QAnalyzer 1.0c

User's Manual

Packard BioScience **QuantArray(QA)** microarray analysis software is a powerful microarray analysis software that enables researchers to easily and accurately visualize and quantitate gene expression data. Compared to other softwares of this kind, **QuantArray** provides automated analysis of up to five color microarray images without manually drawing grids. However, the built-in analysis function of **QuantArray** is very simple and not enough for further analysis. An efficient statistical analysis method which can be used to analyze microarray data is the analysis of vaiance (ANOVA) method. We can simultaneously get the fold change estimates and some other effects by the ANOVA procedure. Many research show that the most commonly used fluorescent dyes, such as Cy3 and Cy5, are relatively unstable and may have dye effect on channel intensities. This kind of dye effect cannot be accounted for by the simple linear normalization. The most effcient nonlinear normalization method which can deal with the dye effect is Lowess fit method which was originally proposed by William S. Cleveland as a statistical method. This program *QAnalyzer* reflects our efforts to apply the above methods to **QuantArray** output. Furthermore, this program also inherits some features of the program *QAmerge*.

Features

- ### QA Data Modification and Normalization:

  QAnalyzer performs the background subtraction and several normalization functions. Each spot intensity is background subtracted. Then two simple methods of normalization are available. They are the normalization to mean which is equivalent to the normalization to total and the normalization to median. After these steps, users have an option to further normalize the QA data using a nonlinear normalization procedure–Lowess fit method. This method of normalization has been shown to be effective in removing the variation caused by the differences in labeling efficiency between the two fluorescent dyes. The background subtracted & normalized channel intensities and related individual ratios for each data set are listed in an output data file so that users can use them for further studies. The user will be prompted to give the file name of this output data file.

- ### ANOVA Procedure:

  The users have options to perform ANOVA procedure to analyze their QA data. The ANOVA Table together with the estimated fold changes and the corresponding P-values based on T-tests for each gene and each pair of treatments selected are listed in an output file with default filename AnovaTable.txt. Users are able to select their own name for this output file.

- ### Simple Data Analysis:

  QAnalyzer calculates all the CASE/CONTROL ratios. Then based on all useful replicated spots obtained from multiple data sets, the mean of ratios and Median of ratios are calculated. We listed the mean of ratios, the standard deviation of (mean) ratios, the median of ratios, the mean of Log-Ratios(base 2) and the standard deviation of (mean) Log-Ratios(base 2) of multiple data sets in the output files. The user will be prompted to give the file name of this output file. In the output files you can also find the Gene Number, Gene Name, the mean of ratios for each data set, the number of spots actually used in the calculation of the overall mean and Median of ratios, and the number of outliers.

- ### Outliers Detecting

  A very simple outlier searching algorithm was incorporated in *QAnalyzer*. Spots which leads to extreme large difference between the mean of ratio and the median of ratio are defined as outliers and eliminated from the calculation. The number of outlier spots for each gene is listed in the output merge files. Moreover, for some spots the channel intensities after background subtraction may be less than 0. We eliminate these spots also from the overall ratio calculation. The number of spots actually used in ratio calculation for each gene is also listed in the output merge files.

- Alternative Ratio Selection

  With *QAnalyzer* a user can choose which ratios should be considered(Ch.1/Ch.2 or Ch.2/Ch.l and Ch.1/Ch.3 or Ch.3/Ch.l). The related channel(fluor) information is listed in the header of the output merge files.

Input File Preparation and Usage

- Convert all QuantArray (QA) output files into tab delimited ASCII text format (if they are not in this format).

  The input files of this program requires are the original QuantArray (QA) output files in tab delimited ASCII text format. However, QuantArray's default output file is in tab delimited UNICODE text format, so it must be converted to a tab delimited ASCII text file. *QAnalyzer* will detect the input file format automatically and will remind the user if the files need to be converted. There are three ways to do so.

(a) Open QA output files using Windows® Notepad, select Save as from the File menu, choose ANSI rather than Unicode Encoding, then click on the Save button.

(b) Sometime if a file has very large size, you may fail to open it with Windows Notepad. In this case, open QA output files with Windows Wordpad. Select Save as from the File menu. In the Save as type box, choose Text Document or Text Document-MS-DOS Format rather than Unicode Text Document, then click on the Save button.

(c) Or open QA output files with MS Excel. Delete the two weird symbols (ÿp) at the very beginning of the file manually (important! don't delete any valid characters). Then save them as tab delimited text files (they are now in ASCII format). If you are using Unix operating system, try Emacs or other Unix programs to convert a unicode file to an ASCII file. (Warning: Be careful when you save a large file using EXCEL. Sometimes you may lose part of the data values in the resaved files. Therefore if you encounter problems when running Qanalysis, check your data files to see if they are changed. You should avoid using EXCEL to save the data files.)

- Put QAnalyzer and all files to be analyzed in the same directory or folder.

(a) If you are using windows 95/98/2000/NT, just click QAnalyzer.exe, then the program will be activated. If you prefer to run the program under DOS command line, just open a DOS window from start menu of your computer, go to the folder where your files and program are, enter *QAnalyzer* and press Return or Enter key, the program will be activated.

(b) If you are using a Linux operating system, just go to the directory where your files and program are, enter *./qanalyzer.out* and press Return or Enter key, the program will be activated. Follow the program and enter the parameters, in general the total computing takes only a few seconds.

- Output Files.

  QAnalyzer has several output files: one data file, one ANOVA results file (optional), and two or more merge files depending upon how many channels in the input files. All files are in tab delimited text format. You can open them with MS Excel for further studies.

- Parameters

  The MOST IMPORTANT requirement for the QA output files is that the same treatment (variety, control and cases) must be assigned to the same channel through all the data sets (arrays). Otherwise, the user have to fill in the "Status" column in the header of the input data file in the "Image Info" region. For example, the following Table 1 is the part of a formal header QuantArray output data file, each of the arrays contains three channels. The three treatments "Control, Case1 and Case2" are assigned to channels 2, 1 and 3 respectively for this array. Formally, the QuantArray output data file contains the information about the dye assignments in the "image info" portion of QuantArray output data file. In Table 1 the dye assignment is "Cy3, Cy5 and Cy2" for Channels 1, 2 and 3 respectively. *QAnalyzer* will automatically detect the dye assignments and use this information to do the dye effects adjustment. If the QuantArray output data file does not contain this information, we recommend the user to add it in the step of data file preparation. Otherwise, *QAnalyzer* will prompt the user to type in this information

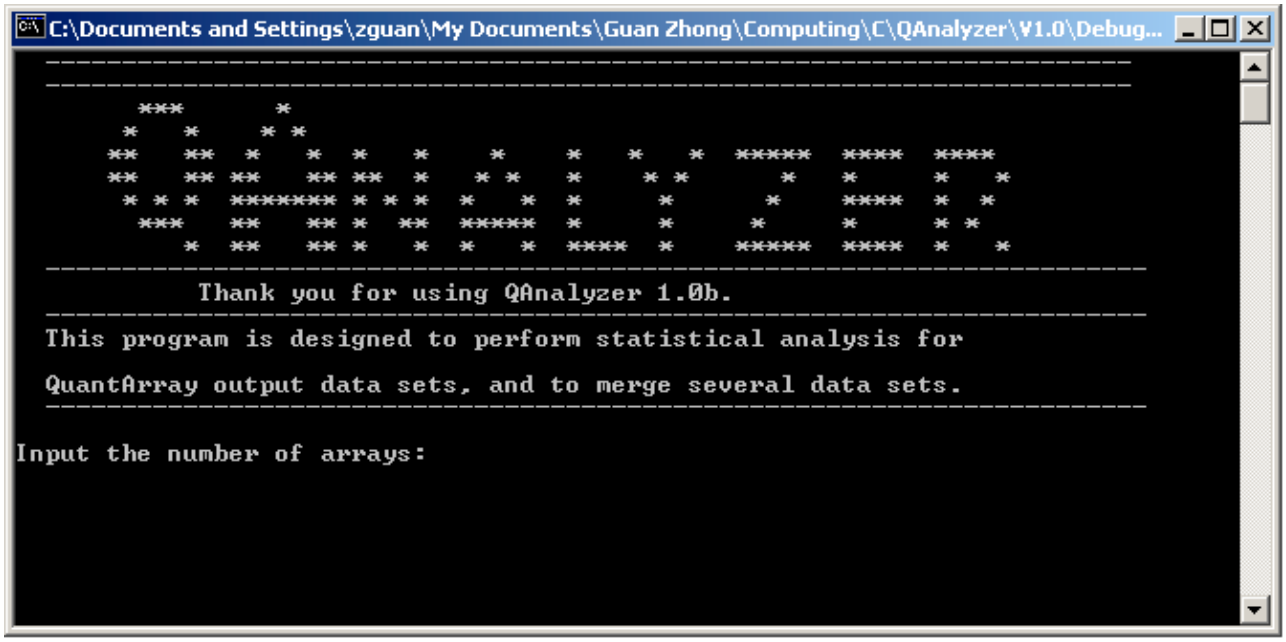
during the running of *QAnalyzer*. One more thing for the user to do before running *QAnalyzer* is to convert the QA data files from Unicode format to ASCII format(see the above description).

Table 1: Portion of Header of QuantArray output File

27										
28	Begin Image Info									
29	Channel	Image	Fluorophor	Barcode	Units	X Units Pe	Y Units Pe	X Offset	Y Offset	Status
30	ch1	E:Image1.TIF	Cy3		Microns	10	10	0	0	Case1
31	ch2	E:Image2.TIF	Cy5		Microns	10	10	0	0	Control Image
32	ch3	E:Image3.TIF	Cy2		Microns	10	10	0	0	Case2
33	End Image Info									
34										

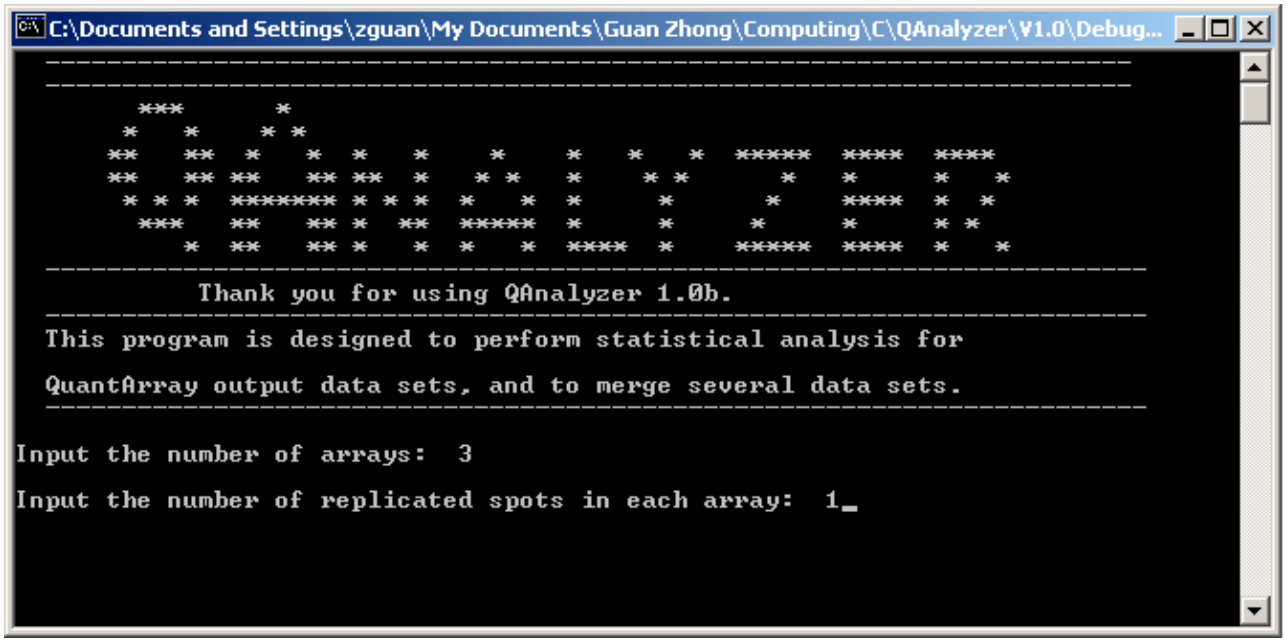
- Parameters.

After starting the program, besides the file names of the input and output data sets to be given by the users, the program requires only a few parameters to be entered. They are:
(1). **The number of data sets (arrays)** to be analyzed;



Step 1: Input the NUMBER OF ARRAYS.

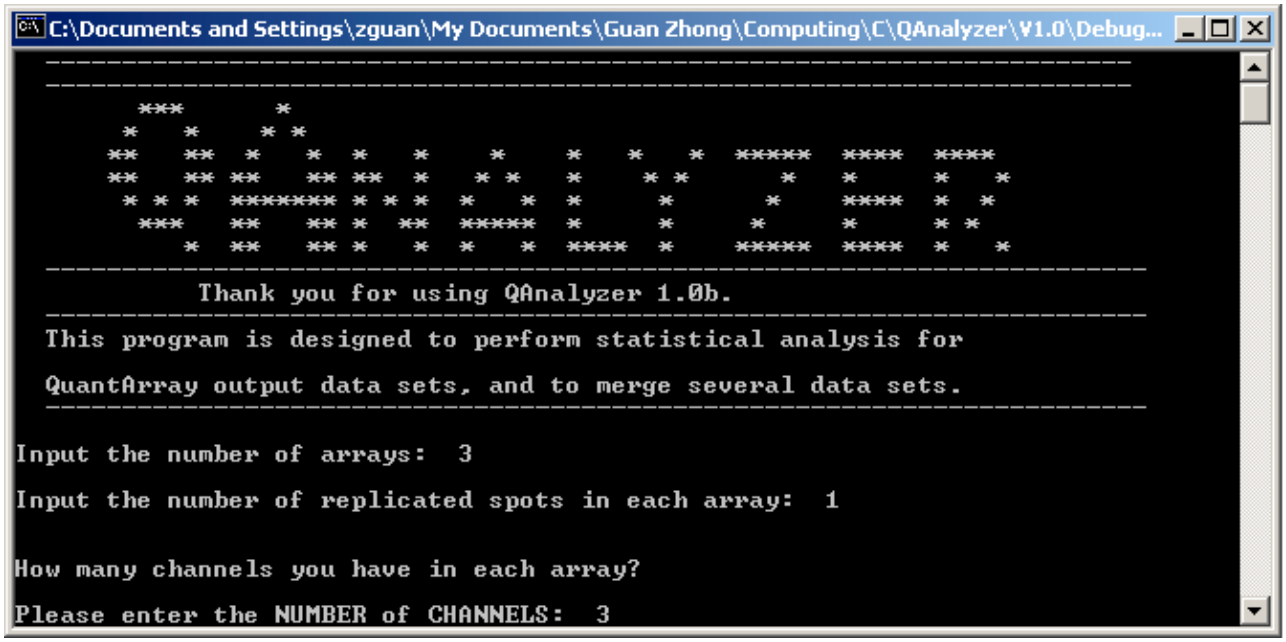
(2). The **number of replicate spots** in each data set. In most arrays, this number is 1. In some arrays, there may be equal number of consecutive spots represent the same gene. For example, in the following portion of an array (Table 2), each gene is spotted consecutively 4 times. Therefore the number of replicate spots in this array is 4.



Step 2: Type in the NUMBER OF REPLICATED SPOTS

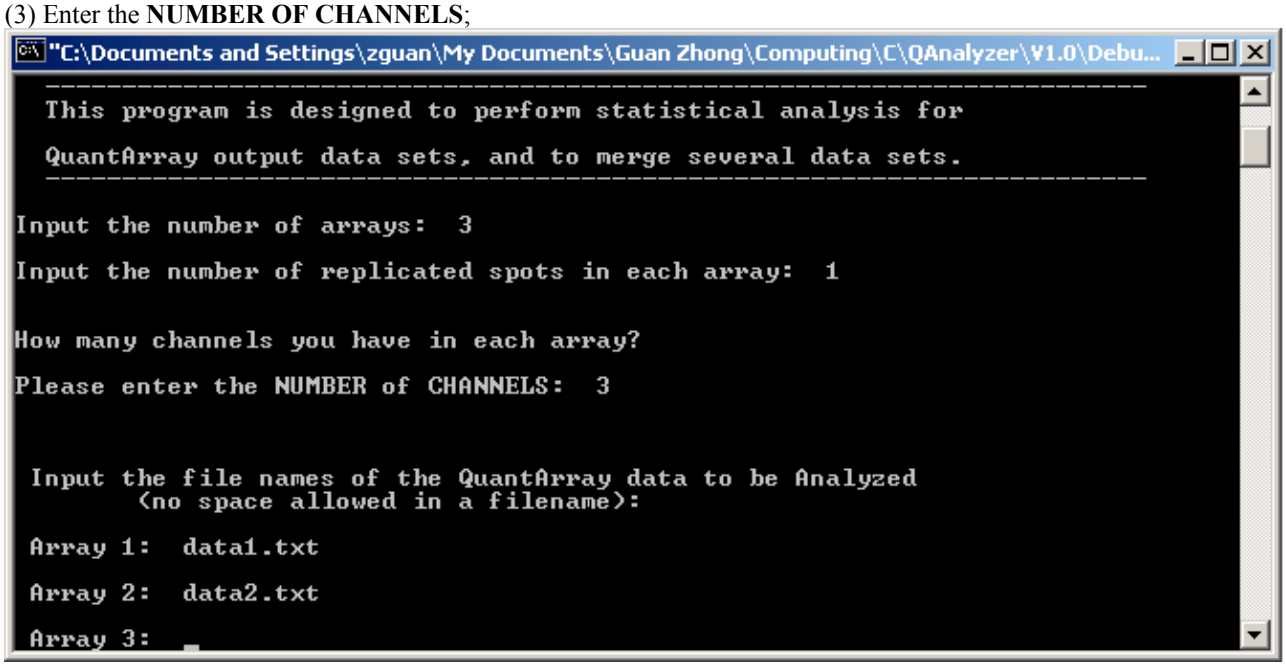
34												
35	Begin Measurements											
36	Number	Array Row	Array Colu	Row	Column	Name	ch1 Ratio	ch1 Perce	ch2 Ratio	ch2 Perce	ch3 Ratio	c
37	1	1	1	1	1	Gene1	1	26.1363	2.071965	54.15349	0.754132	
38	2	1	1	1	2	Gene1	1	24.93018	2.398732	59.80083	0.61247	
39	3	1	1	1	3	Gene1	1	36.78017	1.218921	44.83212	0.499935	
40	4	1	1	1	4	Gene1	1	55.4336	0.597636	33.12911	0.206324	
41	5	1	1	1	5	Gene2	1	40.56815	1.117146	45.32054	0.347842	
42	6	1	1	1	6	Gene2	1	27.99769	2.005585	56.15175	0.566138	
43	7	1	1	1	7	Gene2	1	26.30497	2.162444	56.88303	0.639119	
44	8	1	1	1	8	Gene2	1	25.64349	2.279736	58.46039	0.619889	
45	9	1	1	1	9	Gene3	1	28.0485	1.997009	56.01311	0.568244	
46	10	1	1	1	10	Gene3	1	31.90022	1.524105	48.6193	0.610669	
47	11	1	1	1	11	Gene3	1	37.66502	1.008957	38.00239	0.646026	
48	12	1	1	1	12	Gene3	1	26.94248	1.982181	53.40488	0.72943	
49	13	1	1	1	13	Gene4	1	19.51109	3.464838	67.60276	0.660453	
50	14	1	1	1	14	Gene4	1	19.52023	3.504503	68.4087	0.618388	
51	15	1	1	1	15	Gene4	1	29.92155	1.622832	48.55766	0.71924	
52	16	1	1	1	16	Gene4	1	20.12275	3.316726	66.74165	0.652773	
53	17	1	1	1	17	Gene5	1	19.08142	3.595259	68.60265	0.645441	
54	18	1	1	1	18	Gene5	1	21.96272	2.942104	64.6166	0.611067	

Table 2: Portion of a Typical QuantArray Output with Number of Replicated Spots Equal to 4.



Step 3: Enter the NUMBER OF CHANNELS

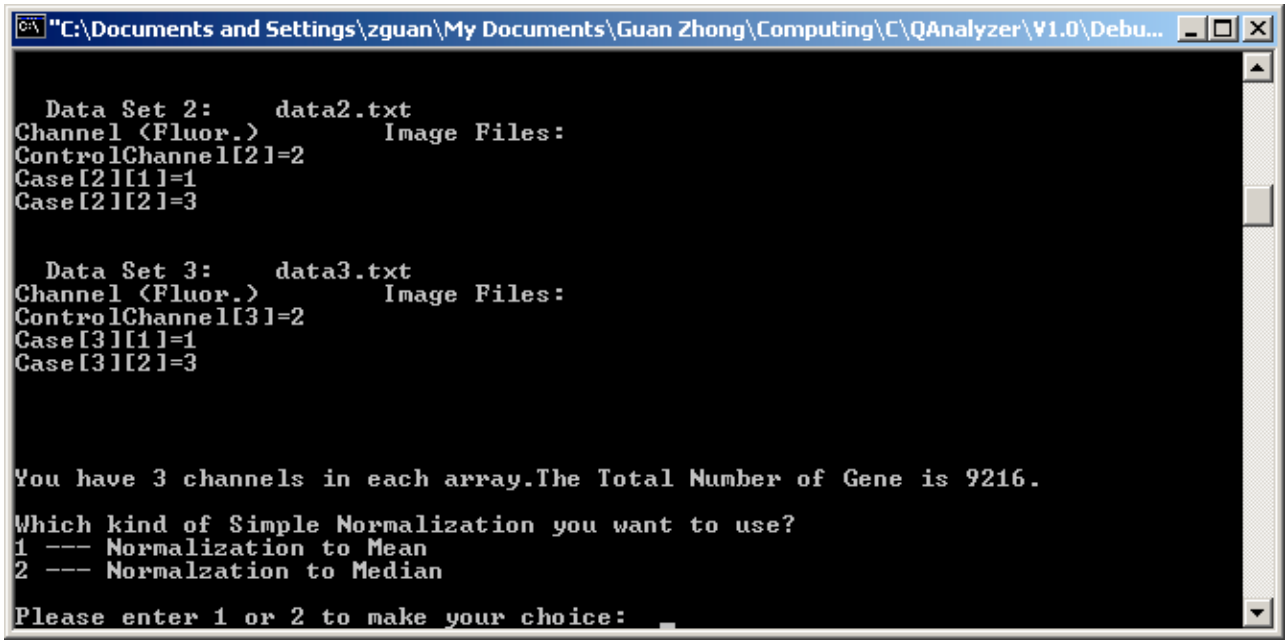
(3) Enter the **NUMBER OF CHANNELS**;



Step 4: Type in the FILE NAMES OF THE QA DATA SETS;

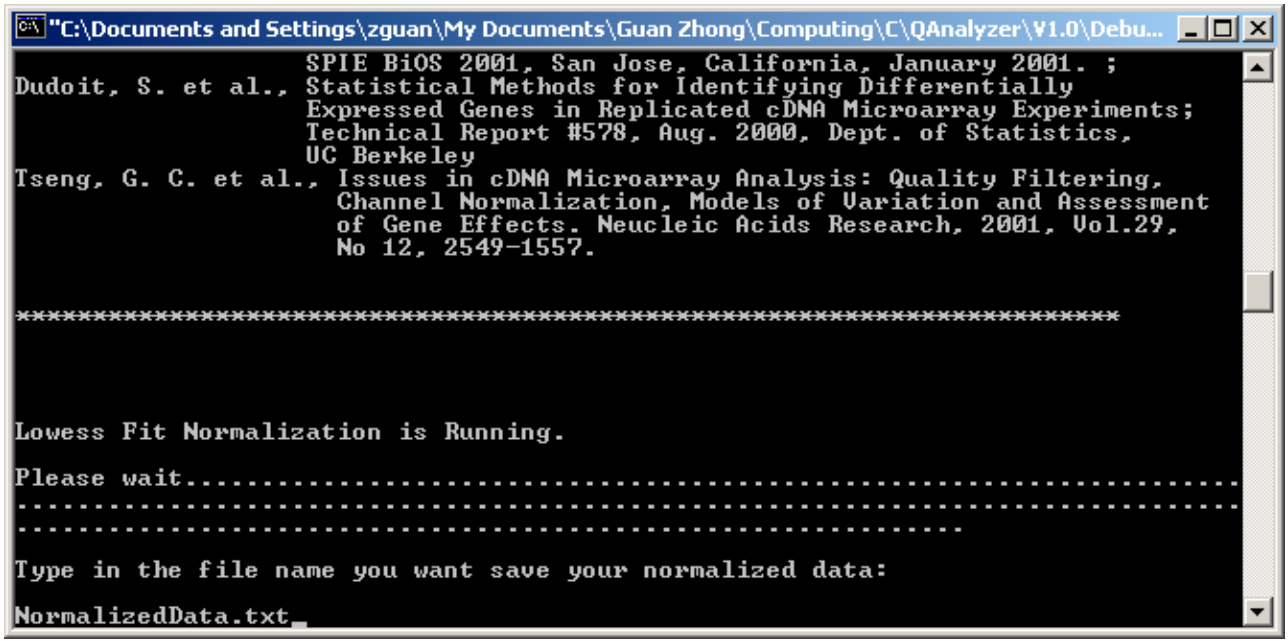
(4) Type in the **FILE NAMES OF THE QA DATA SETS**;

(5). Which kind of **Simple Normalization** you want to use? 1–Normalization to mean, 2–Normalization to median.



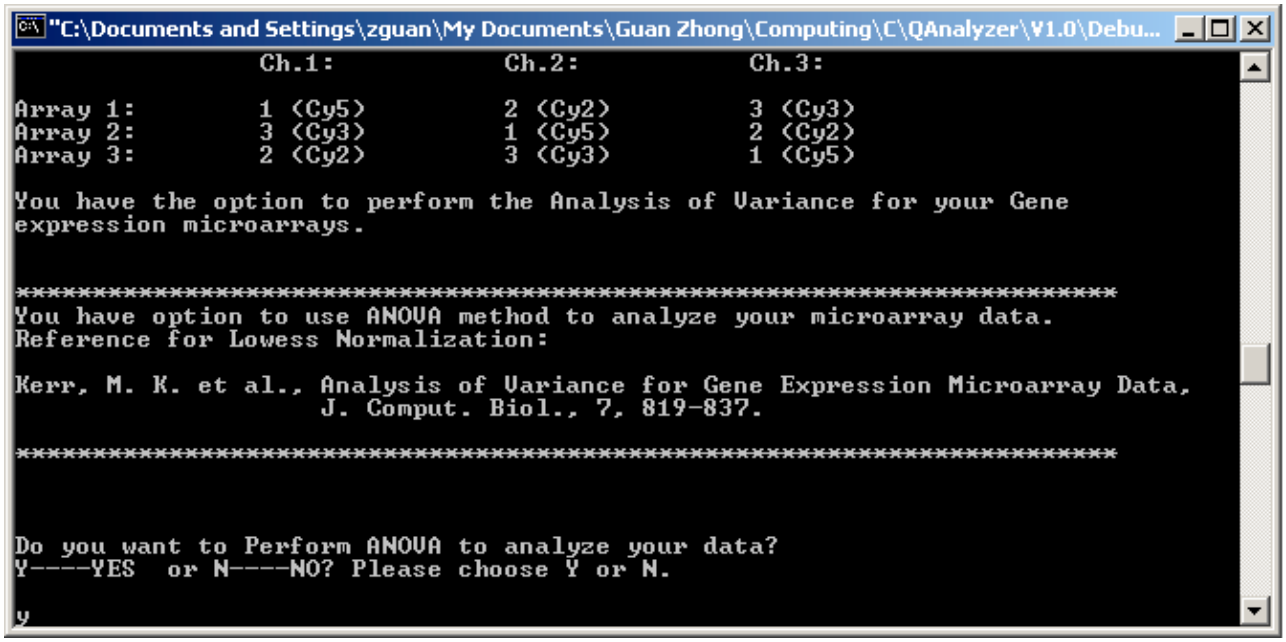
Step 5: SIMPLE NORMALIZATION

(6). Provide a File Name for the **Normalized Data** Set



Step 6: Type in the Normalized Data File Name

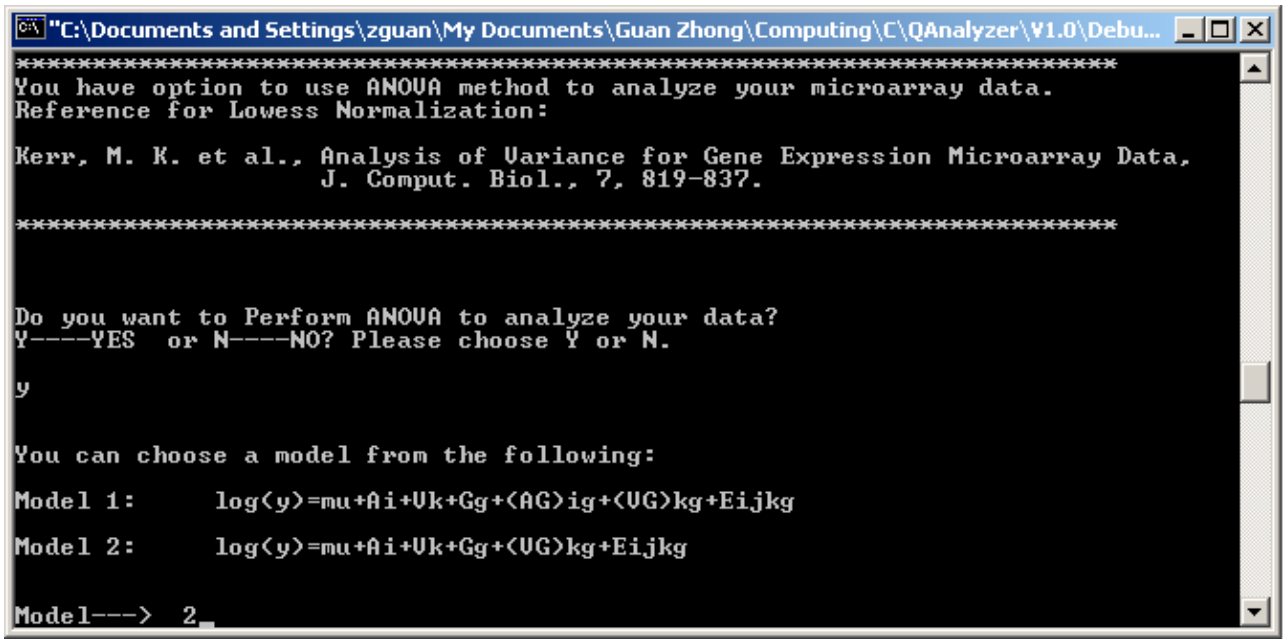
(7) Do you want to Perform the analysis of variance–**ANOVA**– to analyze your data?



Step 7: ANOVA

(8). If you choose ANOVA in step 7, then you need to select one of the following **models**:

- Model 1: $\log(y) = \mu + A_i + V_k + G_g + (AV)_{ik} + (VG)_{kg} + \varepsilon_{ikg}$;
- Model 2: $\log(y) = \mu + A_i + V_k + G_g + (VG)_{kg} + \varepsilon_{ikg}$.



Step 8: ANOVA Moddel Selection

The other steps can be finished by following the prompts.

Release and Version

The current version is *QAnalyzer* 1.0c, which was released on July 15, 2002. *QAnalyzer* was designed for the convenience of YMD(Yale Microarray Database) users to merge replicated data sets. No warranty is expressed or

implied. Compared to the previous version 1.0a, this version simplifies the use of this program. For examples, the Lowess fit normalization as a method of dye effects adjustment has been set to be a non-optional procedure in the program, because if there were no dye effects at all, then this procedure would not change the data very much. Since Lowess fit normalization will be always performed, we can reduce the number of ANOVA models. The user needs only to choose one or both of two ANOVA models. Compared to version 1.0b, in this version we require the user to specify the control- and caes-channels in the IMAGE INFO portion of the header of QuantArray data file. Only the Case/Control ratios are calculated.

QAnalyzer was released as a package in zip format which can be downloaded from http://zhao.med.yale.edu/. The package includes *QAnalyzer* win32 executive program, a User's Manual in PDF format, two replicated data sets with two (2) channels: *test1.txt* and *test2.txt*, and three data sets with three (3) channels *data1.txt, data2.txt* and *data3.txt*. User can use these files to test the program. In each of the first two data sets, there are 12 replicated spots. In each of the other three data sets, the genes are singly spotted.

Contacts

Please send bugs report and source code request to zhong.guan@yale.edu.
For comments, suggestions and critics please contact hongyu.zhao@yale.edu, kenneth.williams@yale.edu, janet.hager@yale.edu and zhong.guan@yale.edu
Last Update
July 15, 2002