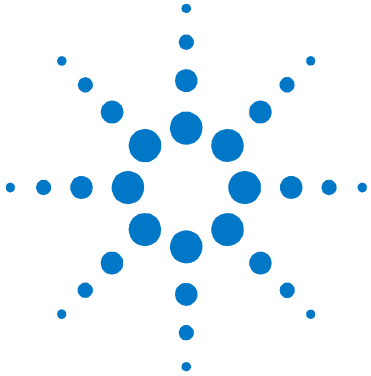


RMA Probe Summarization

GeneSpring GX 7.3.1

And

GeneSpring GX 9.0

Contents

Probe Summarization Algorithms

Definition and Applications

Probe summarization algorithms perform the following 3 key tasks:

- Background Correction
- Normalization
- Probe Summarization (i.e. conversion of probe level values to probeset expression values in a robust, i.e., outlier resistant manner)

The order of the last two steps could differ for different probe summarization algorithms.

For probe intesnsity measurements from Affymetrix Gene expression chips, one of the algorithms used in both GeneSpring GX 7.3.1 and GeneSpring GX 9.0 is RMA

RMA

RMA (Robust Multi-array Analysis) is a method for normalizing and summarizing probe-level intensity measurements from Affymetrix Gene Chips ®. Starting with the probe-level data from a set of Gene Chips, the perfect-match (PM) values are background-corrected, normalized and finally summarized resulting in a set of expression measures. The three steps of the process are outlined below.

Background Correction

It has been argued that background correction is the most crucial step for probe level processing.
The background correction used in RMA is a non-linear correction, done on a per-chip basis. It is is based on the distribution of PM values amongst probes on an Affymetrix array. PM values are a mixture of a background signal, caused by optical noise and non-specific binding, plus a signal, which is what we are trying to detect. The background is estimated as expectation of the signal (S) conditioned on observed PM values (O), using a kernel density estimation in both GeneSpring GX 7.3.1 and GeneSpring GX 9.0. However,, however GeneSpring GX 7.3.1 uses direct convolution while GeneSpring GX 9.0 uses Fast Fourier Transformation.

Normalization

Normalization is necessary so that multiple chips can be compared to each other, and analyzed together. The normalization procedure is aimed at making the distributions identical across arrays. The normalization used in RMA is quantile normalization. This usually gives very sharp normalizations.

Both GeneSpring GX 7.3.1 and GeneSpring GX 9.0 use quantile normalization. Note that, in this procedure, all the arrays are used and no chip is discarded based on extreme value considerations.

Summarization

Once the probe-level PM values have been background-corrected and normalized, they need to be summarized into expression measures, so that the result is a single expression measure per probe-set, per chip. The summarization used is motivated by the assumption that observed log-transformed PM values follow a linear additive model containing a probe affinity effect, a gene specific effect (the expression level) and an error term. For RMA, the probe affinity effects are assumed to sum to zero, and the gene effect (expression level) is estimated using median polishing. Median polishing is a robust model fitting technique, that protects against outlier probes.

Both GeneSpring GX 7.3.1 and GeneSpring GX 9.0 use same methodology for summarization.

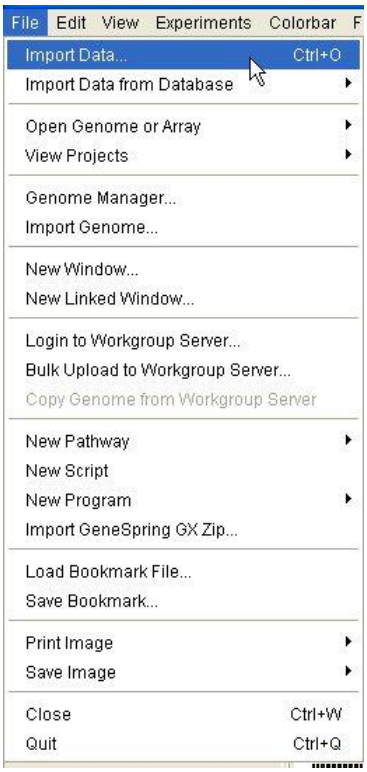
Analyzing Affymetrix Expression Data

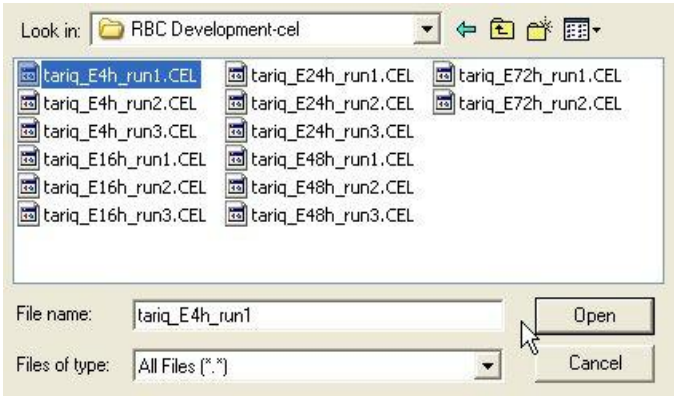
GeneSpring GX 7.3.1

The following steps need to be performed in GeneSpring GX 7.3.1 to analyze Affymetrix gene expression chips :

Step 1 : Import Data

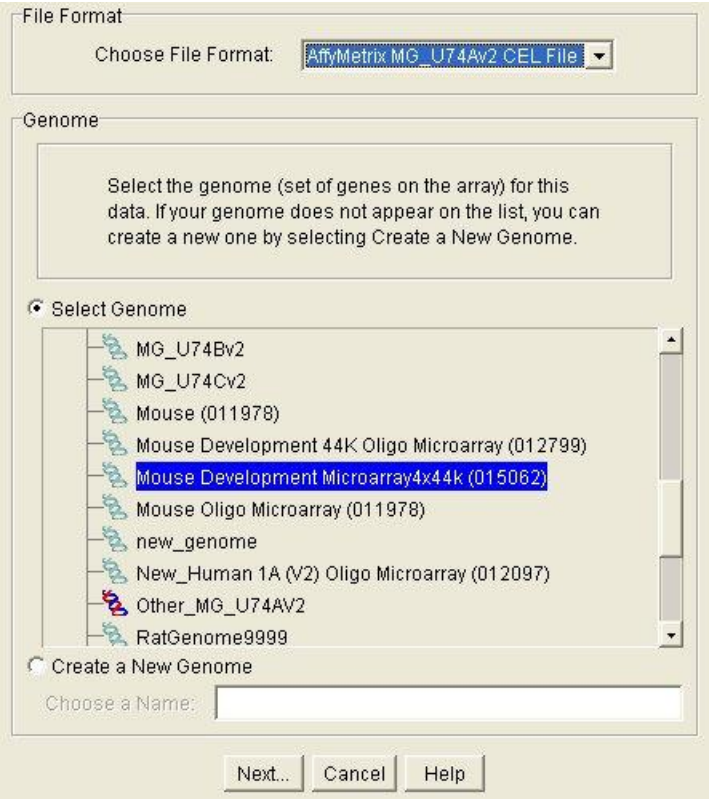
Select the data file you want to import in GeneSpring GX 7.3.1 using *File* > *Import Data*





Step 2 : Choose File format and select the appropriate genome

GeneSpring GX 7.3.1 automatically recognizes the file format and displays it for standard Affymetrix expression; Agilent one color and two color; Illumina; and Codelink chips.

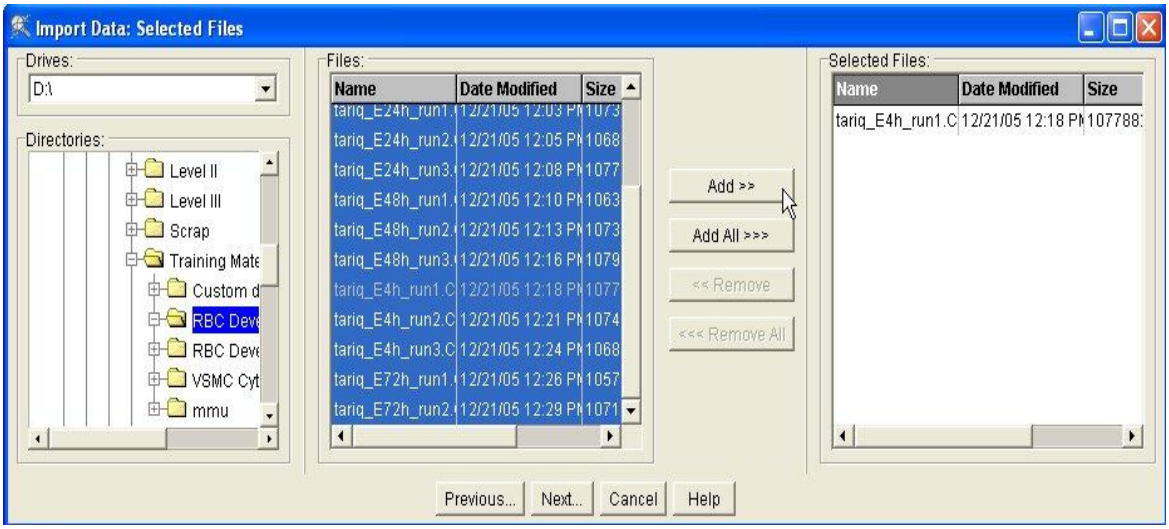


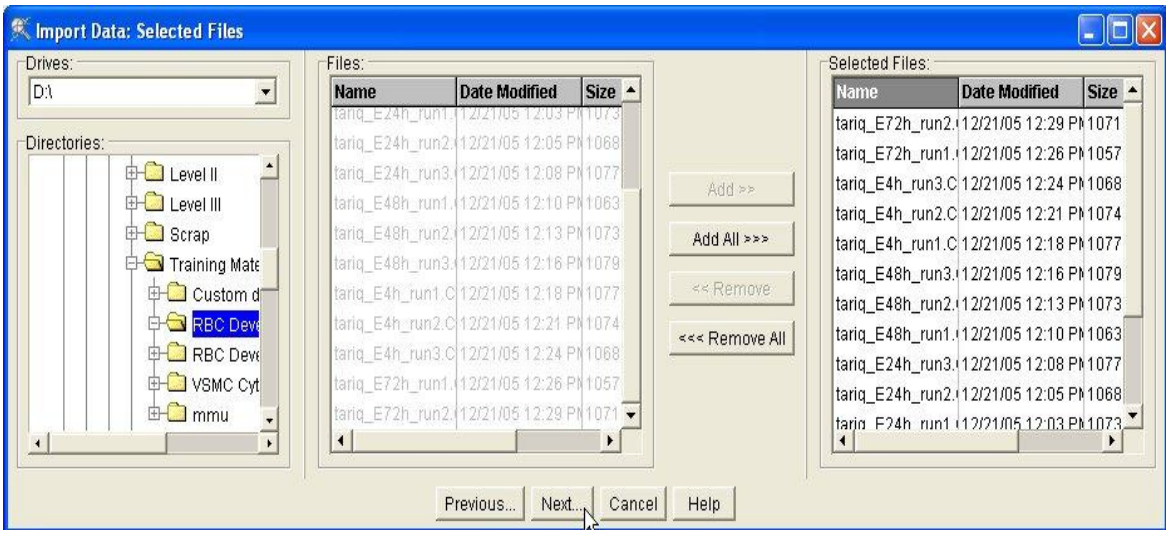
Step 3 : Choose the Preprocessor

Select the appropriate preprocessing algorithm – *'RMA'* or *'GC RMA'*. You might be asked to define the location of the CDF file or Array Definition file.

Step 4 : Choose more data files

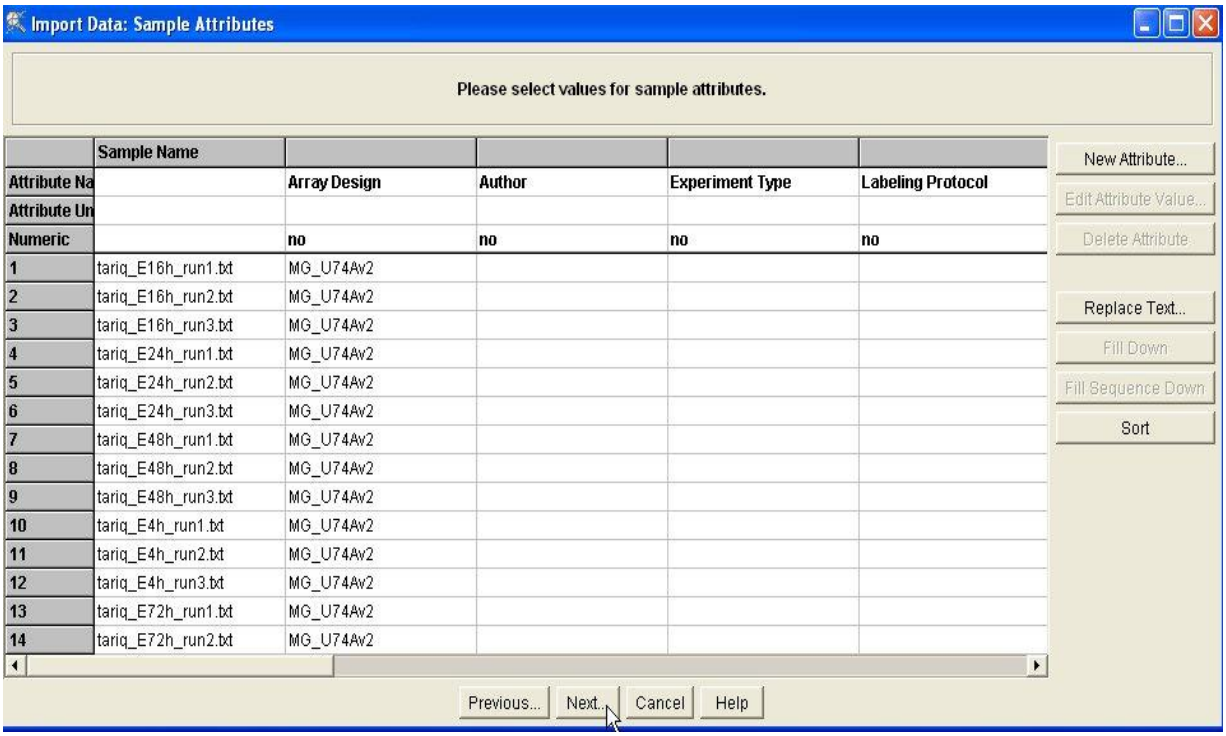
This window allows you to add more files of the same type to add to your experiment during the import process.





Step 5 : Sample Attributes window

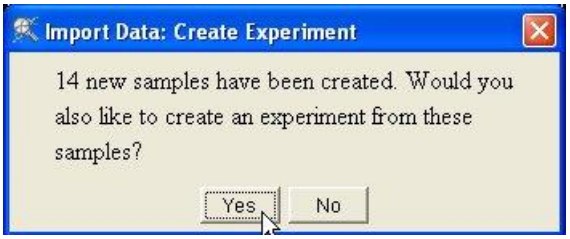
This window allows you to add sample attributes, which are required for MIAME compliance. This is an optional step and can be performed at a later stage as well.



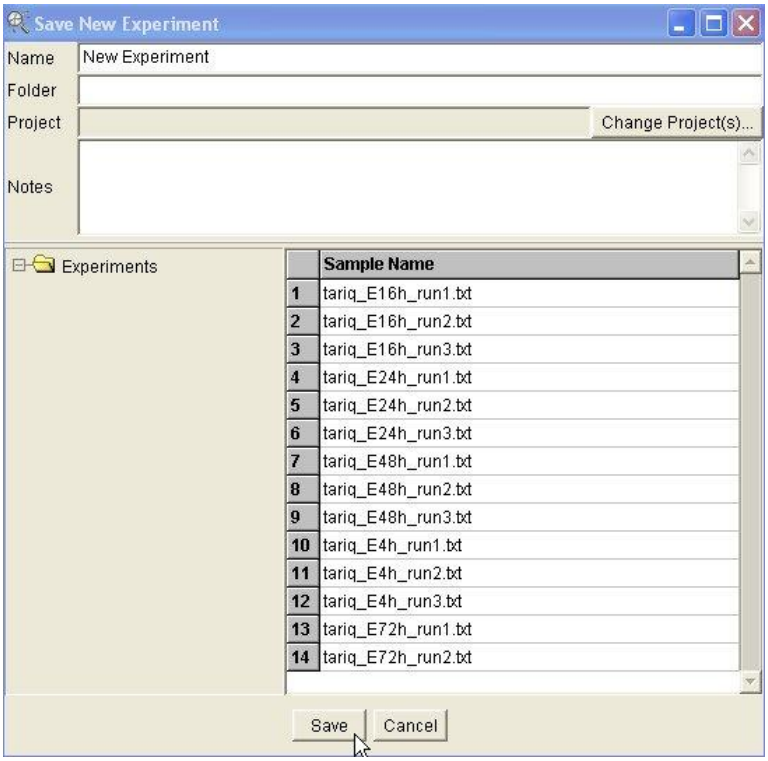
	Sample Name				
Attribute Na		Array Design	Author	Experiment Type	Labeling Protocol
Attribute Un					
Numeric		no	no	no	no
1	tariq_E16h_run1.txt	MG_U74Av2			
2	tariq_E16h_run2.txt	MG_U74Av2			
3	tariq_E16h_run3.txt	MG_U74Av2			
4	tariq_E24h_run1.txt	MG_U74Av2			
5	tariq_E24h_run2.txt	MG_U74Av2			
6	tariq_E24h_run3.txt	MG_U74Av2			
7	tariq_E48h_run1.txt	MG_U74Av2			
8	tariq_E48h_run2.txt	MG_U74Av2			
9	tariq_E48h_run3.txt	MG_U74Av2			
10	tariq_E4h_run1.txt	MG_U74Av2			
11	tariq_E4h_run2.txt	MG_U74Av2			
12	tariq_E4h_run3.txt	MG_U74Av2			
13	tariq_E72h_run1.txt	MG_U74Av2			
14	tariq_E72h_run2.txt	MG_U74Av2			

Step 6 : Experiment Creation

After the data files have been successfully imported and samples have been created, GeneSpring GX 7.3.1 prompts you to create an experiment from these samples.

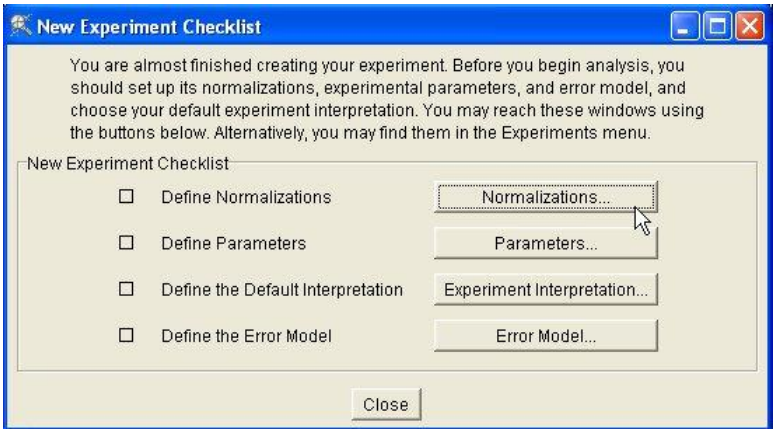


Provide an appropriate name for the New Experiment



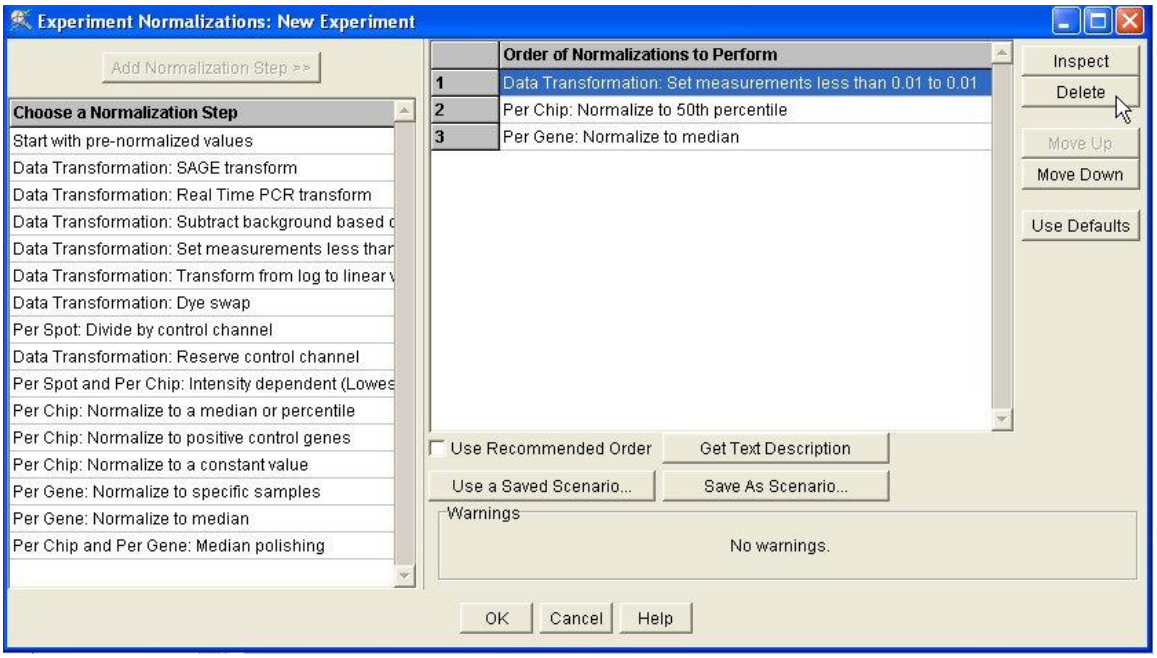
Step 7 : New Experiment Checklist

After the Experiment is created, you get the option to define Experiment Normalizations, Parameters, interpretation and Cross Gene Error Model.

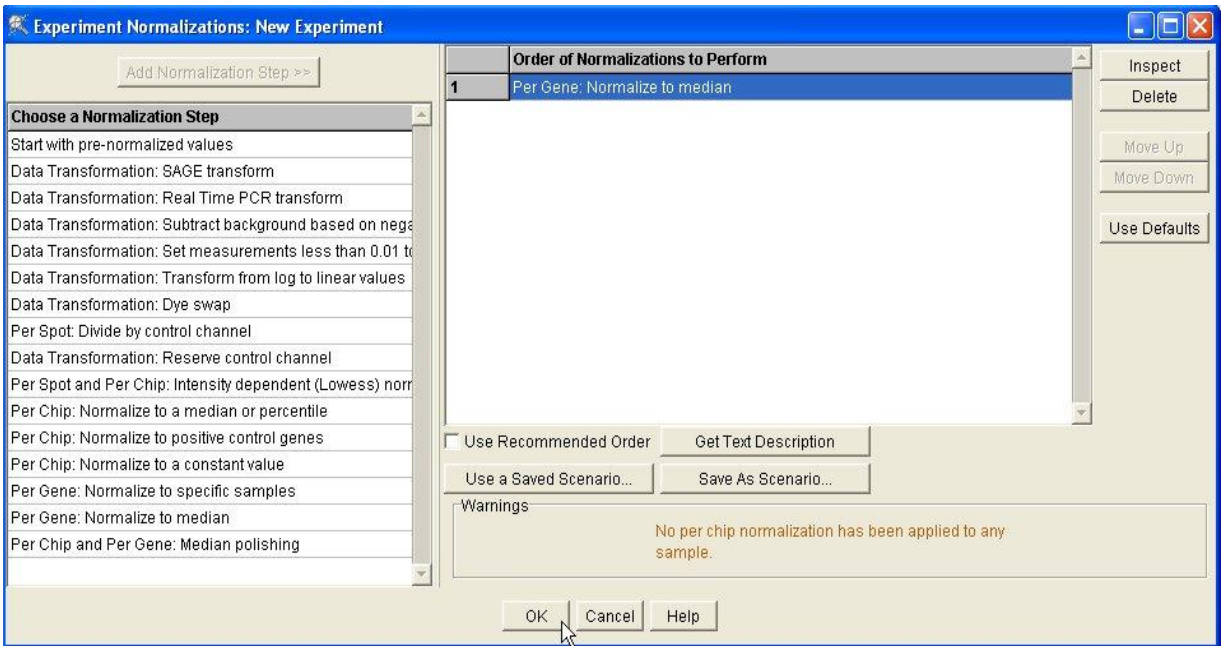


Step 8 : Experiment Normalizations

This window allows you to define what normalization(s) need to be performed on your data.



For affymetrix data preprocessed using RMA or GC RMA preprocessor, 'data transformation' and 'per chip' normalization needs to be deleted at this step, as these normalization steps have already been performed during preprocessing.

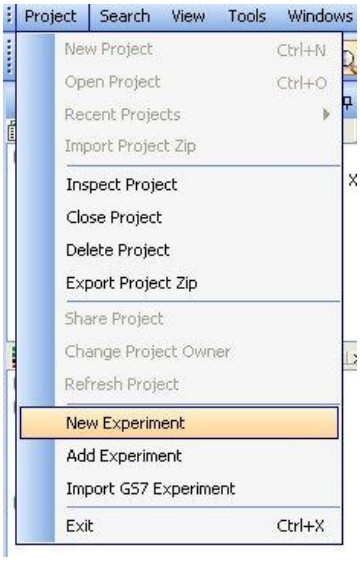


GeneSpring GX 9.0

The following steps need to be performed in GeneSpring GX 9.0 to analyze Affymetrix gene expression chips :

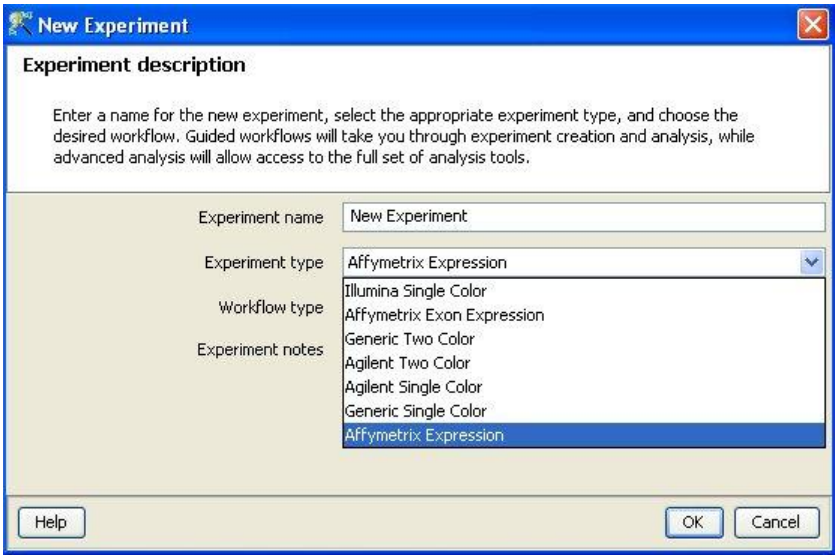
Step 1 : Create New Experiment

Create a new experiment using Project > New Experiment



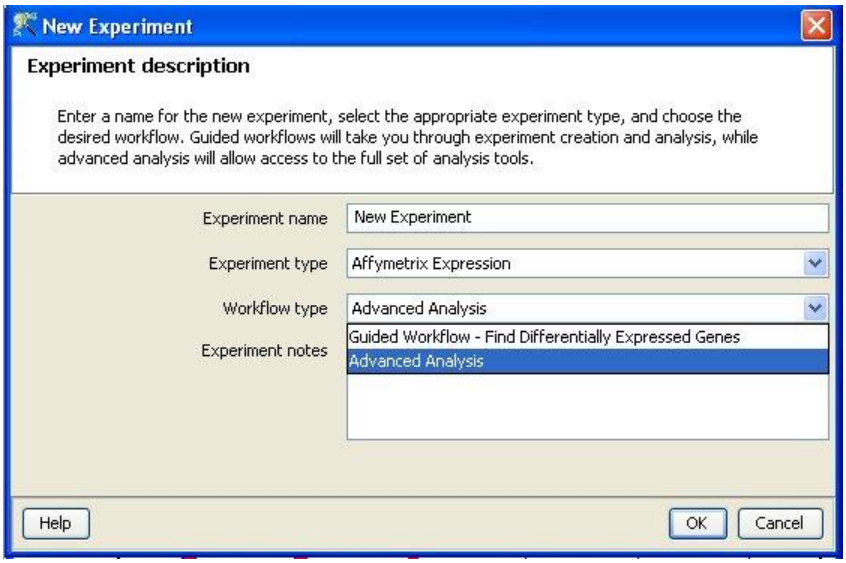
Step 2 : Experiment Description

Provide an appropriate Name and Experiment type (or, chip type) for the new experiment



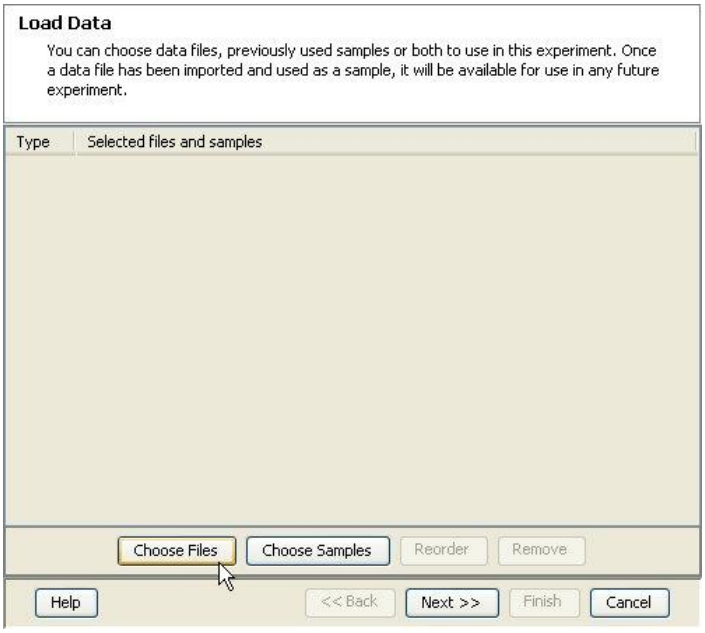
You can also define the Workflow type – *'Guided Workflow'* or *'Advanced Analysis'*

Guided workflow is designed to assist the user throughout the creation and analysis of an experiment with a set of default parameters, while in the *Advanced Analysis*, the parameters can be changed to suit individual requirements.

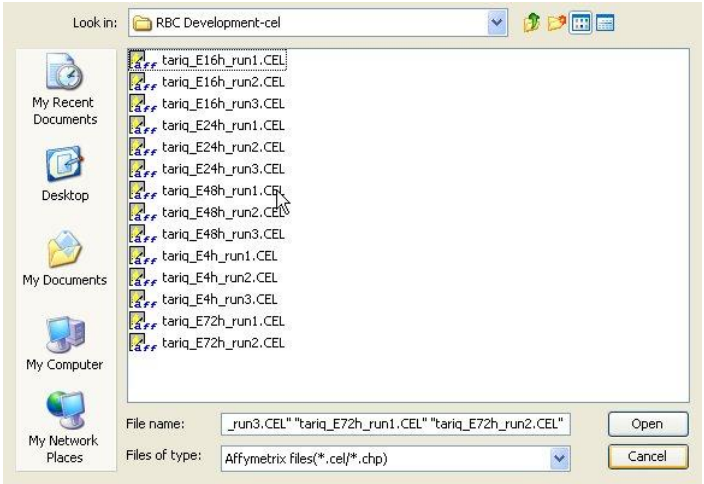


Step 3 : Load Data

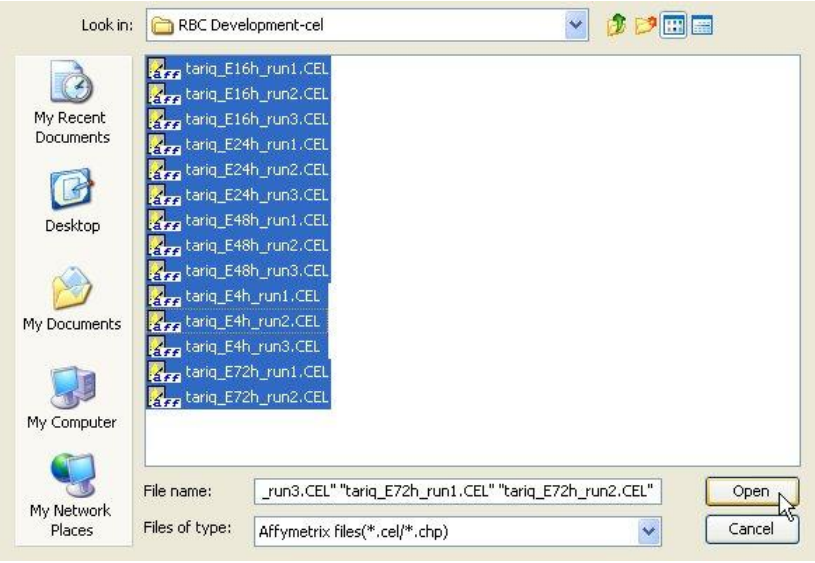
An experiment can be created using either the data files or else using samples. Upon loading data files, GeneSpring GX associates the files with the technology (see below) and creates samples. These samples are stored in the system and can be used to create another experiment via the *Choose Samples* option. For selecting data files and creating an experiment, click on the *Choose File(s)* button.

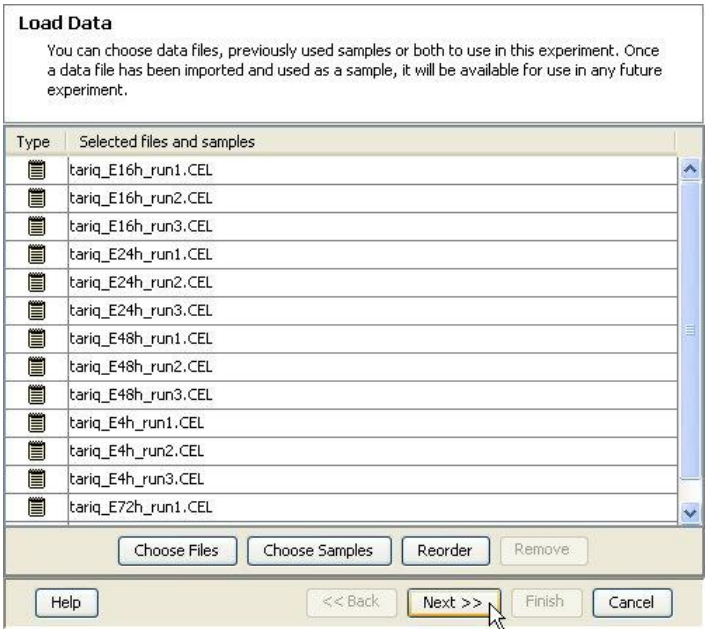


Navigate to the appropriate folder



Select the files of interest and select *Open* to proceed.

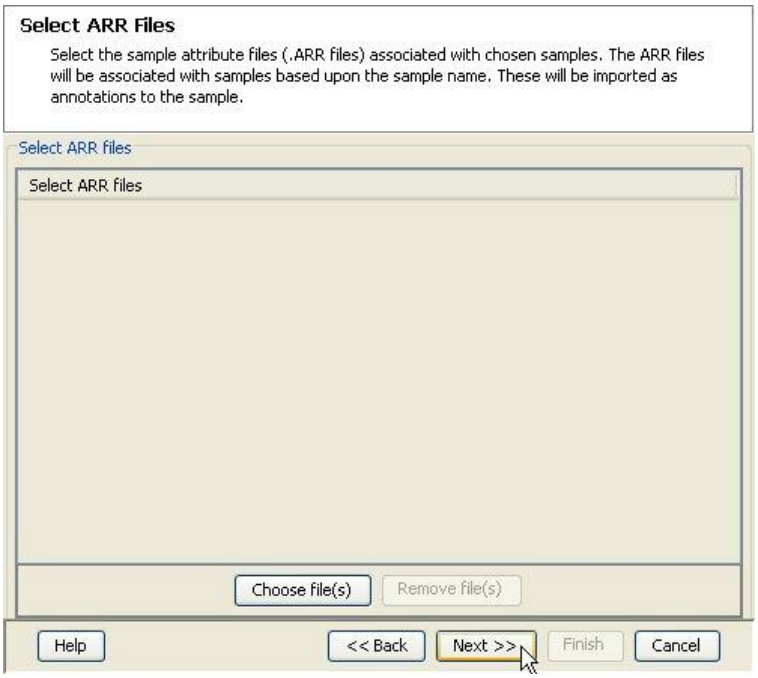




There are two things to be noted here. Upon creating an experiment of a specific chip type for the first time, the tool asks to download the technology from the GeneSpring GX update server. If an experiment has been created previously with the same technology, GeneSpring GX then directly proceeds with experiment creation.

Step 4 : Select ARR files

ARR files are Affymetrix files that hold annotation information for each sample CEL and CHP file and are associated with the sample based on the sample name. These are imported as annotations to the sample.



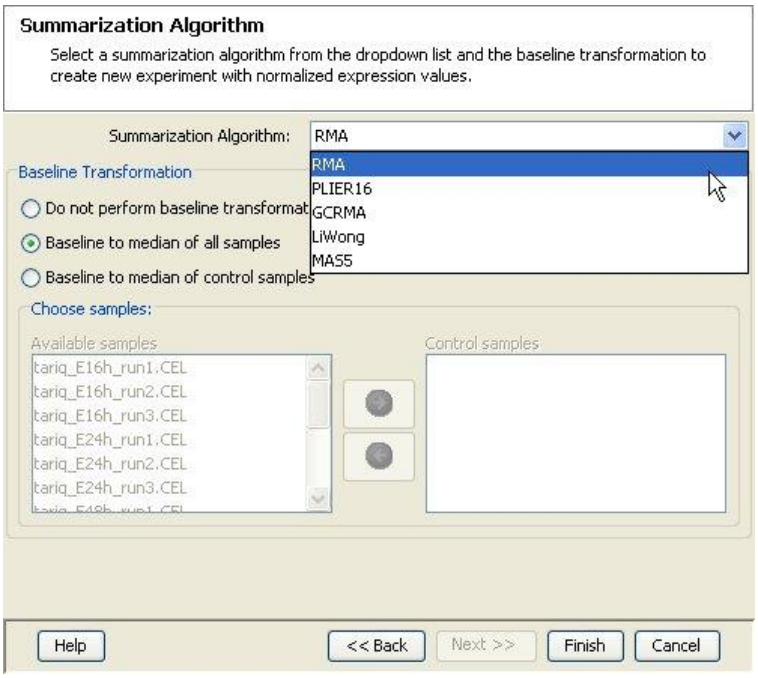
Step 5 : Select Probe Summarization and Normalization options

Select RMA as the Probe Summarization algorithm from the drop down list.

Subsequent to probe set summarization, baseline Transformation of the data can be performed. The baseline options include:

- Do not perform baseline
- Baseline to median of all samples
- Baseline to median of control samples

Note : 'Baseline Transformation' in GeneSpring GX 9.0 is equivalent to 'per gene normalization' in GeneSpring GX 7.3.1



Clicking Finish creates an experiment, which is displayed as a Box Whisker plot in the active view. Alternative views can be chosen for display by navigating to View in Toolbar.