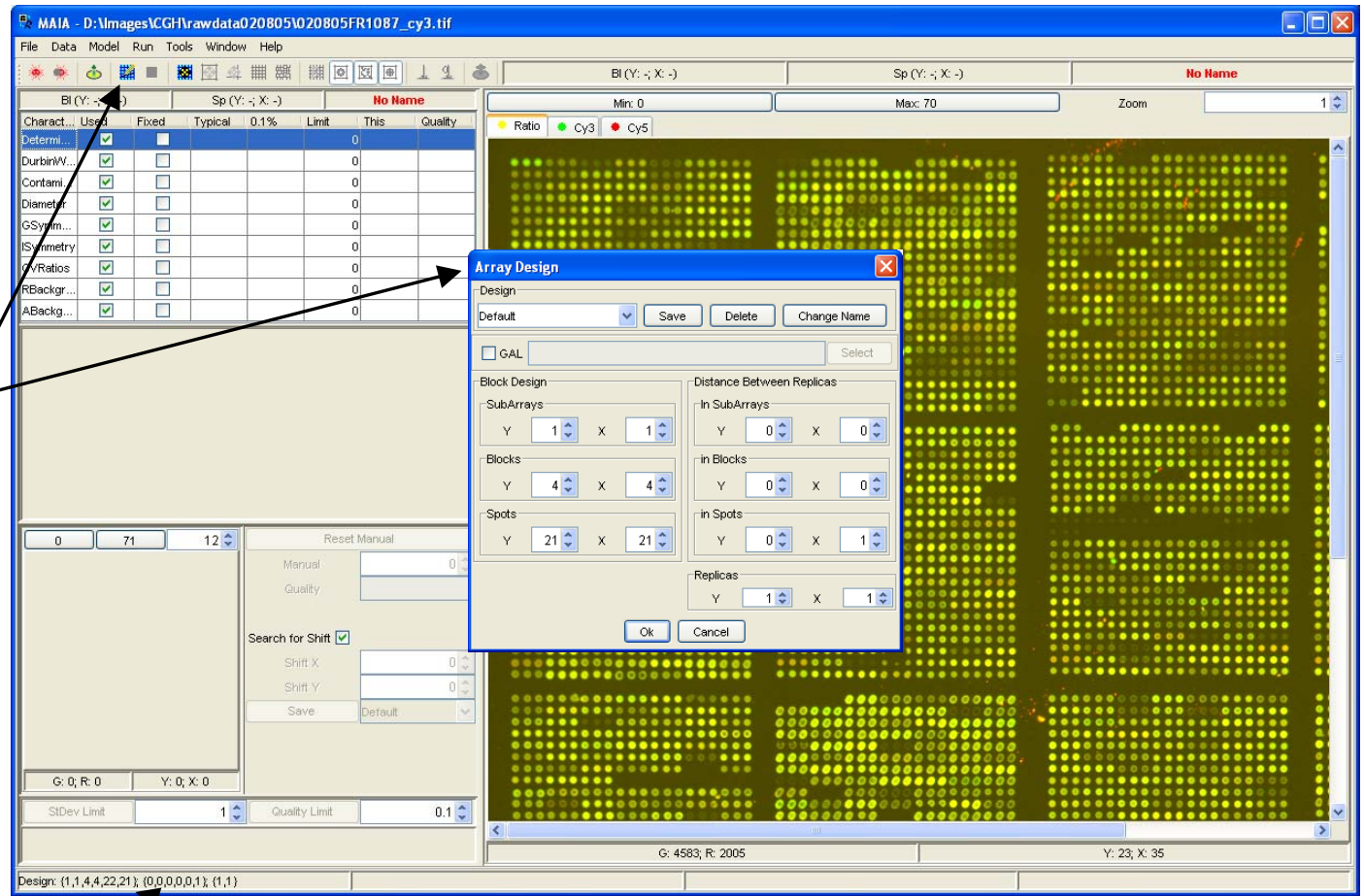
Negative values of the zoom indicate contraction; positive values indicate stretching. Original image is obtained with either 1 or -1 zoom. (Zoom does not influence the analysis.)



Array Design

To start image processing, array design has to be properly defined: use the “Array Design” button from the Toolbar or select the Menu Item “Tools|Array Design” (F2).

See next page for details.

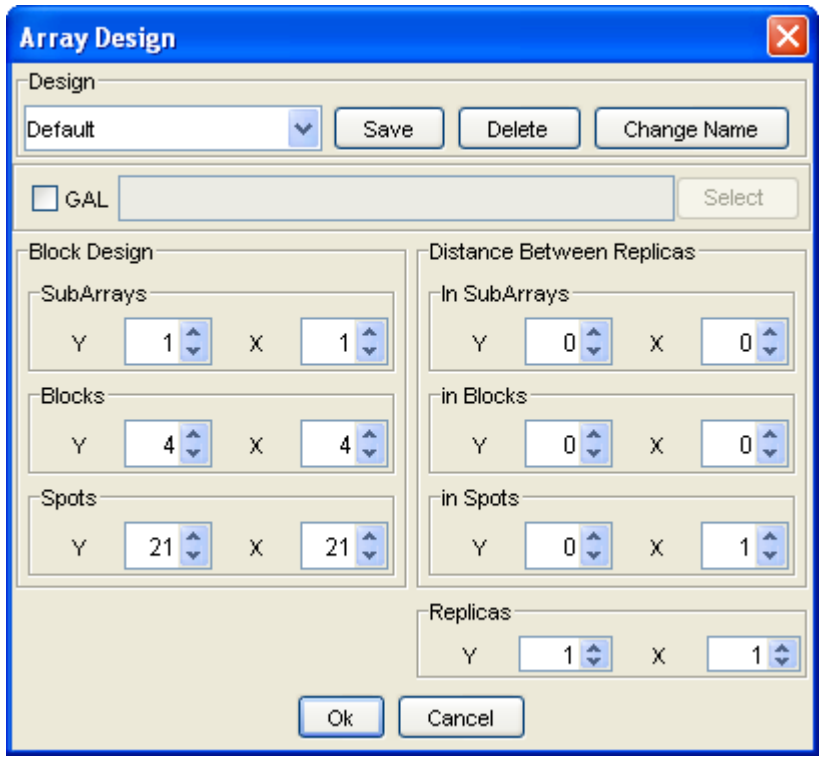


Description of the currently used Array Design.

Array Design in Detail

One may use several microarray designs under different names to be able to switch quickly from one design to another one.

Amount of sub-arrays, blocks (per sub-array) and spots (per block) in Y and X directions of the array.



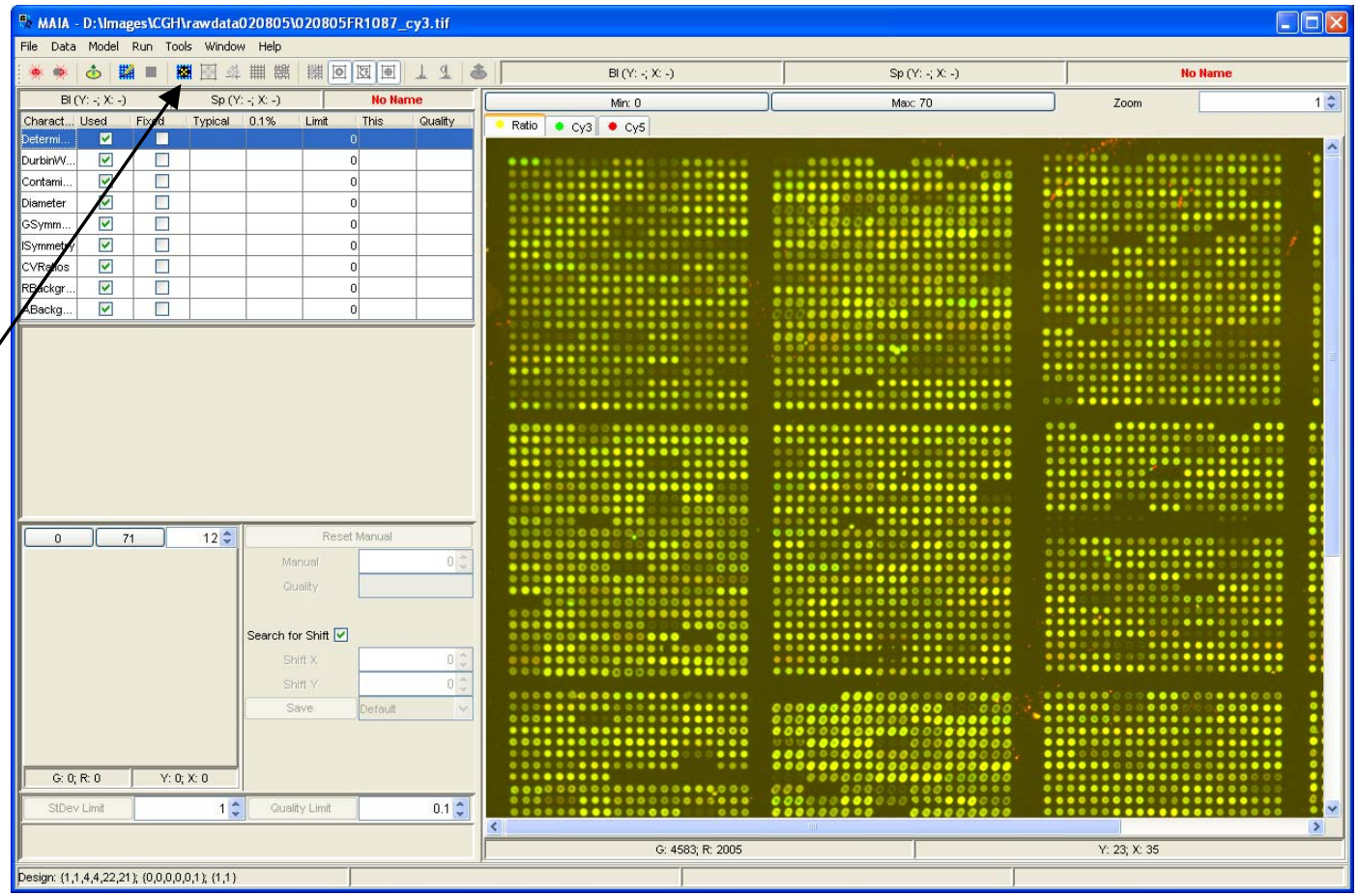
One may specify optional GAL file (Axon Instruments, Inc. 2003. GenePix Pro 5.0. http://www.axon.com, User's Guide and Tutorial). The purpose of the GAL file in this program is to provide clone names for the spots to be displayed and exported.

Relative coordinates of the replicated spots: it defines the position of the replicated spot with respect to the current one.

Amount of the replicated spots in the Y and X directions.

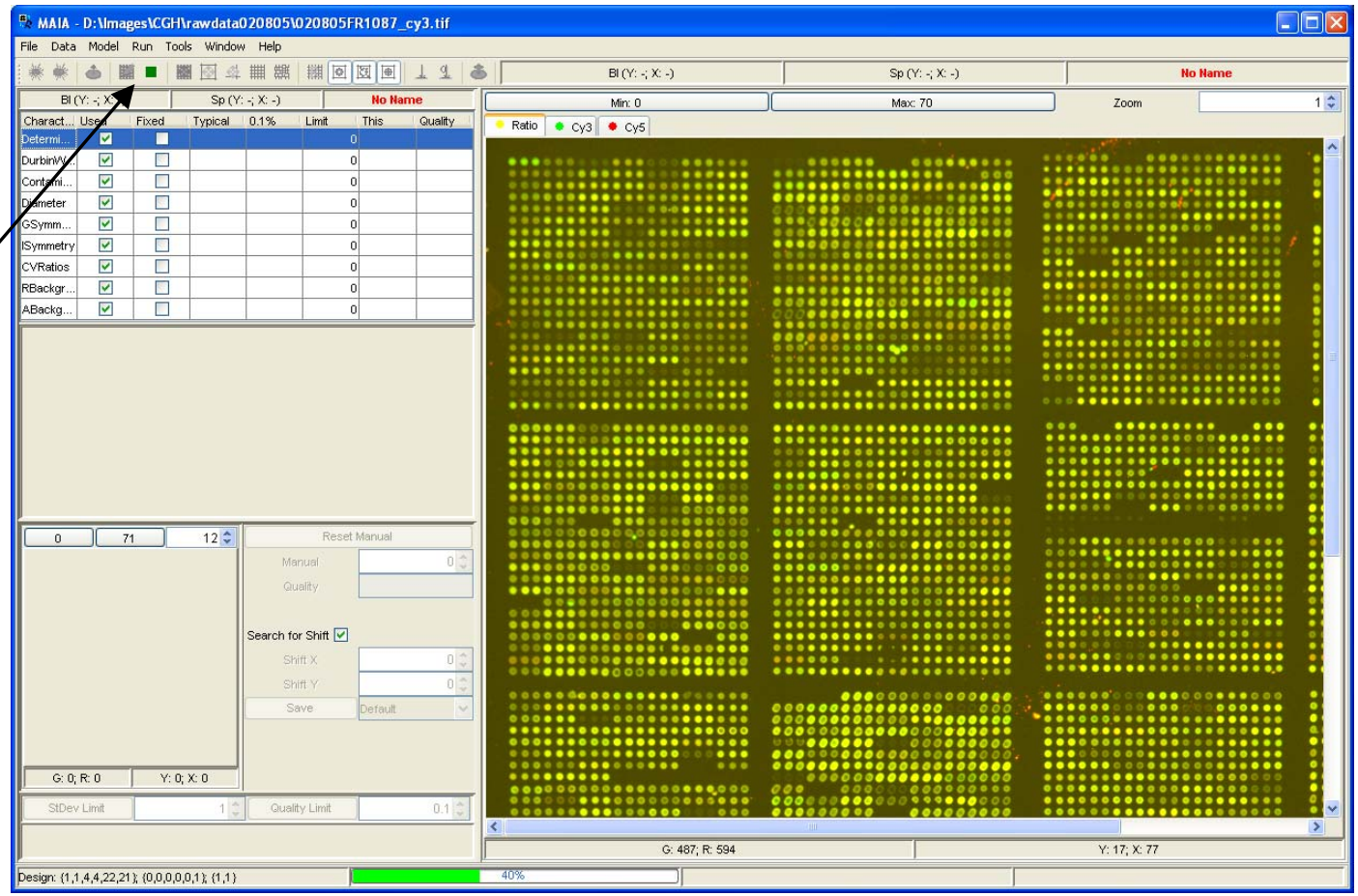
Spot Localization

To start Spot Localization (or grid finding) use the "Spot Localization" button from the Toolbar or select the Menu Item "Run|Spot Localization" (F3).



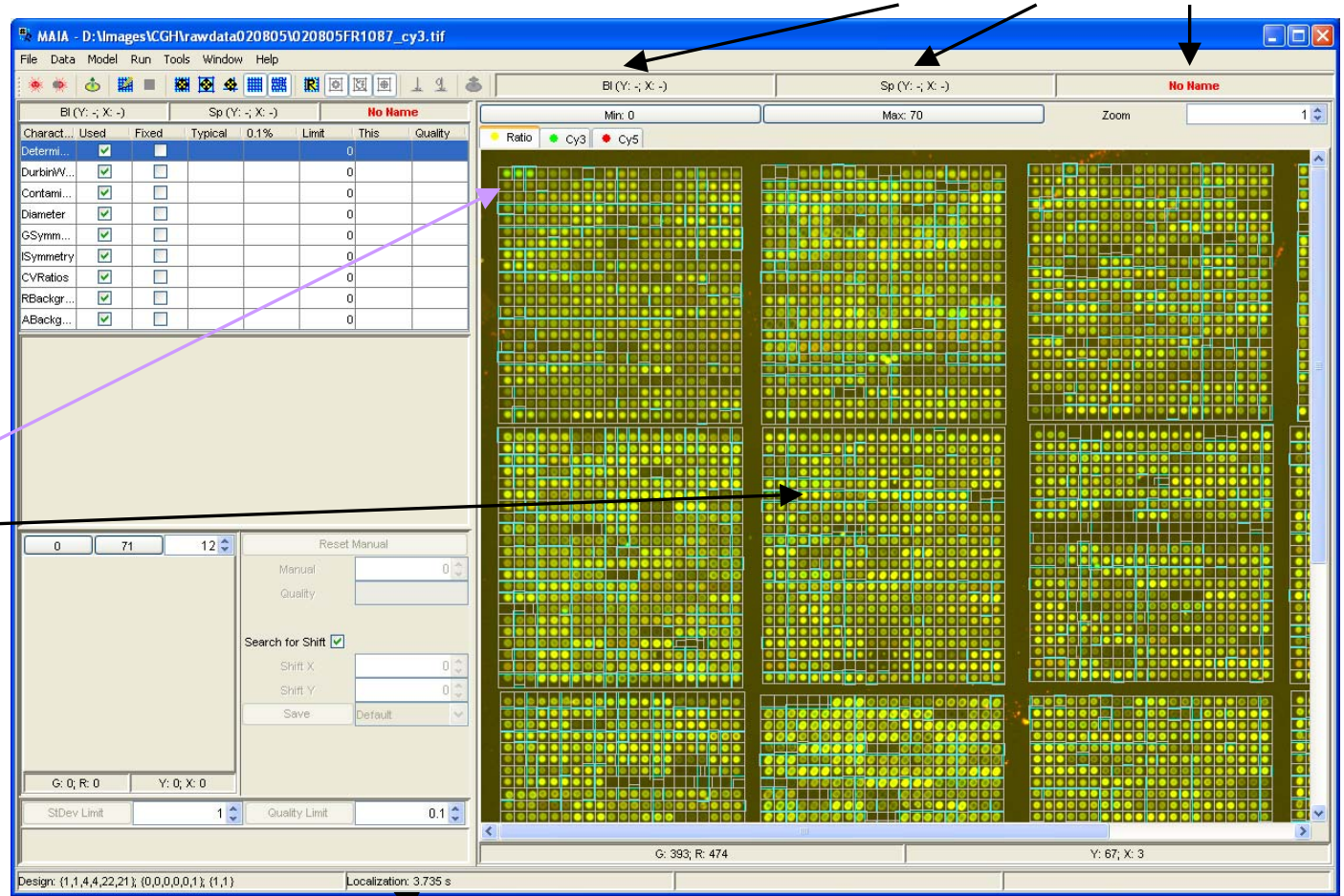
Terminate Processing

Any processing can be stopped by pressing the "Stop Button" on the Toolbar.



Spot Localization Output

"Under-mouse" coordinates of the block (Bl), spot (Sp) and clone name.



Typical result of the Spot Localization procedure: two grids are imposed over the image:

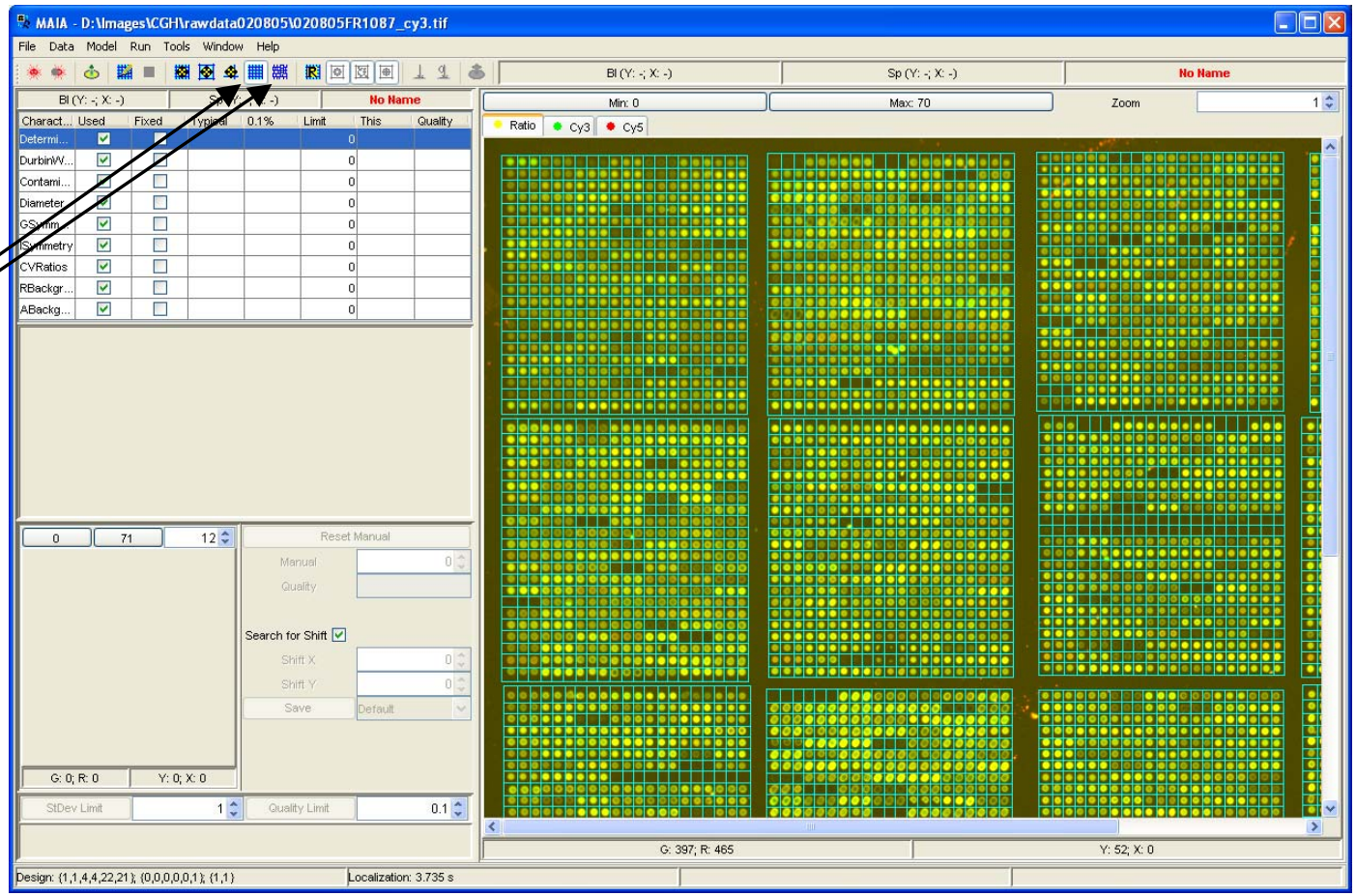
- *Main Grid* is composed of the straight lines separating neighborhood spot rows or columns;
- *Adjusted Grid* is composed of the piecewise lines providing refined borders between the neighborhood spots.

Status of the Spot Localization procedure.

Spot Localization Output: Main Grid

Using the Toolbar buttons "Show/Hide Main grid" or "Show/Hide Adjusted grid" one can mask either of two spot localization grids.

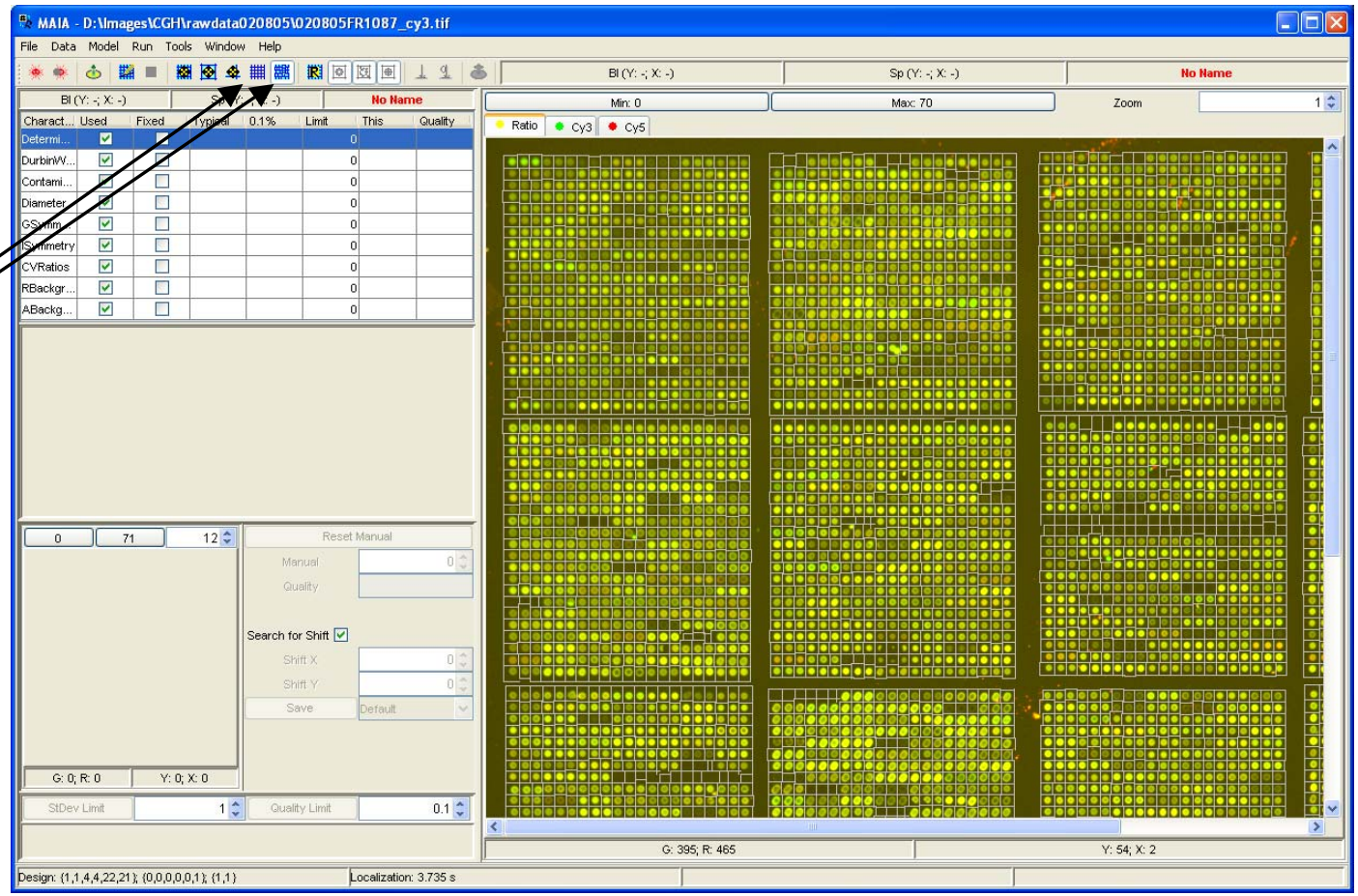
Main Grid is shown.



Spot Localization Output: Adjusted Grid

Using the Toolbar buttons "Show/Hide Main grid" or "Show/Hide Adjusted grid" one can mask either of two spot localization grids.

Adjusted Grid is shown.



Manual Correction of the Main Grid: Grid Movements

If the block grid is corrupted, one can shift the selected grid on the discrete number of spot rows/columns or to move smoothly the selected grid over the image.

Select a grid and iterate through the grids:

Shift + Mouse Click
Shift + Home
Shift + End
Shift + PgUp
Shift + PgDn

Move Selection by Pixel :

Shift + Drag
Shift + {↑,↓,→,←}

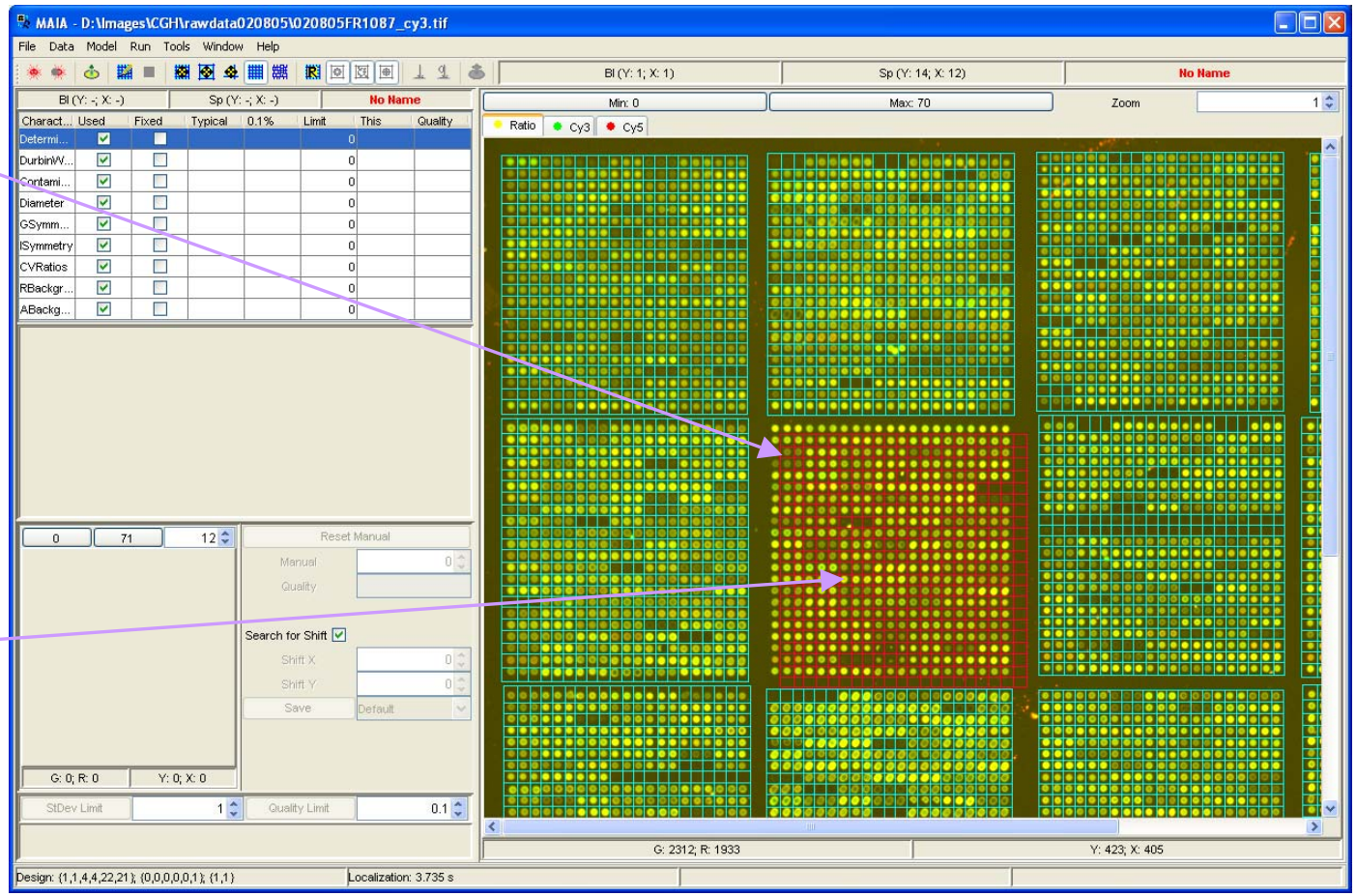
Move Selection by Spot :

Shift + Ctrl + {↑,↓,→,←}

Undo Moving:

Ctrl + Del

Upon selection the grid changes the color.



Manual Correction of the Main Grid : Line Movements

If the block grid is corrupted, one can perform manual correction of the positions of the line separations in the *Main Grid*.

Select a line and iterate through the lines:

Ctrl + Mouse Click
Ctrl + Home
Ctrl + End
Ctrl + PgUp
Ctrl + PgDn

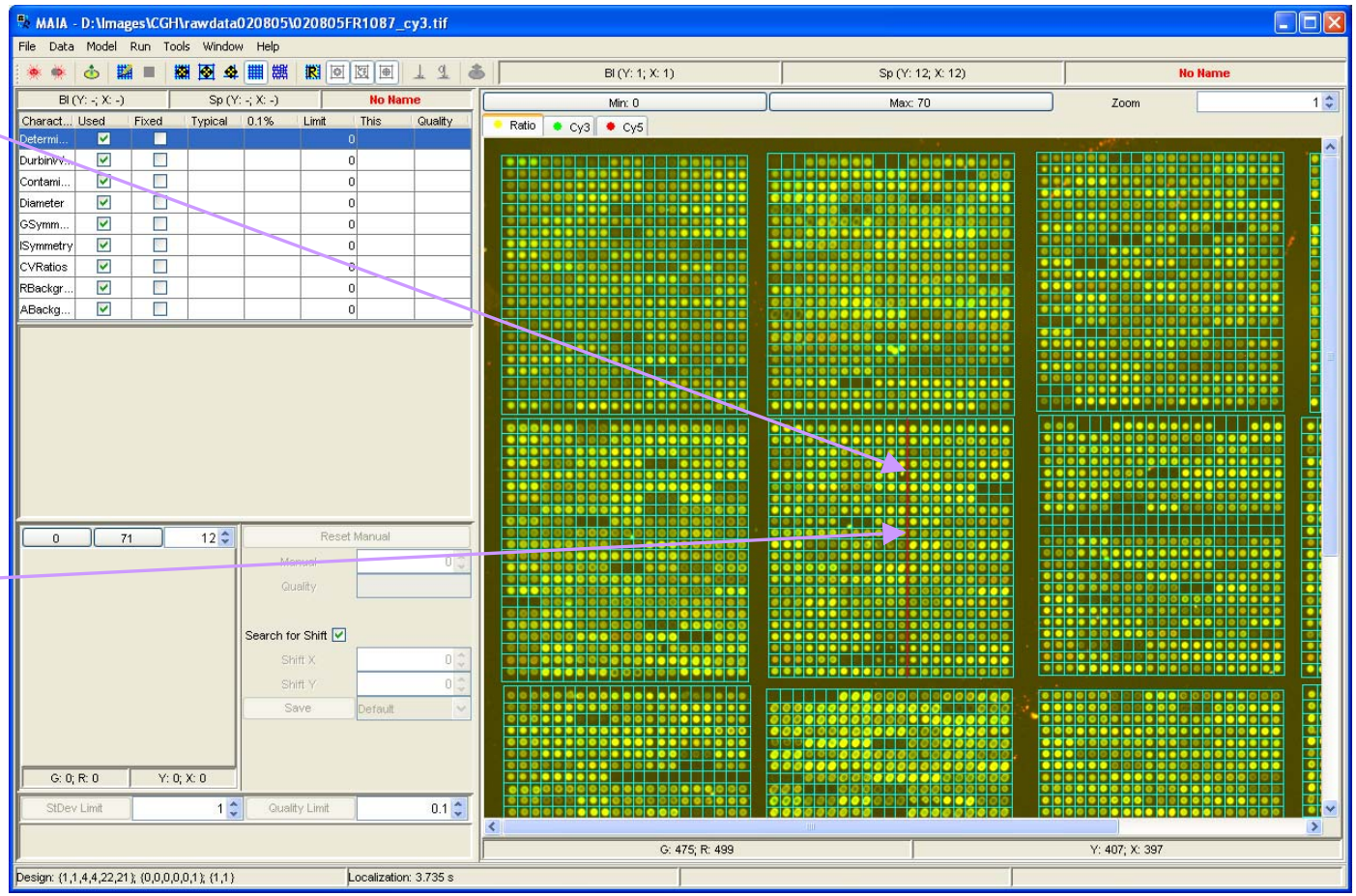
Move Selection by Pixel:

Ctrl + Drag
Ctrl + {↑,↓,→,←}

Undo Moving:

Ctrl + Del

Upon selection the line changes the color.



Manual Correction of the Adjusted Grid

If a separation (cut) between the neighborhood spots is erroneous, one can perform manual correction of the selected cut position.

Select a cut and iterate through the cuts:

Alt + Mouse Click
Alt + Home
Alt + End
Alt + PgUp
Alt + PgDn

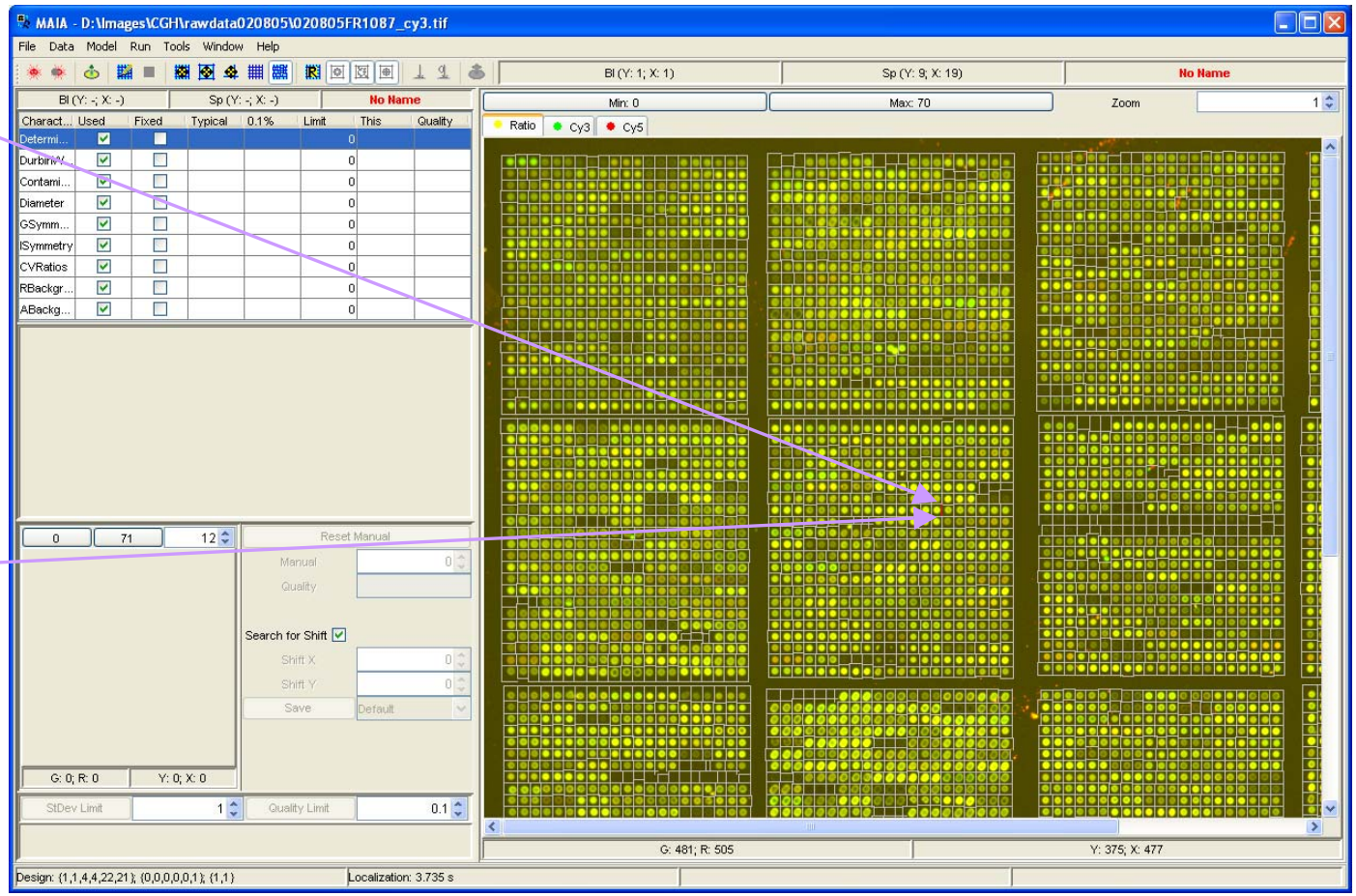
Move Selection by Pixel:

Alt + Drag
Alt + {↑,↓,→,←}

Undo Moving:

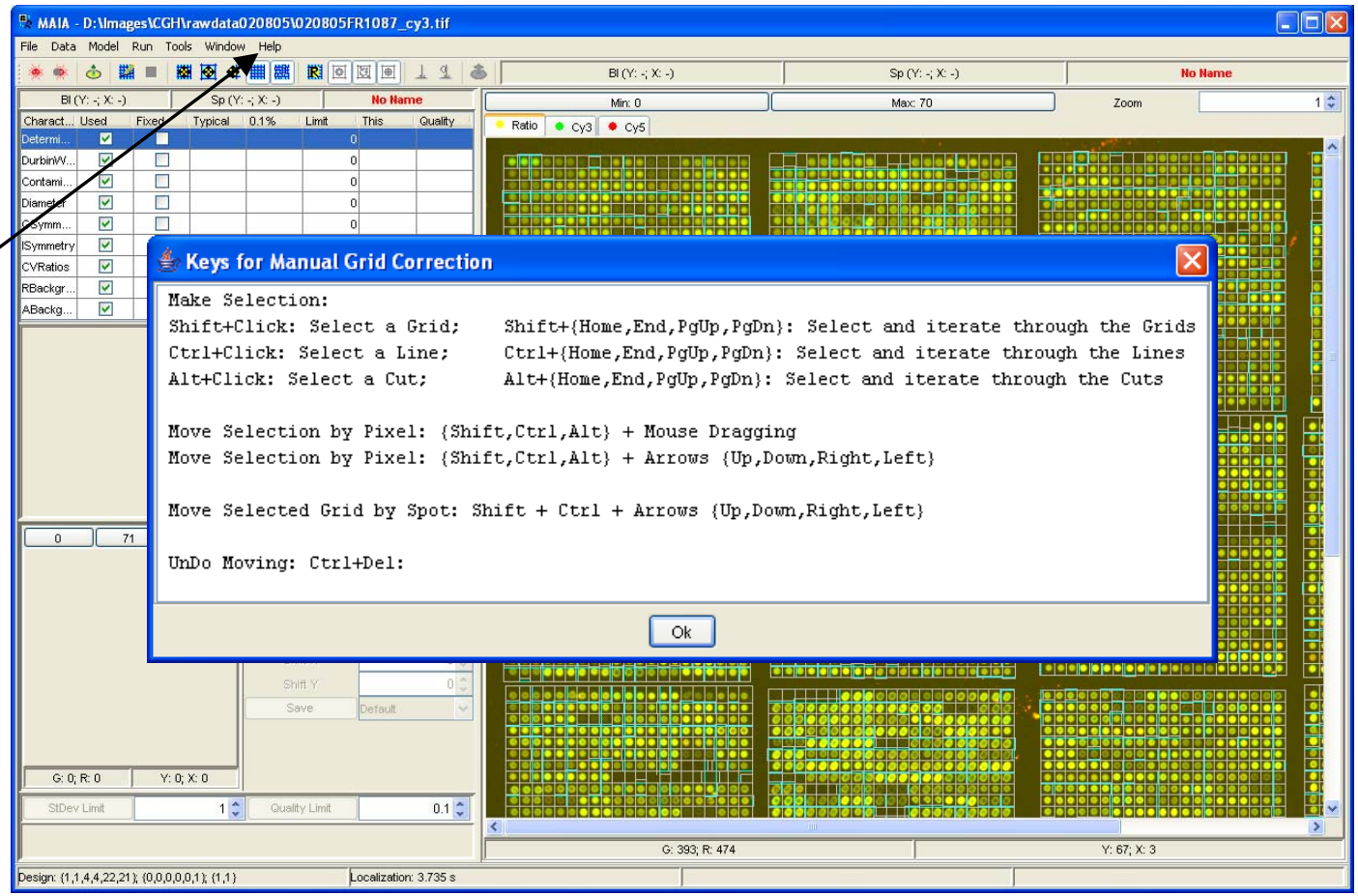
Ctrl + Del

Upon selection the cut changes the color.



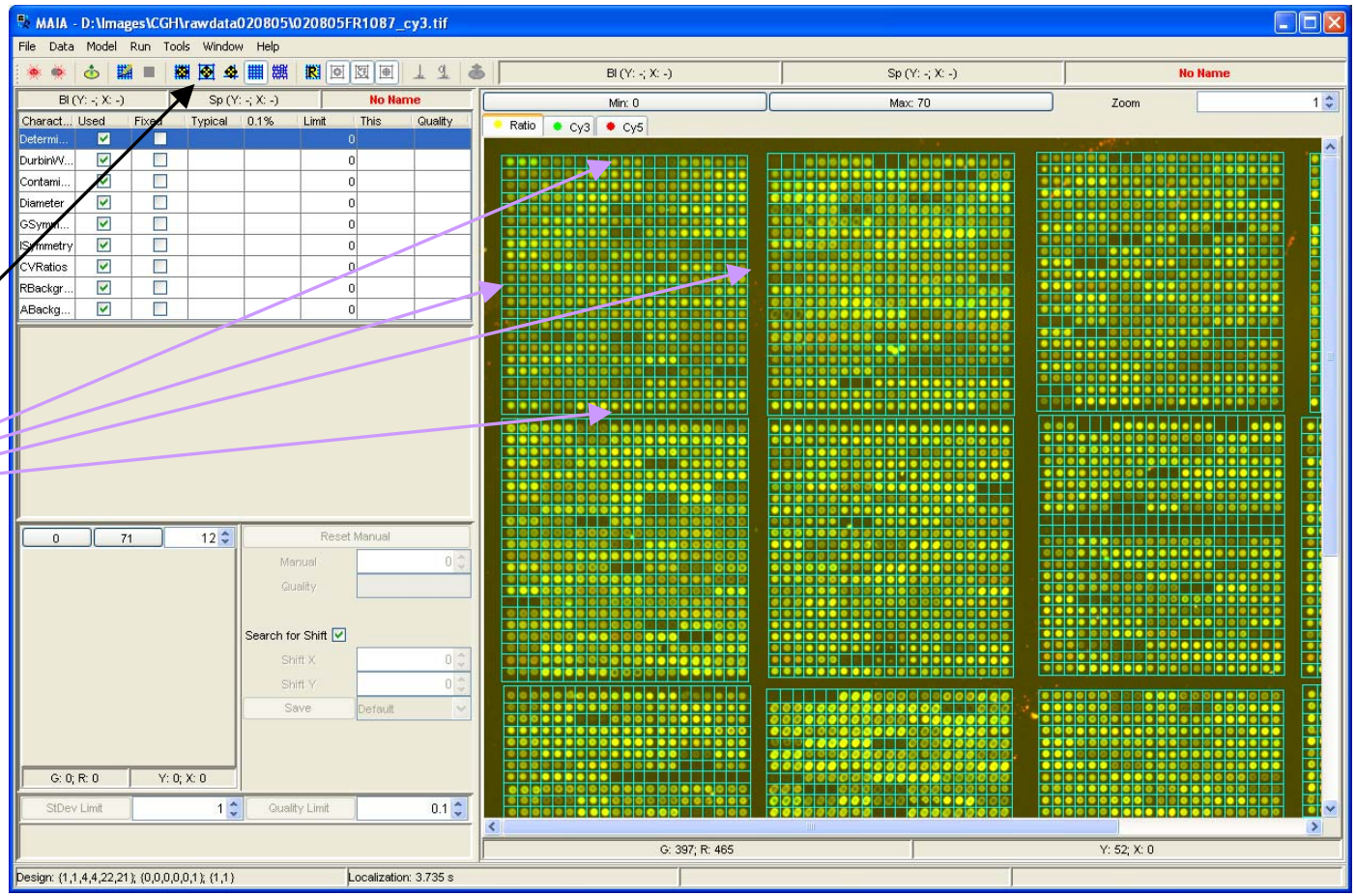
Brief Help on Manual Correction

To get brief help on the manual correction possibilities one may select the Menu Item “Help|Manual Grid Info”.



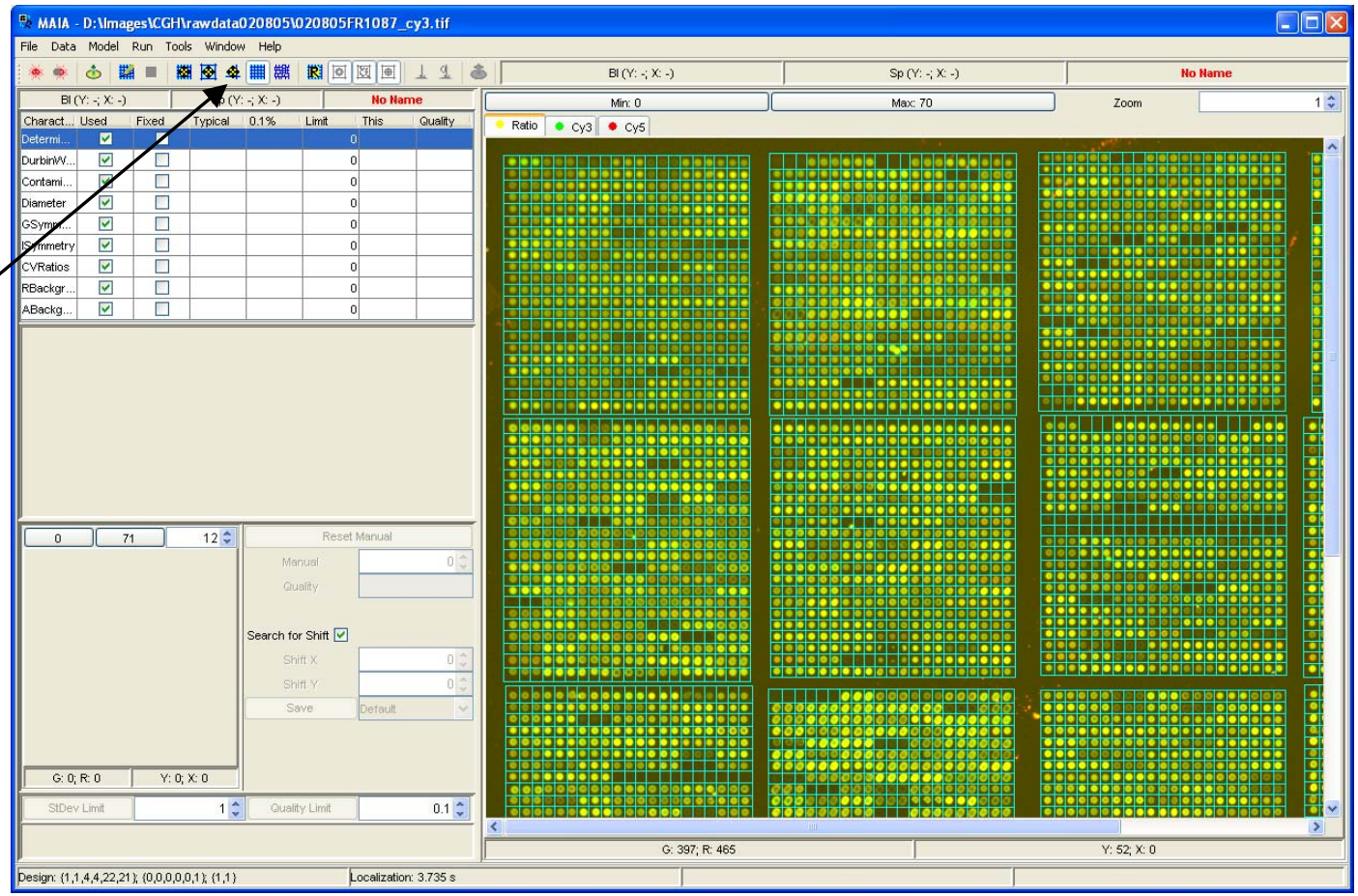
Further Main Grid Refinement: Find Grids in Blocks

Manual correction can be done only for the borders of the blocks (in the *Main Grid*). The other “internal lines” of the grids are found automatically using the “Grids in Blocks” button from the Toolbar or the Menu Item “Run|Grids in Blocks” (Ctrl+F3).



Further Main Grid Refinement: Lines Refinement

When the *Main Grid* is "almost" good, further refinement procedure will try to place the grid lines in the positions with the minimal inter-spot intensity more precisely (use the "Lines Refinement" button from the Toolbar or the Menu Item "Run|Lines Refinement" (Ctrl+Shift+F3)).



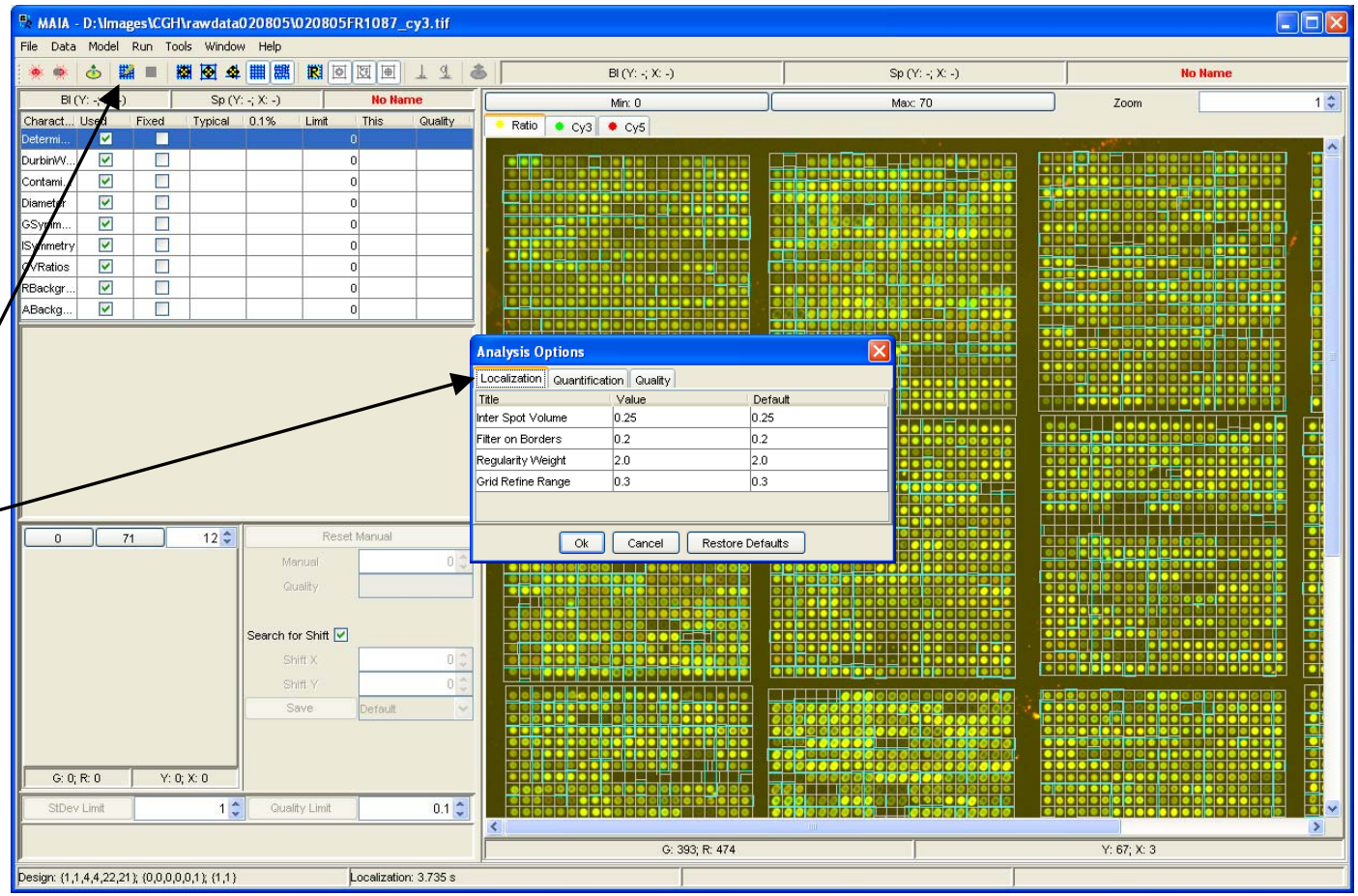
Save/Restore Grids

The generated grid can be saved on the disk (using the Menu Item "File|Save|Grid …") to be able to apply it (using the Menu Item "File|Load|Grid …") in the future to analyze other images with the similar design.

Localization Settings

Several settings that may influence the localization procedure are available through the Menu Item “Tools|Analysis Options” (Ctrl+A), tab “Localization”.

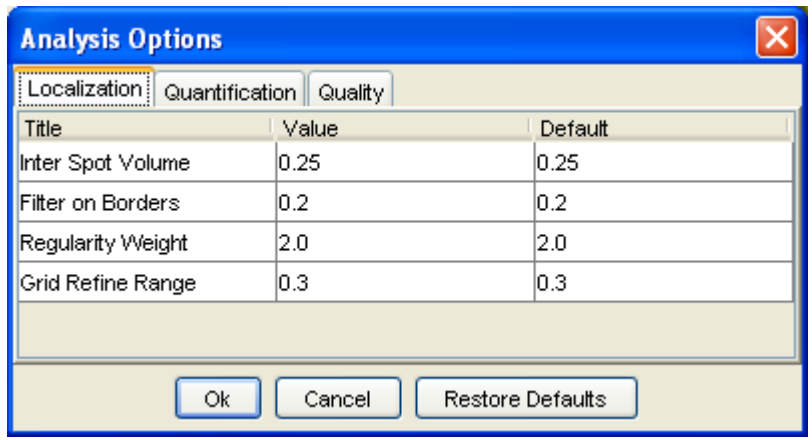
See next page for details.



Localization Settings in Detail

Inter Spot Volume represents (roughly) the ratio of the inter-spot gap to the inter-spot distance.

Filter on Borders defines filtering properties at the edges of the array. Higher this value, less sensitive the algorithm to the bright regions at the edges of the array.



Regularity Weight controls contribution of the regularity components with respect to the intensity component in the regularity parameter. With the weight equals to 0 the regularity components will be ignored.

Grid Refine Range defines the range (related to the inter-spot distance) for the final grid lines adjustments.

The default values of these parameters are suitable for a broad variety of experimental designs.

Spot Quantification

To start Spot Quantification use the "Spot Quantification" button from the Toolbar or the Menu Item "Run|Spot Quantification" (F4).

Spot Quantification Output

Each spot is contoured.

Quality characteristics of the spots.

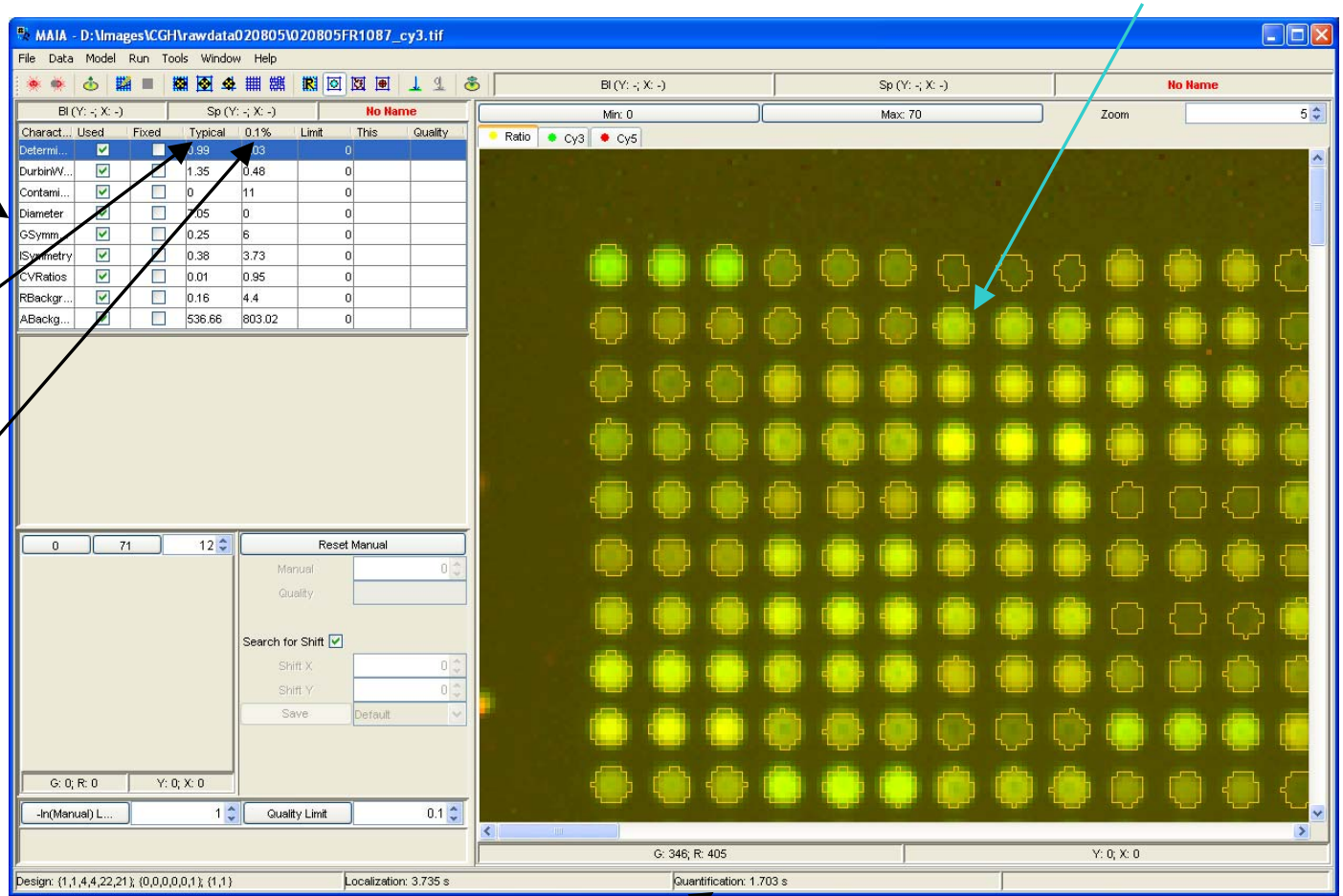
See next page for details.

Typical (median) value for each characteristic over all spots on the current array.

0.1% percentile (low-tail or high-tail, depending on the characteristic) over all spots on the current array.

The percentile (0.1%) can be modified using the the Menu Item "Tools|Analysis Options" (Ctrl+A), tab "Quality".

See page Quality Settings in Detail.



Status of the *Spot Quantification* procedure.

Quality Characteristics

Coefficient of determination (*CD*) of the linear regression indicates the degree of linear relationship between the intensities in Cy3 and Cy5 channels. For higher quality spots relatively high values of determination coefficient ($\rightarrow$ 1) are expected. Much lower values would point on either strong contribution of statistical noise, which normally characterizes low-level (or absent) spots, or presence of a relatively bright but non-correlated contamination. $q_1(CD) = CD^*$.

Durbin-Watson statistic (*DWS*) controls the presence of first-order autocorrelation in the residuals of the linear regression fit. It ranges from 0 to 4, 0 meaning positive correlation and 4 – negative correlation. $DWS \cong 2$ leads to the conclusion that the residuals are uncorrelated and the model is appropriate. Large departures from 2 suggests that this spot can not be modeled in terms of simple linear regression. $q_2(DWS) = 1-|DWS-2|/2^*$.

Spot contamination is a number of aberrant pixels (within the spot contours) flagged out by the filtering procedure (*N*). $q_3(N) = 1-N/S$, where *S* is the size of the correspondent spot, i.e. the number of pixels within the spot contour*.

Diameter of the spot: $D = 2(S/\pi)^{1/2}$. Since it is hard to impose *a priory* an exact ideal value for the diameter, the median diameter over all spots on the array is taken as a typical one. Spots with exceptionally small diameters should normally be penalized. $q_4(D) = exp\{D-T\}$, if $D < T$ and $q_4(D) =1$, if $D > T$, where *T* is the typical diameter*.

Geometrical symmetry parameter measures deviation of the contoured spot from the ideal circle. Both the real spot and the ideal circle are divided into 8 sectors and for each sector the number of pixels belonging to the spot (N_{si}, i = 1,…,8) and to the circle (N_{ci}, i = 1,…,8) is counted. Then the quality characteristic is defined as $GS = \sum|N_{si}- N_{ci}|/ N_{ci}$. For ideal circular spots *GS* must approach 0, whereas highly un-circular spots should give relatively high *GS* values. $q_5(GS) = exp(-GS)^*$.

Intensity symmetry of the spot is defined as $IS = \sum|I_i-I|/I$, where I_i, i = 1,…,8 are the mean intensities for the same 8 sectors and *I* is the mean intensity for the whole spot A spot may have perfect circular shape, but within this circle very bright (or dark) and highly concentrated groups of pixels originated from the pieces of dust or other contamination may occur. *IS* is calculated for each of two channels (Cy3 and Cy5) and the worst (i.e. highest) value is taken as a final estimate. $q_6(IS) = exp(-IS)^*$.

Coefficient of variation of two ratio estimates: $CVR = 2^{1/2}|RR-RS|/(RR+RS)$. Despite the differences in the estimation, the variation between the two obtained ratios *RS* and *RR* should be as small as possible. Large variation would indicate a problematic spot. $q_7(CVR) = exp(-CVR)^*$.

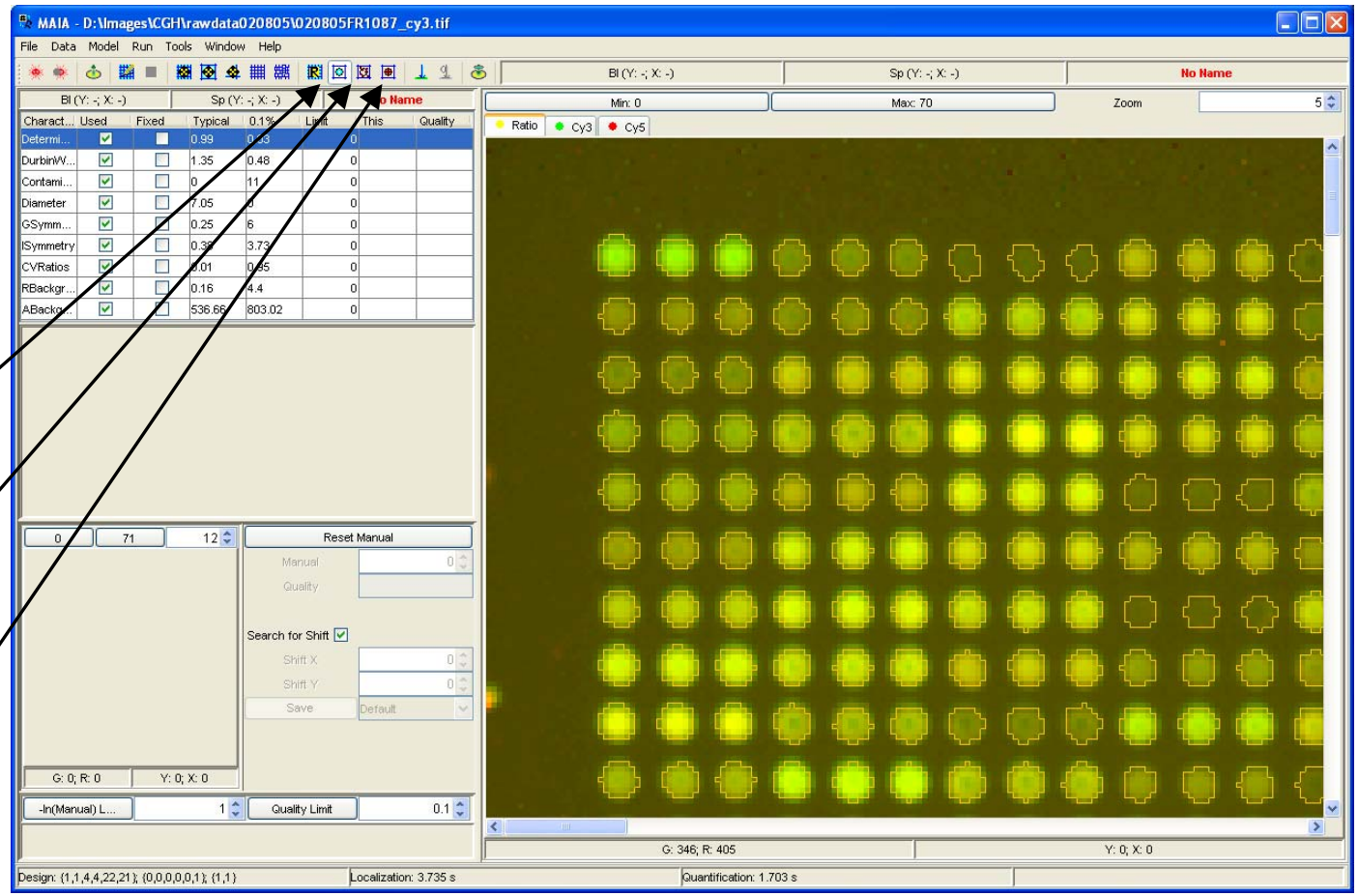
Uniformity of the background (*UB*) around the spot, i.e. along the grid lines separating neighborhood spots, is defined as $UB = \sum|B_i-B|/B$, where B_i, i = 1,…,8 are the mean intensities in 8 sectors of the grid line around the spot, and *B* is the mean intensity for the whole grid line around the spot. Extremely small values may be due to relatively bright contamination around the spot, large variability in the background or merged neighborhood spots. $q_8(UB) = exp(-UB)^*$.

Absolute level of background (*AB*) calculated in the proximity of each particular spot is compared to the typical level of the local background estimates for a given array. Large deviations from the typical state may indicate the presence of the contamination areas, which are larger than the size of the spot. $q_9(AB) = exp(T_{AB}- AB)$, if $AB>T_{AB}$ and $q_9(AB) = 1$, if $AB<T_{AB}$. where T_{AB} is the typical background level*.

**For the purposes of further quality analysis, functions q_i, i = 1….9 rescale quality characteristics to fit the range between 0 ("bad" spot) and 1 ("good" spot).*

Visualization of the Spot Quantification Elements

Using the Toolbar buttons one can control, which elements of the spot quantification results are visible:



Show/Hide Spot Contours.

Show/Hide Outlier Pixels
(*see following pages for details*).

Show/Hide Quality Markers
(*see following pages for details*).

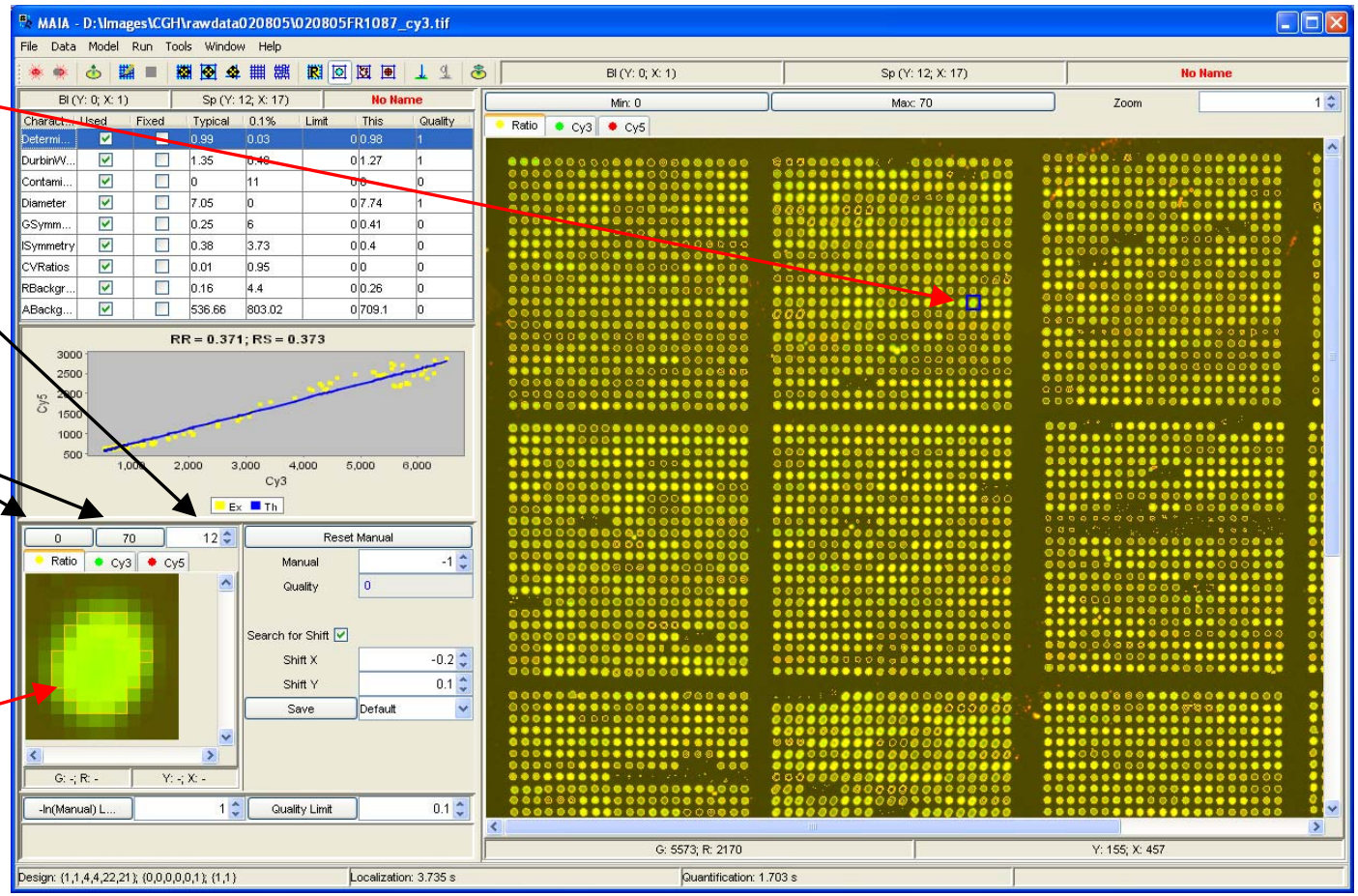
Spot Selection

Select a spot.

Spot zoom can be adjusted using either the “Zoom” spinner box or the mouse wheel.

“Min” and “Max” controls can be used to adjust brightness and contrast of the selected spot.

Magnified image of the selected spot with the contour.



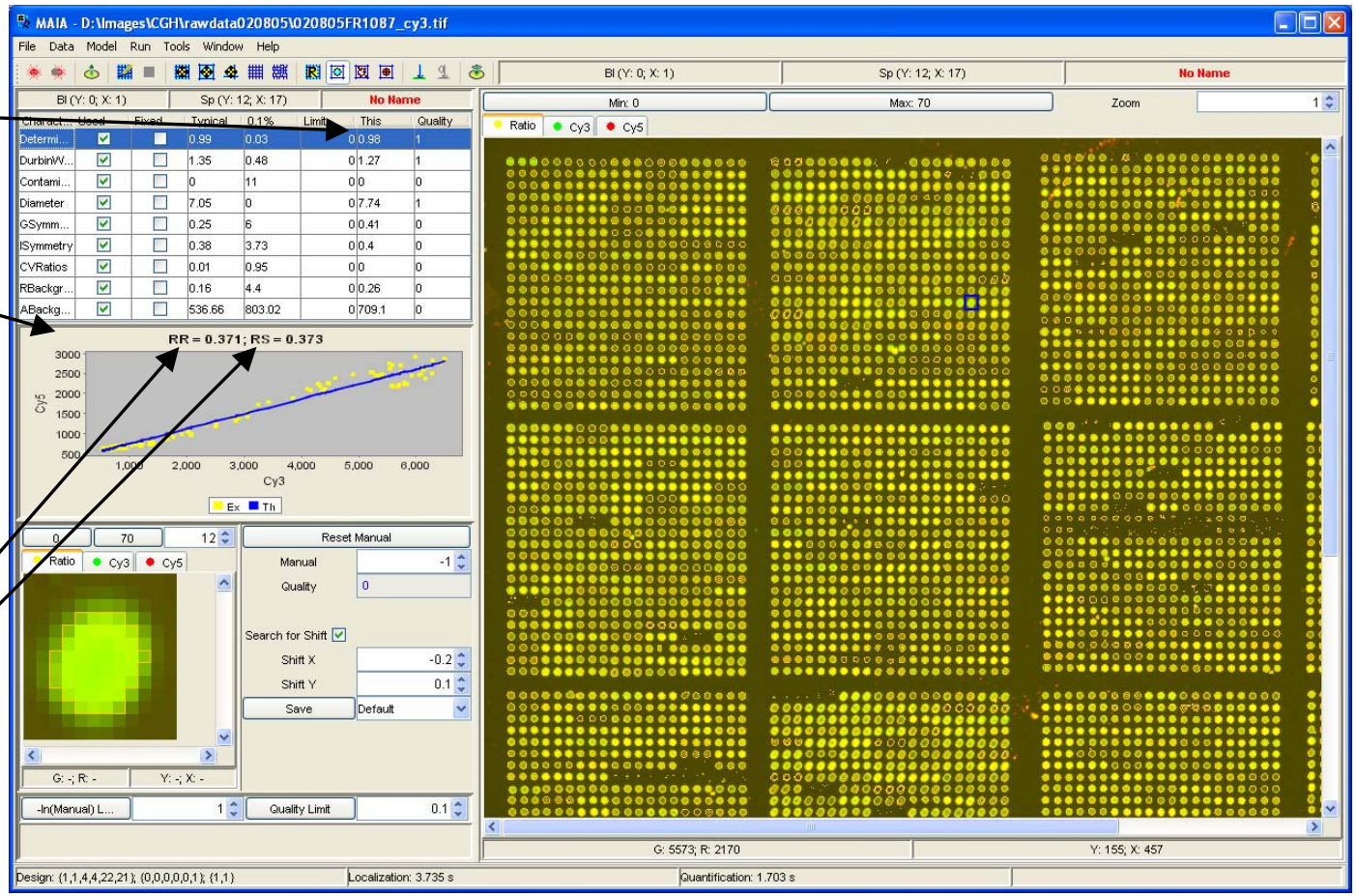
Spot Characteristics

Quality characteristics of the selected spot.

Linear regression plot for the selected spot.

Two ratio estimates:
RR is based on the slope of the linear regression;
RS is based on the segmentation of the spot area

See next page for details.



Ratio Estimation

Segmentation Ratio. This approach is based on isolation of the spot pixels from the background pixels surrounding the spot. Once this is done, the quantification procedure is fairly straightforward: one can compose the following ratio:

$$R=\frac{S_{Cy5}-B_{Cy5}}{S_{Cy3}-B_{Cy3}}$$

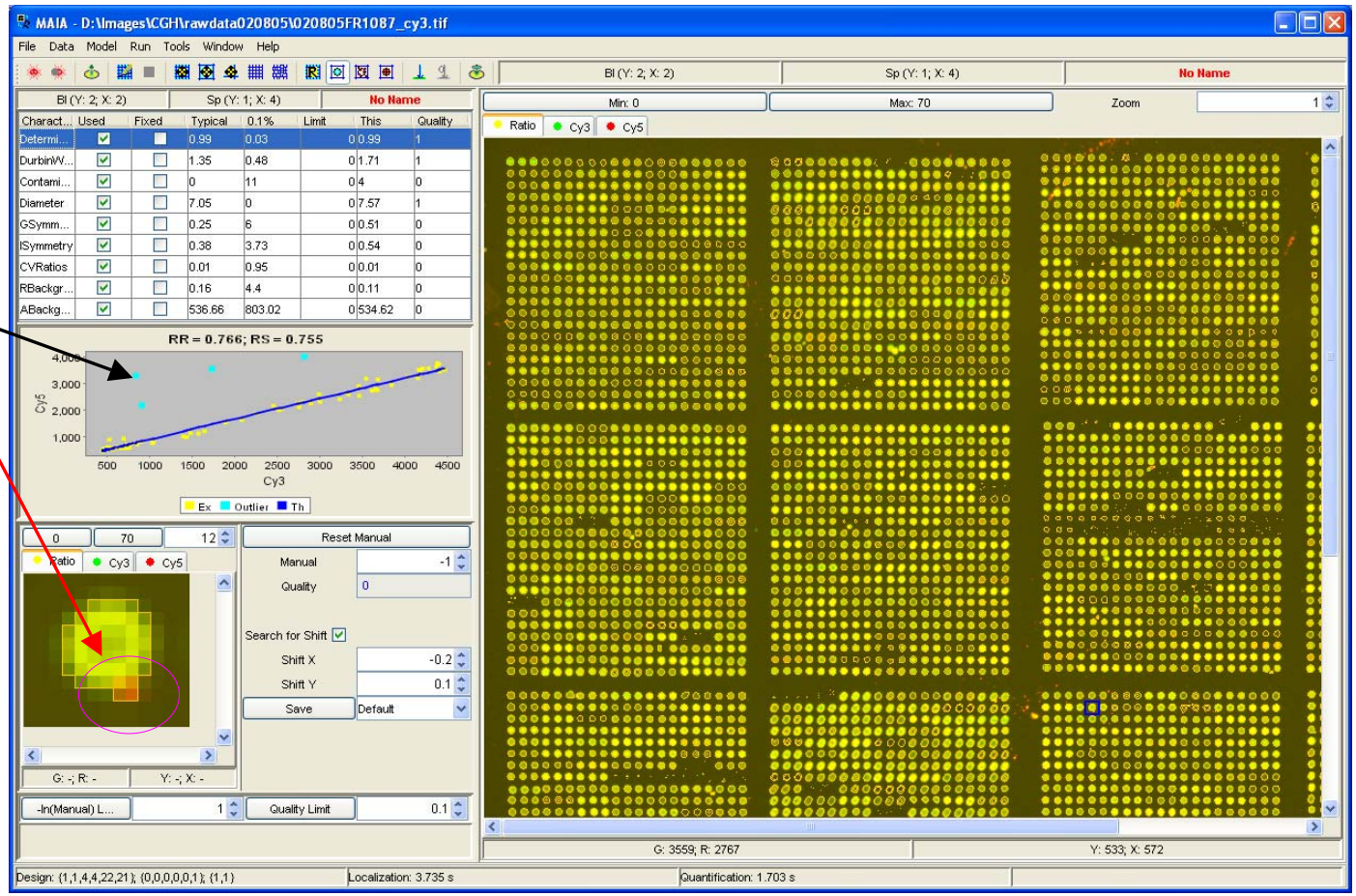
where $S_{Cy5}(S_{Cy3})$ is the mean estimate of the intensity within the contoured spot in the Cy5(Cy3) channel, and $B_{Cy5}(B_{Cy3})$ is the mean estimate of the background level in the Cy5(Cy3) channel. Mean estimates are known to be more precise, but they can be very much affected by the outliers. Since regression filtering eliminates outliers, we can safely use mean estimates for the spots.

Regression Ratio. In this approach a ratio can be represented as a slope of the linear regression line of the pixel intensities in, say, Cy5 channel versus Cy3 channel. The main advantage of this method is that the obtained ratio is directly delivered from the regression analysis, thus making the procedure of spot segmentation unnecessary. Background pixels are concentrated at the initial part of the linear regression and do not influence the slope of the regression line. However the linear regression approach suffers from the presence of the outlier or aberrant pixels within the spot cells. These pixels, occurring even in small quantities, can distract the regression line and strongly bias the regression ratio. With the aim to fully exploit the advantages of the linear regression approach we have reinforced this procedure by systematical filtering out aberrant pixels

See page Pixel Regression Outliers.

Pixel Regression Outliers

Pixel regression outliers.



All Pixel Regression Outliers

Using the Toolbar button "Show/Hide Outlier Pixels" one can control whether the pixel regression outliers are visible.

Pixel regression outliers.

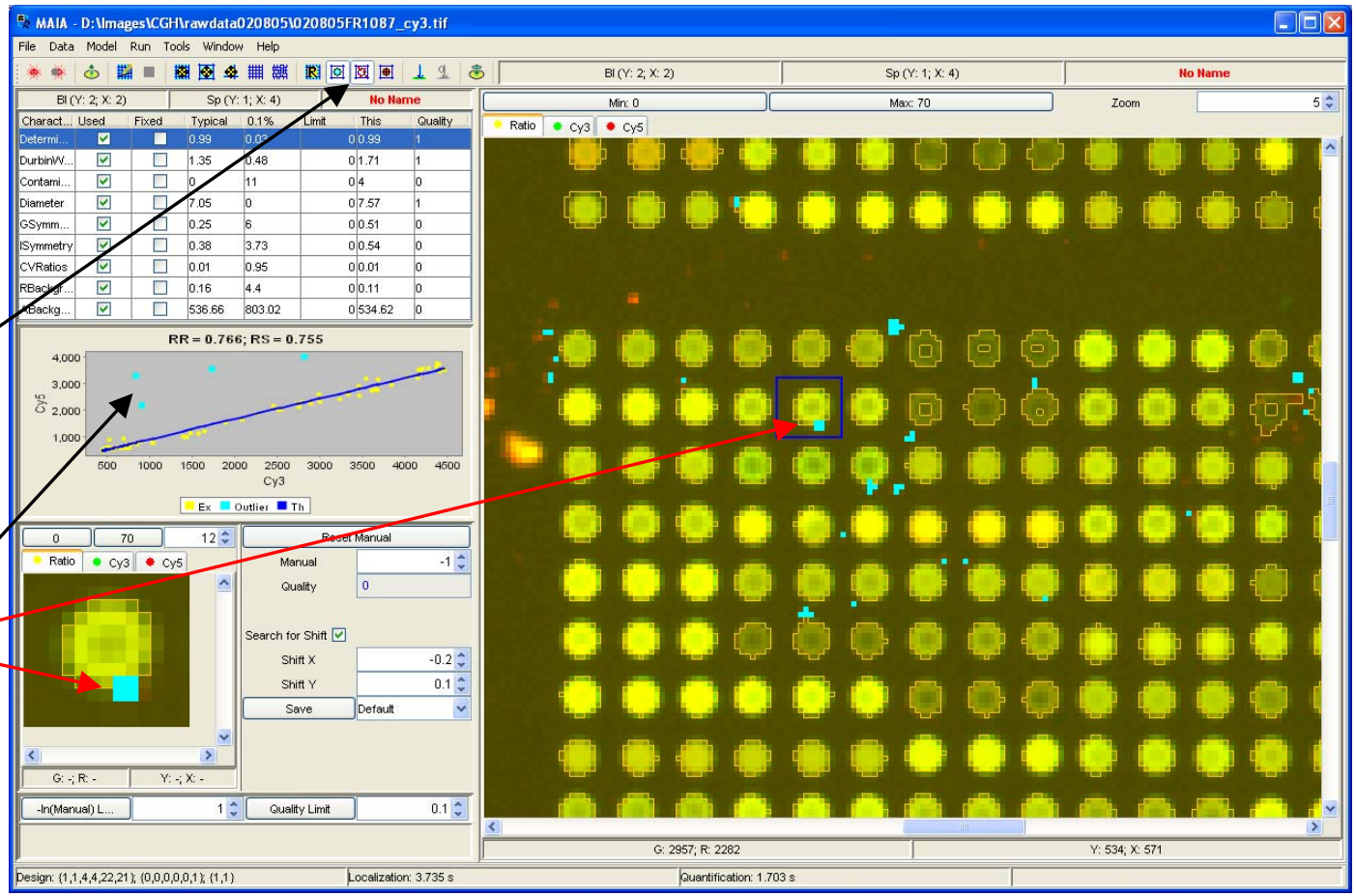
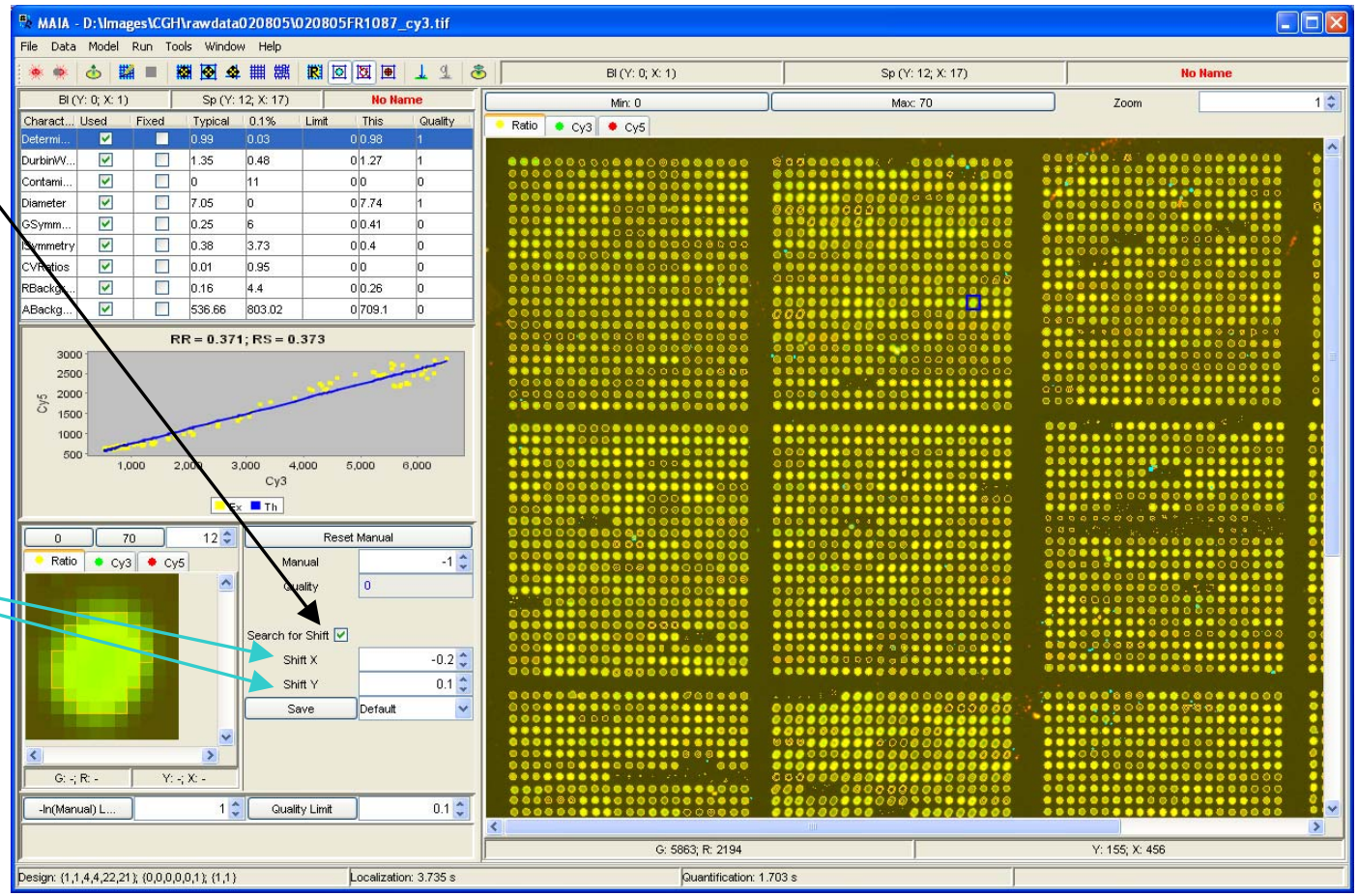


Image Shift

If the “Search for Shift” checkbox is selected, the quantification procedure tries to identify the relative shift between the Cy3 and Cy5 images before any other processing.

Relative shift (in pixels) in the horizontal (X) and vertical (Y) directions between the Cy3 and Cy5 images.

This shift is visualized only for the selected spot and not for the whole image.

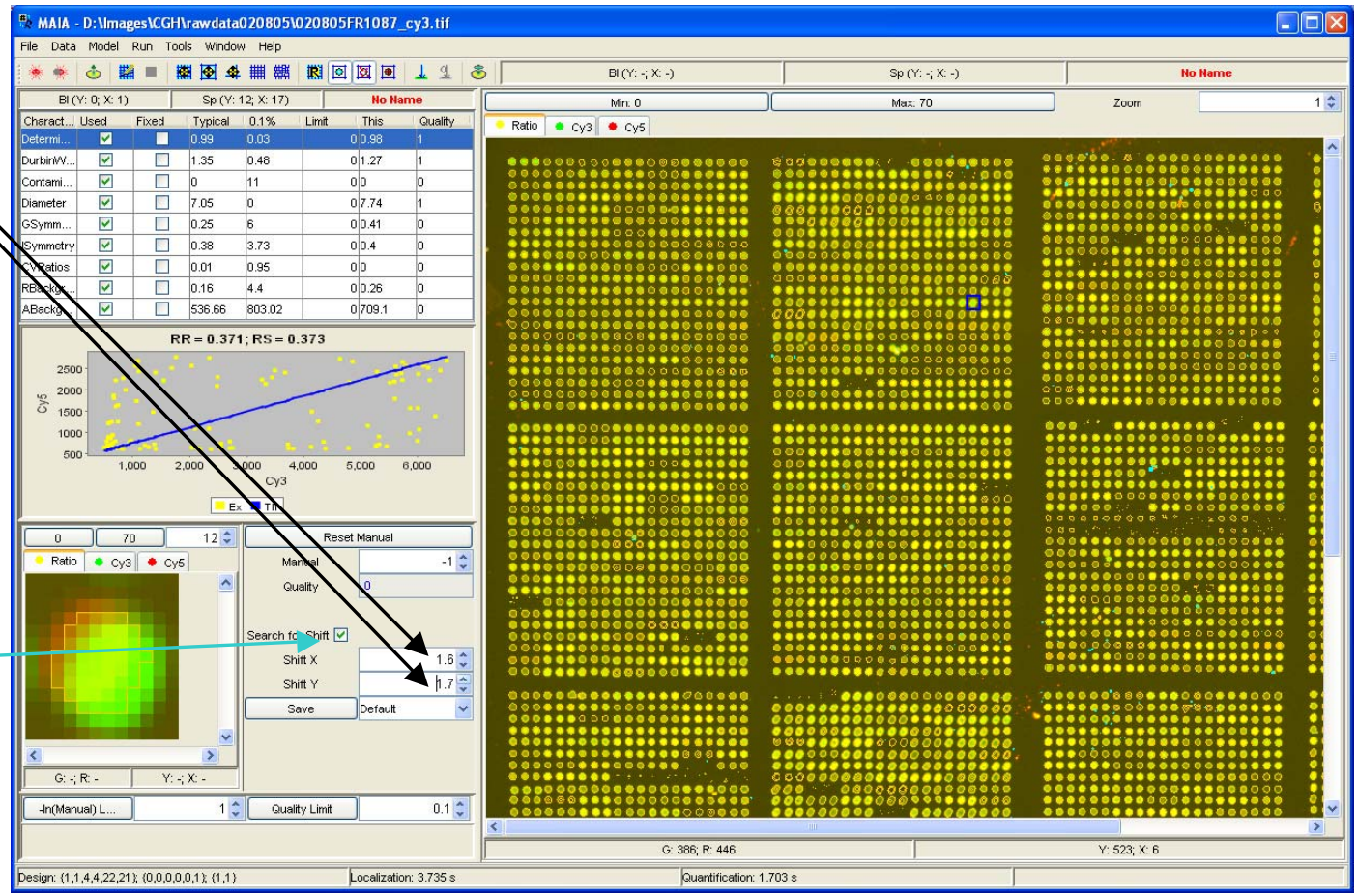


Manual Adjustment of the Image Shift

Using the “Shift” spinners one can adjust, if necessary, the values of the shift.

The new values will be valid for all spots from the given block.

To perform quantification with the new shift, one has to uncheck the “Search for Shift” checkbox and start the quantification procedure again.

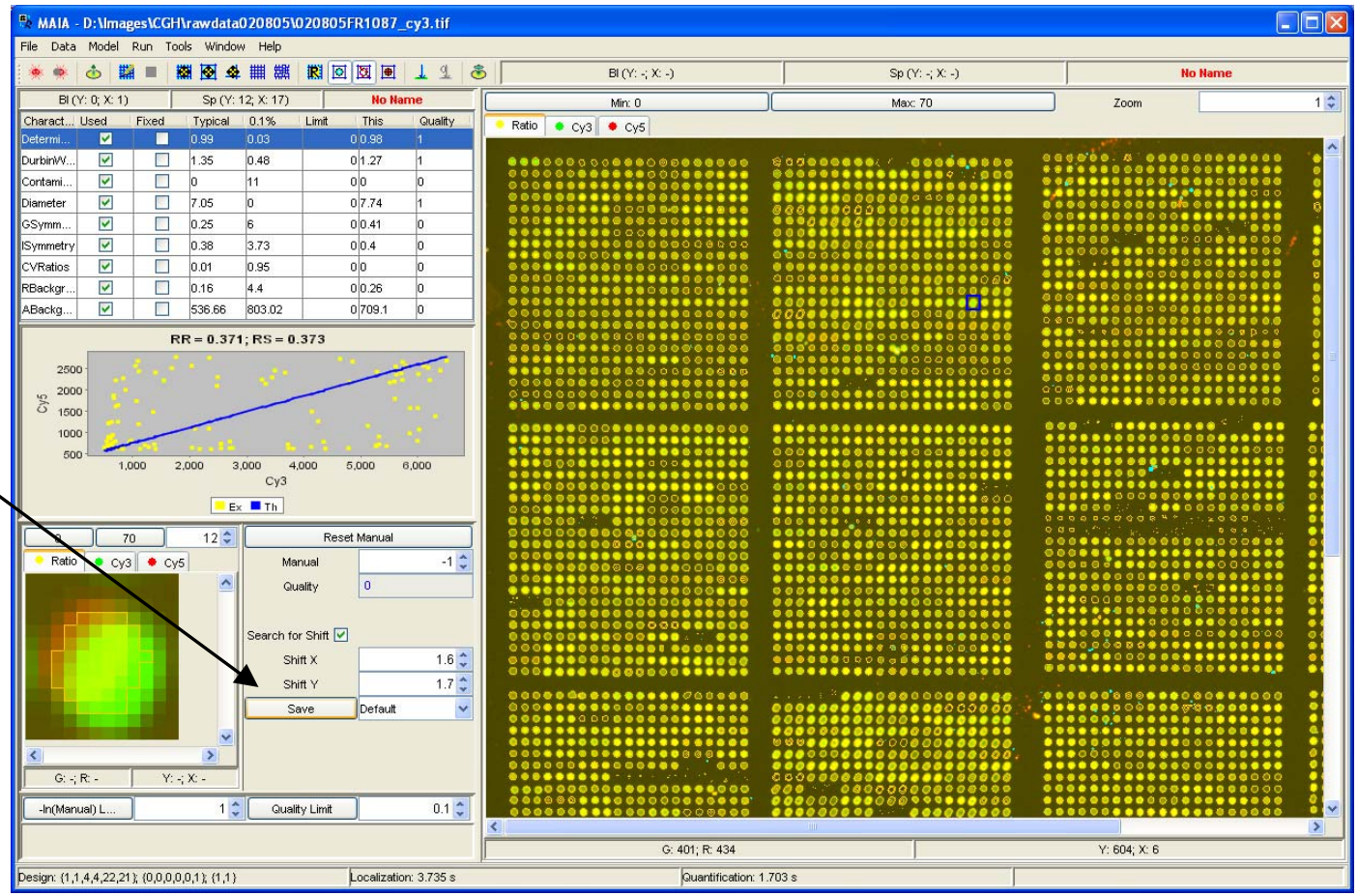


Compare Different Image Shifts

The new values of the shift can be saved (using the button “Save”) and used for comparison with the automatically generated (Default) and zero (=0) shift.

Proper identification and correction for the image shift is important in order to increase the efficiency of the linear regression filtering.

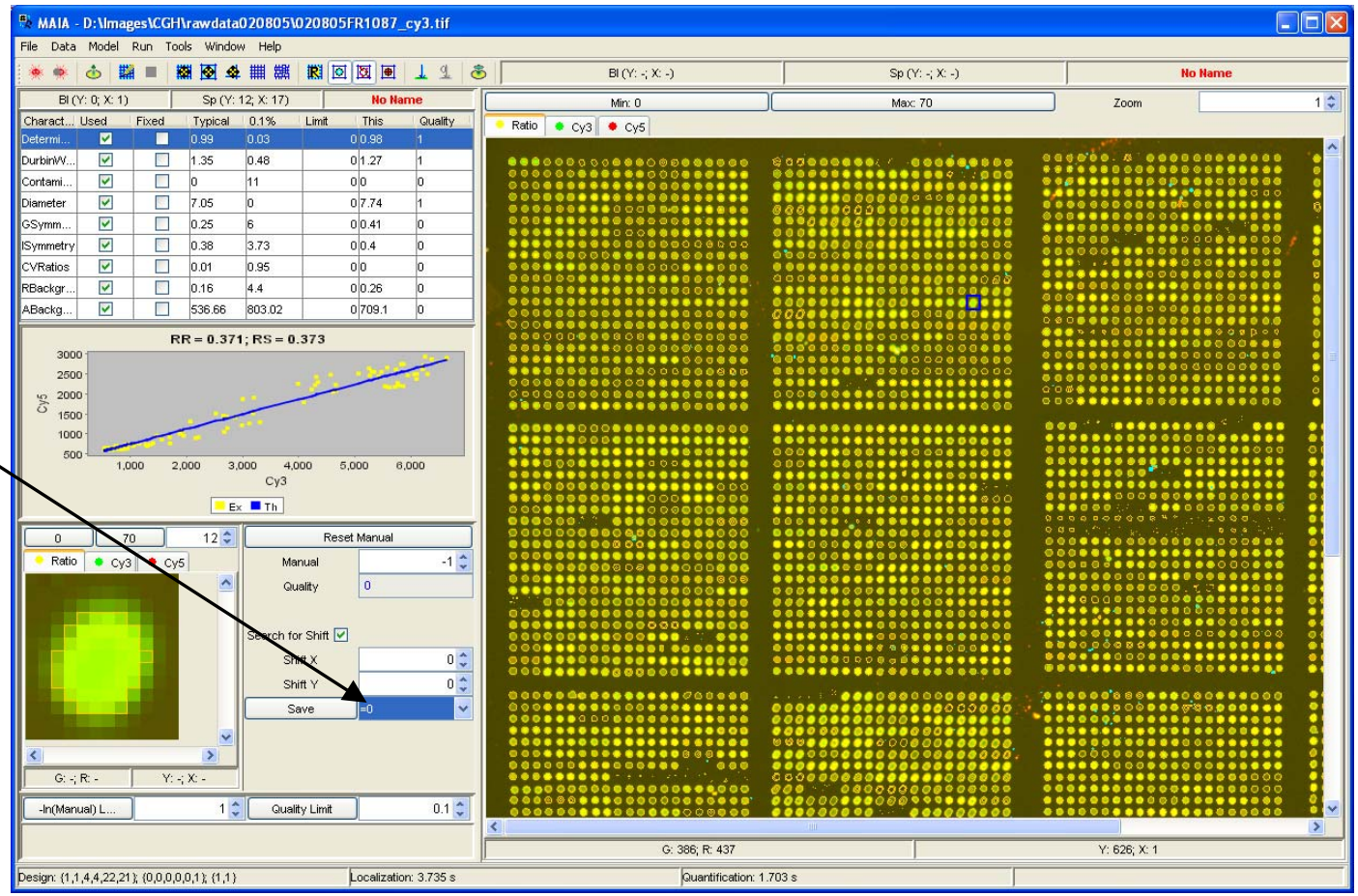
Removal of the shift enhances the correlation between the two channels thus making uncorrelated pixels easier detectable.



Zero Shift

To switch between different shift values one can use the “Shift” combo box.

Zero shift is selected.

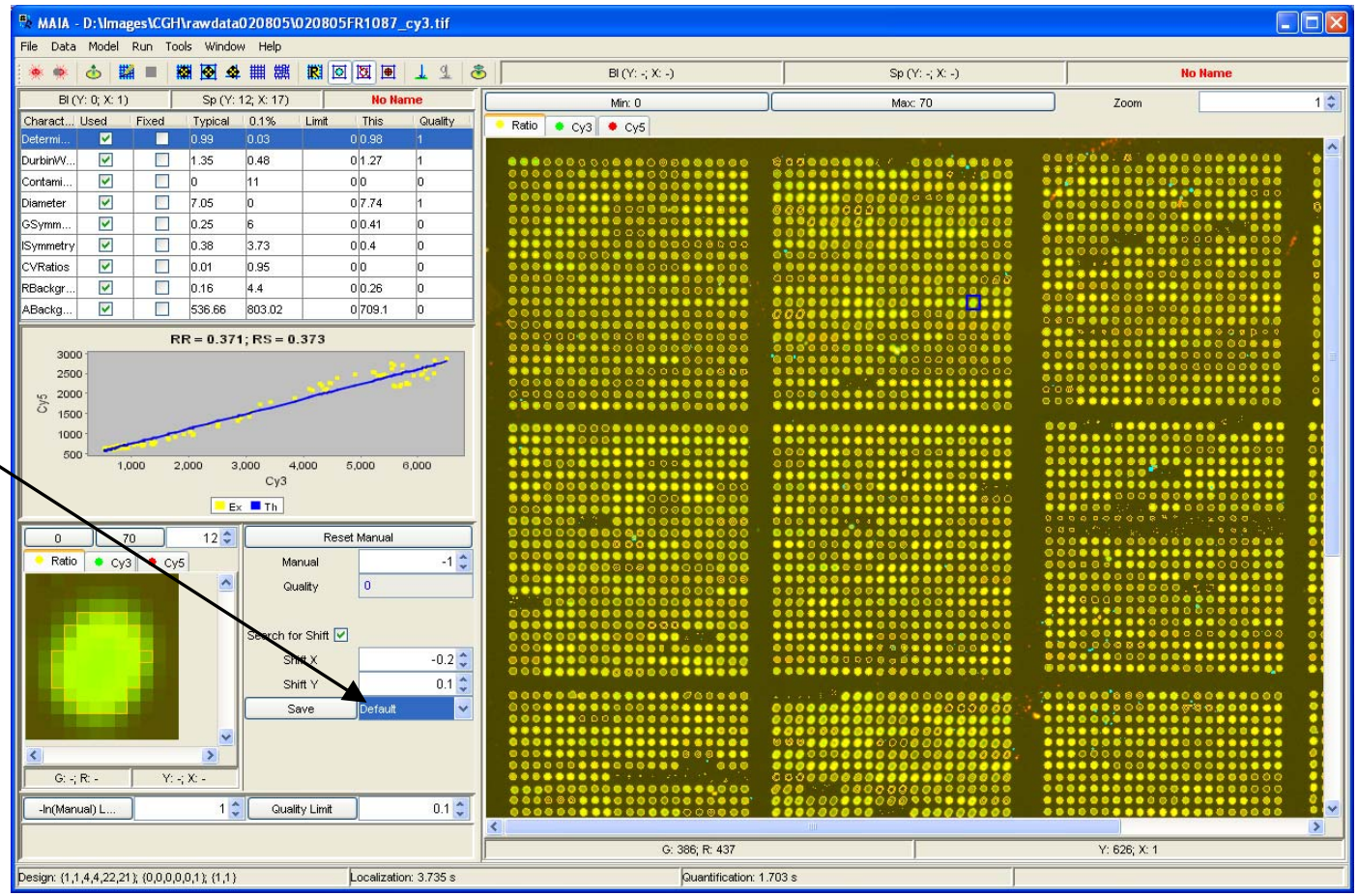


Default Shift

To switch between different shift values one can use the “Shift” combo box.

Default shift is selected.

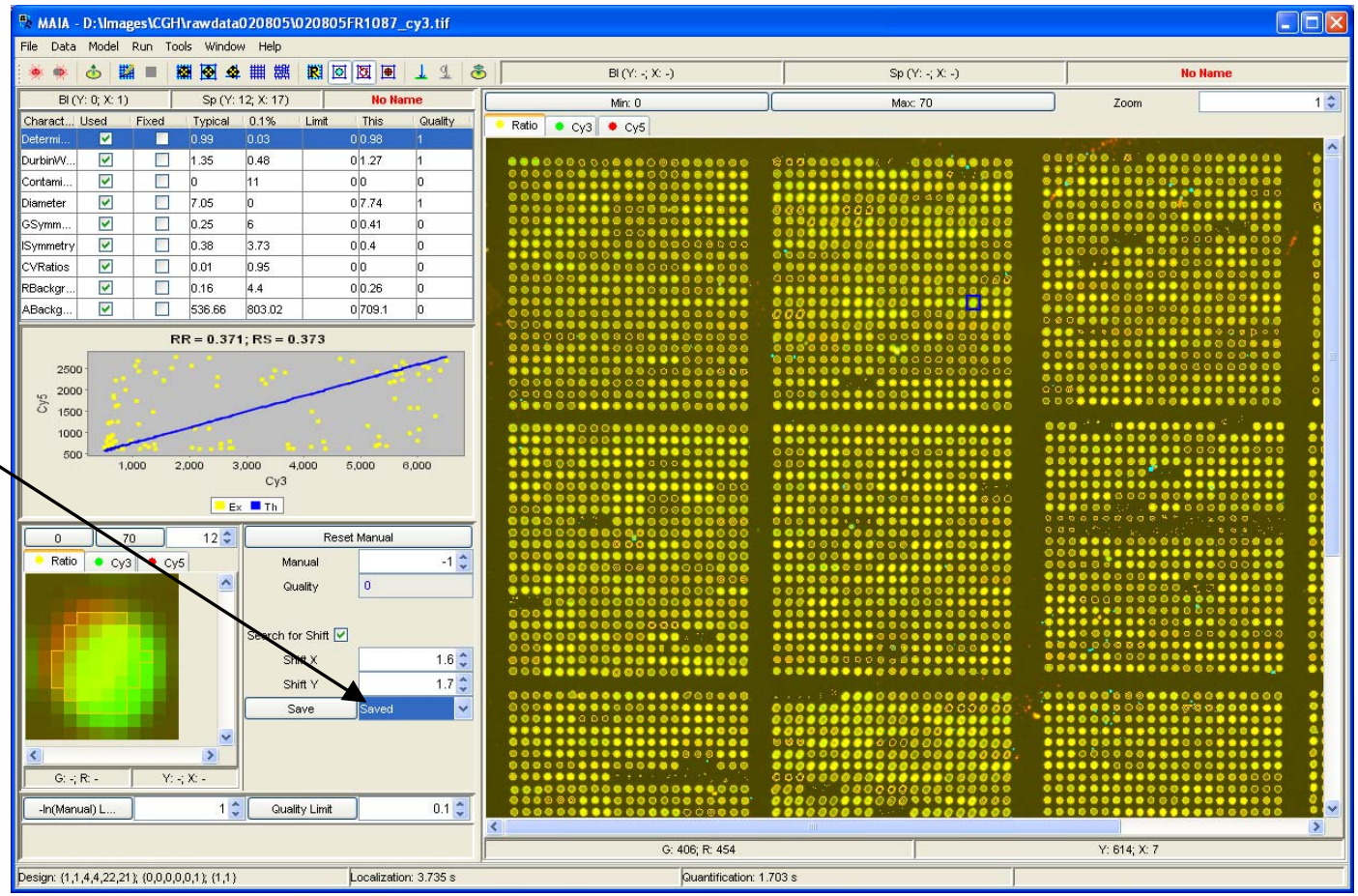
Note the difference in the linear regression plot as compared to the Zero shift.



Saved Shift

To switch between different shift values one can use the “Shift” combo box.

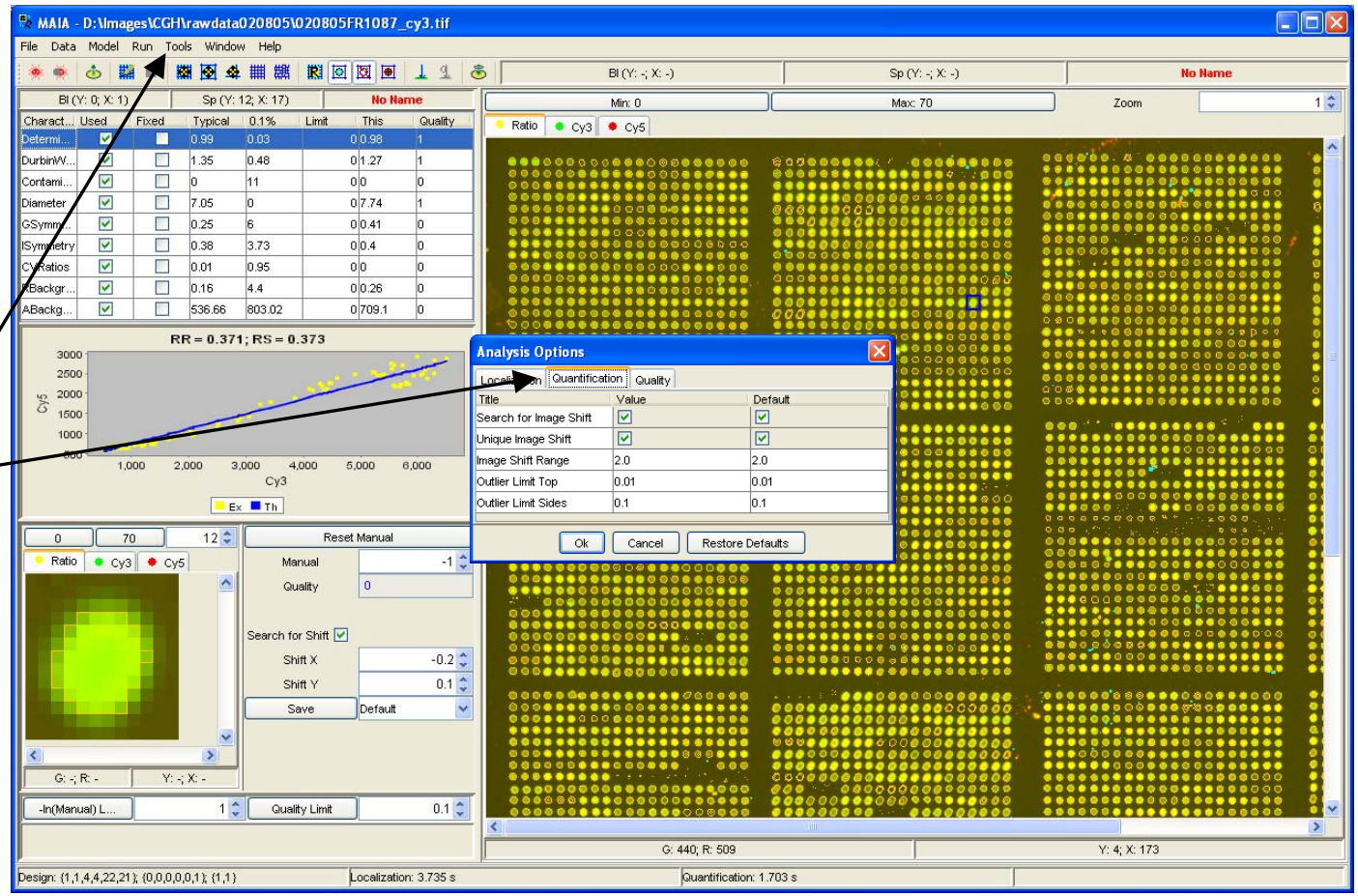
Saved shift is selected.



Quantification Settings

Several settings that may influence the quantification procedure are available from the Menu Item “Tools|Analysis Options” (Ctrl+A), tab “Quantification”.

See next page for details.



Quantification Settings in Detail

Search for Image Shift switches on/off the automatic procedure for the relative (Cy3/Cy5) shift identification.

Unique Image Shift defines whether the shift is the same for all blocks on the array (on) or it is different for each block (off).

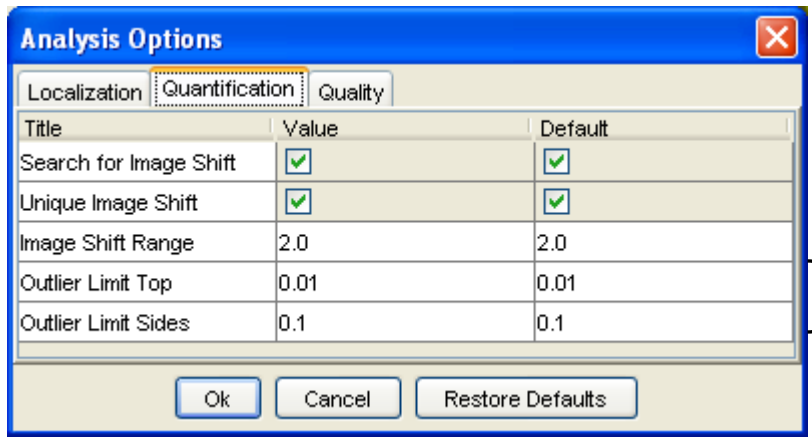


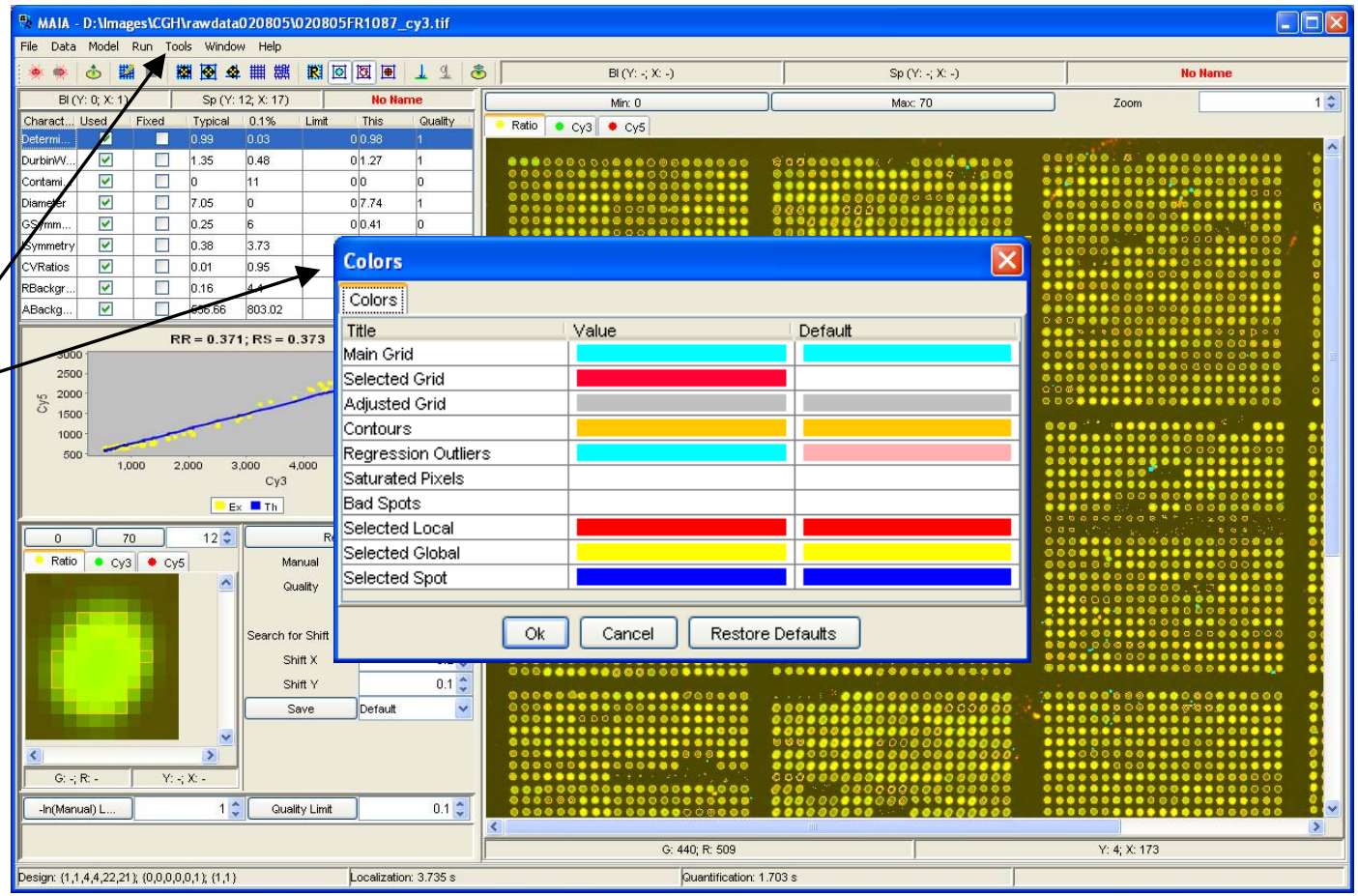
Image Shift Range establishes the boundaries (in pixels) for the maximally possible relative shift between the two images (2 pixels, by default).

Outlier Limit Top/Sides defines critical *p*-values of the *F*-statistics in the detection of the pixel outliers selected from the top of the intensity ranges and from the sides of the linear regression fit.

The default values of these parameters are suitable for a broad variety of experimental designs.

Colors

To change the color of some elements of the localization and quantification outputs use the Menu Item “Tools|Colors”.

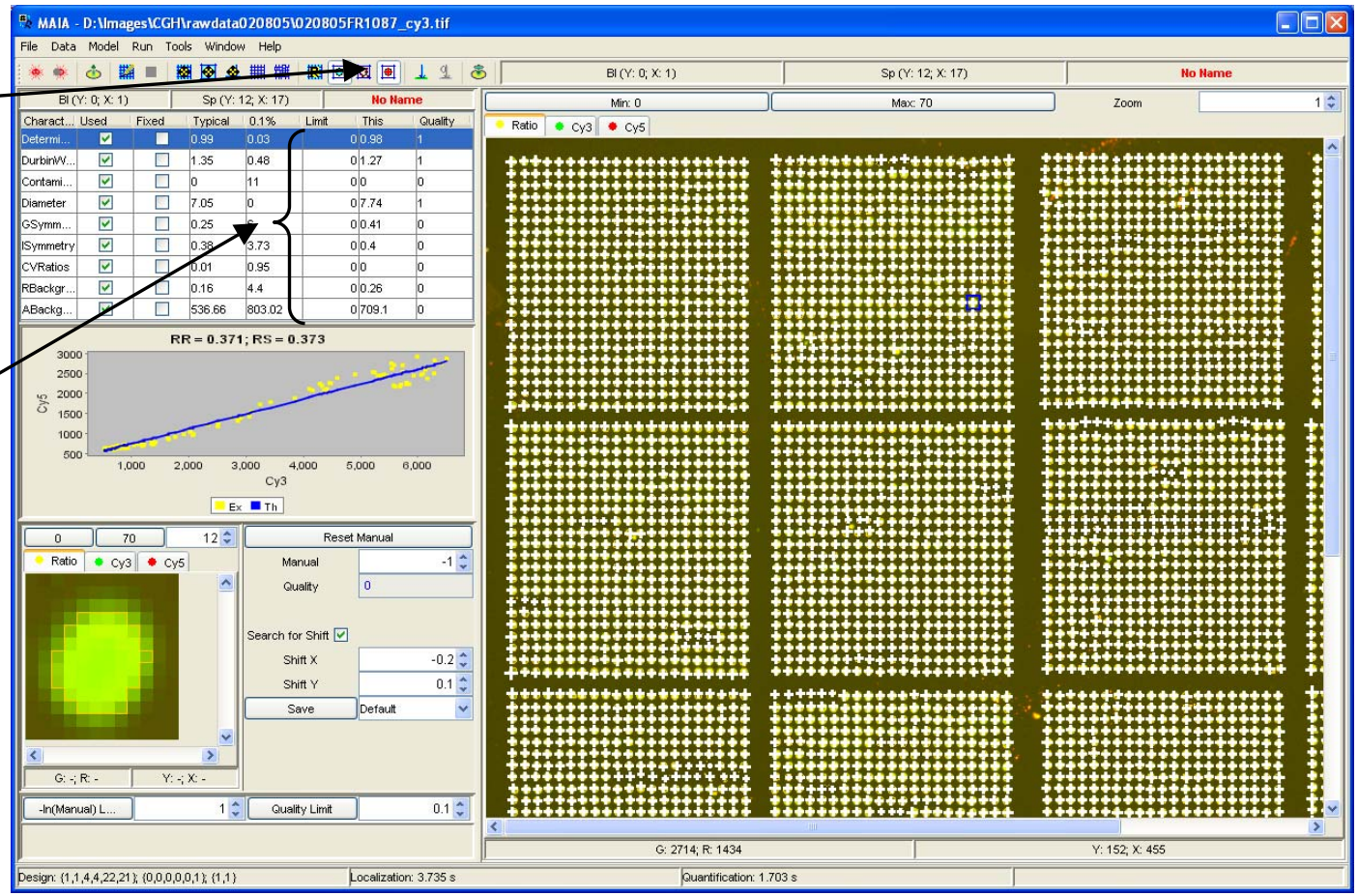


Quality Analysis (I): Without Replicated Spots

Switch-on the toolbar button “Show/Hide Quality Markers”.

For each used quality characteristic a reasonable critical level (limit) must be selected.

If one of the quality characteristics of a spot exceeds the correspondent limit, this spot will be indicated by a cross.



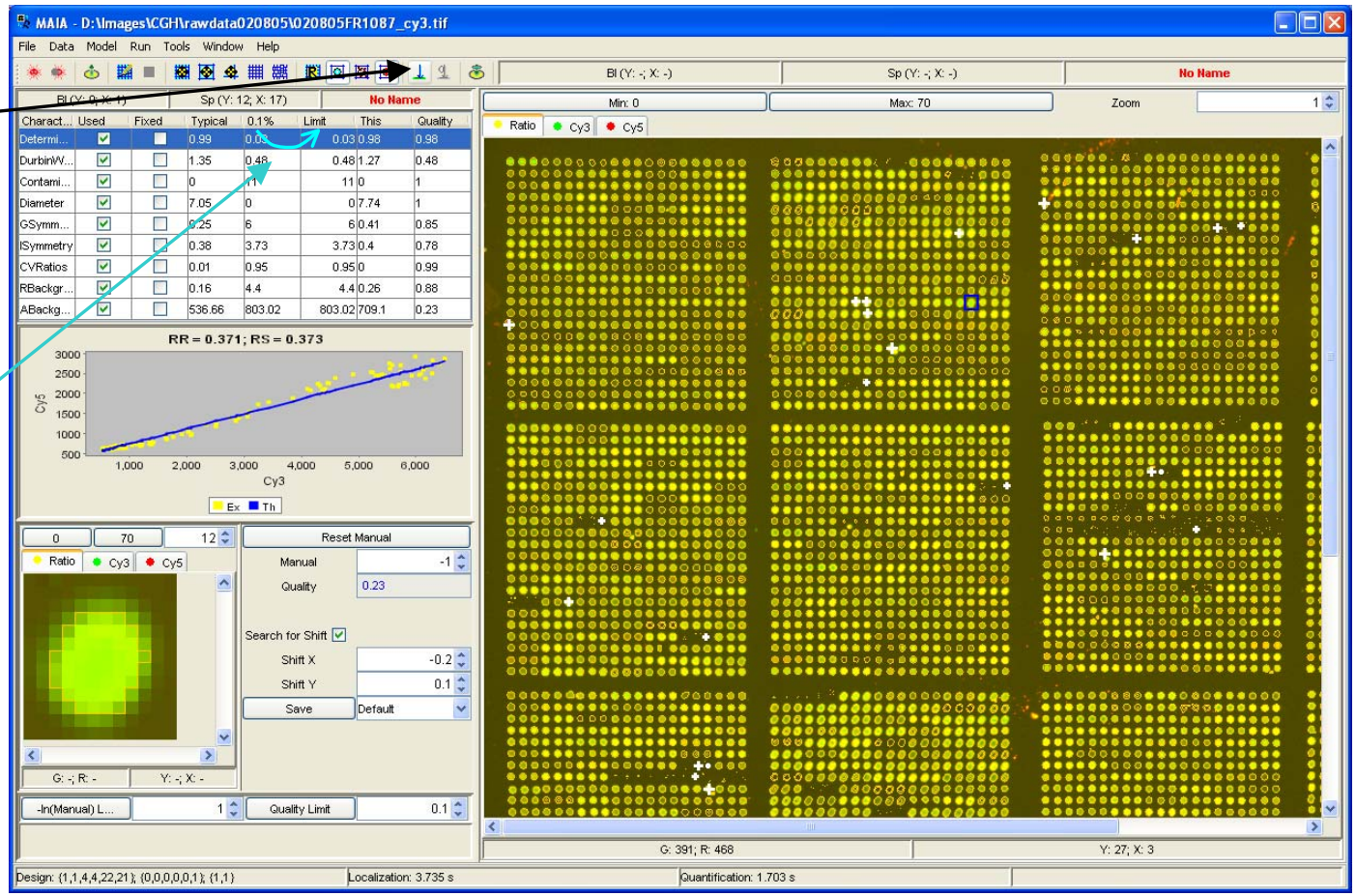
Automatic Limits Initialization

Limits can be automatically initialized using the “Init Limits” button from the Toolbar or the Menu Item “Run|Init Limits” (Ctrl+F5).

The procedure simply copies the values from the “0.1%” field into the “Limit” field of the table.

The quantile (0.1%) can be modified using the the Menu Item “Tools|Analysis Options” (Ctrl+A), tab “Quality”.

See page Quality Settings in Detail.

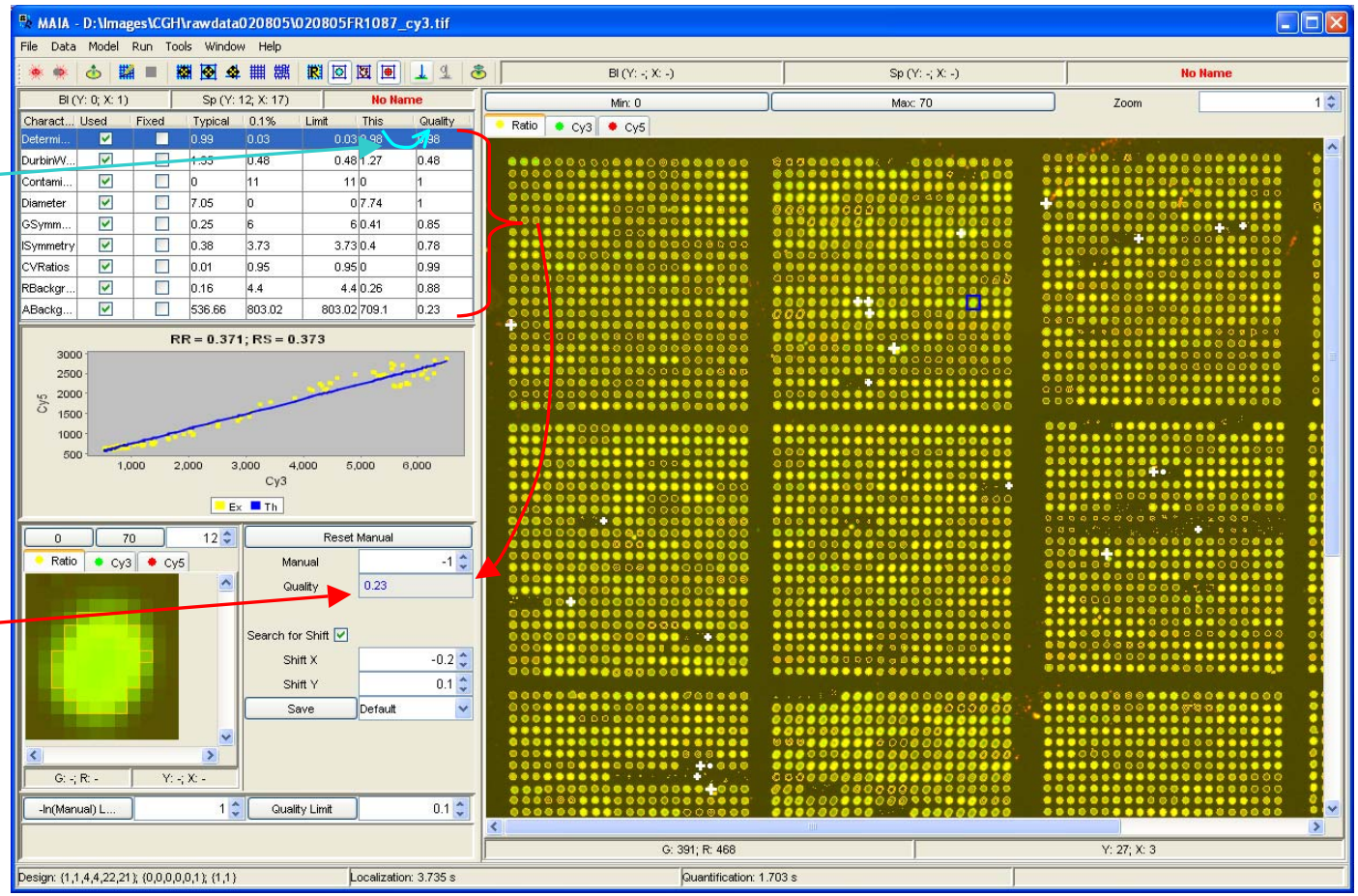


Quality Parameter

Using the limit value each quality characteristic is rescaled into the correspondent marginal quality parameter ∈[0;1].

See page Quality Characteristics.

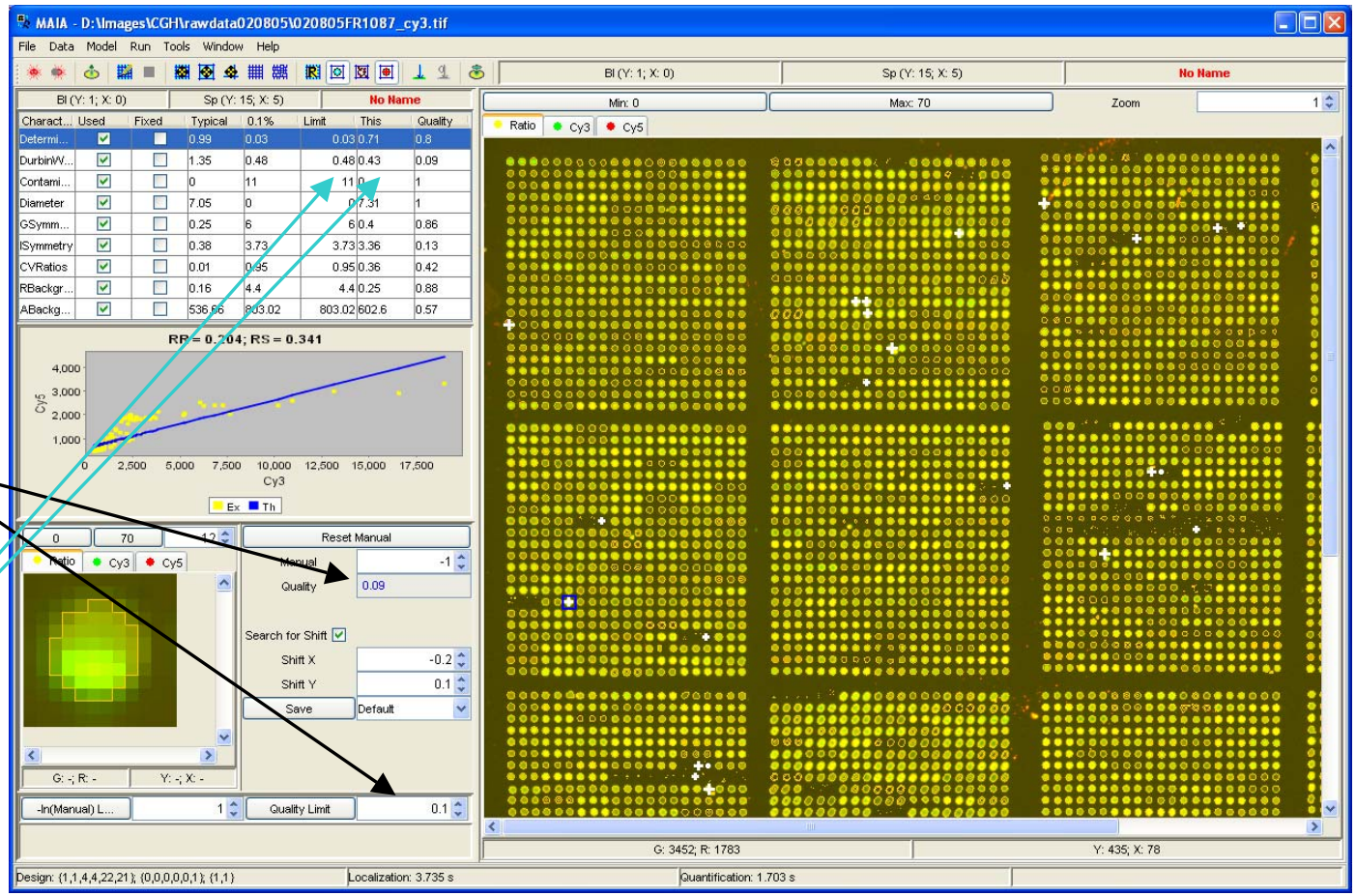
The minimal quality value from a set of marginal quality parameters is taken as an overall quality value.



“Bad” Spots

White crosses indicate “bad” spots, i.e. spots whose overall quality value is below the *Quality Limit* as defined by the “Quality Limit” spinner …

or, equivalently, if one of the quality characteristics of a spot exceeds the correspondent limit.

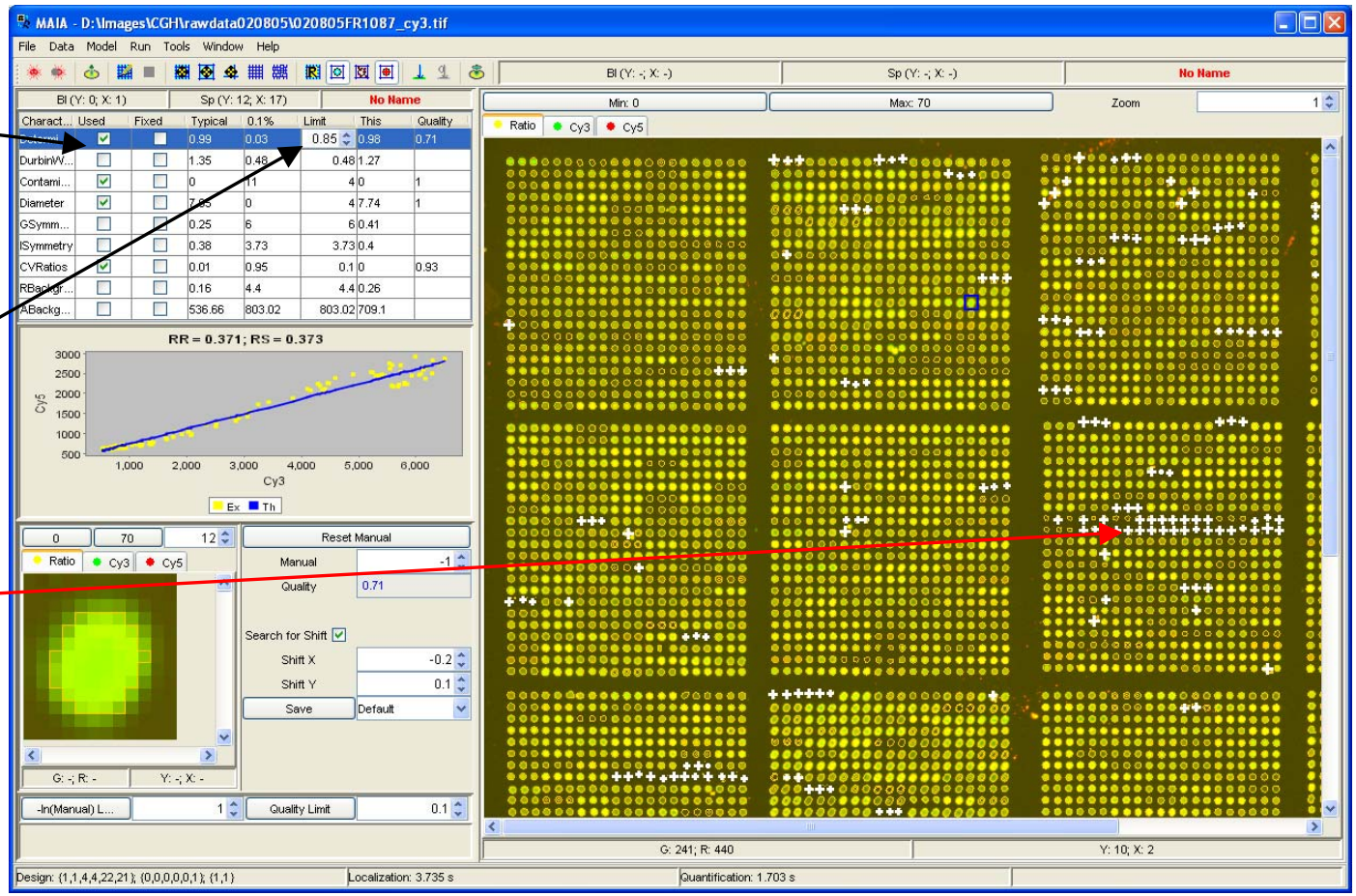


Manual Adjustment of the Limits

Using the table field "Used" one can select a set of quality characteristics, which seem relevant for this particular image.

For each used quality characteristics critical level (limit) can be further adjusted.

Limit adjustment should be continued until all spots, visually classified as "bad" spots, are flagged out.



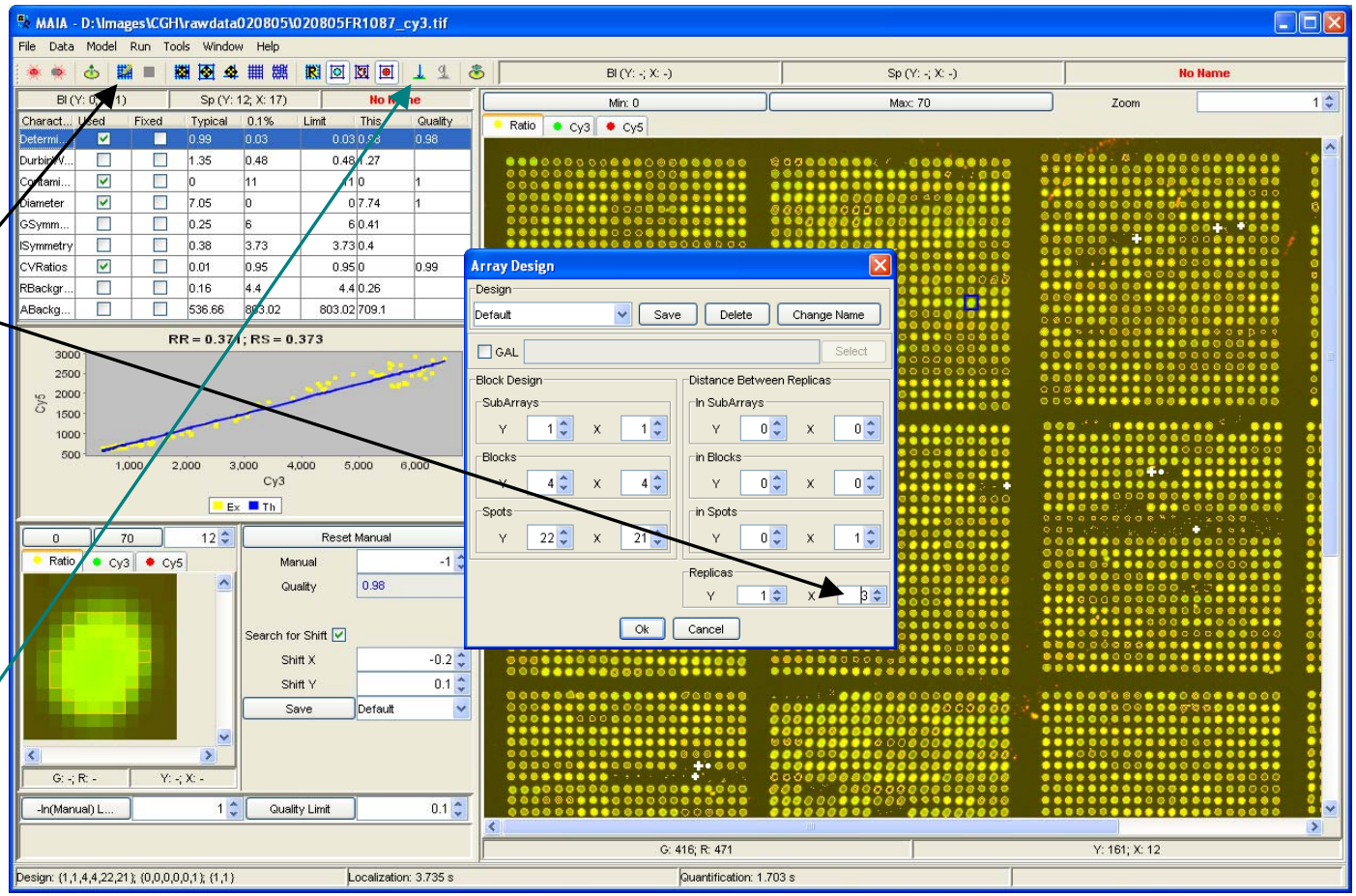
Quality Analysis (II): With Replicated Spots

On this image, three replicated spots are placed as neighbors in a row.

This can be signaled using the Array Design dialog (click the “Array Design” button from the Toolbar or select the Menu Item “Tools|Array Design” (F2)).

See page Array Design in Detail.

Using the “Init Limits” button from the Toolbar or the Menu Item “Run|Init Limits” (Ctrl+F5) the default Limits can be reconstructed.



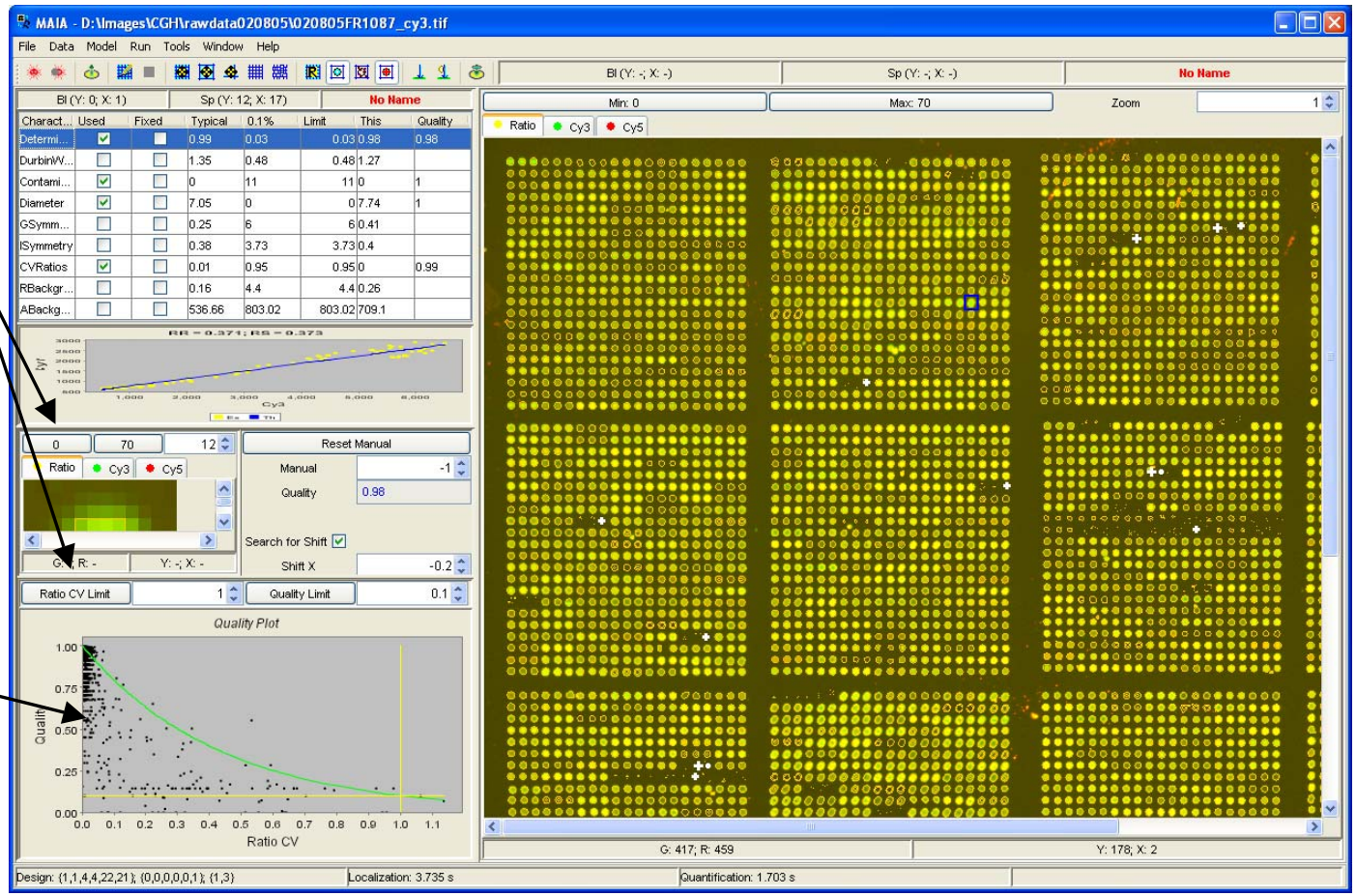
Quality Plot

Slide up the bars separating the panels and open up the quality plot:

y-axis is the overall quality value;
x-axis is the ratio variation coefficient (CV) of the replicates on the current array.

Each dot represents a replicate with the overall quality value at *y*-axis and ratio CV at *x*-axis.

See page Spot Quality Fit.

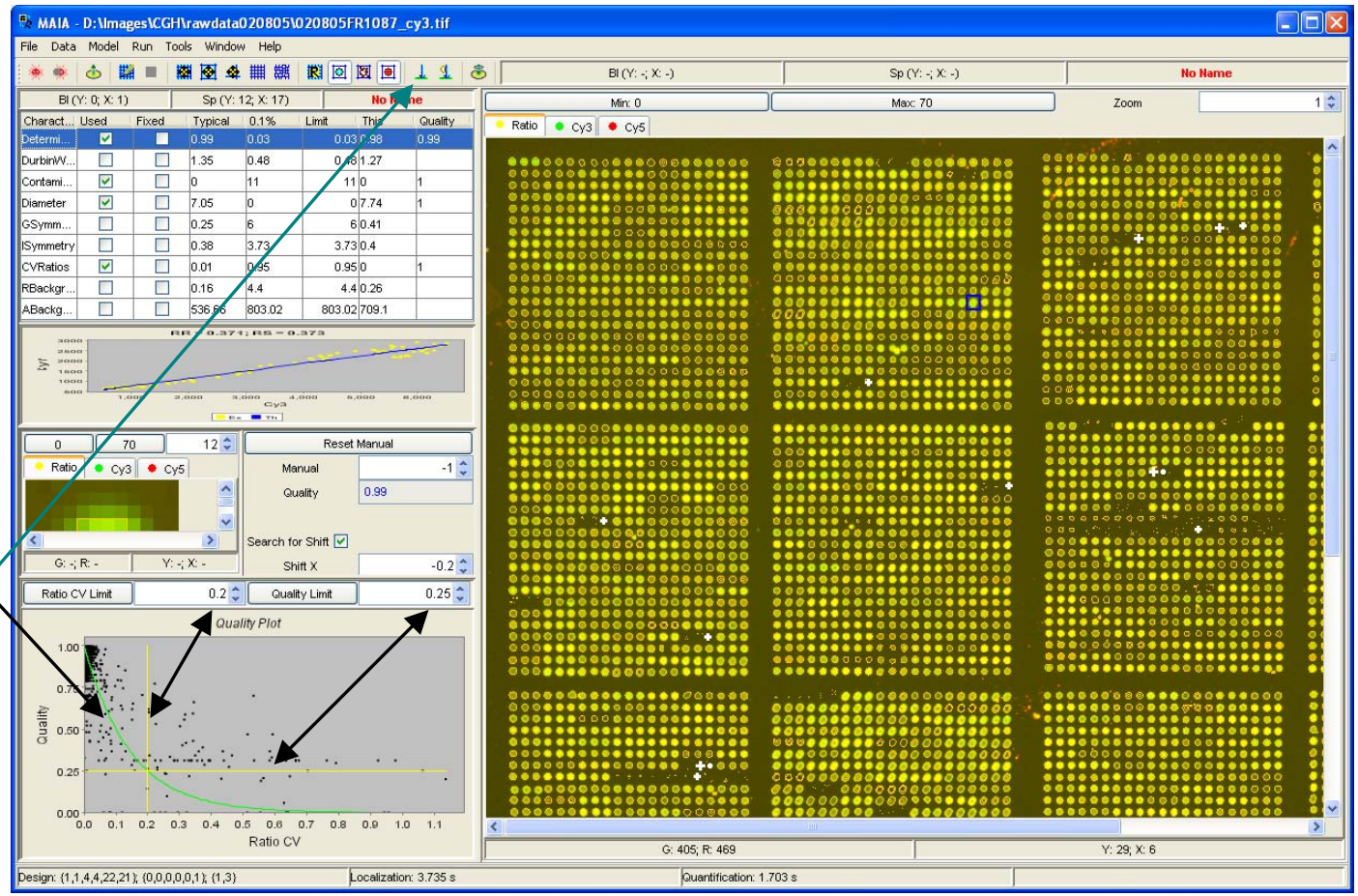


Quality Curve

Use the mouse pointer or the spinners “Ratio CV Limit” and “Quality Limit” to define the quality curve (green line).

See next page for details.

Before Limits Fitting it is advisable to reconstruct the default Limits (the “Init Limits” button from the Toolbar or the Menu Item “Run|Init Limits” (Ctrl+F5)).



Spot Quality Fit

The overall quality value is defined as:

$$Q = min_{i=1,\ldots,9}\{q_i^{wi}\}, \quad (1)$$

where $q_i = q_i(x_i) \in [0;1]$, $i = 1,\ldots,9$ are the marginal scaled quality parameters defined on page Quality Characteristics for $x_1 = CD$, $x_2 = DWS$, $x_3 = N$, $x_4 = D$, $x_5 = GS$, $x_6 = IS$, $x_7 = CVR$, $x_8 = UB$, $x_9 = AB$, and w_i are the weights that control the input of the correspondent quality components into the overall quality value. For the user-provided overall quality threshold $Q^{lim} \in [0;1]$, one can establish a link between the weight w_i and the critical value x_i^{lim} for each quality characteristic $i = 1,\ldots,9$:

$$w_i = log\{Q^{lim}\}/log\{q_i(x_i^{lim})\}, \text{ or } x_i^{lim} = q_i^{-1}(\{Q^{lim}\}^{1/wi}), \quad (2)$$

where $q_i(x_i^{lim})$ is the scaled quality parameter calculated for x_i^{lim}. The critical value x_i^{lim} sets up the limit such that if a certain characteristic i exceeds this limit, the correspondent quality parameter $q_i(x_i^{lim})$ will become lower than Q^{lim}.
The experimental quality parameters q_i, $i = 1,\ldots,9$ are obtained from the quantification procedure, whereas the weights w_i (or the critical values x_i^{lim}) are yet unknown. The problem of spot quality analysis is therefore converted into the problem of weights (w_i) estimation, which can be solved only if additional information is provided, for example, from the replicated spots on the same array or over a set of replicated arrays. The high-quality spots belonging to the same replicate are expected to demonstrate very close to each other ratio value. Relatively big difference between the observed ratios in the same replicate will signal that some of the spots from this replicate are irregular. To formalize this approach, we first define the quality value for the replicate:

$$Q_k = min_{j=1\ldots n}\{Q_{kj}\}, \quad (3)$$

where k enumerates the replicates, n is the number of spots in a replicate, and Q_{kj} is a spot quality value given by Eq. (1).
Substituting Eq. (1) into (3) yields

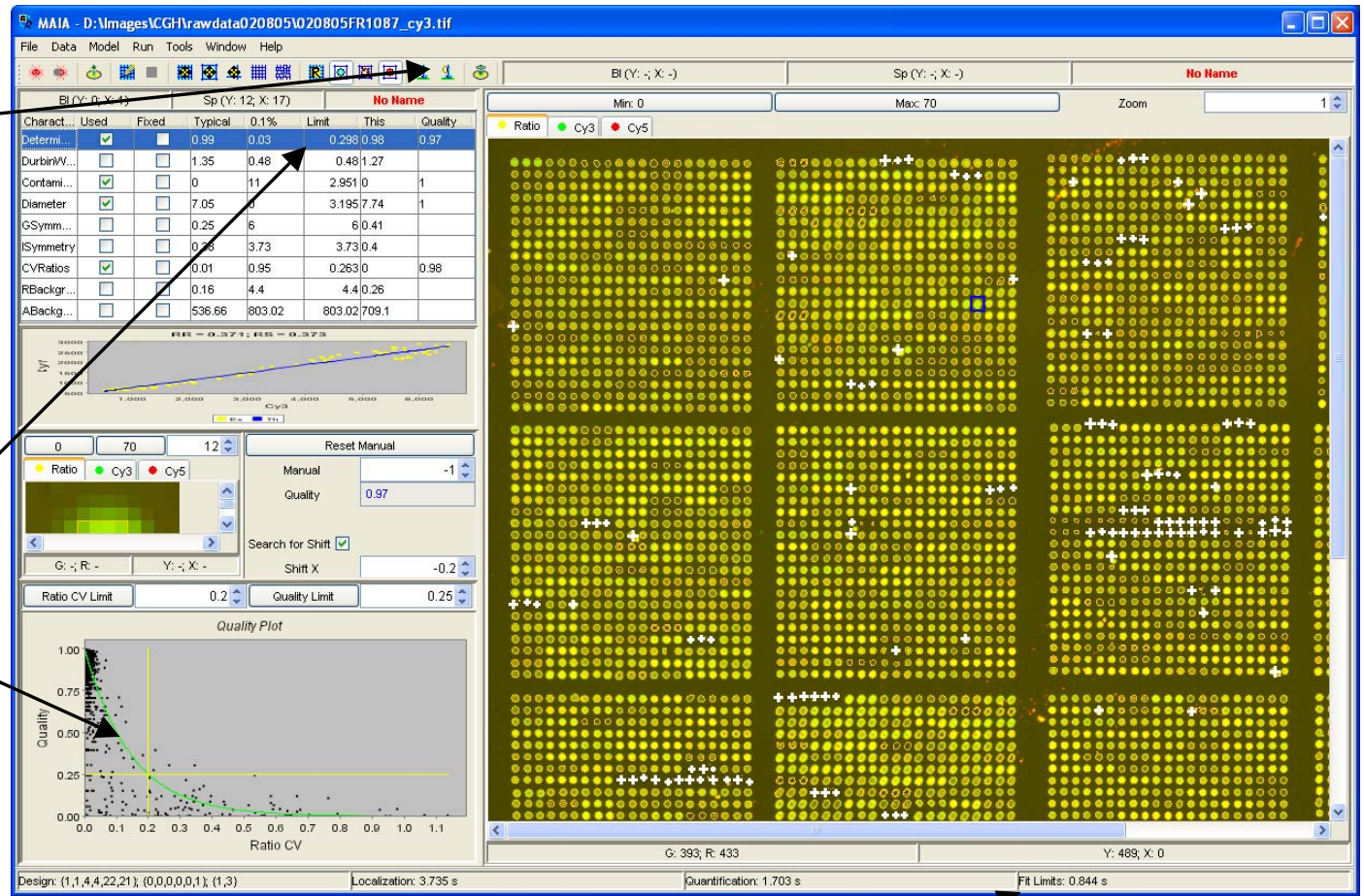
$$Q_k = min_{j=1\ldots n}\{min_{i=1,\ldots,9}\{q_{kji}^{wi}\}\} \quad (4)$$

where q_{kji} is the i-th scaled quality parameter of the j-th replicated spot in the k-th replicate.
The weights w_i, $i = 1,\ldots,9$ can be determined as the parameters ensuring the best fit of the obtained experimental quality values (Q_k versus V_k) to the user-defined (ideal) quality curve $f(V_k)$, where V_k is the ratio variation coefficient in the k-th replicate. $f(V_k)$ defines how fast the overall quality of the replicates must decrease with the increase of the ratio variation. The shape of the user-defined quality curve $f(V_k)$ should demonstrate monotonic decay. We always use the exponential function $f(V_k) = exp\{\}-V_k/V\}$, and in this case only the expected (typical) ratio variation coefficient V must be predefined.

Fit the Limits

The quality limits are fitted using the “Fit Limits” button from the Toolbar or the Menu Item “Run|Fit Limits” (F5).

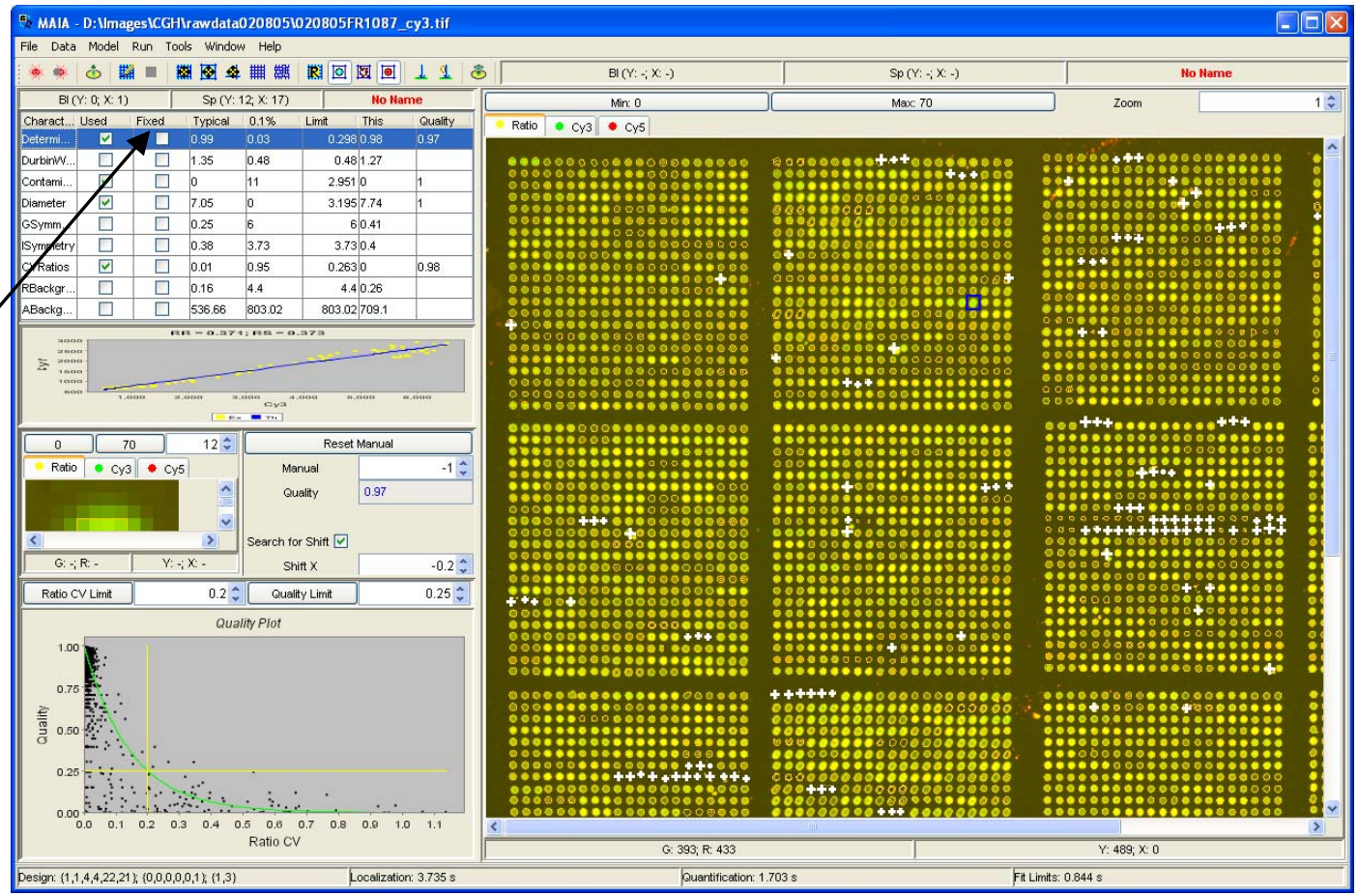
Quality fit gives the limits estimates for quality characteristics, such that the experimental quality dots are aligned along the user-defined quality curve.



Status of the Fit Limits procedure.

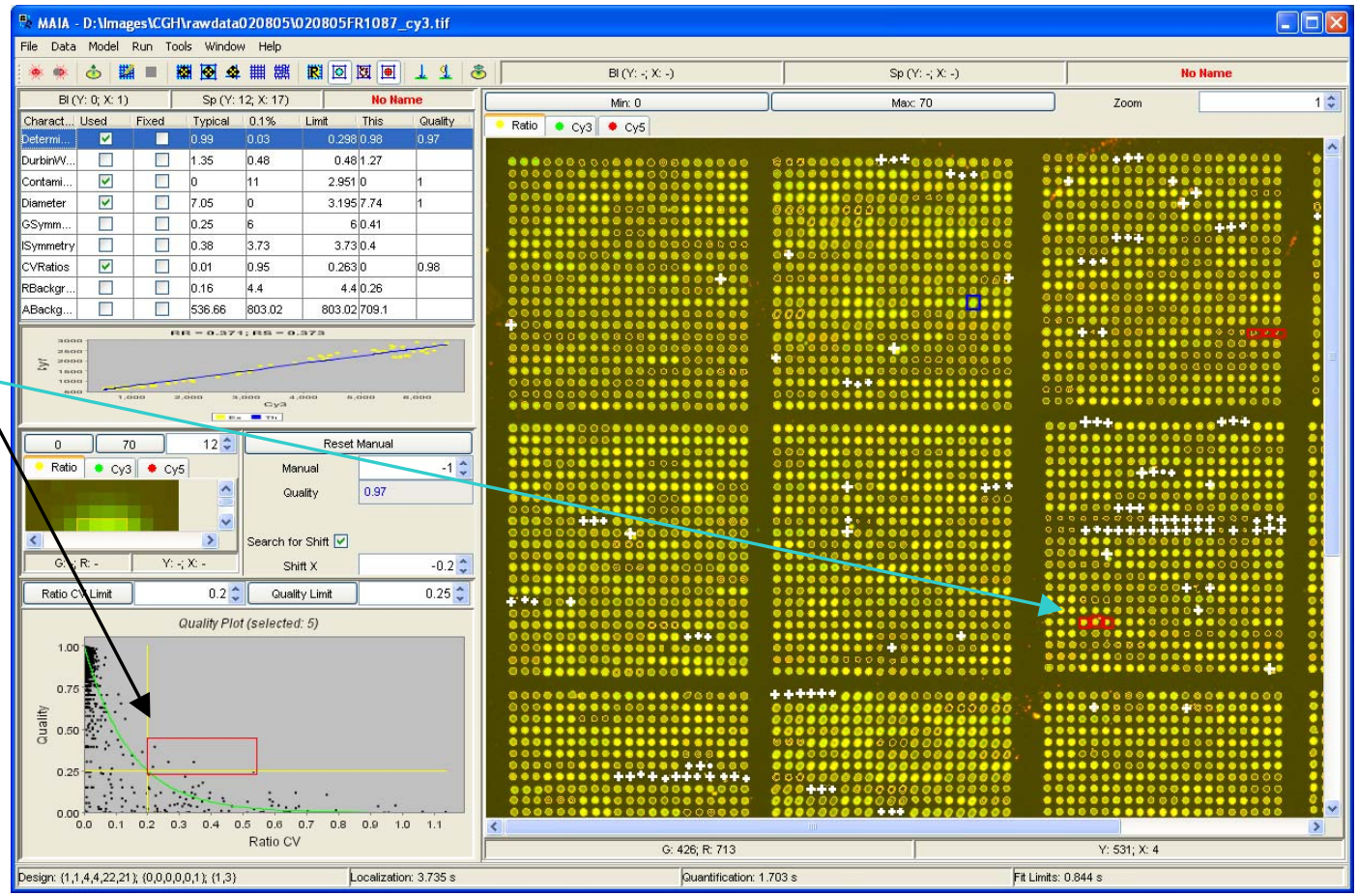
Fix the Limits

Certain limits, which are proved to be reasonable from previous experience, can be fixed, i.e. they are not changing during the fit.



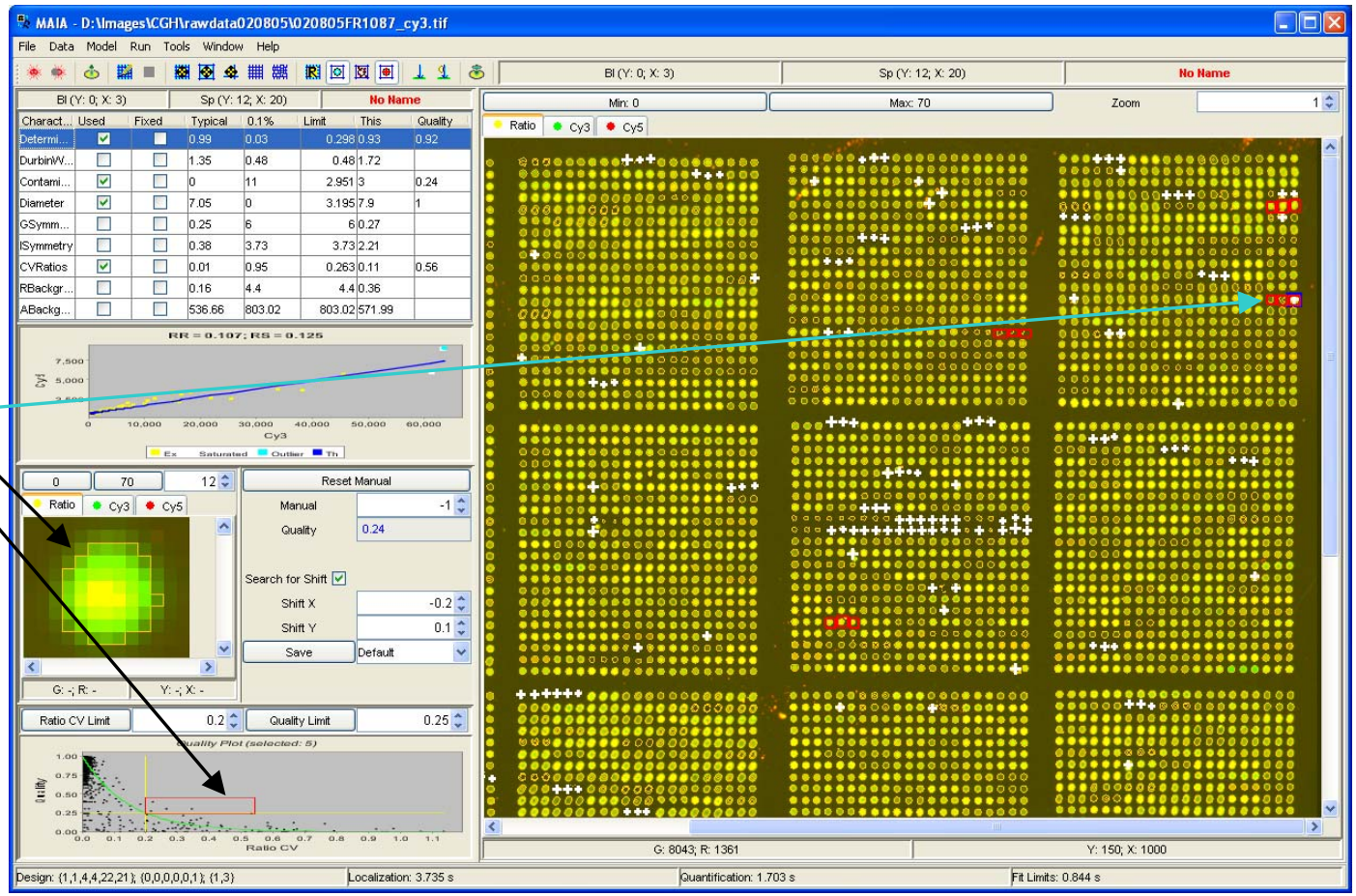
“Bad” Replicates

Left Click – Drag – Right Click on the Quality plot to select the replicates to be able to find them on the image.



Problematic Spots

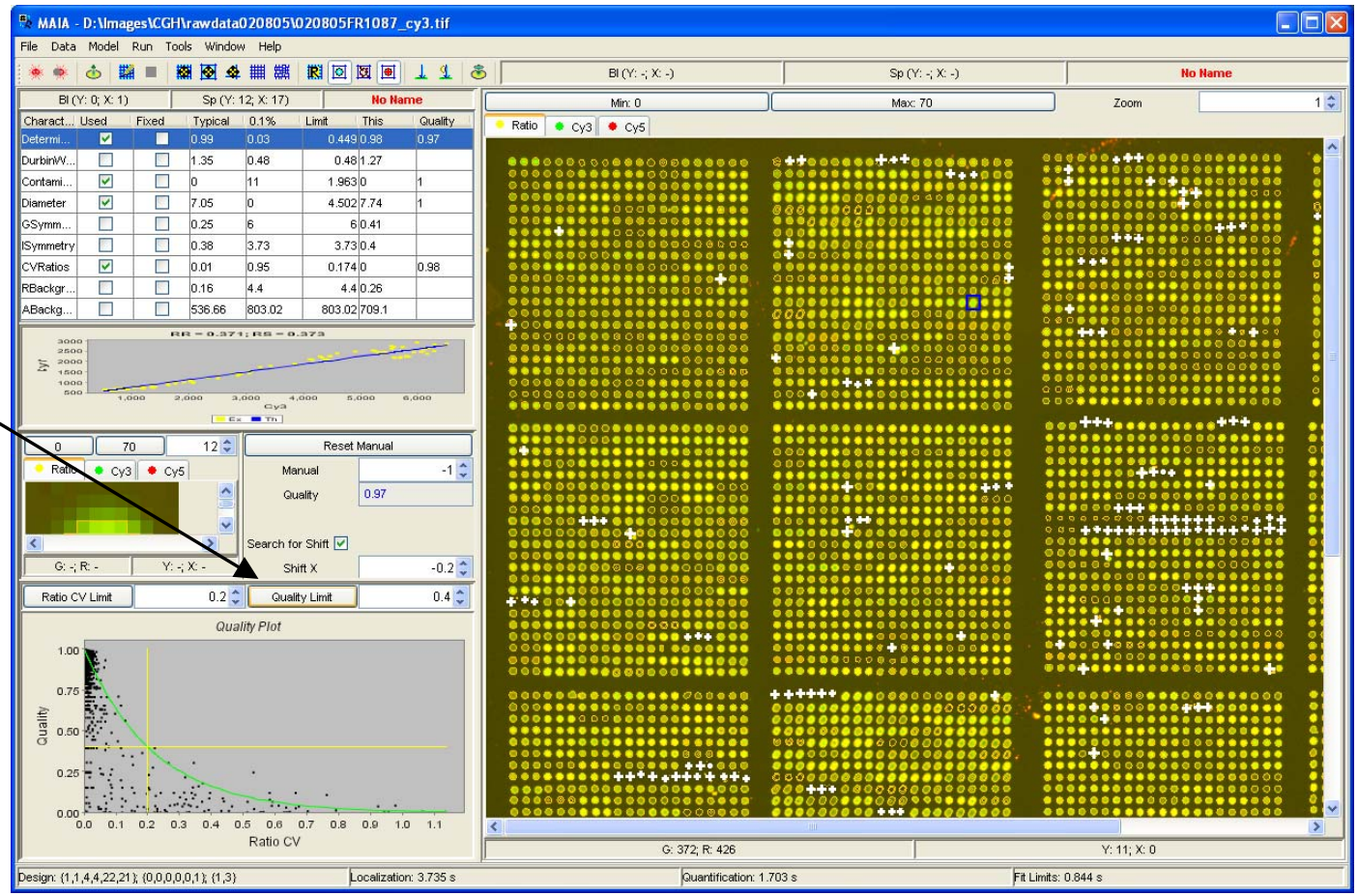
Closer look at the problematic spot may figure out the reason, why the quality value is not as low as we would expect.



Optimize the Quality Limit

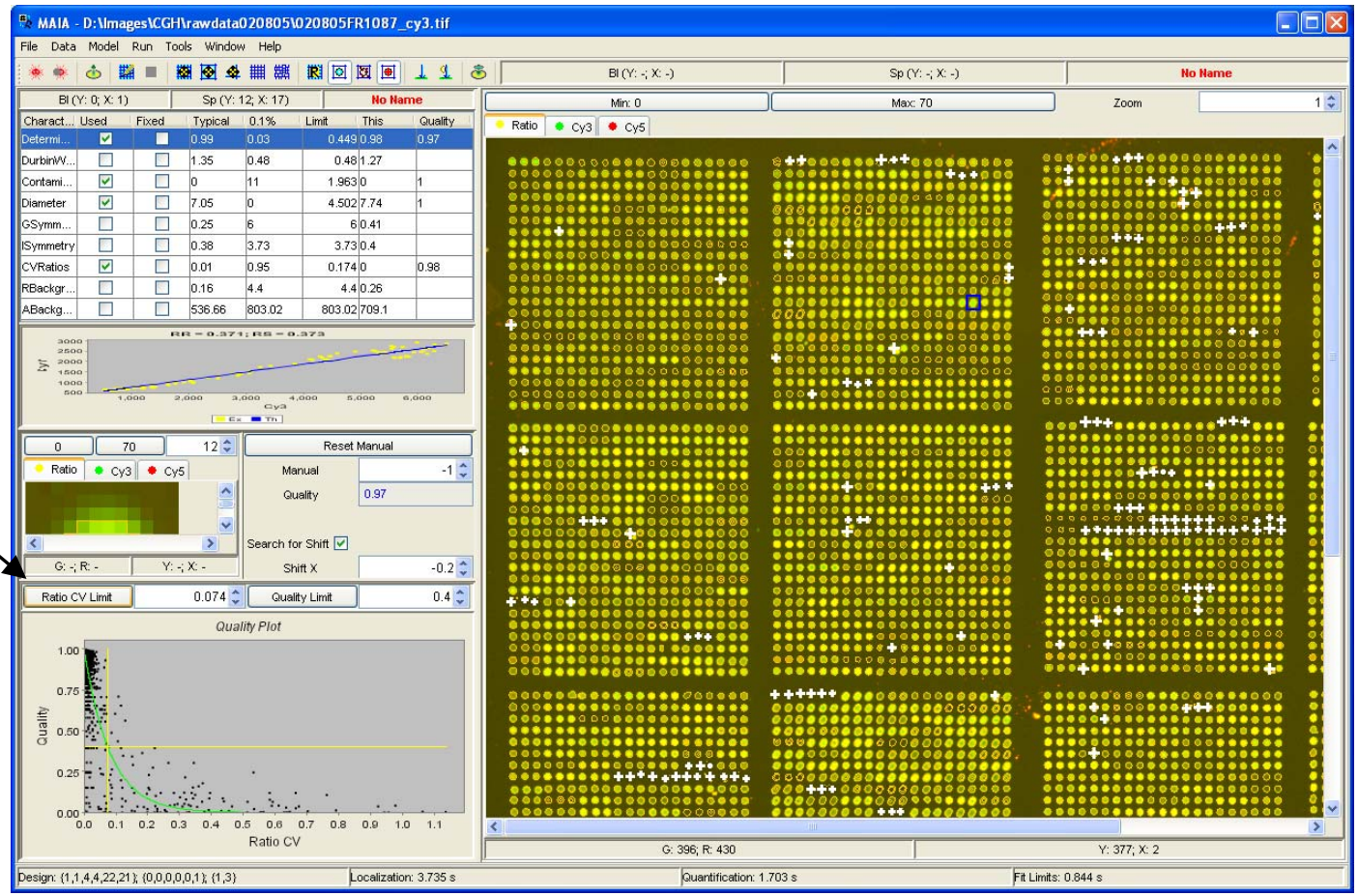
To optimize the position of the Quality Limit press the button “Quality Limit”.

A special procedure searches for the limit value such that the number of replicates in the “Bottom-Left + Top-Right” quadrants of the quality plot should be as small as possible, whereas in the “Bottom-Right+Top-Left” quadrants - as big as possible.



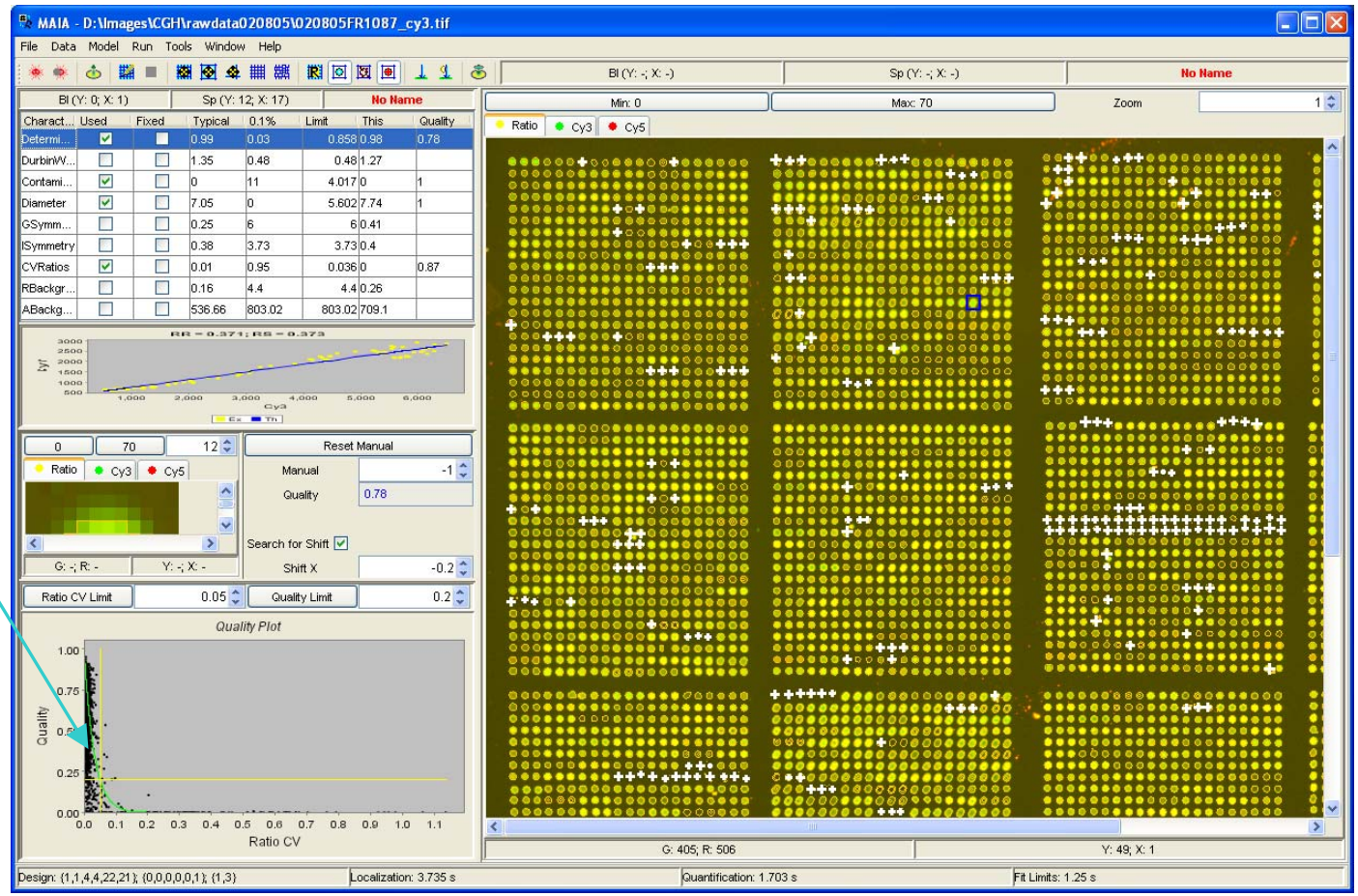
Optimize the Ratio CV Limit

Using the button “Ratio CV Limit” the correspondent limit is set into a value ensuring the best exponential approximation for the “cloud” of replicates (black dots).



Quality Plot

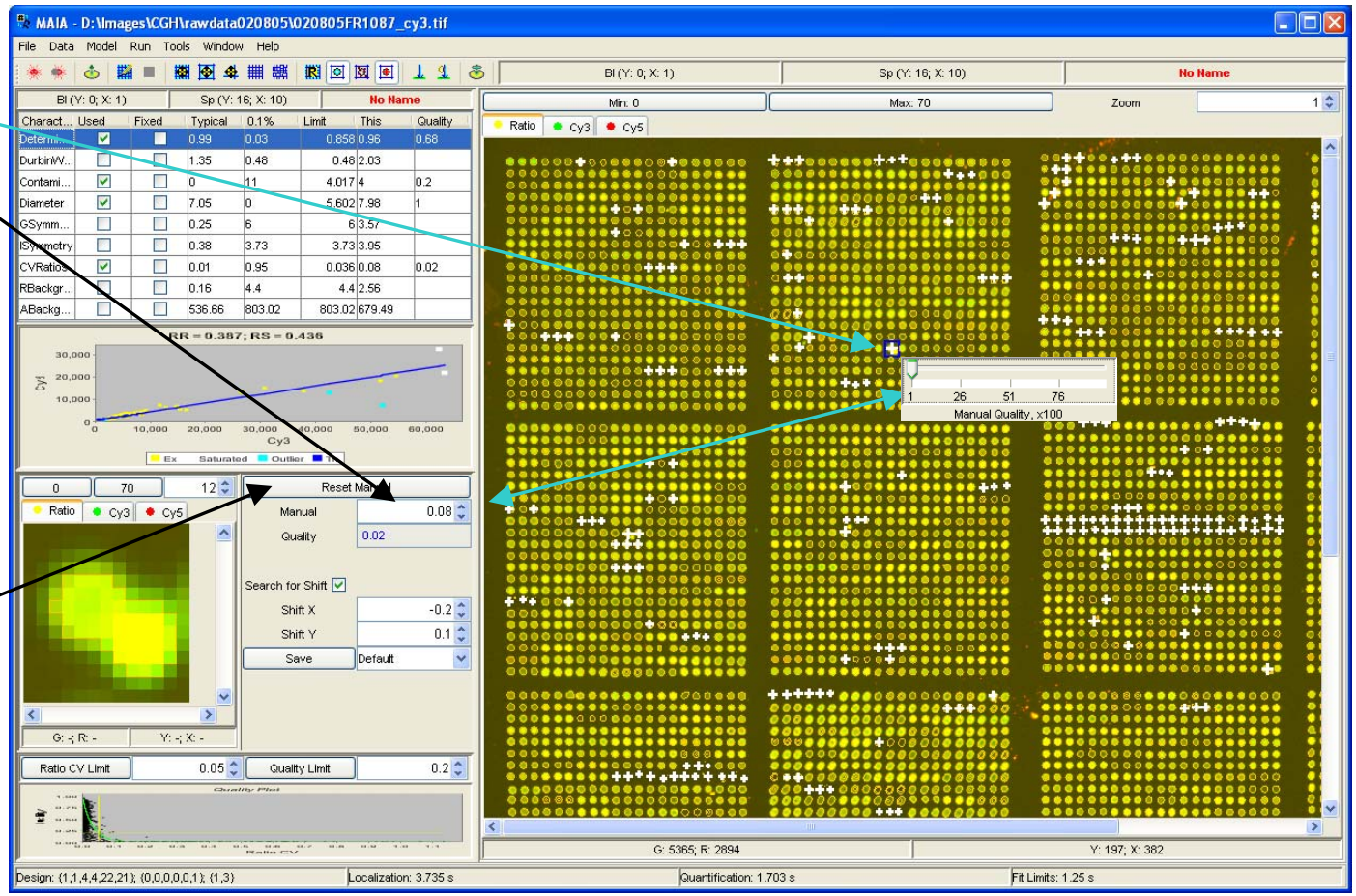
A somewhat more stringent quality curve is applied.



Manual Spot Characterization

Using the mouse right button or the spinner "Manual", any spot can be assigned a certain value from the interval [-1;1], which can further be used, for example, as an additional parameter of quality.

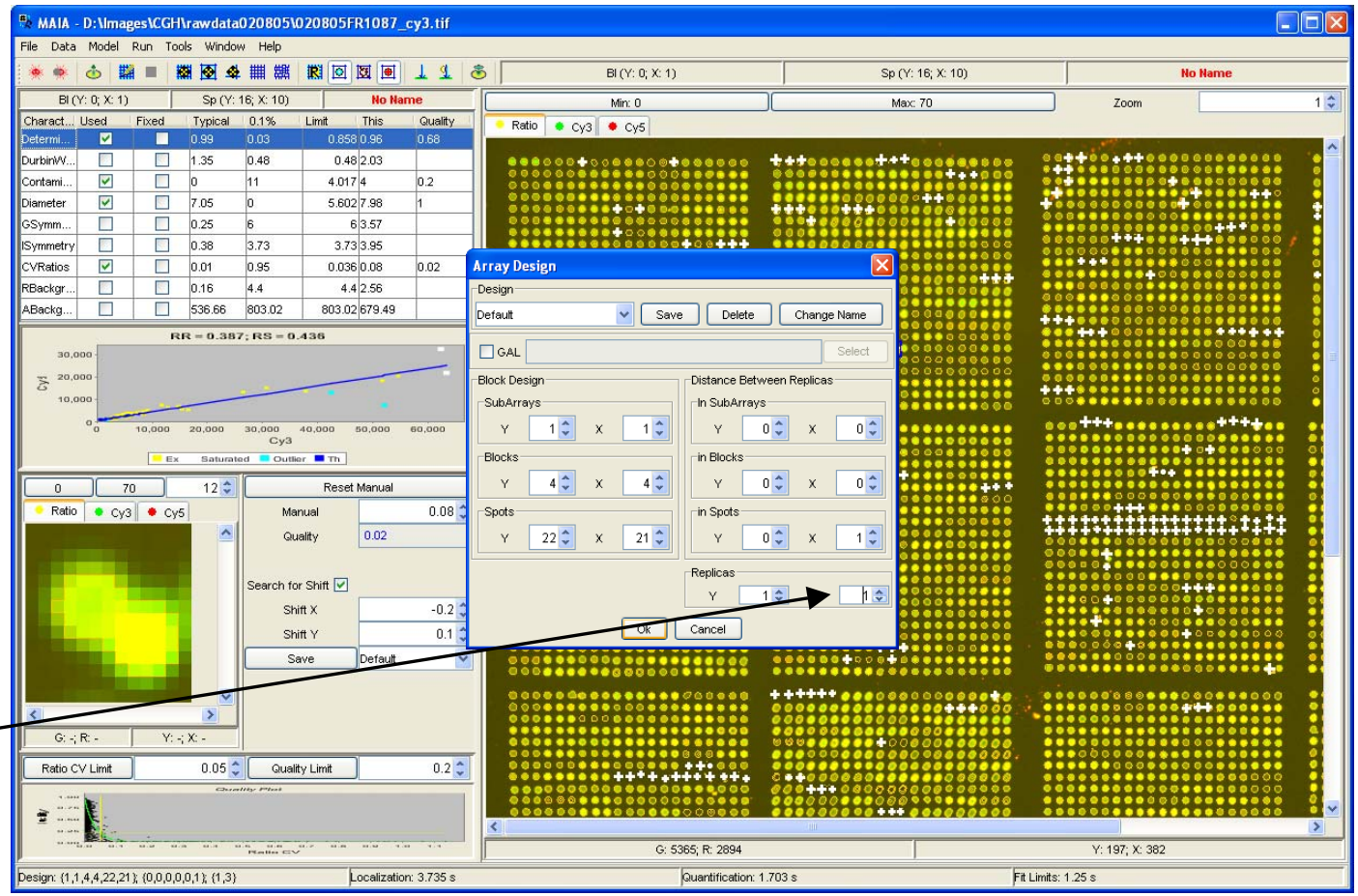
The "Reset Manual" button sets the Manual parameter for all spots on the array in –1.



Quality Analysis (III): Manual Classification of the Spots

To demonstrate possibilities of the quality analysis based on used classification of the spots, we come back to the case, where replicated spots are not available.

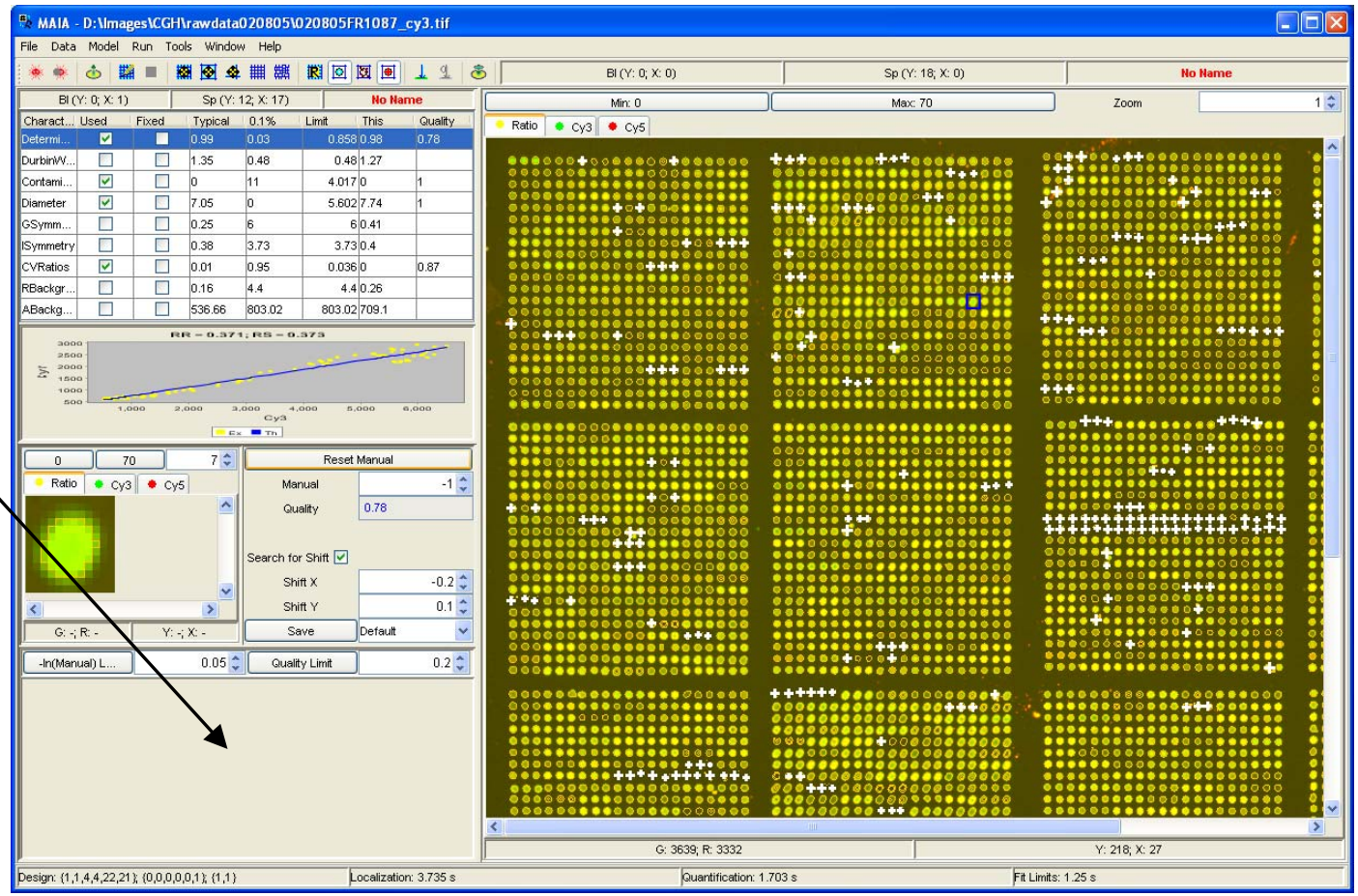
Use the Array Design dialog by clicking the "Array Design" button from the Toolbar or selecting the Menu Item "Tools|Array Design" (F2)).



See page Array Design in Detail.

Quality Plot Removed

Disregarding the replicates clears up the quality plot.

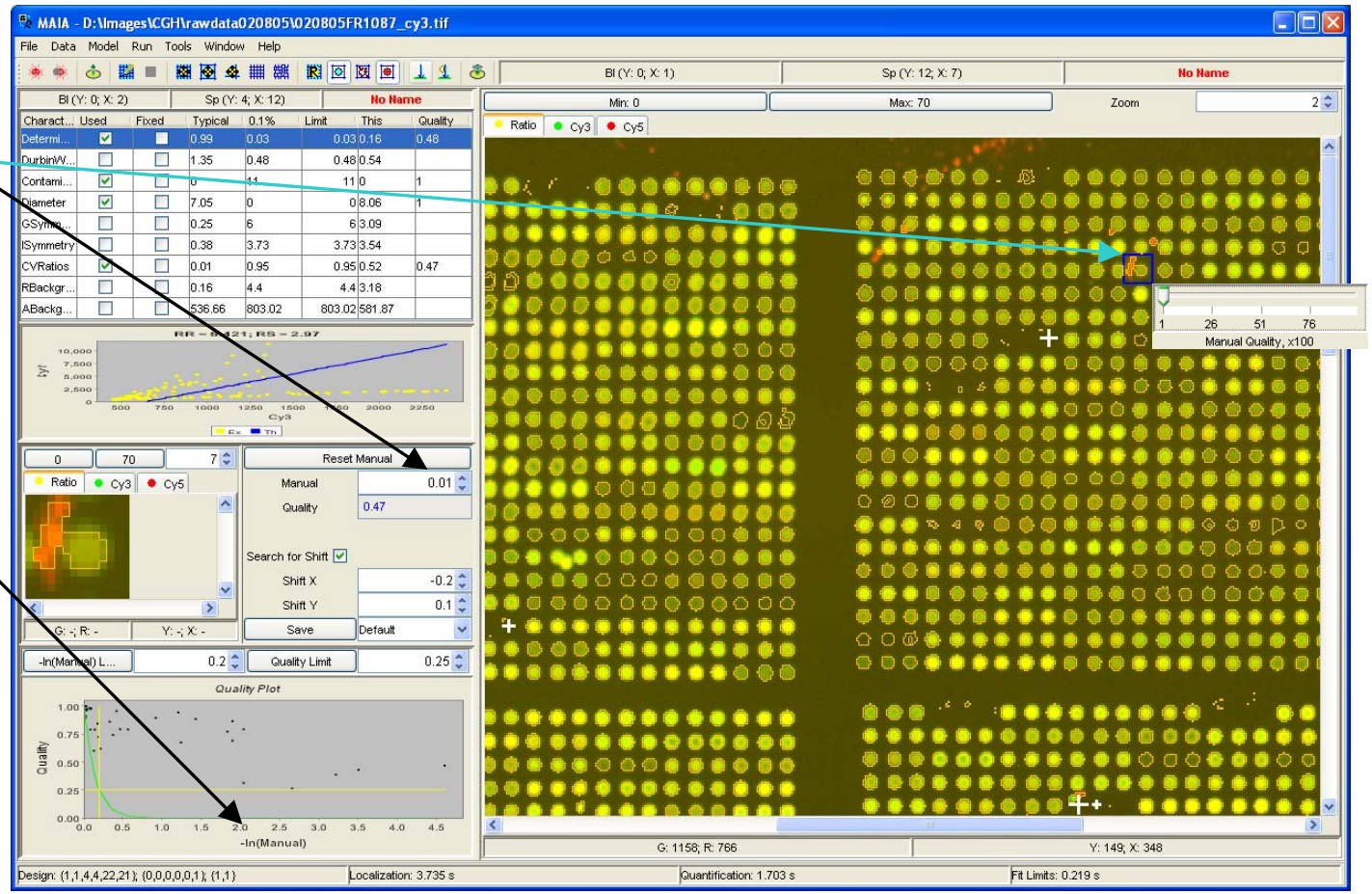


Manual Spot Quality Quantification

Using the mouse right button or the spinner "Manual", user can assign to any spot a certain value from the interval [0;1], reflecting the user appreciation of the quality of the spot.

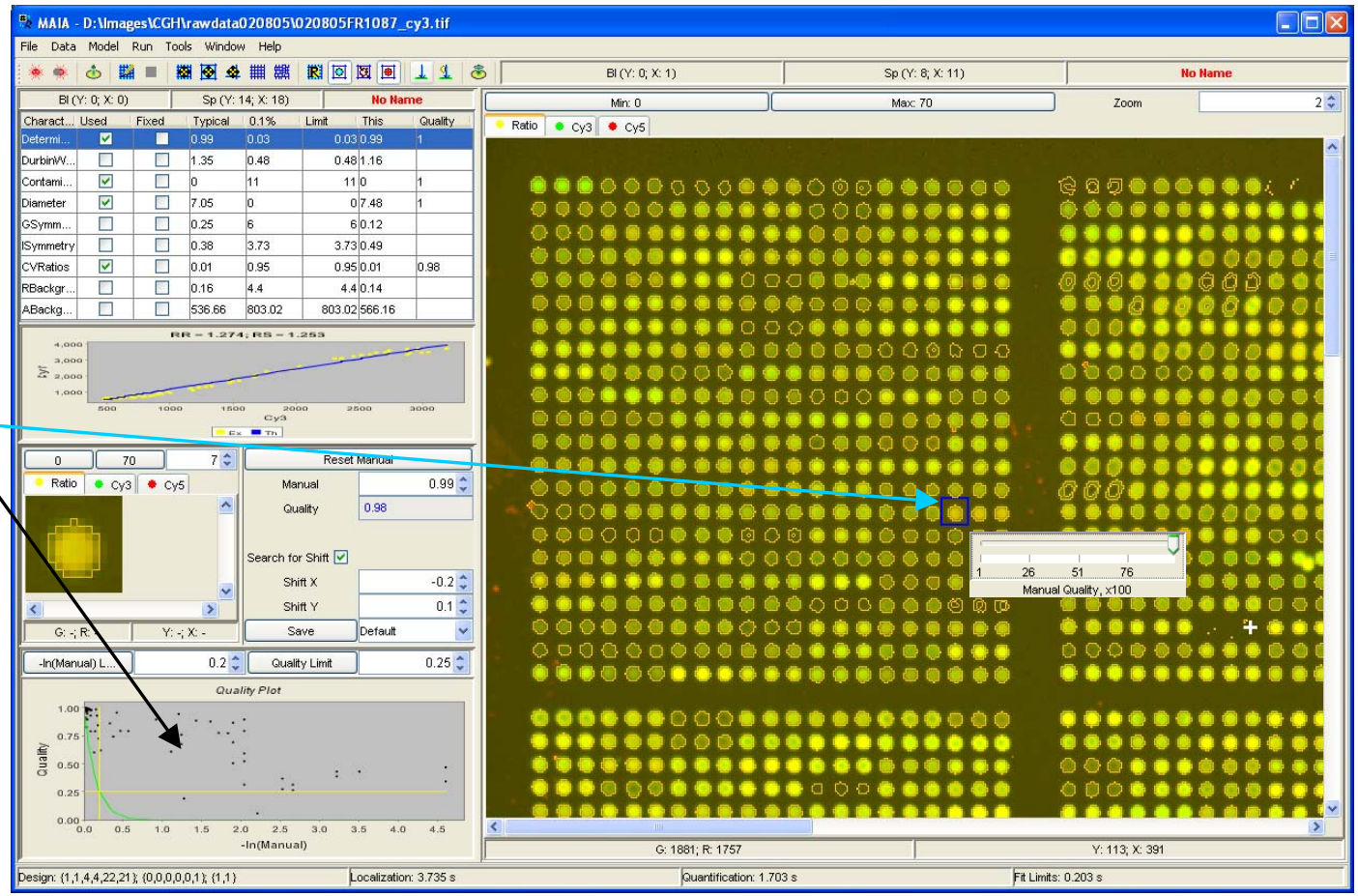
These manually defined values (z) are converted as –ln(z) to create the *x* axis of the quality plot. *y-axis: the overall quality parameter as before.*

Negative values of the Manual parameter, admissible in the spinner "Manual", will be ignored in the quality plot.



Representative Set of Spots is Required

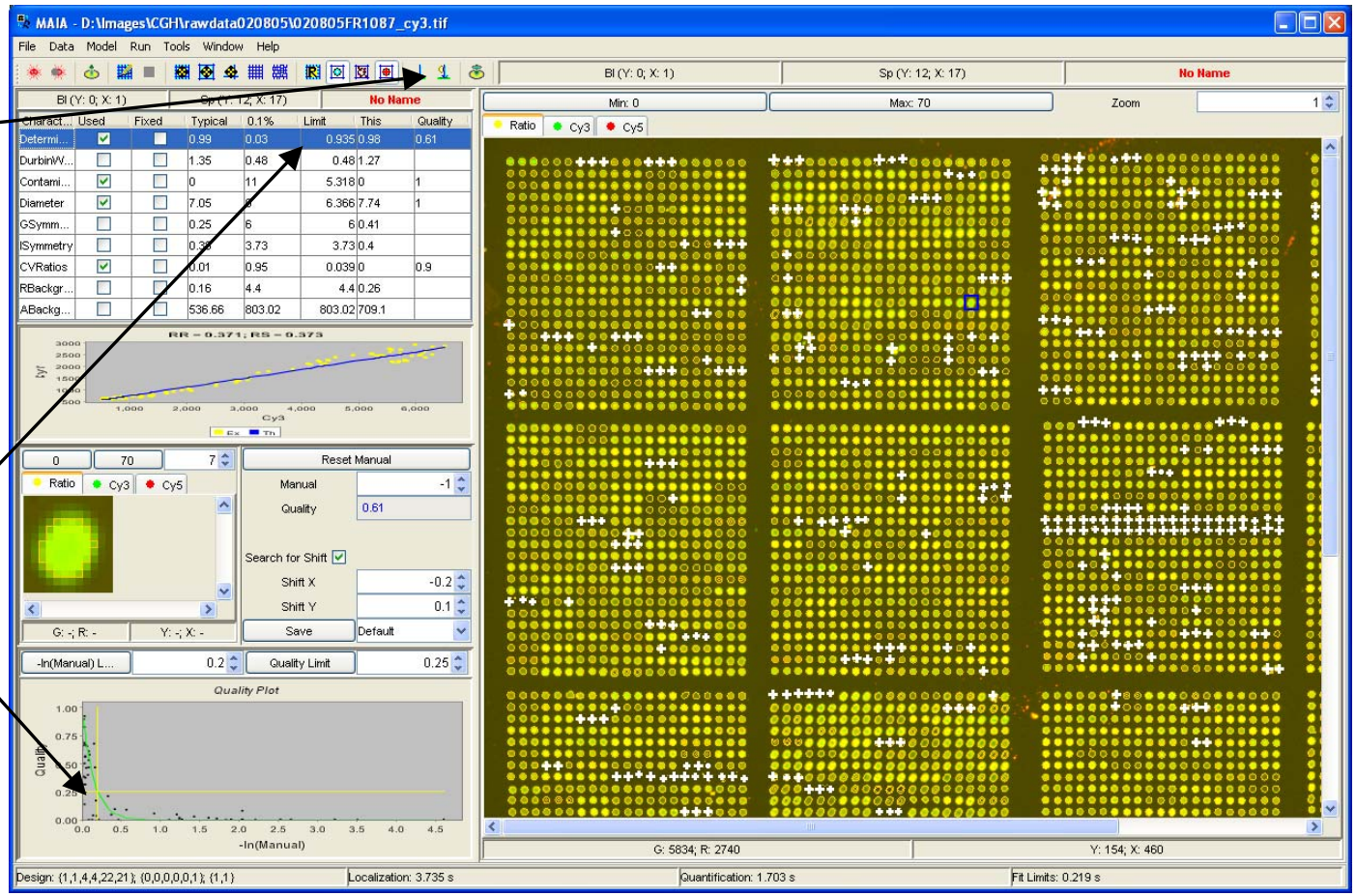
It is important to supply sufficient amount of the representative spots, i.e. spots covering wide quality range and containing all sorts of deficiencies, relevant for the used experimental setup/design.



Fit the Limits

Fit the quality limits by the "Fit Limits" button from the Toolbar or by the Menu Item "Run|Fit Limits" (F5).

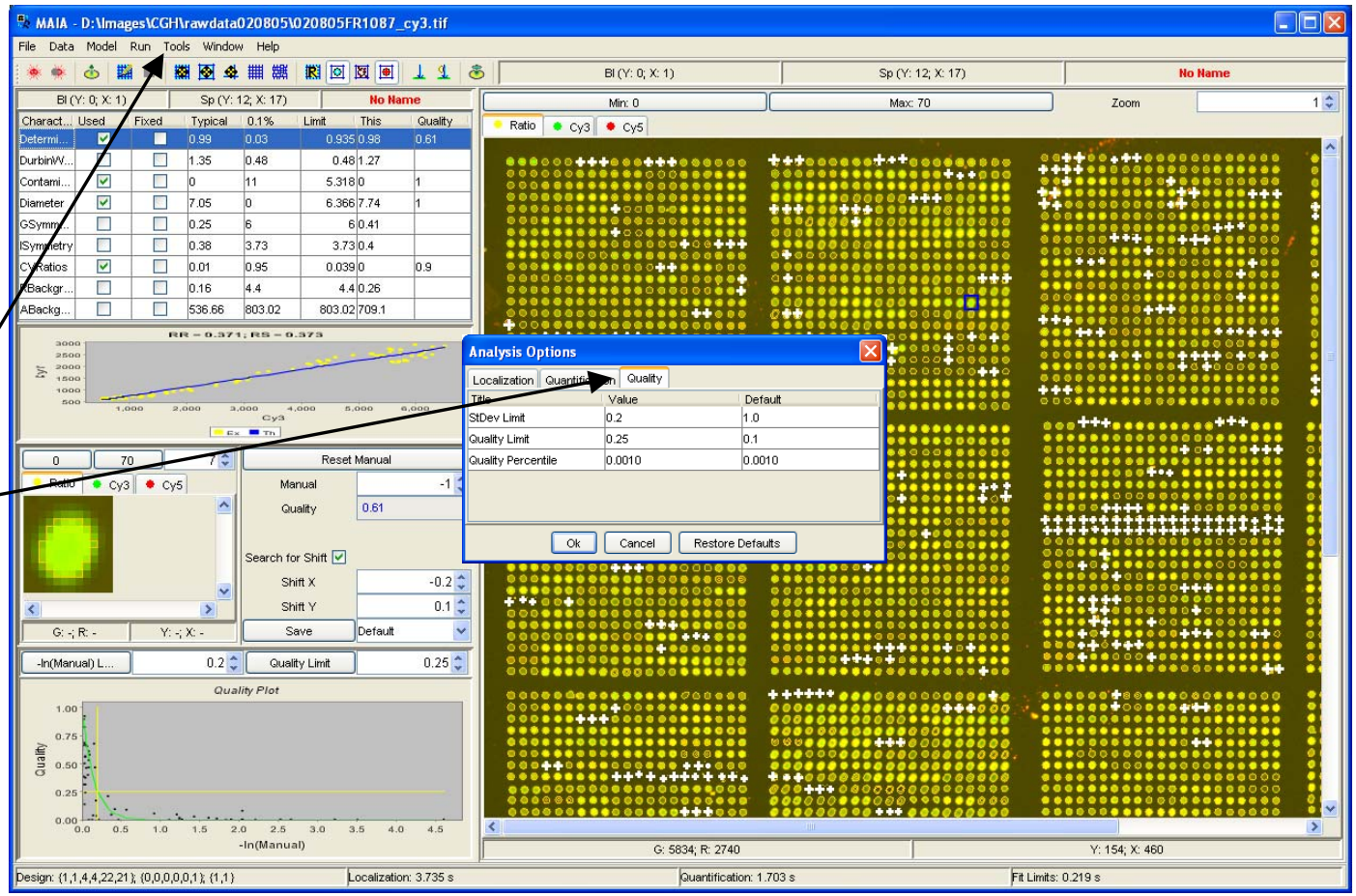
Quality fit gives the limit estimates for quality characteristics such that the experimental quality dots are aligned along the user-defined quality curve.



Quality Settings

Several settings that may influence the quality analysis are available through the Menu Item “Tools|Analysis Options” (Ctrl+A), tab “Quality”.

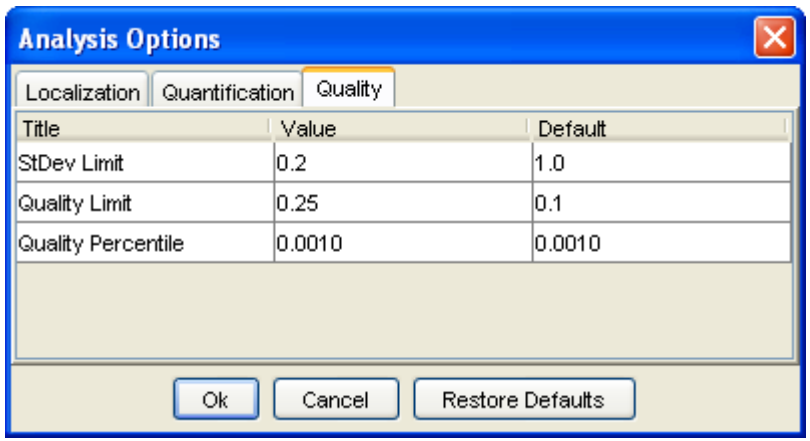
See next page for details.



Quality Settings in Detail

StDev Limit is a characteristic value of the user-defined (ideal) quality curve.

Quality Limit is the limit such that the spots with the overall quality values below this limit will be indicated by a cross.



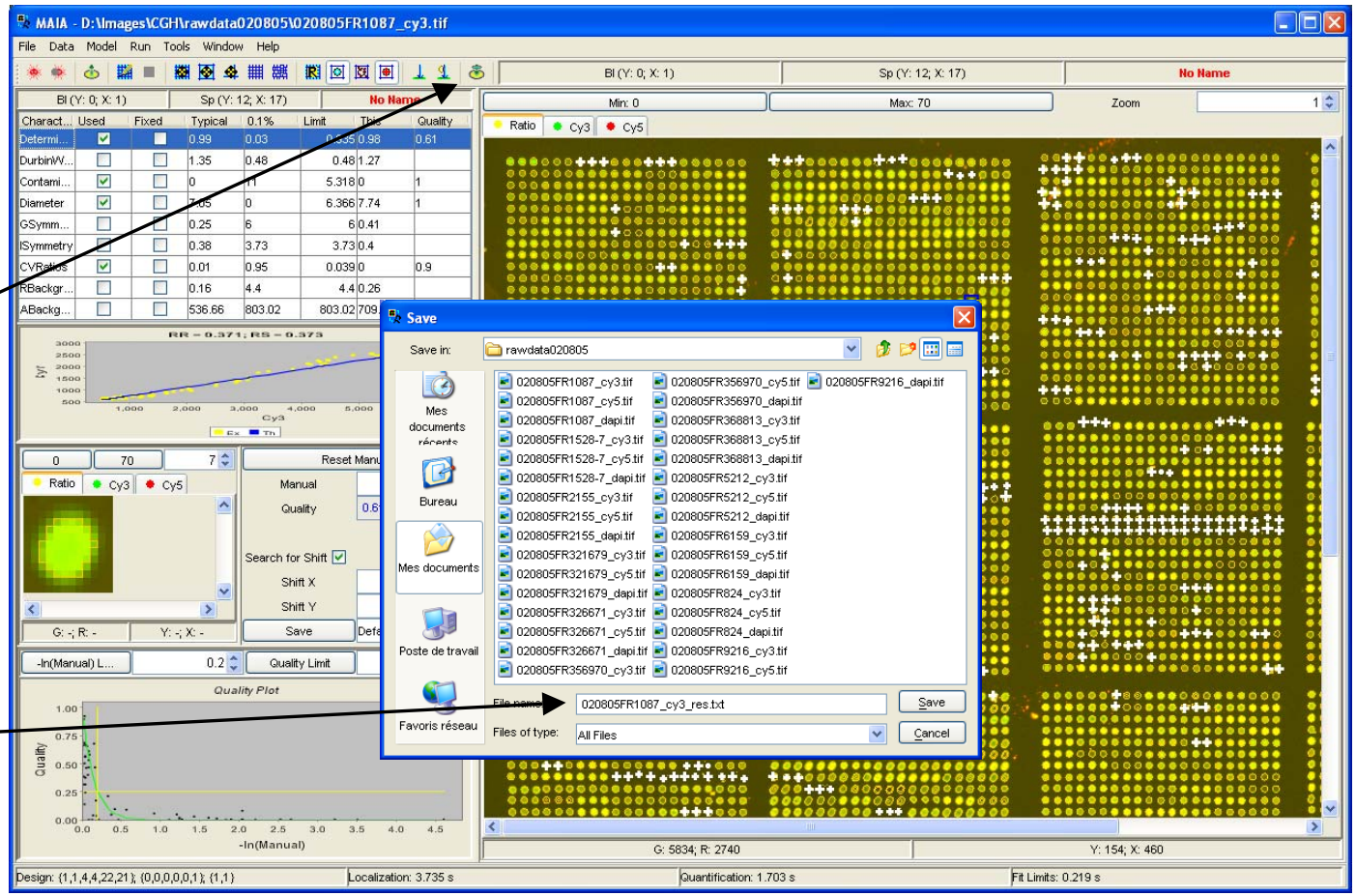
Quality Percentile establishes the values of the quality characteristics in the sorted lists of the quality characteristics (built up based on the results for all spots from the array) that will be displayed in the correspondent field of the spot characteristics table and eventually will be used to initialize the quality limits.

Save the Results

To save the results of quantification and quality analysis use the “Save Results ...” button from the Toolbar or the Menu Item “File|Save|Results …” (Ctrl+S).

The results are saved as a table in the text file (importable into Microsoft Excel).

A list of fields of the exported table can be found on the next page.



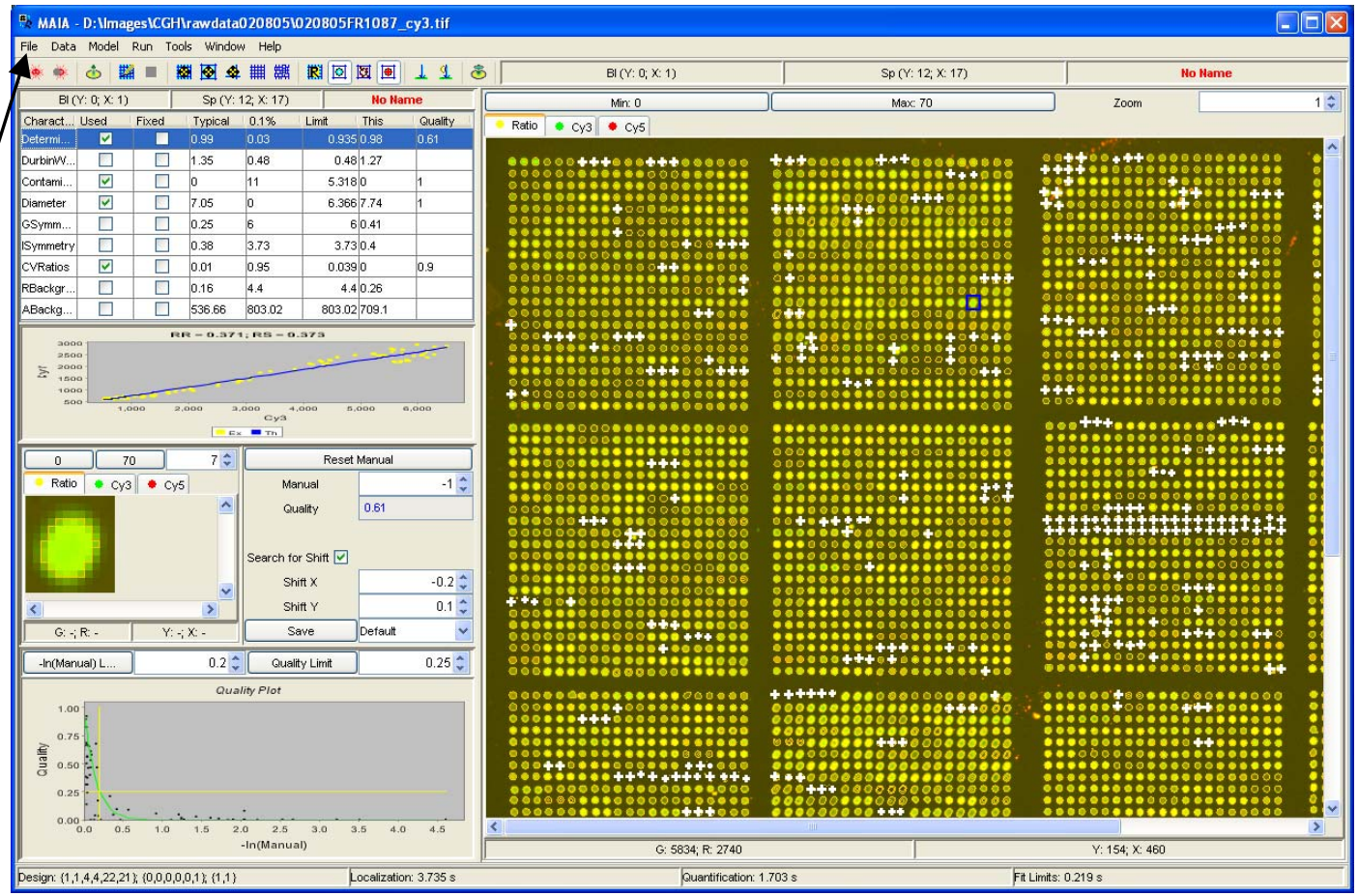
Output Table Fields

Block	Microarray block index
Column	Column coordinate (within the block) of the current spot
Row	Row coordinate (within the block) of the current spot
ID	Clone ID
Name	Clone Name
X	X coordinate of the spot center (in pixels)
Y	Y coordinate of the spot center (in pixels)
RR	Regression Ratio
RS	Segmentation Ratio
Overall Quality	Overall quality value
Manual	User-defined quality value
Determination*	Coefficient of determination of the linear regression
DurbinWatson*	Durbin-Watson parameter for the residuals of the linear regression fit
Contamination*	Amount of aberrant pixels flagged out by the filtering procedure
Diameter*	Diameter of the spot
GSymmetry*	Geometrical symmetry
ISymmetry*	Intensity symmetry
CVRatios*	Coefficient of variation of two ratios, one is based on the segmentation approach and the other one is based on the linear regression approach
RBackground*	Uniformity of the background around the spot
ABackground*	Absolute level of the background in the proximity of the spot
S Cy3	Mean intensity within the spot (Cy3 channel)
S Cy5	Mean intensity within the spot (Cy5 channel)
S Cy3 Sd	Standard deviation of the spot intensity (Cy3 channel)
S Cy5 Sd	Standard deviation of the spot intensity (Cy5 channel)
S Cy3 Pixels	Number of pixels within the spot (Cy3 channel)
S Cy5 Pixels	Number of pixels within the spot (Cy5 channel)
B Cy3	Mean background intensity (Cy3 channel)
B Cy5	Mean background intensity (Cy5 channel)
B Cy3 Sd	Standard deviation of the background intensity (Cy3 channel)
B Cy5 Sd	Standard deviation of the background intensity (Cy5 channel)
B Cy3 Pixels	Number of background pixels (Cy3 channel)
B Cy5 Pixels	Number of background pixels (Cy5 channel)

*) Quality characteristic. For each used quality characteristic the program adds one more field (quality parameter) in the table (as in the field “Quality” of the quantification table). The name of the field is formed by adding the prefix “Q” to the correspondent quality name (e.g. for the Determination quality characteristics the field name will be “Q Determination”). To this name it finally adds in the brackets the correspondent limiting value taken from the field “Limit” of the quantification table.

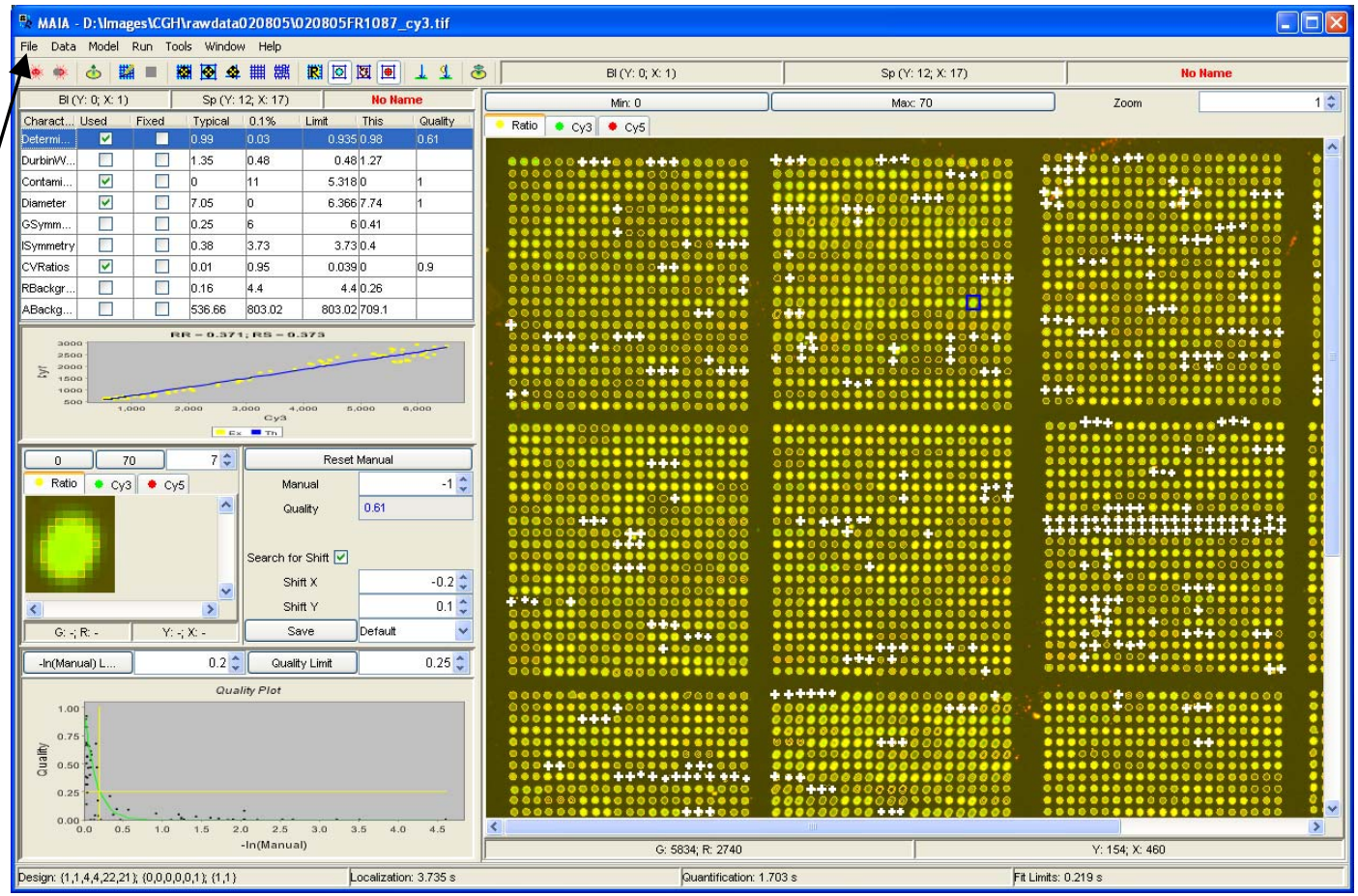
Save the Experiment: Experiment File

The whole experiment (results, parameters, grid, and other settings) can be saved on the disk (using the Menu Item "File|Save|Experiment …" (Ctrl+W)) in the internal (binary) format to be able to restore it (using the Menu Item "File|Load|Experiment …" (Ctrl+R)) in the future to reanalyze the data.



Set Batch Options

Using the Menu Item “File|Set Batch Options”, all settings from the *Main Processing Window* can be sent to the *Batch Processing Window* to be applied to the other images from the same batch.



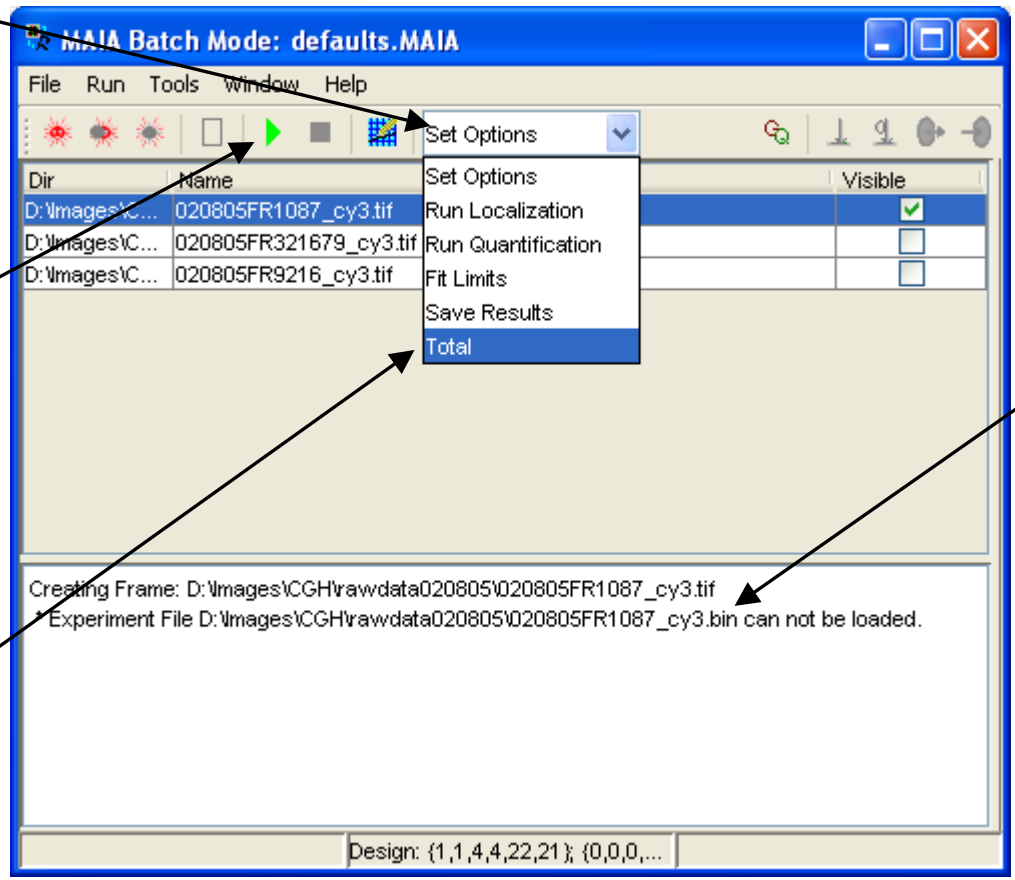
Batch processing

To run batch processing one have to define the action to be applied to all files in the batch.

The batch processing can be started using the "Run Batch" button from the Toolbar or the Menu Item "Run| Run Batch" (F9).

The whole procedure (all steps, one by one) can be launched by choosing the action "Total".

Normally it is advisable to check the spot localization step to be sure that automatically generated grid is not corrupted.



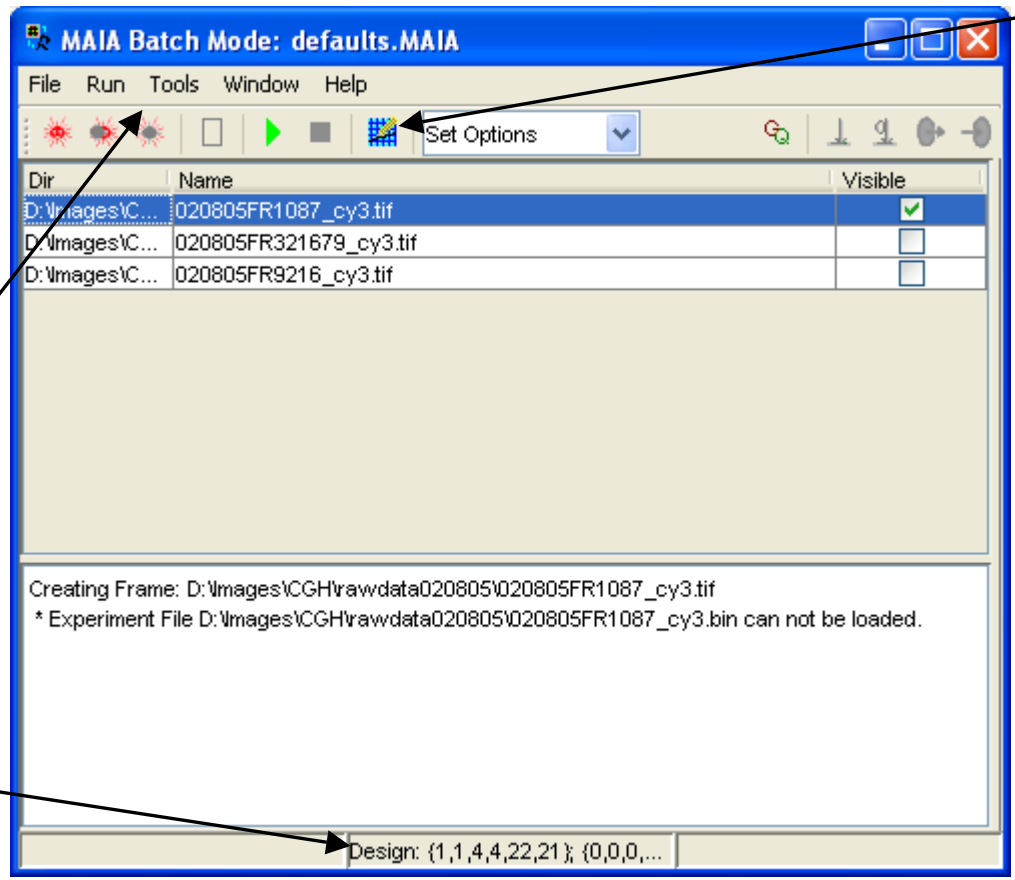
After the first processing, images with the obtained results (grid, parameters, settings, etc) are saved on the disk in the internal (binary) format (experiment files). If the program is unable to find such a file, it opens up the original image and applies the default settings (which can be defined via different items of the Menu "Tools": "Data Options", "Analysis Options", "Colors" and "Array Design").

Modify Batch Settings

Typically all arrays from the batch are of the same array design, and have the same settings.
One may want to define/modify these settings before further processing.

This can be done using the items of the Menu "Tools": "Data Options", "Analysis Options", "Colors" and "Array Design".

Description of the current Array Design



To open the Array Design window click the "Array Design" button from the Toolbar or select the Menu Item "Tools|Array Design" (F2).

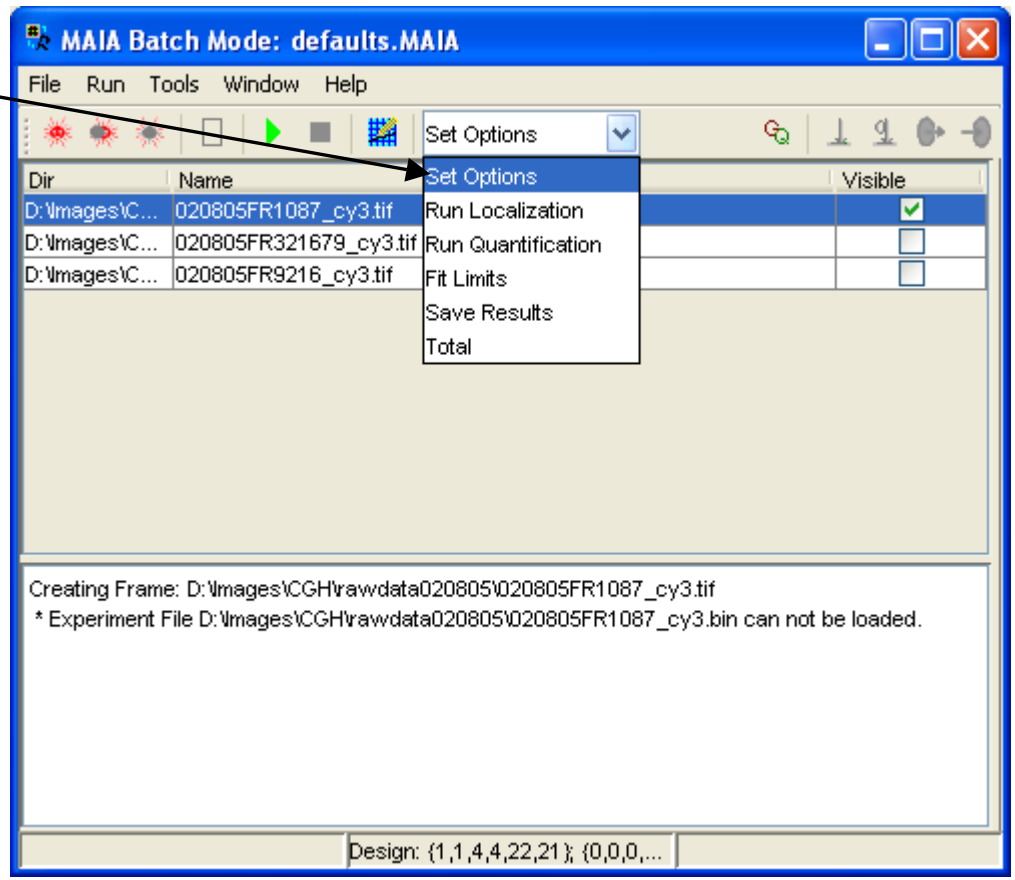
See page Array Design in Detail.

Yet another possibility to modify settings is to open (visualize) one of the images and modify settings for that image. Then the Menu Item "File|Set Batch Options" of the *Main Processing Window* will send the new settings into the *Batch Processing Window*.

Apply Setting to the Batch

To send the modified settings to all images of the batch one needs to run the batch with the task “Set Options”.

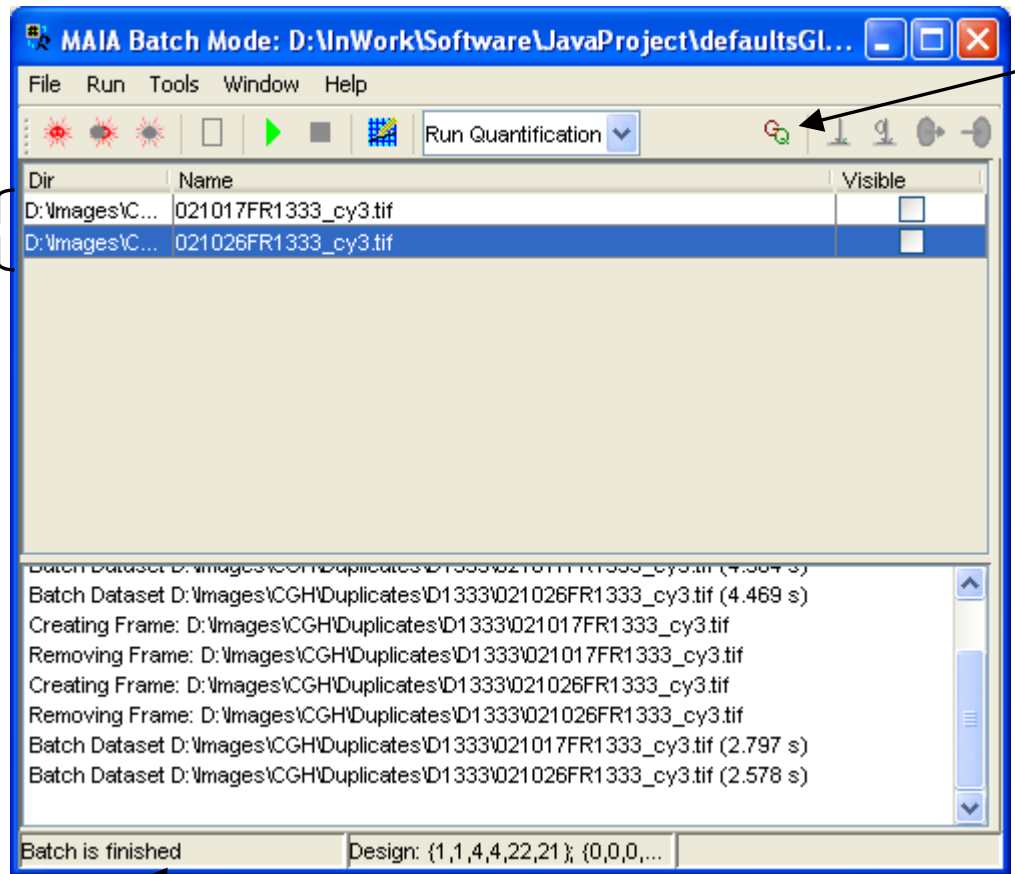
This is required only if the binary files have already been generated.



Otherwise new settings will be applied as defaults in the processing of each new image from the batch.

Global Quality Analysis

To start global quality analysis two, or more, arrays have to be selected and quantified.



Press the Toolbar button “Global Quality” to open the panel for identification of the global *Quality Limits*.

Status of the Batch processing.

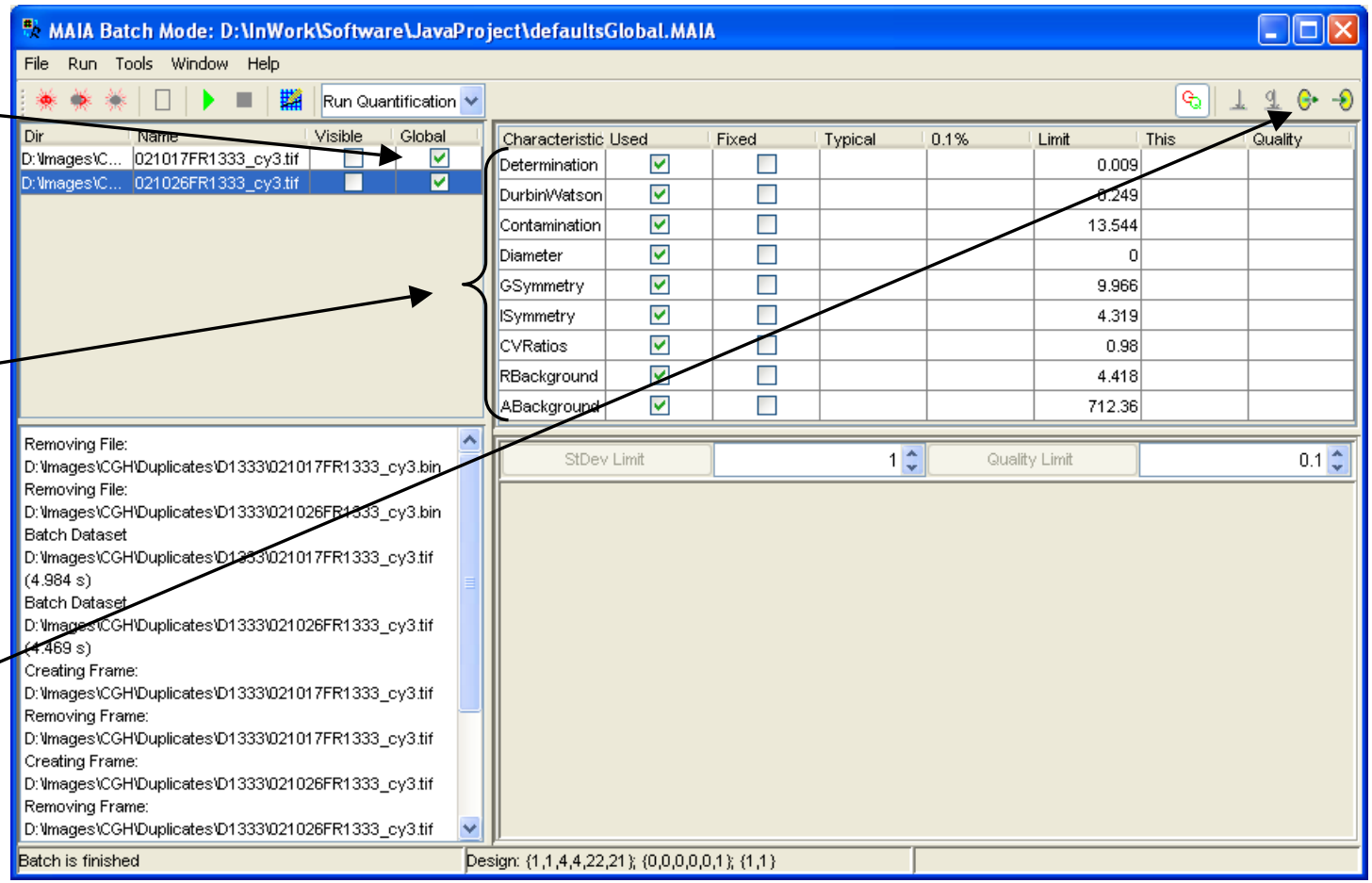
Global Quality Analysis: Main Window

Check the field “Global” to specify which arrays will be used for global quality analysis.

Global quality analysis panel shows up with the same set of quality characteristics as for each particular image.

See page Spot Quantification Output.

Press the Toolbar button “Get Experiments” to copy quantification results from all selected arrays into the global quality analysis window.

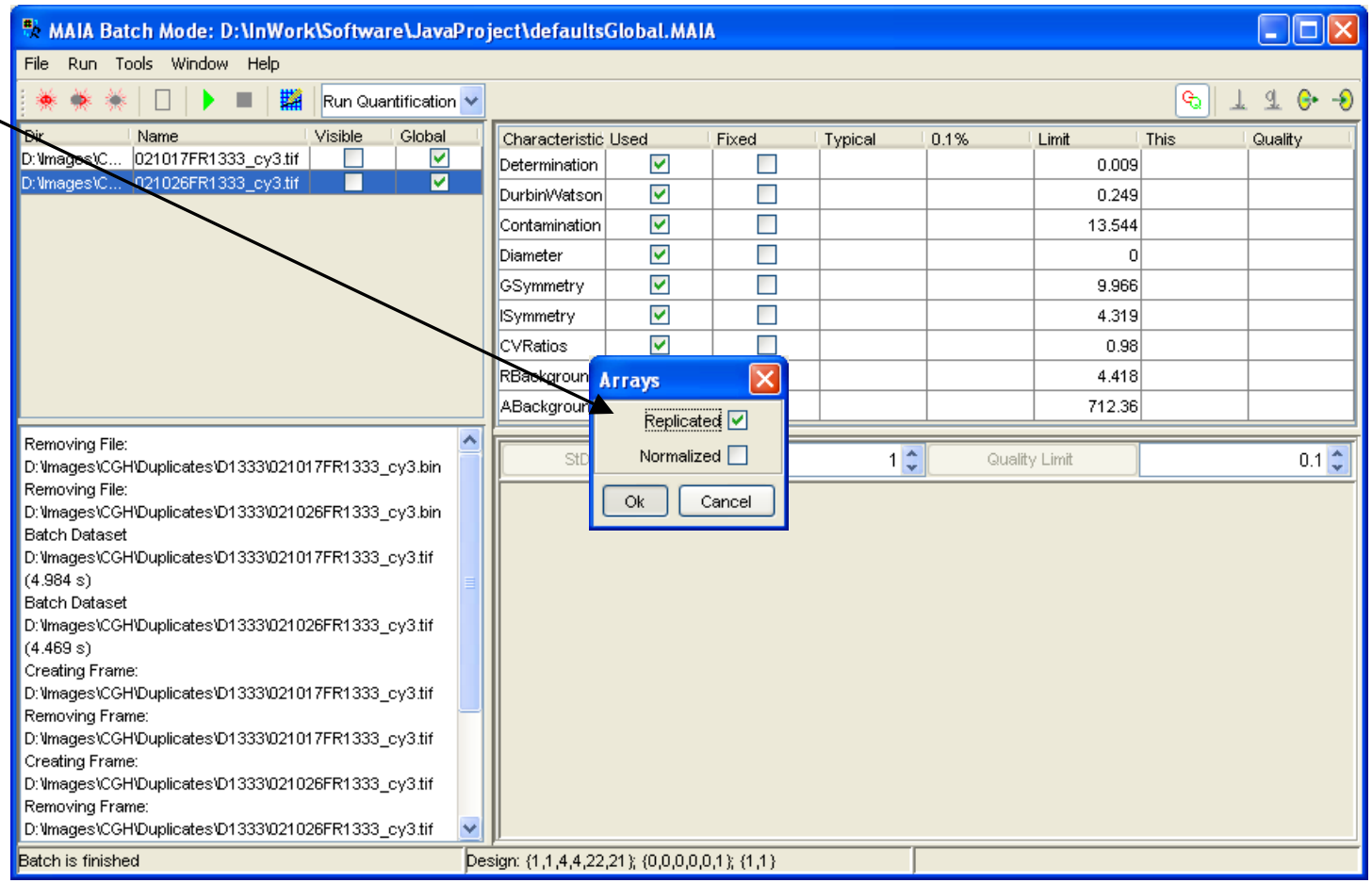


Options for Replication

Global quality analysis can be performed assuming that the selected arrays are either replicates or not.

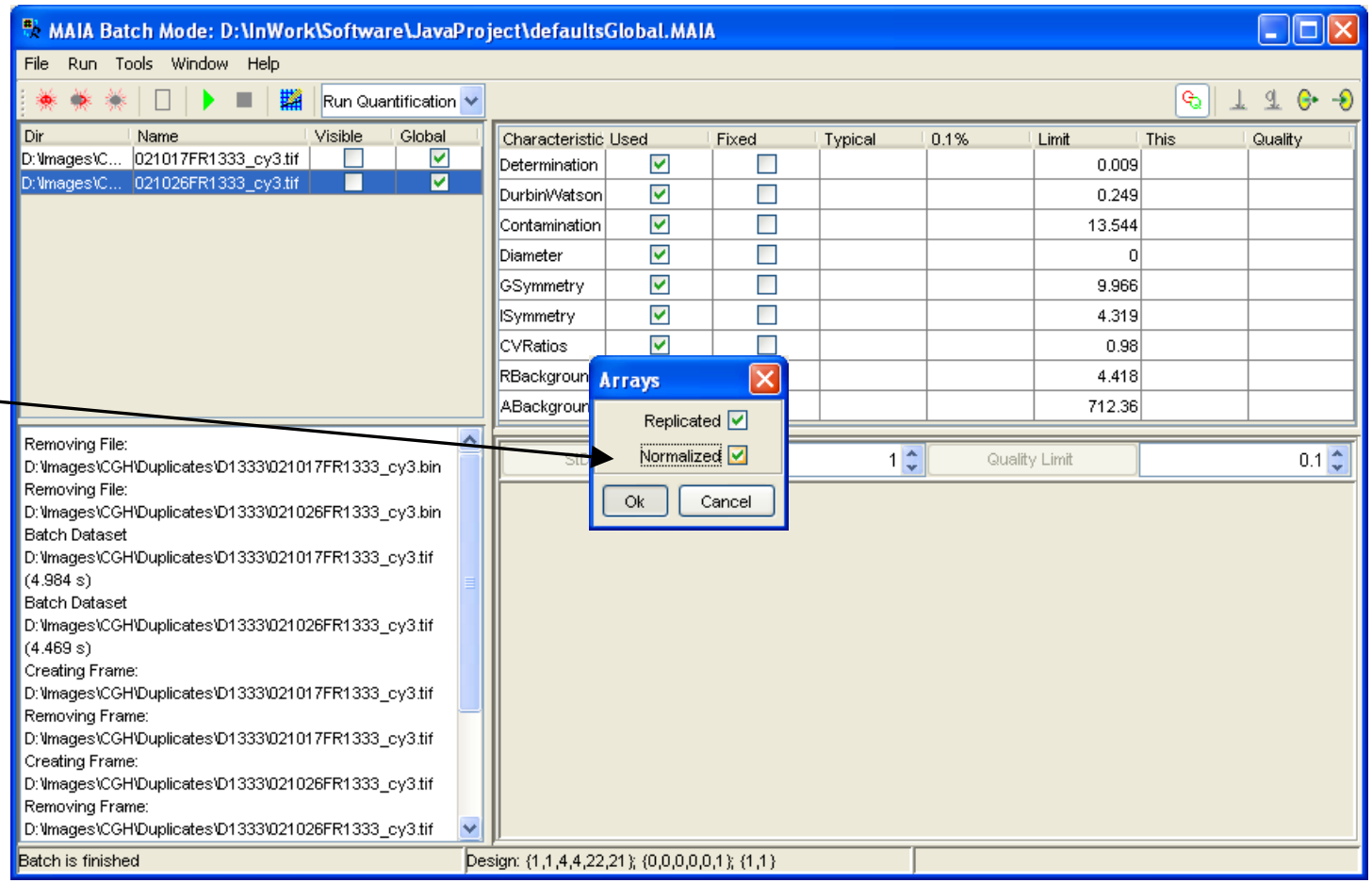
If they are replicates, then all locally replicated spots from different arrays are combined, and a unique overall quality value and a unique ratio CV are calculated for each replicated clone.
If the selected arrays are not replicates, then local spot replicates from different arrays are treated independently in the overall quality plot.*

*) In this case, to have local spot replicates is essential for quality analysis.



Options for Normalization

If the selected arrays are replicates, then before combining locally replicated spots from different arrays into a unique overall quality value and a unique ratio CV, one may want to align arrays, so that the averaged log ratio is equivalent for all arrays in the selection.

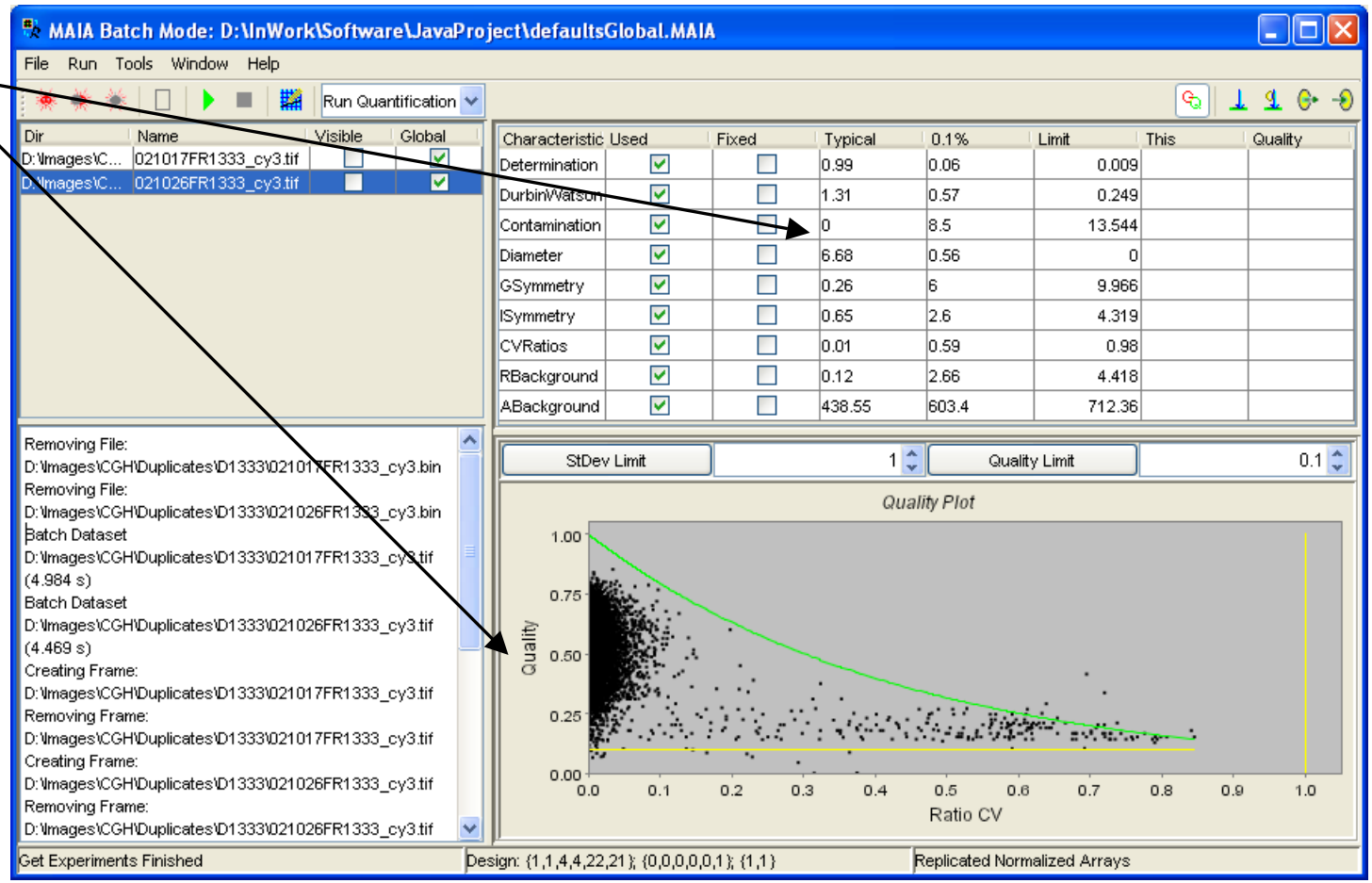


Results Downloaded

The quantification results have been downloaded.

The following quality analysis procedure is equivalent to the quality analysis performed for each particular image.

See page Quality Analysis (II).

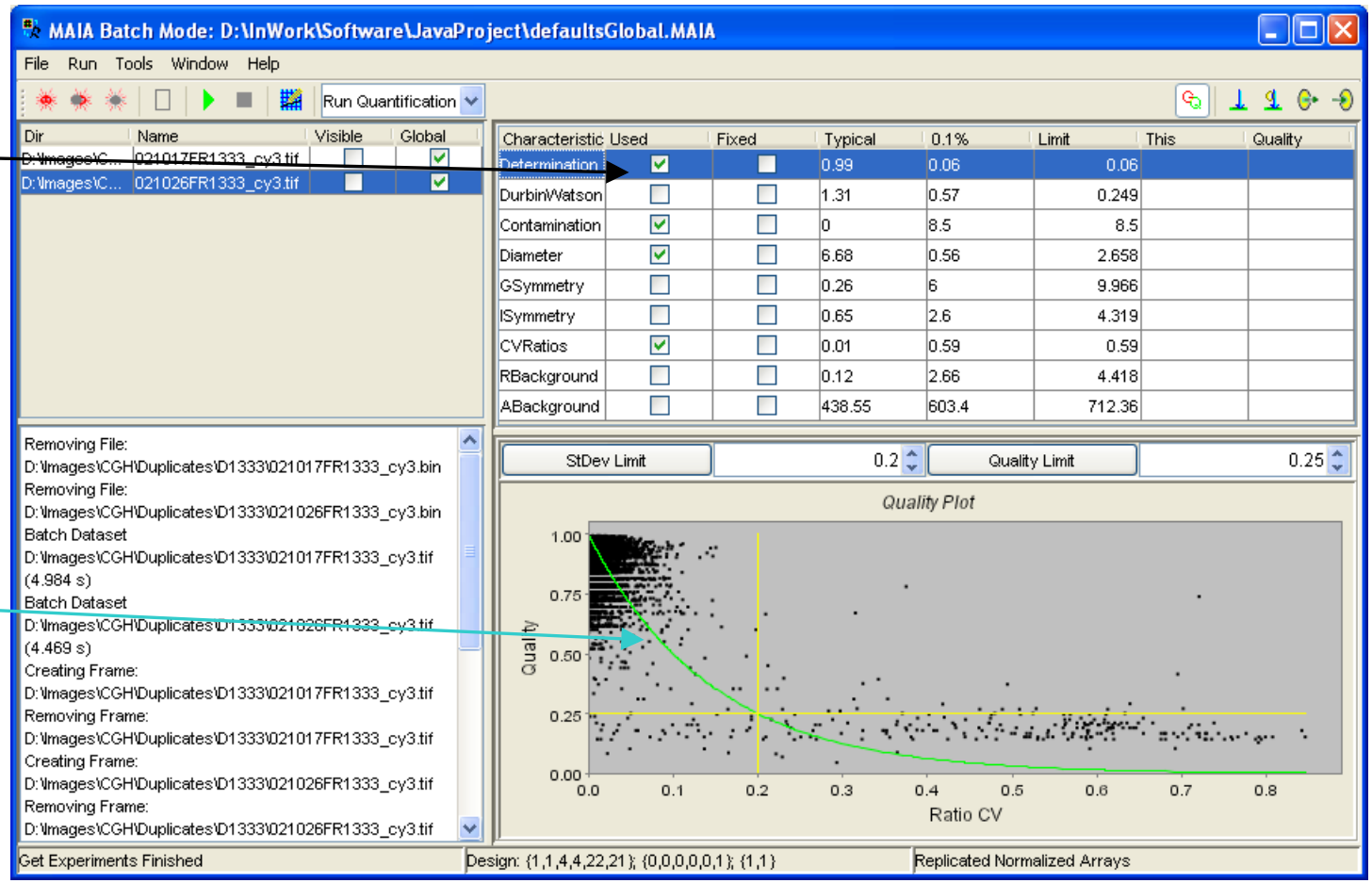


Global Quality Plot

Using the table field "Used" one can select a set of quality characteristics, which seem relevant for this particular batch of images.

To identify the shape of the quality curve one can use the same tools as for each particular image.

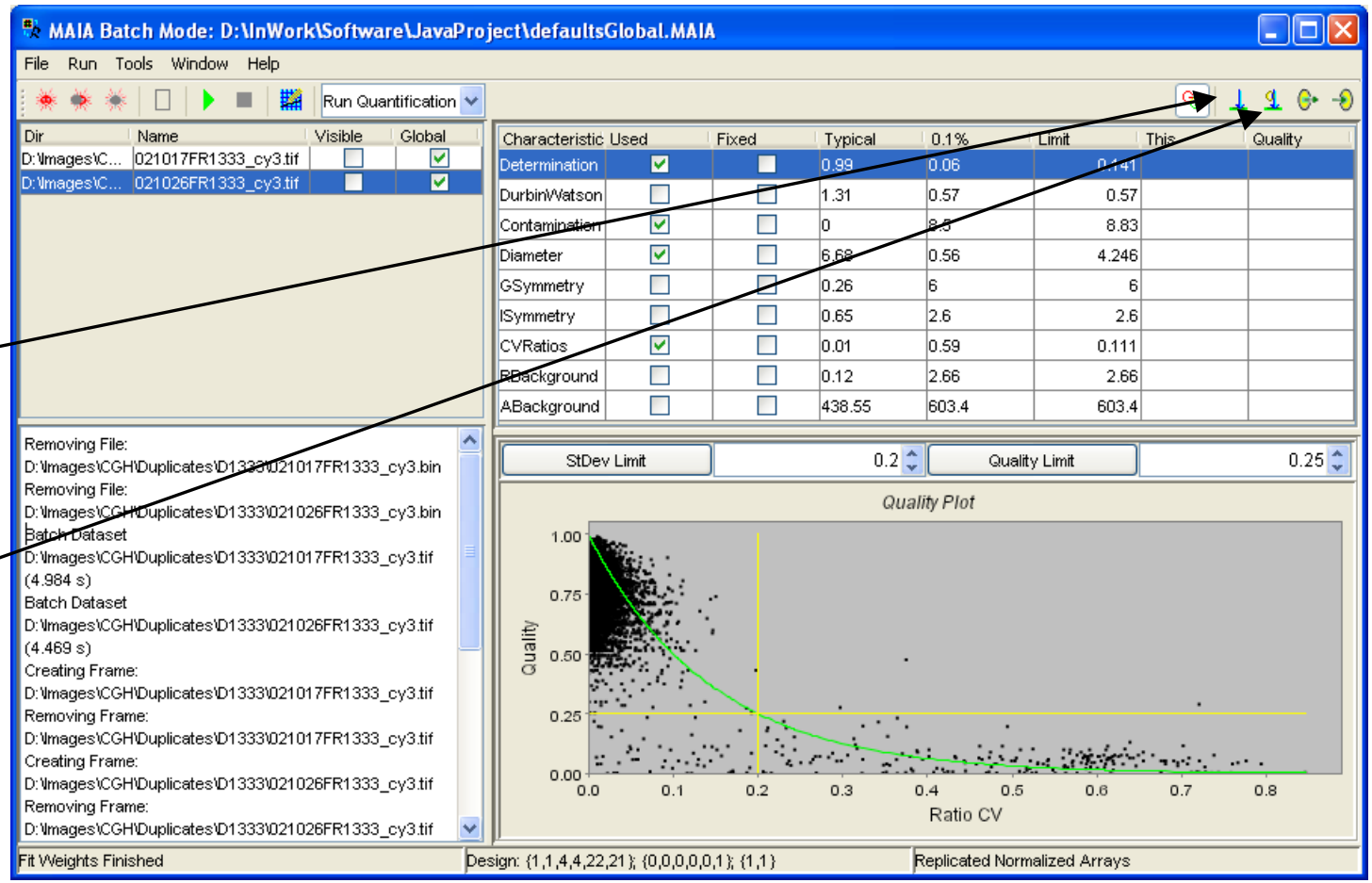
See page Quality Analysis (II).



Fit the Limits

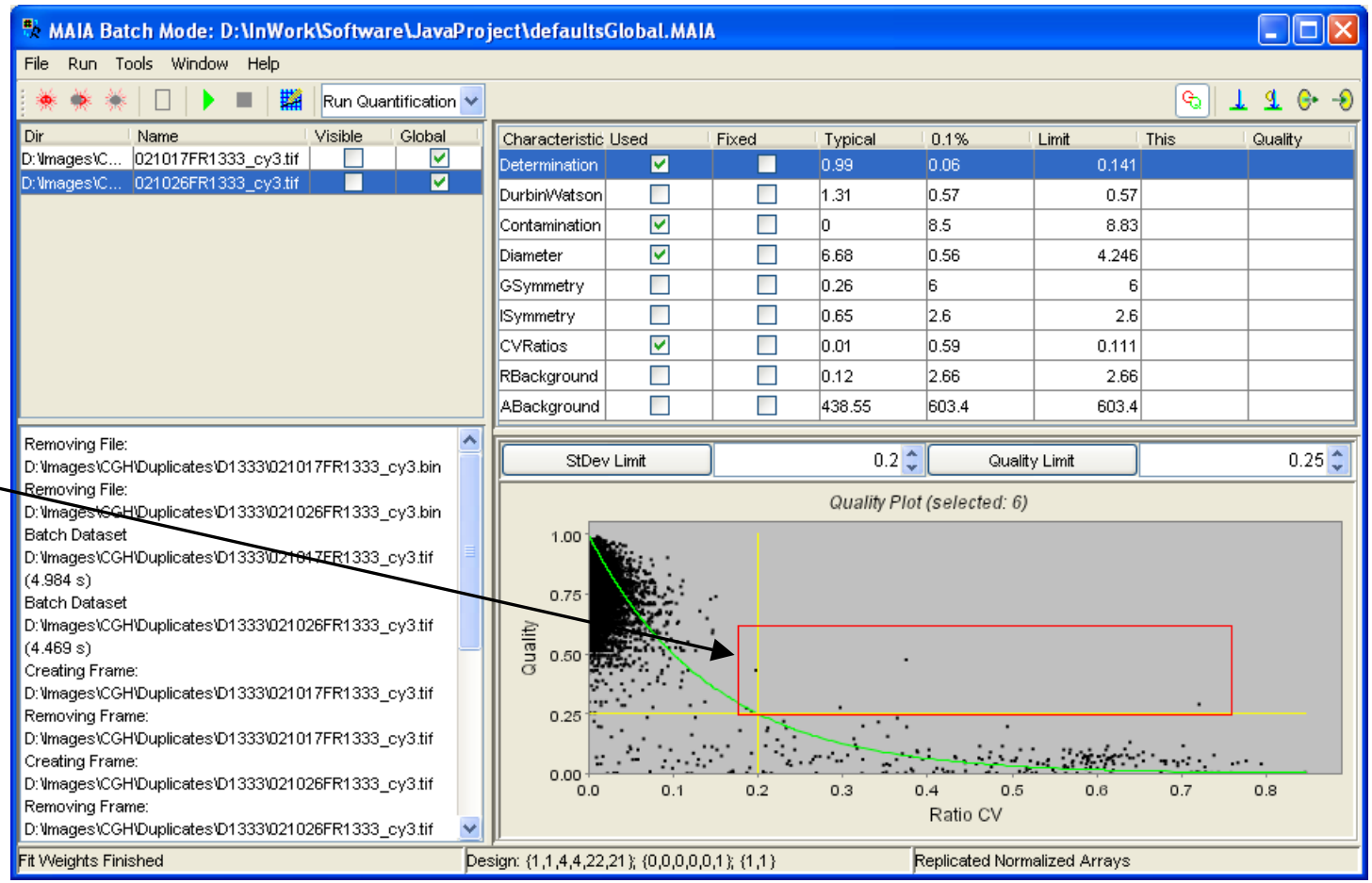
To initialize the Limits use the “Init Limits” button from the Toolbar or the Menu Item “Run|Init Limits” (Ctrl+F5).

To run fitting procedure use the “Fit Limits” button from the Toolbar or the Menu Item “Run|Fit Limits” (F5).



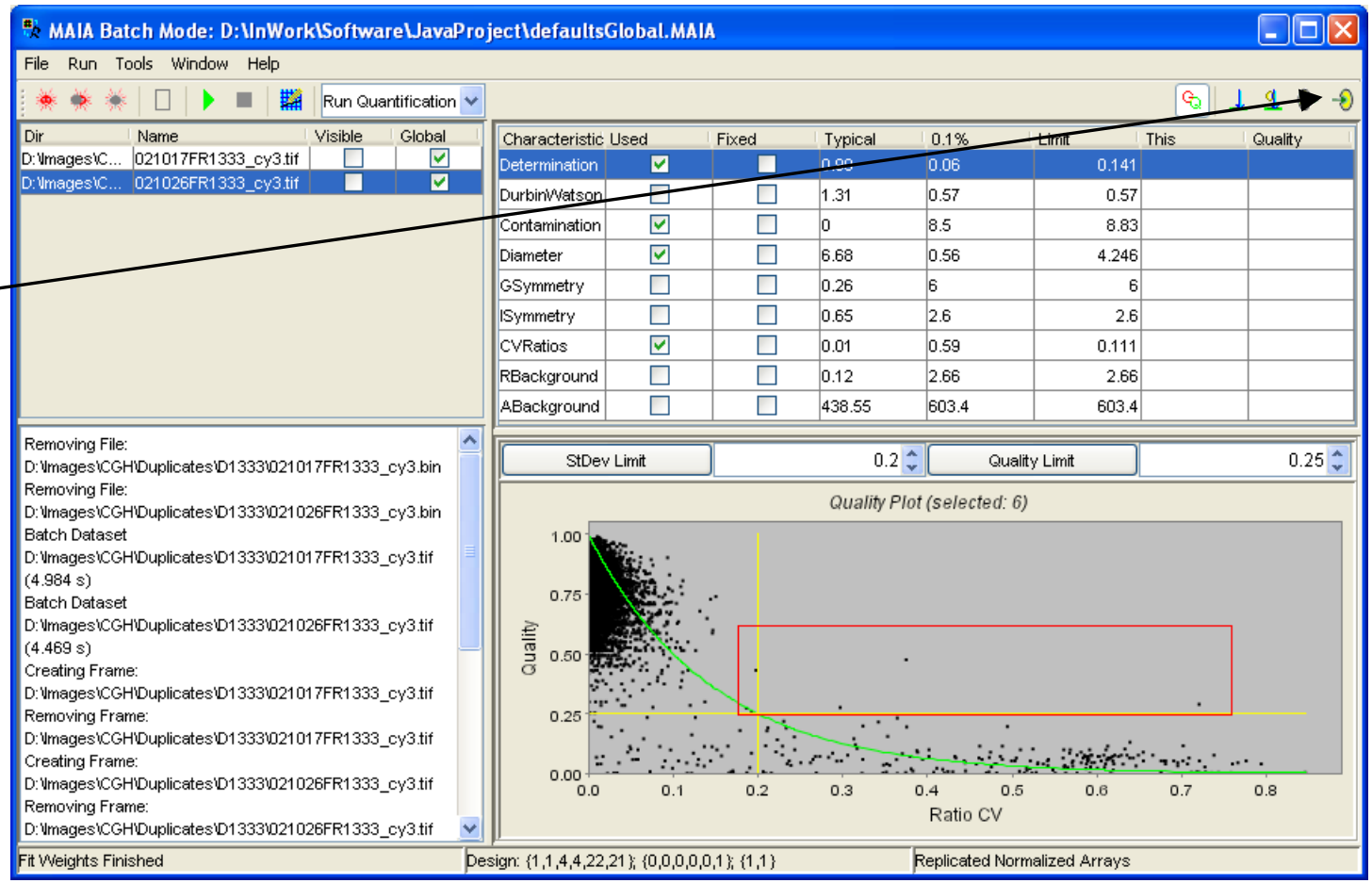
Select "Bad" Replicates

Left Click – Drag – Right Click on the *Quality Plot* to select the replicates to be able to find them on the arrays from the globally analyzed selection of arrays.



Export Quality Limits

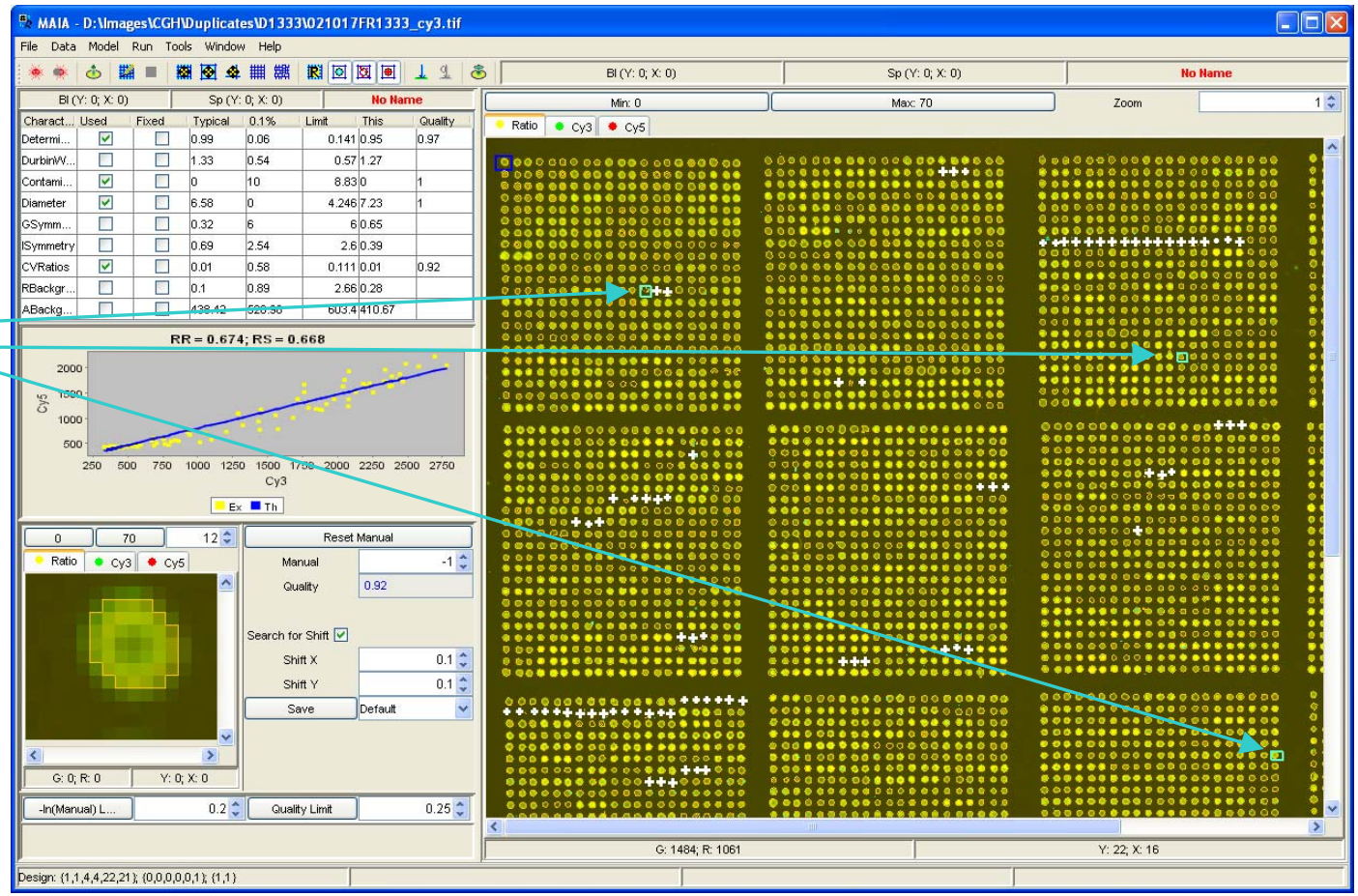
To send the obtained quality limits and selected replicates to each array file from the given selection press the Toolbar button “Set Limits”.



The Selected "Bad" Replicates (I)

The selected replicates are indicated by cyan rectangles on both arrays.

The first array "021026".



The Selected "Bad" Replicates (II)

The selected replicates are indicated by cyan rectangles on both arrays.

The second array "021017".

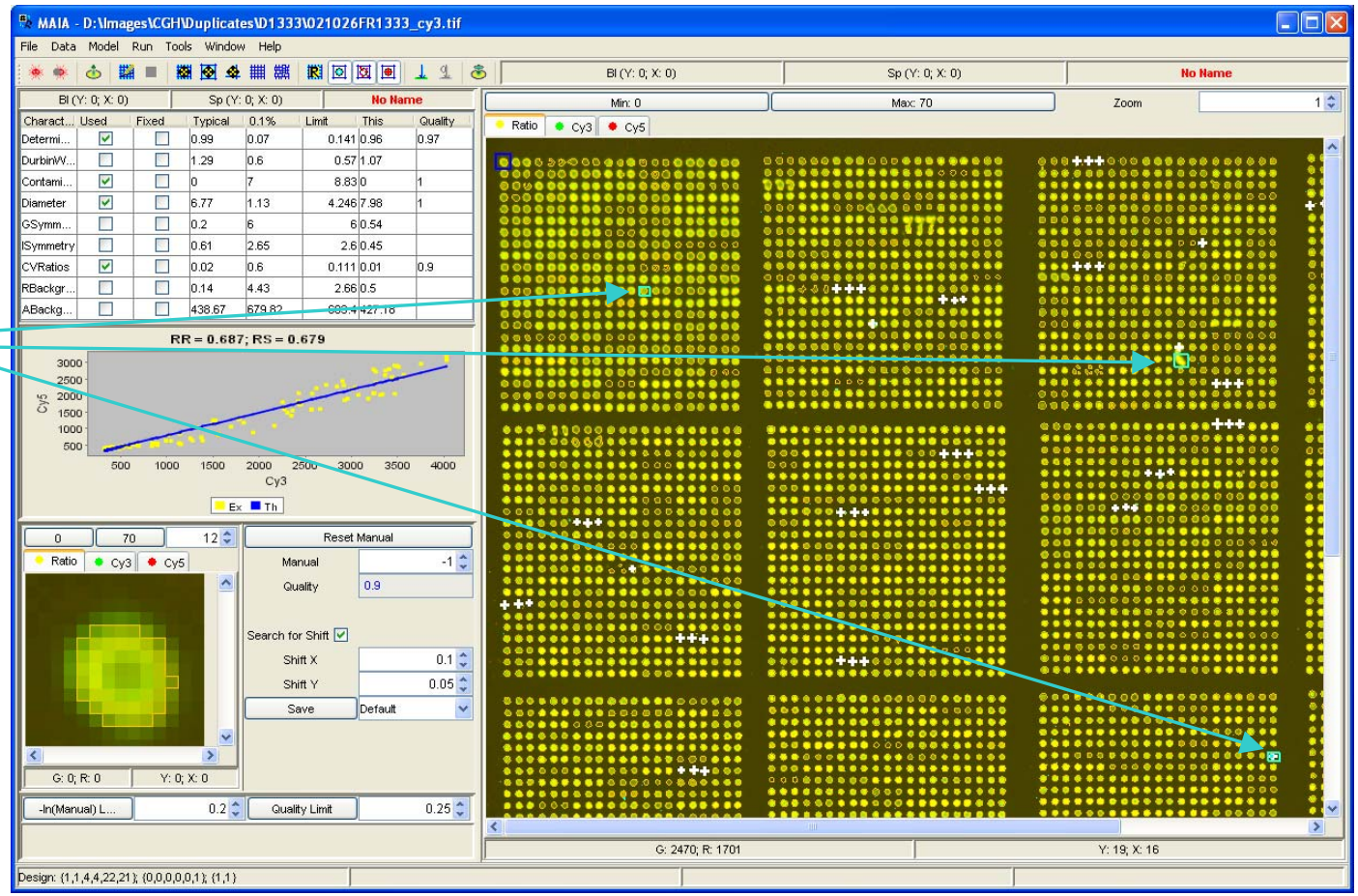
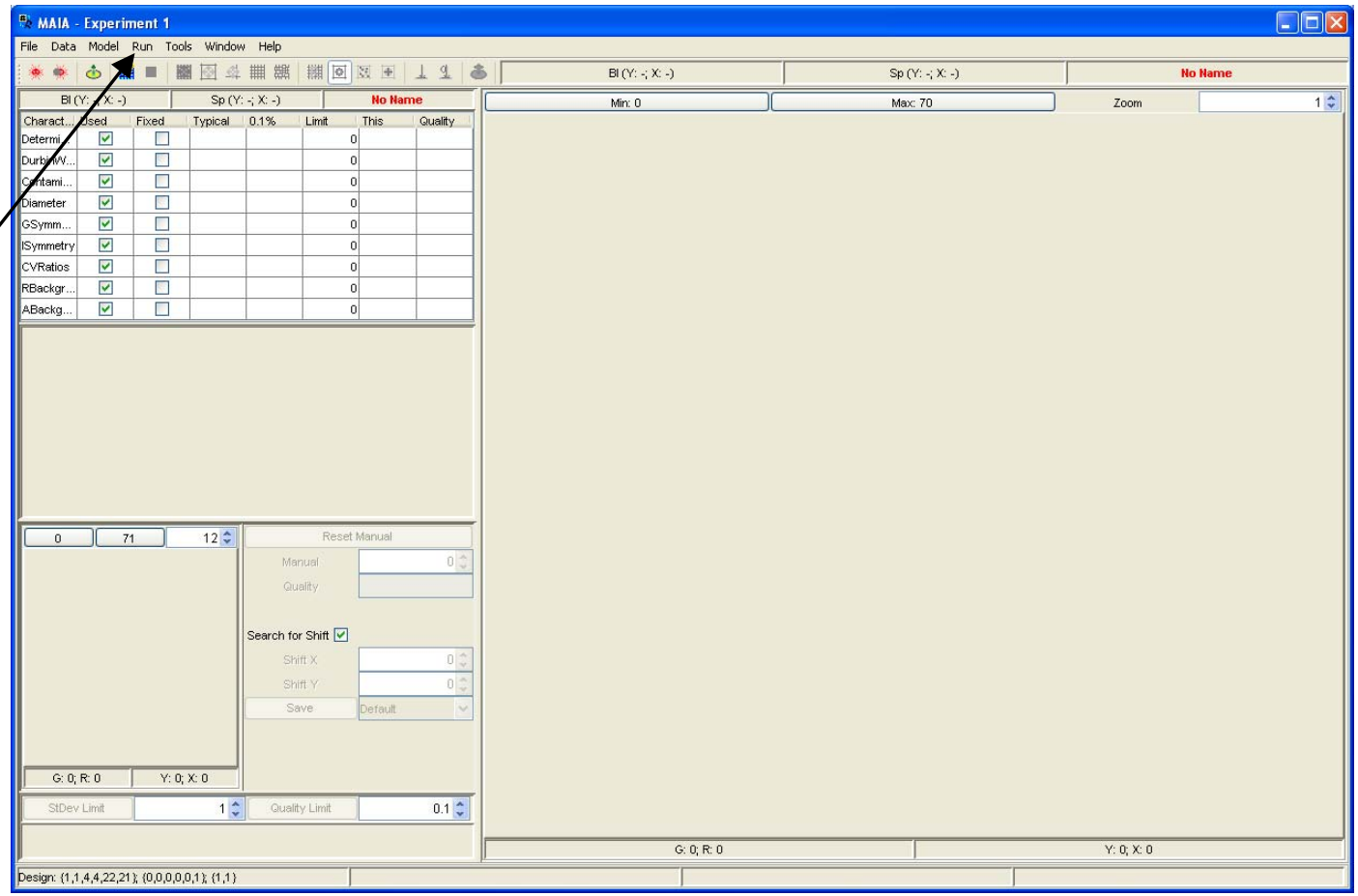


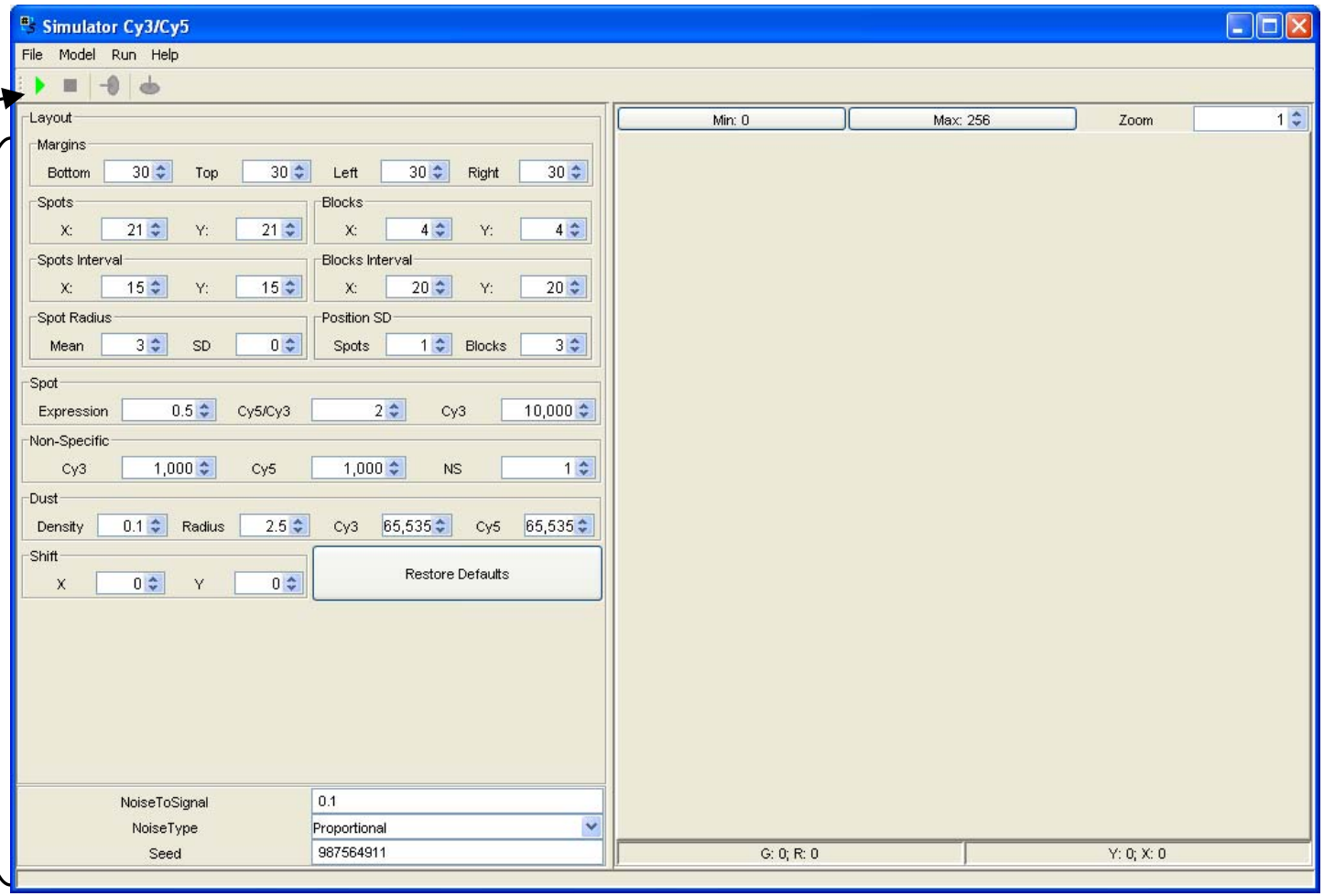
Image Simulator

To open Image Simulator Window select the Menu Item “Run|Simulator”.



Main Simulator Window

To start simulations press the “Run Simulations” button from the Toolbar or select the Menu Item “File|Run Simulations”.



To simulate an image the following parameters should be defined.

See page Artificial Images.

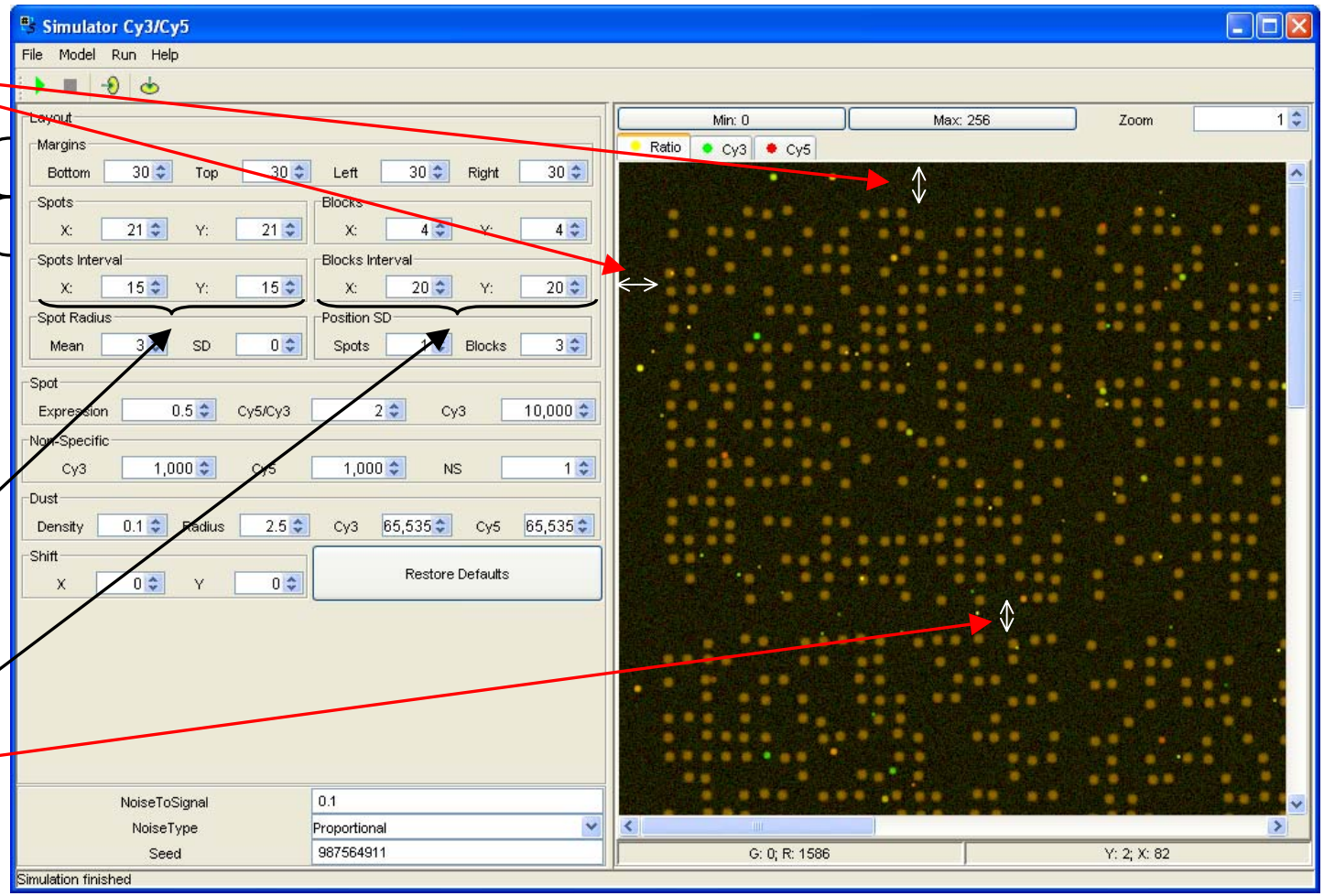
Array Layout (I)

The distance in pixels from the edges of the array to the spotting area.

Amount of spots (per block) and amount of blocks on the array.

Distance in pixels between the spots in the blocks.

Distance in pixels between the blocks.



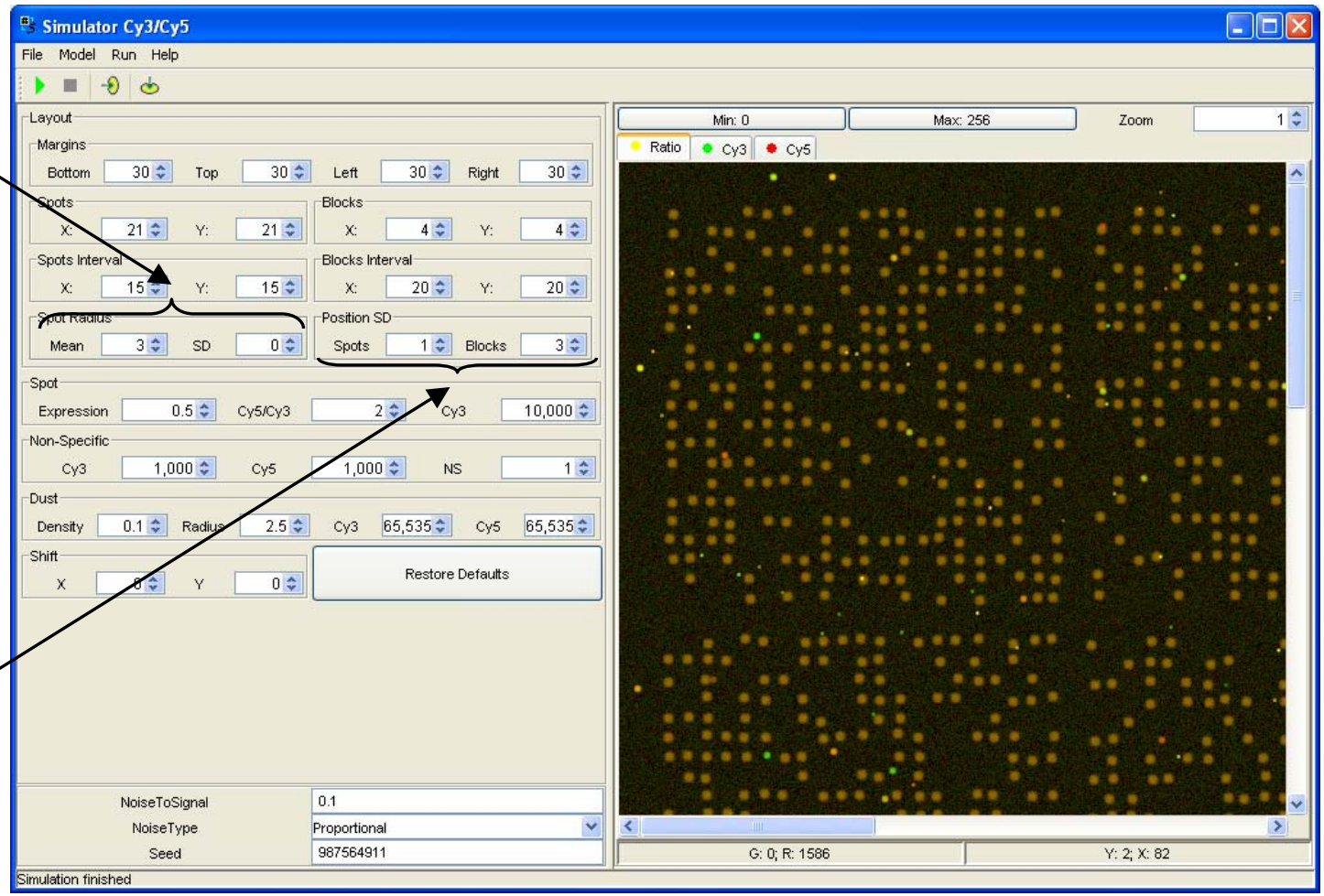
Array Layout (II)

Mean and standard deviation of the Spot Radius.

If *SD*>0, spots will be generated with randomly selected (around Mean) radius.

Standard deviation of the positions of the spots and blocks with respect to the ideal alignment.

Larger *SD* value, larger deviation of the positions of the spots/blocks from the ideal spot/block alignment.

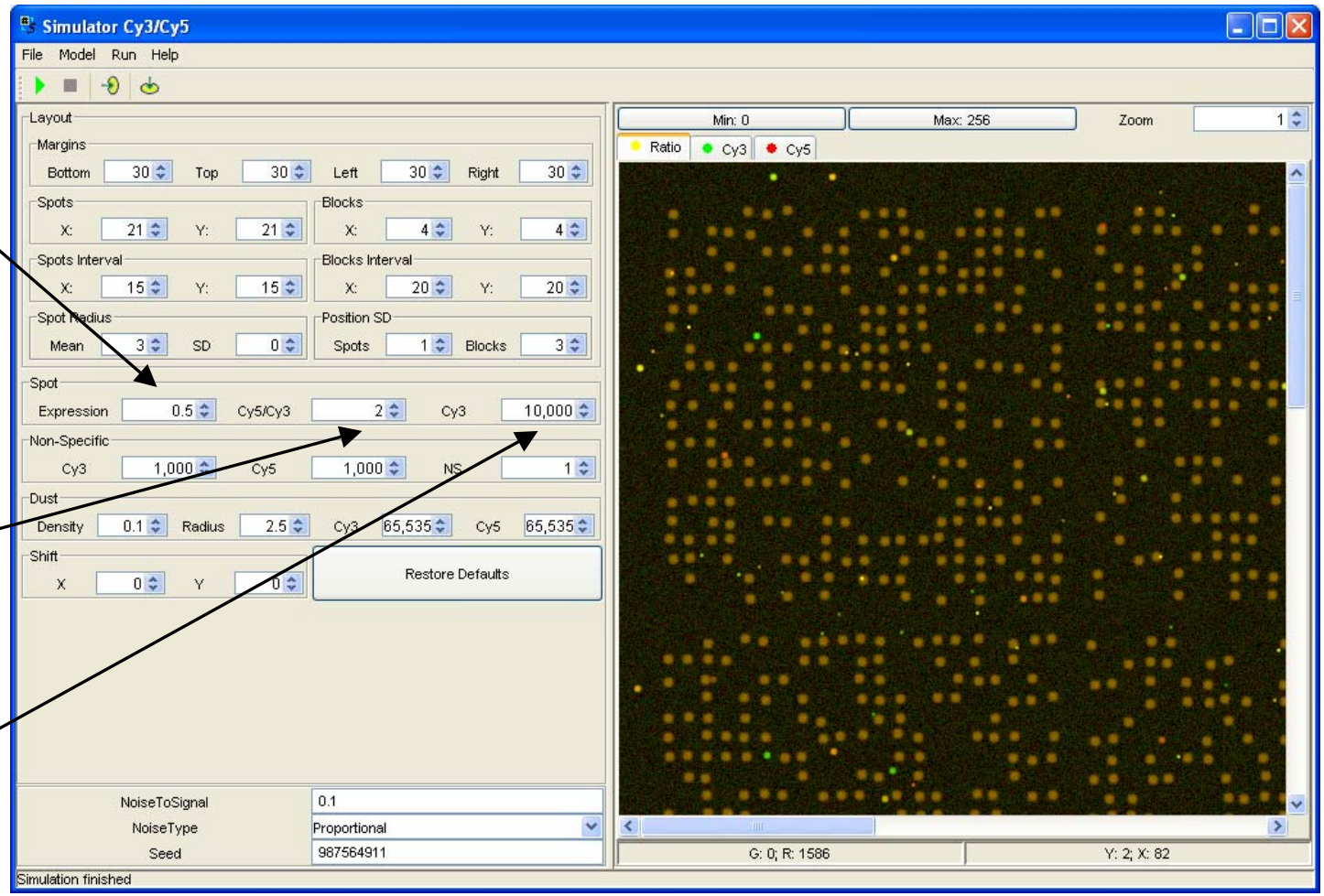


Spot Characteristics

Rate of the bright (visible) spots on the array: 0 – no visible spots are generated, 1 – all spots are visible.

Ratio of Cy5/Cy3 color channels, the same for all spots on the array.

Maximal spot intensity in the Cy3 channel (i.e. fluorescence intensity in the center of the spot).

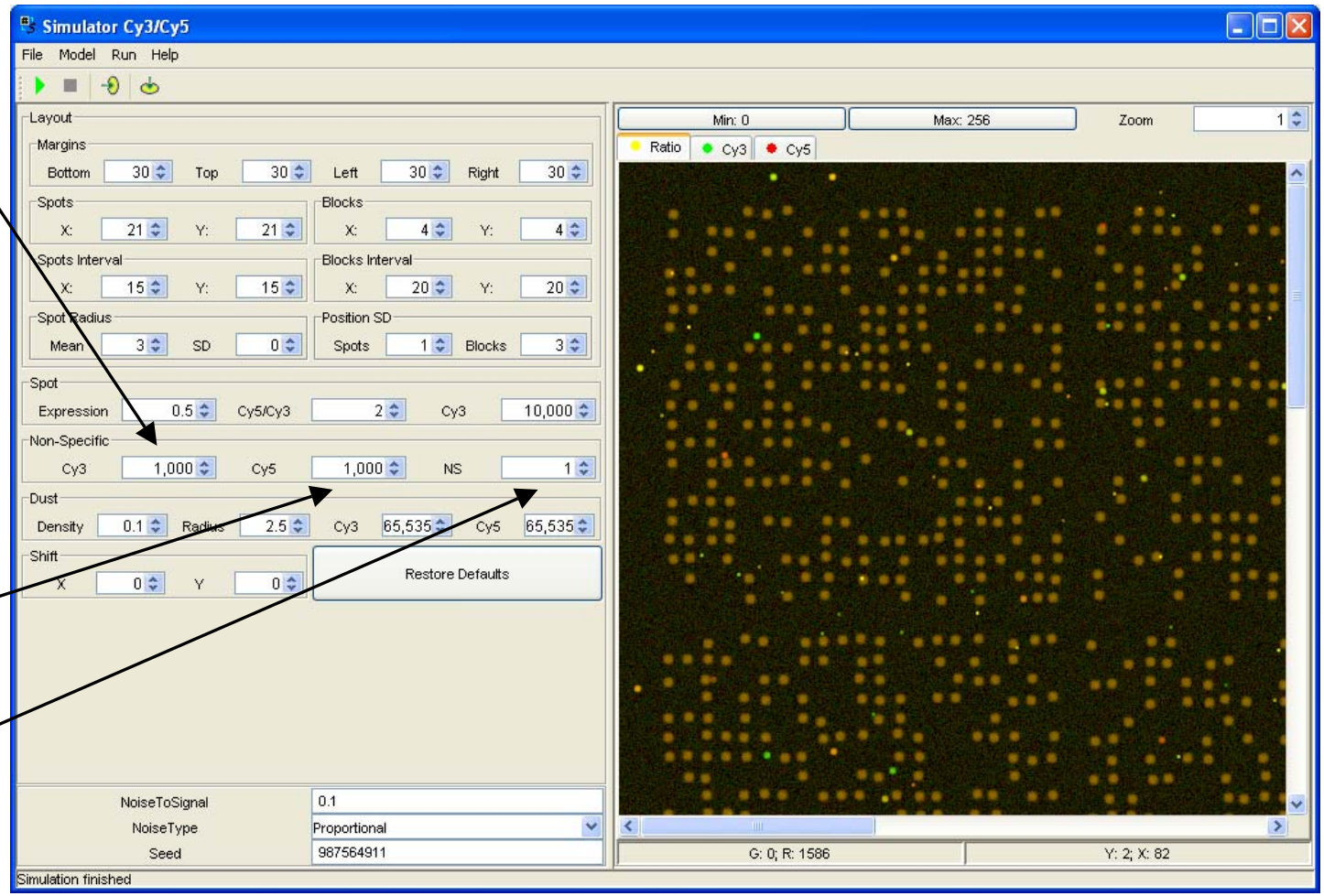


Non-Specific Hybridization

Average intensity of non-specific hybridization in the Cy3 color channel.

Average intensity of non-specific hybridization in the Cy5 color channel.

Noise to signal ratio for non-specific hybridization for both color channels.



Dust

Maximal dust radius.
The radius of the dust spot is randomly chosen from the interval from 0 to the given value.

Density of dust is defined with respect to the number of “good” spots on the array:
0 – no dust spots, 1 – the number of dust spots corresponds to the number of “good” spots.

Maximal intensity* of dust in the Cy3 color channel.

Maximal intensity* of dust in the Cy5 color channel.

* Real intensity will be randomly chosen from the interval from 0 to the given value.

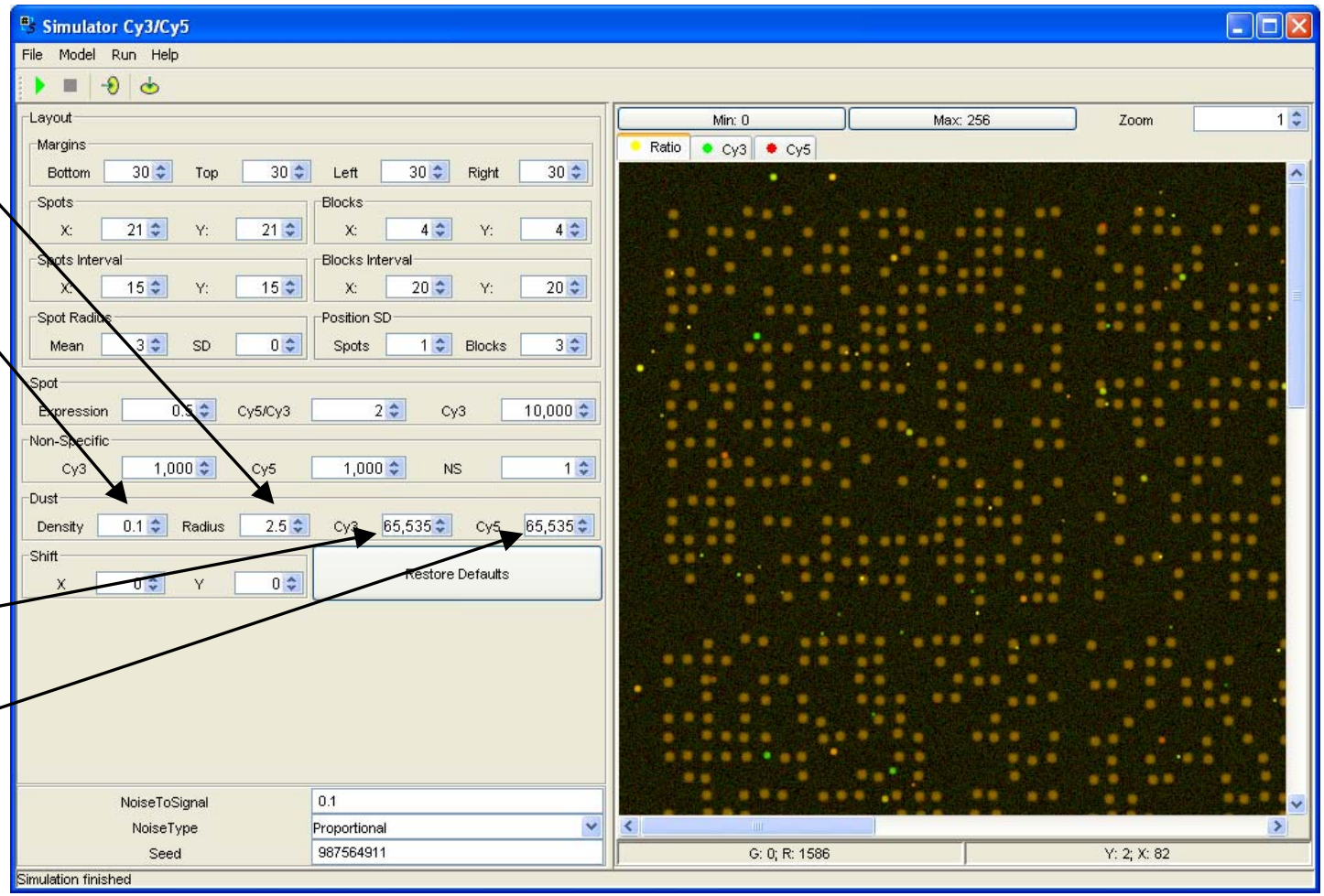
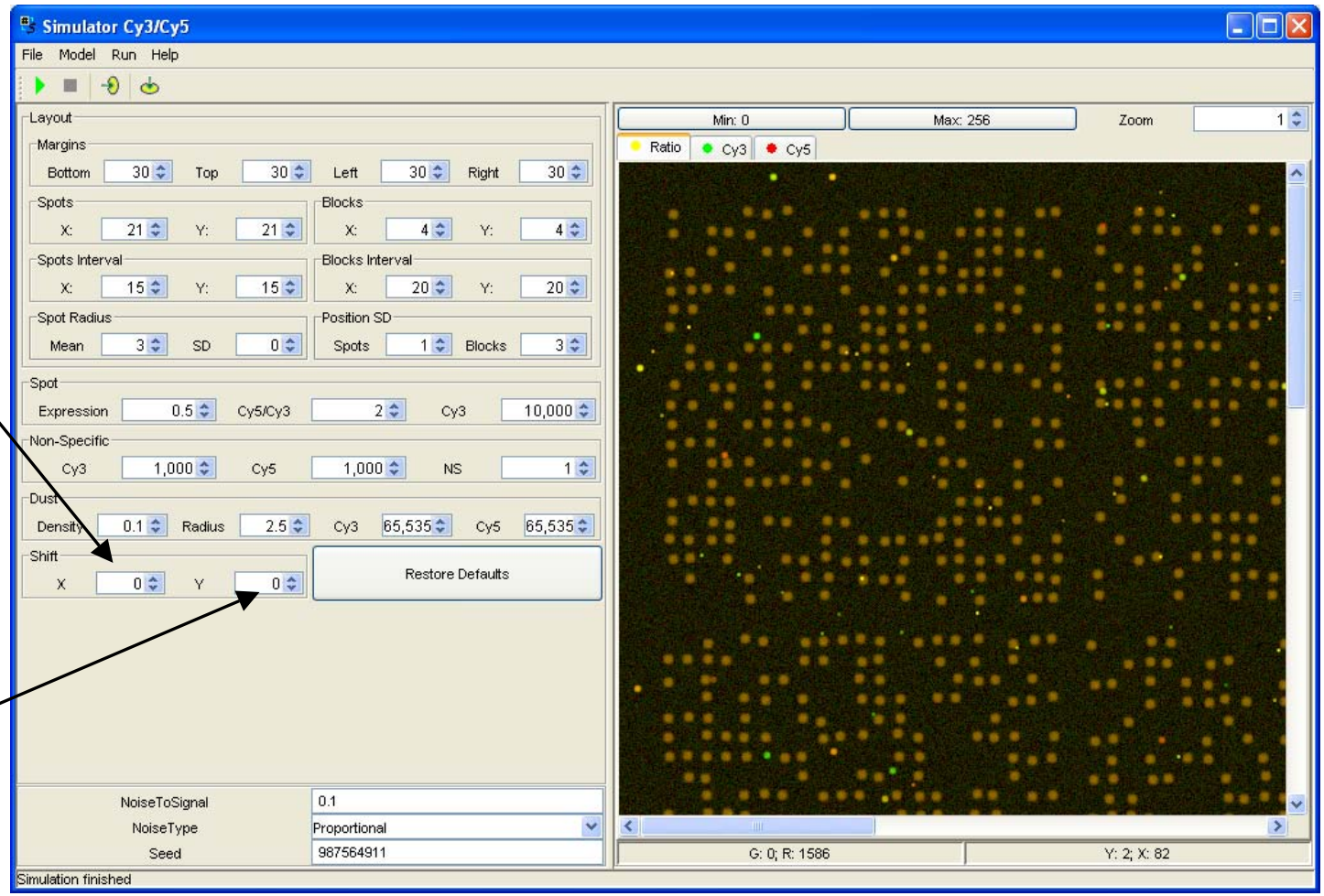


Image Shift

Image shift in pixels in horizontal direction.

Image shift in pixels in vertical direction.



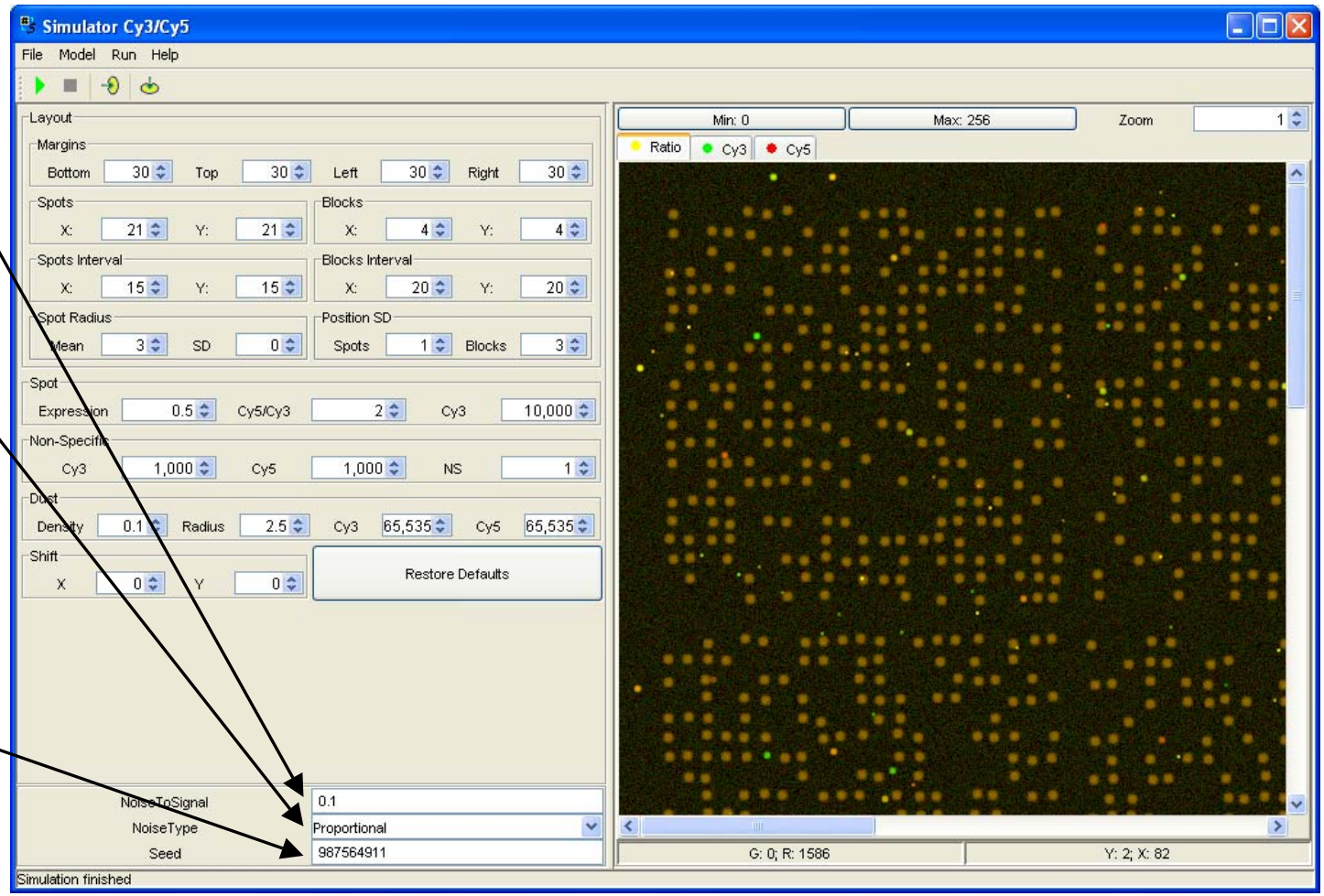
Non-integer pixel shifts are possible.

Additive Statistical Noise

Noise to signal level for the additive statistical noise. This noise is finally added to each pixel of the array.

Model for the standard deviation of the additive noise. It can be constant, proportional to signal, or proportional to the square root of signal.

Seed for random number generator (selection –1 as a seed will initiate the random generator with automatically (or randomly) chosen seed).

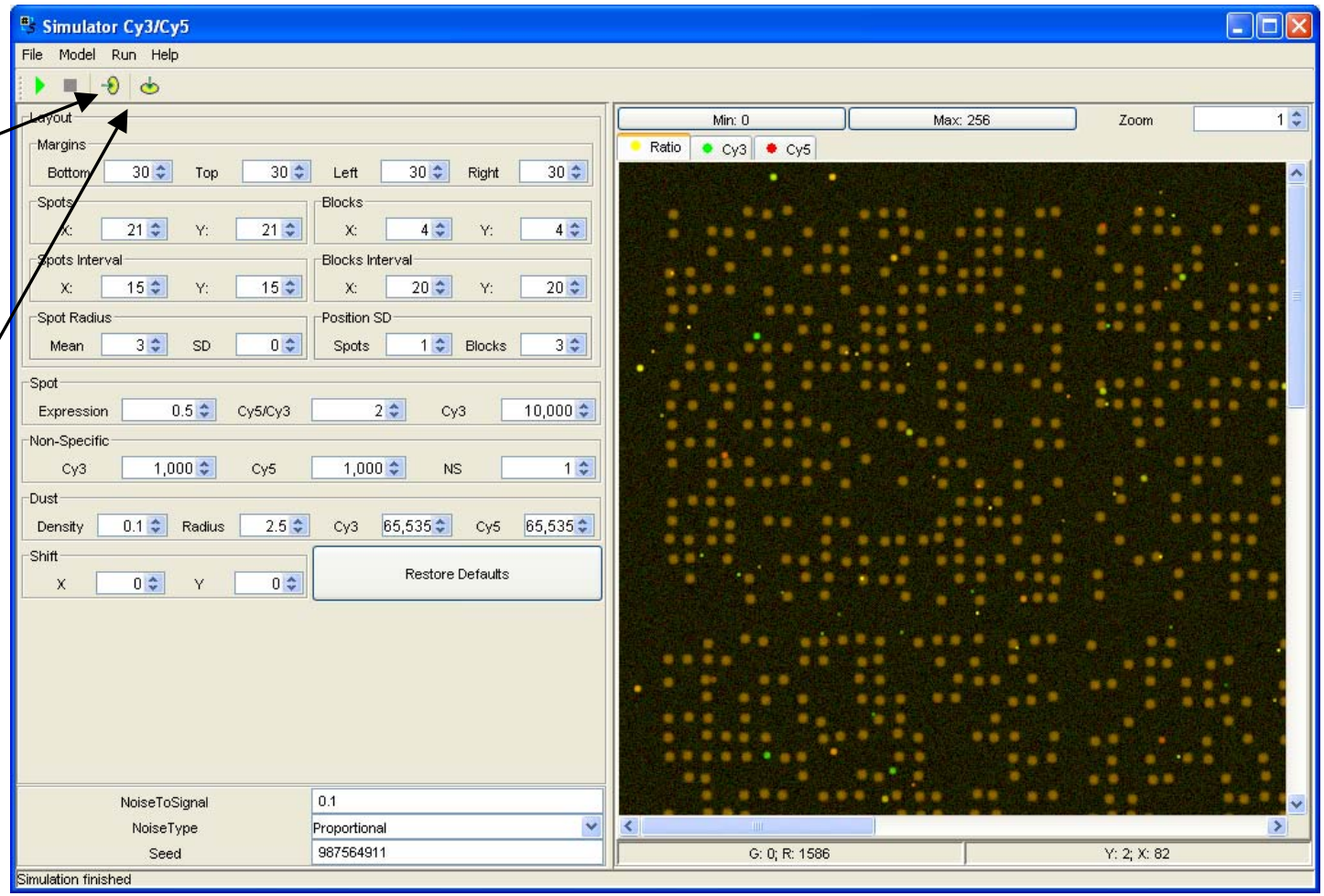


Export of the Generated Image

To send the generated images in the Main Processing Window, use the “Send Data” button from the Toolbar or the Menu Item “File|Send Data”.

To save the generated images in the TIFF files use the “Export Image” button from the Toolbar or the Menu Item “File|Export Image”.

Only single-page TIFF files are currently supported.



Artificial Images

Model for a spot. The generated spots must have more or less circular contours (in the horizontal projection) and relatively sharp edges (in the vertical projection):

$$f_{Cy3}(x,y)=I\exp\left(-\left\{\left[\frac{x-x_c}{r}\right]^4-\left[\frac{y-y_c}{r}\right]^4-\left[\frac{x-x_c}{r}\right]^2\left[\frac{y-y_c}{r}\right]^2\right\}\Big/2\right)$$

where x_c and y_c are the coordinates of the center of the spot, r is its approximate radius and I is the fluorescence intensity in the center of the spot in the Cy3 color channel. Fluorescence intensity in the Cy5 color channel is defined as:

$$f_{Cy5}(x,y)=Rf_{Cy3}(x,y)$$

where R is the ratio of the test and control samples for each spot. The coordinates x_c and y_c, the radius r and the ranges for x and y for each spot cell are defined from the user-established array layout. The intensity parameters I and R should also be provided by the user.

Nonspecific hybridization results in an additional component (B_i) in the detected fluorescence intensity:

$$f_i^B(x,y)=f_i\,(x,y)+B_i$$

The number of non-specific molecules contributing into each scanned fluorescence pixel is a random value:

$$B_i=B_i^*+\sigma_{Bi}B_i^*G$$

where B_i^* and σ_{Bi} are the user-defined average and noise-to-signal ratio of nonspecific fluorescence intensity in the color channel i, and G is a gaussian random variable with zero mean and unit standard deviation.

Dust is represented by randomly distributed over the array more or less bright clusters of pixels, which can hardly be distinguished from the spots. We apply the same profile for the dust clusters as for the spots:

$$d_i\,(x,y)=I_d\exp\left(-\left\{\left[\frac{x-x_{cd}}{r_d}\right]^4-\left[\frac{y-y_{cd}}{r_d}\right]^4-\left[\frac{x-x_{cd}}{r_d}\right]^2\left[\frac{y-y_{cd}}{r_d}\right]^2\right\}\Big/2\right)$$

where x_{cd} and y_{cd} are the coordinates of the center of a dust cluster, r_d is its approximate radius and I_d is the intensity in the center of the cluster. All these parameters are random variables. We use uniform distributions for r_d (in the interval $[0;r_m]$) and I_d (in the interval $[0;I_m]$), where r_m and I_m are user-provided maximal dust cluster radius and maximal dust intensity, respectively. We also assume that the coordinates of the centers of dust clusters x_{cd} and y_{cd} are uniformly distributed over the array. Statistical laws of the dust characteristics can generally be different for two channels (i = Cy3, Cy5). Finally one has to define the number or density of the dust clusters on the array.

The general model for the microarray image takes the form:

$$\bar{f}_i\,(x,y)=\sum_{k=1}^{N}f_{ik}\,(x,y)+B_i+\sum_{k=1}^{M}d_{ik}\,(x,y)$$

where N is the number of spots and M is the number of dust clusters.

Statistical noise is finally added to each pixel of the image:

$$\tilde{f}_i\,(x,y)=\bar{f}_i\,(x,y)+\sigma(x,y)G$$

where $\sigma(x,y)$ is the standard deviation of the pixel noise: $\sigma(x,y)$ can be (i) constant, (ii) proportional to signal, or (iii) proportional to the square root of signal. The type of statistical noise as well as its quantitative characteristics is defined by the user.