GPmerge 4.0a Readme

It is widely accepted that any single microarray output is subjected to substantial variability. By pooling data from replicates, we can provide a more reliable classification of gene expression. Designing experiments with replications will greatly reduce misclassification rates.

The purpose of program ***GPmerge*** reflects our efforts toward to trying a reasonable way to merge replicated data sets obtained from image analysis software **GenePix**, so that a user can use the overall information provided by these replicated experiments.

Features

1. Choosing a method to calculate the channel ratio for a individual spot:
   GenePix calculates the channel ratio for a individual spot from pixels with 5 differerent methods, *i.e.*, *ratio of medians*, *ratio of means*, *median of ratios*, *mean of ratios* and *regression ratio*. ***GPmerge*** prompt the user to choose one of these methods. The ratio obtained with the chosen method was normalized with default **GenePix** normalization factor corresponding to that chosen method.

2. Data sets merging:
   GPmerge calculates the *mean of ratio*, *medium of ratio*, *the mean of log2 ratio* for each gene with all useful replicated spots obtained from multiple data sets. We listed the mean of ratio, the mean of log-ratio and the medium ratio of merged data sets in the output file *merge.txt*. In *merge.txt* you can also find the Gene Number, Gene Name, ID, the mean of ratio for each data set, the standard deviations of overall mean ratio and overall mean log-ratio for each gene, CV, Pvalue, the number of spots actually used to calculate the overall merged mean and medium of ratio, and the number of outliers.

3. Outliers searching:
   A very simple outlier searching algorithm was incorporated in ***GPmerge***, those spots which leads the large difference between the mean of ratio and the medium of ratio are defined as outliers and eliminated from the calculation. The number of outlier spots for each gene is listed in *merge.txt*. Moreover, for some spots the channel intensities after background subtraction may be less than 0, we eliminate these spots also from overall ratio calculation. The number of spots actually used in ratio calculation for each gene is also listed in *merge.txt*.

4. Quality control:
   Any spots which are flagged "Not found" or "Bad" by **GenePix** will be eliminated from the calculation by ***GPmerge***. For a given spot, if the ratio of intensity over background is less than constant C (user can choose C in a range from 2.0 to 5.0) for both channels, then this spot will also not be accounted for by ***GPmerge***.

5. Dye exchange handling:
   In some microarray experiment, dye exchange was conducted, in this case a user can define the *positive* dye configuration, then the dye configuration after exchange can be called *negative*. For those slides with *negative* dye configuration, a user can click *Option* button of **GenePix Pro3.0** to change the ratio formula($W_1/W_2 \rightarrow W_2/W_1$ if *positive* slides consider W_1/W_2; or $W_2/W_1 \rightarrow W_1/W_2$ if *positive* slides consider W_2/W_1). After changing the formula,click *Analysis* button of **GenePix Pro3.0** to extract data from these *negative* slides, the data should be exported as *gpr* files. The *gpr* files obtained in this way from *negative* slides, together with those *gpr* files from *positive* slides, can be used to ***GPmerge*** for pooling purposes.

6. Typing-error Checking:
   GPmerge has a systematical typing-error checking function, if user input a file name which doesn't exist or has been already used, the program will remind the user checking the error and prompt the user to modify the input. If a user enter a unlike parameter, the program will also prompt the user to enter that parameter again.

7. Correlation Coefficients:
   In output file *cor.txt*, different kinds of correlation coefficients are sum-

marized in table formats. These include the correllation coefficients both for intensity and for ratio measurements across different data set or across differents. Users can use these information to examine and compare their data.

8. CV(Coefficient of Variation).
   For each gene, CV(Standard Deviation devided by Mean) for all useful replicated spots was calculated and listed as a column in output file *merge.txt*. Users can use this information to check the quality of their spots or even their sildes.

9. Pvalue:
   For each gene, the *P-value* was evaluated from all useful replicated spots with *t-statistic*. This P-value serves only as an reference of siginifican level of defferential expression.

Usage

1. Put *GPmerge.exe* and all files to be merged in one directory.

   - IF you are using windows 95/98/2000/NT, just click *GPmerge.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 ***GPmerge*** and **Return** key, the program will be activated.
   - If you are using a Unix operating system, just go to the folder where your files and program are, enter ***GPmerge*** and **Return** key, the program will be activated.

   Follow the program and enter the parameters, in general the total computing takes only several seconds.

2. ***GPmerge*** has three output files: *merge.txt*, *cor.txt* and *data.txt*. All three files are in tab delimited text format, you can open them with **MS Excel** for further studies.

Parameters

The ***GPmerge*** is very easy to use, it requires original **GenePix** output *gpr* files, which are in ASCII tab delimited format.

Besides file names of data sets to be merged, the program requires only three parameters, which should be entered by user. They are: 1. The number of data sets to be merged; 2. The number of replicate spots in each data set. 3. The method of calculating ratio from pixels.(1←Ratio of medians; 2←Ratio of Means; 3←Median of ratios; 4←Median of means; 5←Regression ratio.)

Remark

- Always using **original** GPR files to ***GPmerge***.
  Eventhough the current version (4.0a) also works for GPR file which have been opened and resaved with programs like **MS Excel**, we still recommend not to open and resave the GPR file with **MS Excel** before using them to ***GPmerge***. Because **MS Excel** will slightly and silently change the format of GPR file.

- Make sure that your GPR files are from same version of GenePix Pro program before using ***GPmerge***.
  We noticed that the header format of a GPR file of **GenePix Pro** 3.0.0.x-3.0.5.x is different from that of **GenePix Pro** 3.0.6.x. The current version (4.0a) of ***GPmerge*** can automatically detect the different categories of header formats and treat them differently. Please make sure that your GPR files are from same version of GenePix Pro software(or the same category of GPR header format) in order to let ***GPmerge*** work properly.

Release and Version

The current version is ***GPmerge*** 4.0a, which was released on August 2, 2001. ***GPmerge*** was designed for the convenience of YMD(Yale Microarray Database) user to merge replicated data sets obtained from **GenePix Pro** image analysis software. No warranty is expressed or implied.

The ***GPmerge*** 4.0a was released as a package in zip format. The package includes ***GPmerge*** win32 executive program, a *Readme* in PDF format, two replicated data sets *p1.gpr* and *p2.gpr*. User can use these two files to test the program, in each data set there are 1 replicated spots.

The ***GPmerge*** 4.0a was tested for original GPR file outputed from GenePix Pro 3.0.0.98, 3.0.5.56, 3.0.6.52 and 3.0.6.73 (the most recent version) . We believe that it works for all GenePix pro 3.0 versions range from 3.0.0.x to 3.0.6.x. If you have GPR files which don't work with ***GPmerge*** 4.0a, please send email to *jinming.li@yale.edu* and attach your GPR files for the test purpose.

Contact

For bugs report and request for source code please contact *jinming.li@yale.edu*. For comments, suggestions and critics please contact *hongyu.zhao@yale.edu*, *kenneth.williams@yale.edu*, *janet.hager@yale.edu* and *jinming.li@yale.edu*

Last Update

August 2, 2001.