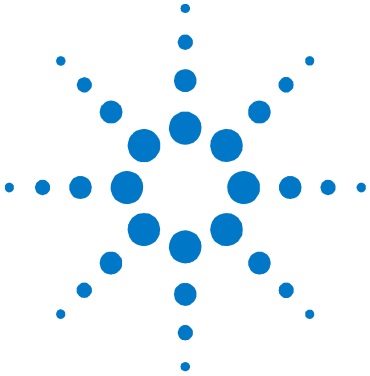


GCRMA Probe Summarization

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 GCRMA.

GCRMA

GCRMA is a method of converting *.CEL* files into expression set using the *Robust Multi-array Average (RMA)* with the help of probe sequence and with GC-content background correction.

It is a method for normalizing and summarizing probe-level intensity measurements from Affymetrix GeneChips. Starting with the probe-level data from a set of GeneChips, 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 given below.

Background Correction:

The background correction used in GCRMA is designed to account for background noise, as well as non-specific binding. Probe affinity is modeled as a sum of position-dependent base effects, and can thus be calculated for each PM and MM value, based on its corresponding sequence information.

The correction is motivated by the assumptions that observed PM and MM values consist of optical noise, non-specific binding noise, and signal. Optical noise is assumed to be normal, and logged non-specific binding noise from PM-MM pairs assumed to be bivariate normal. Using the data on a single array, the corresponding model parameters can be estimated.

The background adjustment in GCRMA consists of three sequential steps:

1. optical background correction
2. probe intensity adjustment through non-specific binding (NSB) utilizing affinity information and optical noise-adjusted MM intensities
3. probe intensity adjustment through gene-specific binding (GSB), where NSB-adjusted PM intensities are further corrected for the effect of PM probe affinities

Optical correction is important, as scanner measuring hybridization strength introduces optical noise. In GeneSpring GX 7.3.1 , background correction is still done using MM but these values are not adjusted for optical noise (i.e we still adjust through non specific binding using affinity and do probe intensity adjustments). Thus, implementation of optical correction in GeneSpring GX 9.0 is an important enhancement.

Each PM value is then adjusted by subtracting a shrunken MM value that has been corrected for its affinity. The affinity of a probe is described as sum of position dependent base affinities; this affinity of probe is calculated from their sequence. Each base at each position contributes to the total affinity of probe in an additive fashion. For a given type of base the positional effect is modeled.

Normalization

Normalization is necessary so that multiple chips can be compared to each other, and analyzed together. It is motivated by the assumption that all n chips should have approximately the same distribution of PM values. The normalization used in GCRMA is quantile normalization. This is a generalization of the idea behind quantile-quantile plots to more than two dimensions. The quantiles for each PM value are plotted in n dimensions, and projected onto the diagonal. The final result is that the PM values on each chip will have the same distribution.

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 GCRMA, 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.

The analysis performed using GCRMA in GeneSpring GX 9.0 shows different results from that of GeneSpring GX 7.3.1. This is primarily due to the difference in the implementations of GCRMA algorithm in both the versions. GeneSpring GX 9.0 uses an additional component of 'Optical Correction' while performing GCRMA, where as GeneSpring GX 7.3.1 does not perform optical correction.

As an example, for the following data set both GeneSpring GX 7.3.1 and GeneSpring GX 9.0 produce different results, when processed using GCRMA.

Experimental design and data set:

Patients with cardiomyopathy have weakened heart pumps which can result in the heart not being able to pump enough blood to the body's other organs- a condition known as congestive heart failure (CHF). Patients with ischemic cardiomyopathy have weakened heart pumps due to insufficient blood and oxygen being delivered to the area. A patient with non-failing hearts was performed.

We have four replicates each for *ischemic cardiopathy* and *Normal*
Ischemic : PAS_3.cel, PAS_6.cel, PAS_7.cel, PAS_8.cel
Normal : PA-N_249.cel, PA-N_300.cel,PA-N_322.cel, PA-N_326.cel

Welch T-test Result Matrix

Total Number of genes = 54675

Application	P value less than 0.05 without MTC
R without optical correction	11862
R with optical correction	33706
GeneSpring GX 7.3.1	12867
GeneSpringGX 9.0	33964

The above data set can be downloaded from the following ftp site
ftp://cg2.med.harvard.edu/pub/proj1/expts/Hs/Affy/

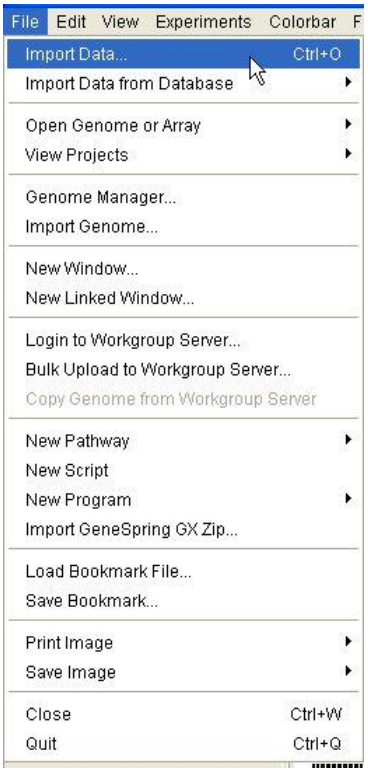
Analyzing Affymetrix Expression Data

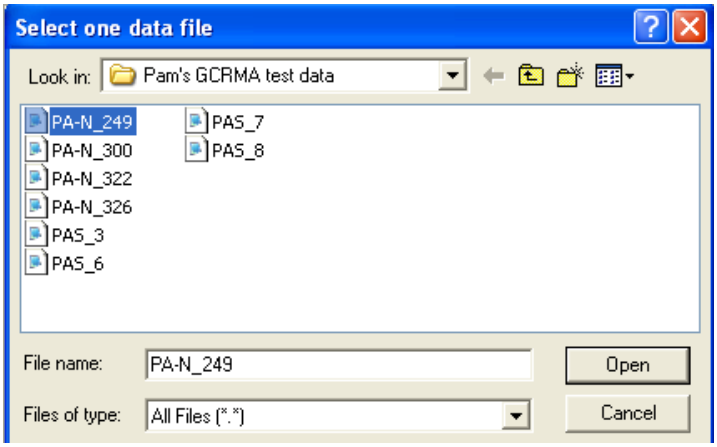
GeneSpring GX 7.3.1

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

Step 1 : Import Data

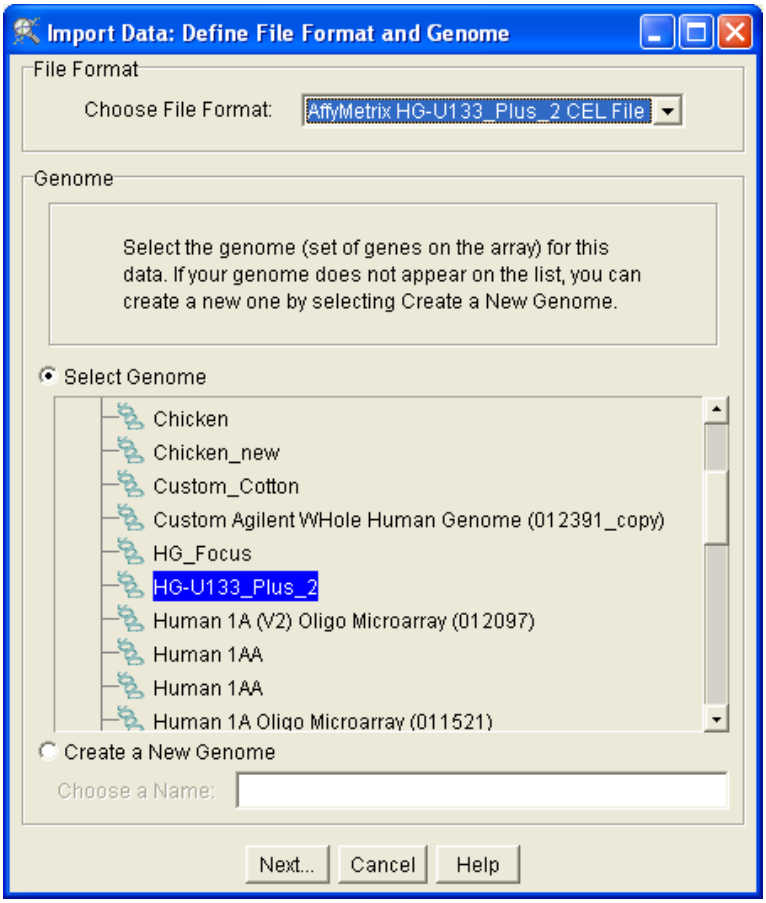
Select the data file you want to import in GeneSpring GX 7.3.1 using *File* > *Import Data*





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.

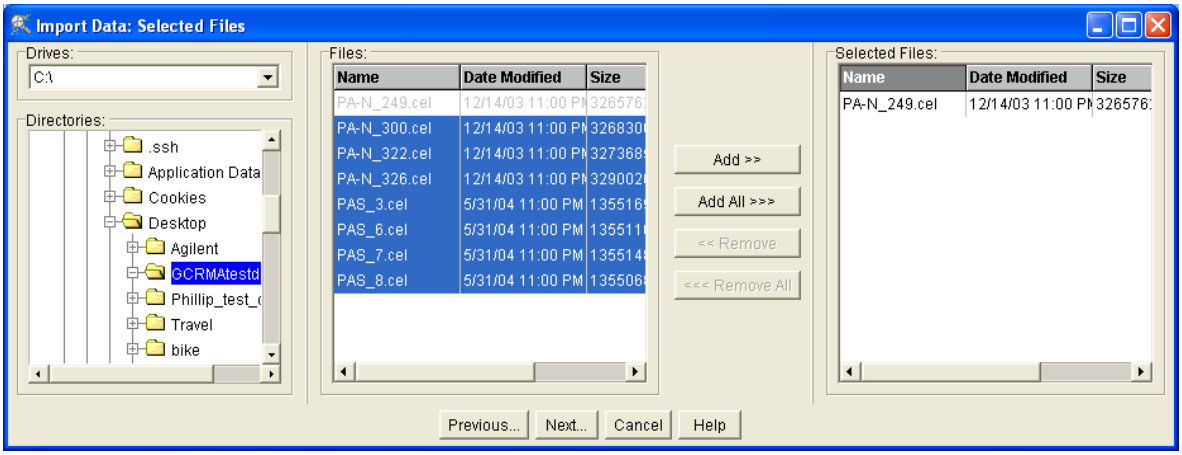


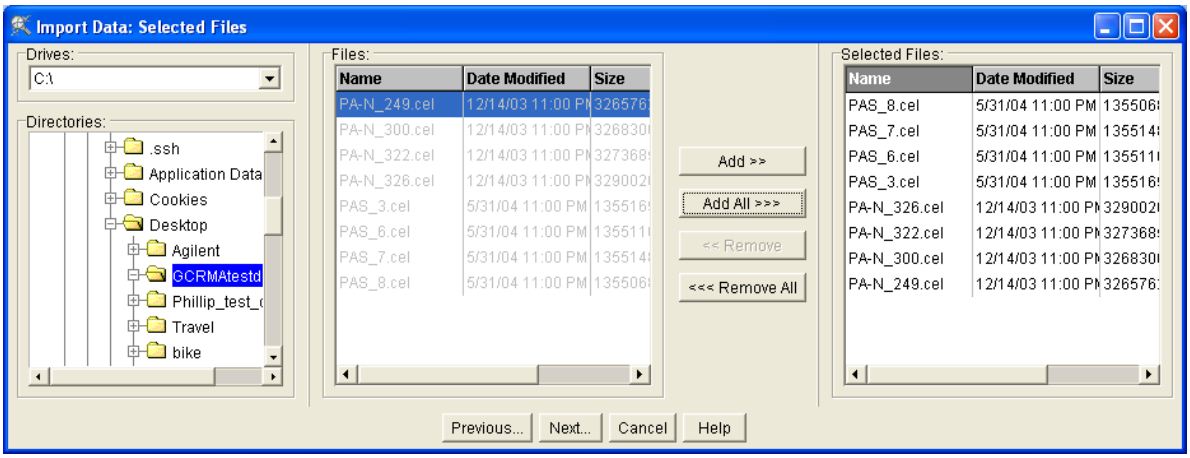
Step 3 : Choose the Preprocessor

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

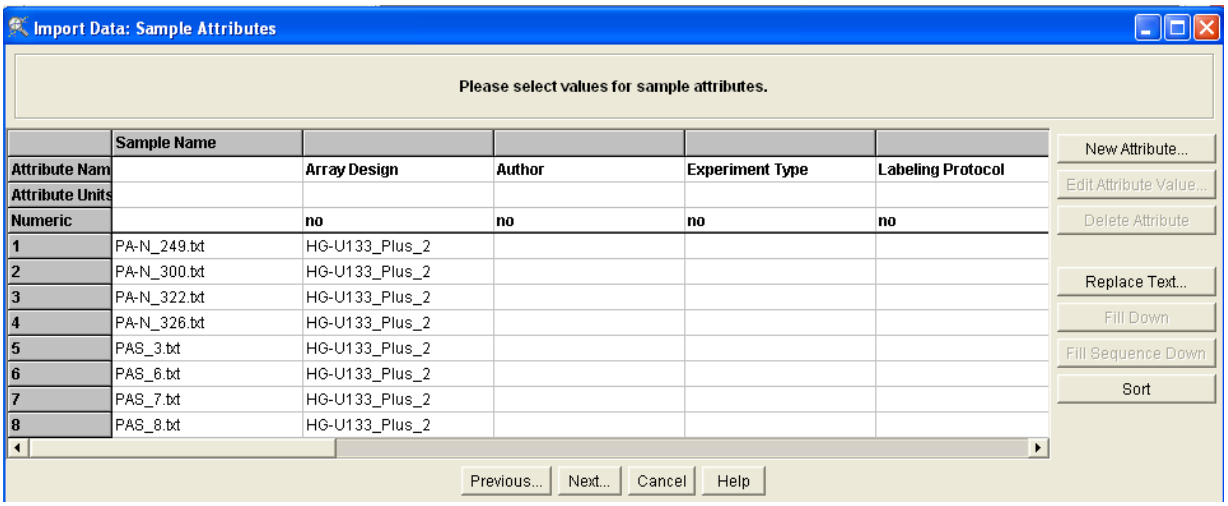
This window allows you to add more files of the same type to add to your experiment during the import process.





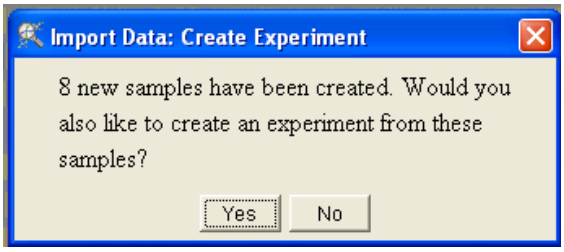
Step 5 : Sample Attributes window

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.

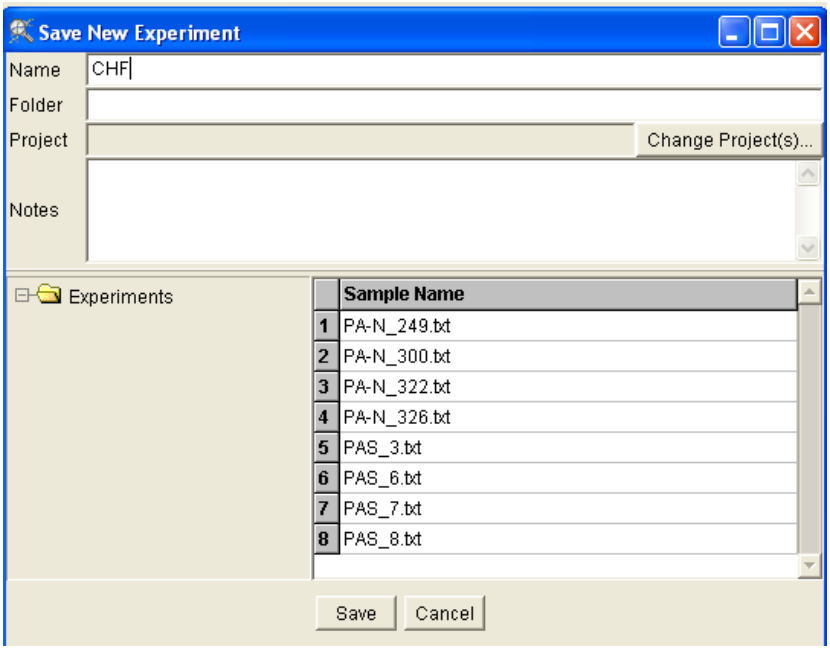


Step 6 : Experiment Creation

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.

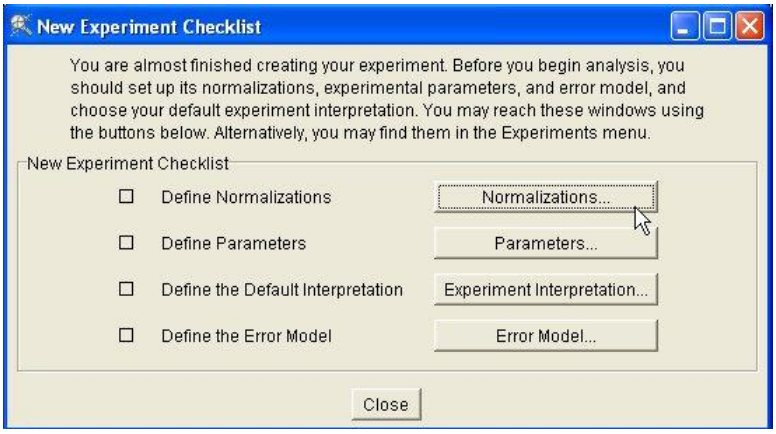


Provide an appropriate name for the New Experiment



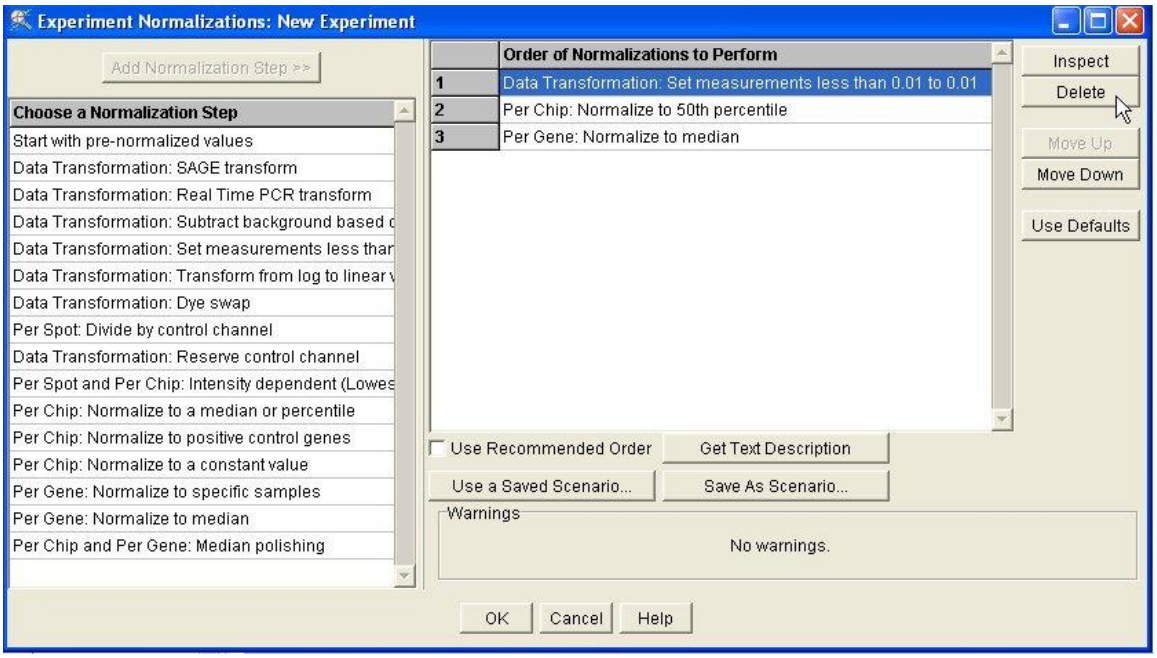
Step 7 : New Experiment Checklist

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

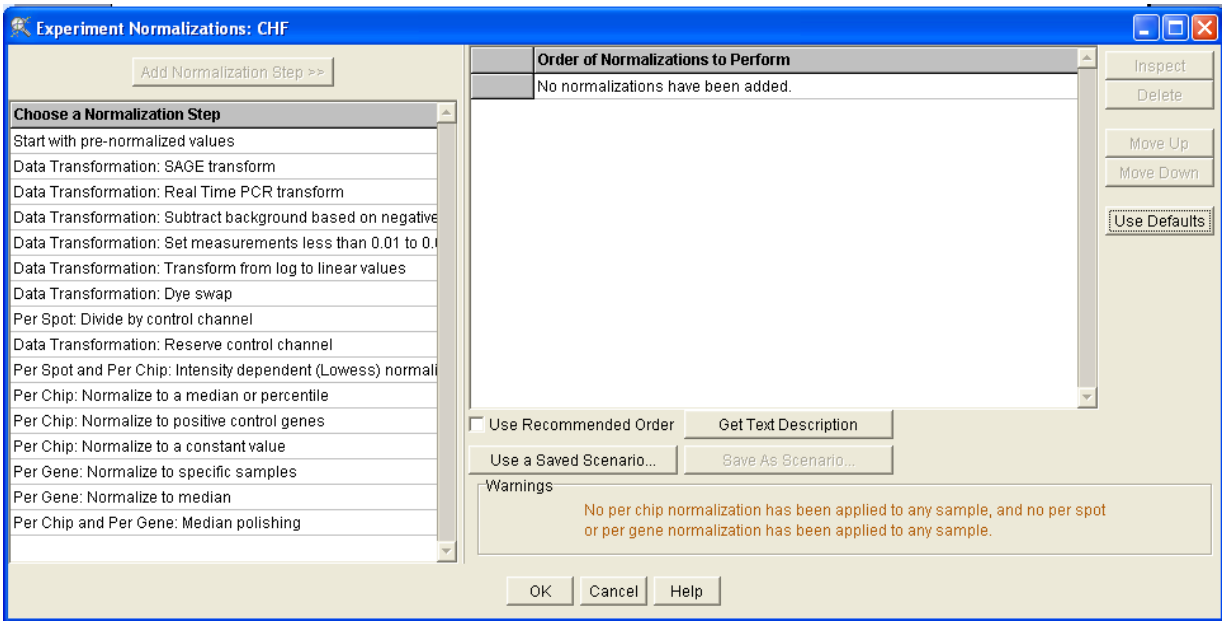


Step 8 : Experiment Normalizations

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. Applying 'Per Gene' normalization is optional.

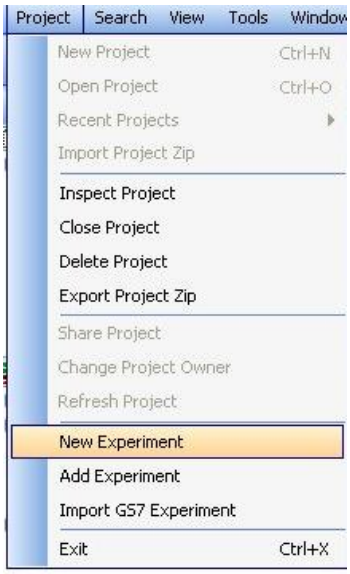


GeneSpring GX 9.0

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

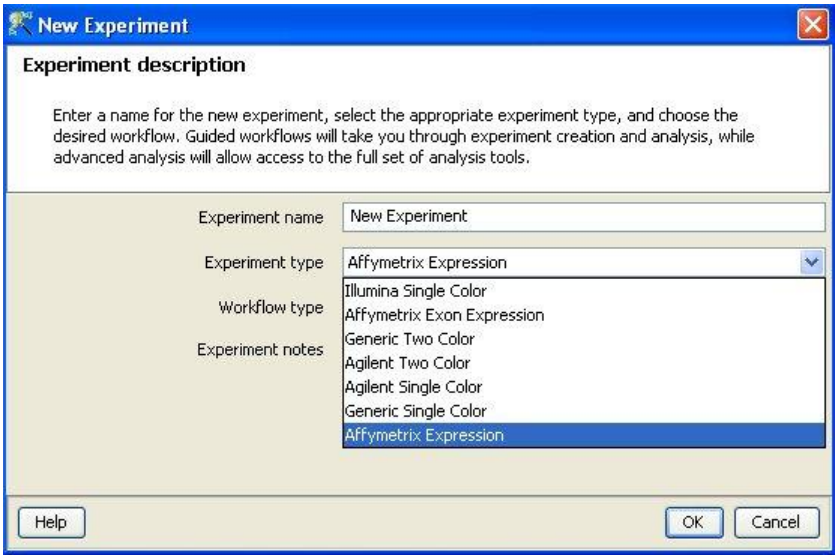
Step 1 : Create New Experiment

Create a new experiment using Project > New Experiment



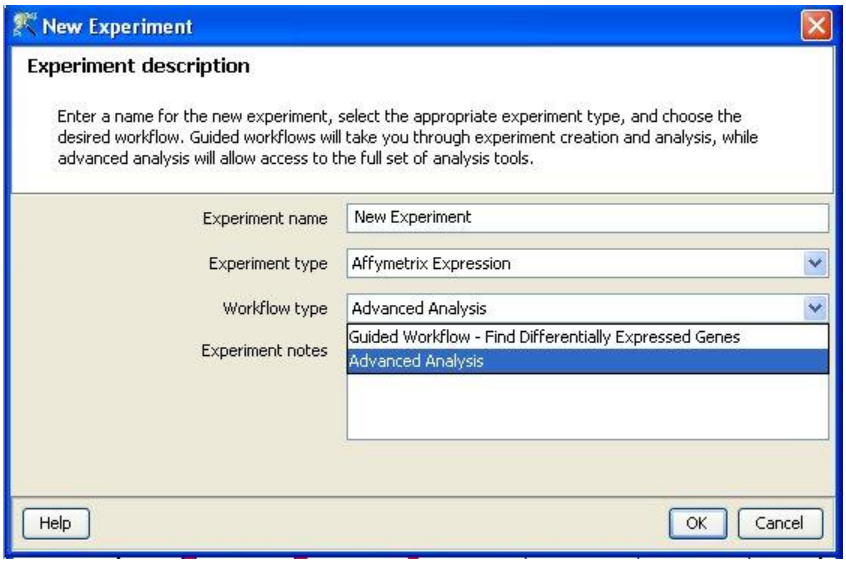
Step 2 : Experiment Description

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



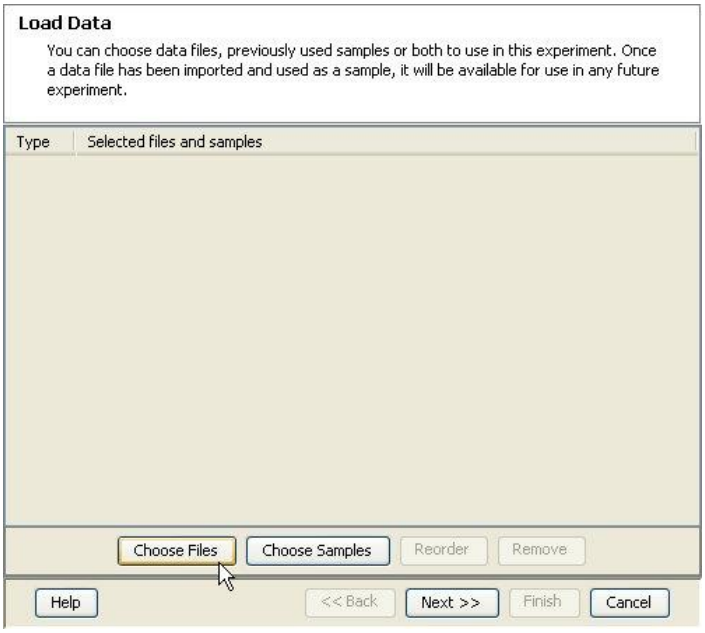
You can also define the Workflow type – *'Guided Workflow'* or *'Advanced Analysis'*

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.

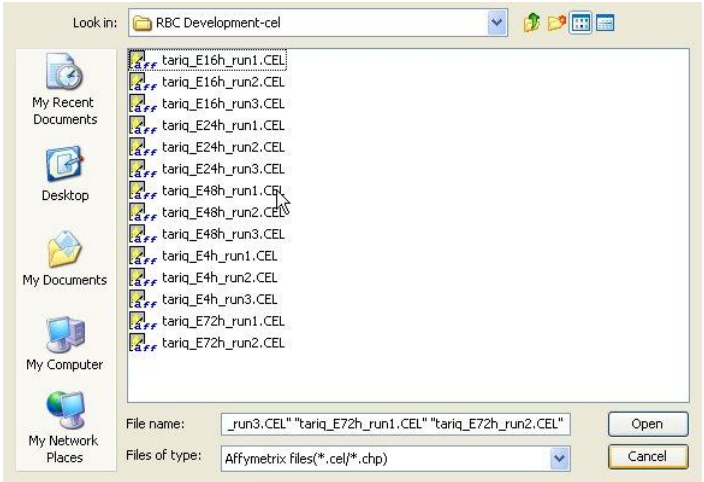


Step 3 : Load Data

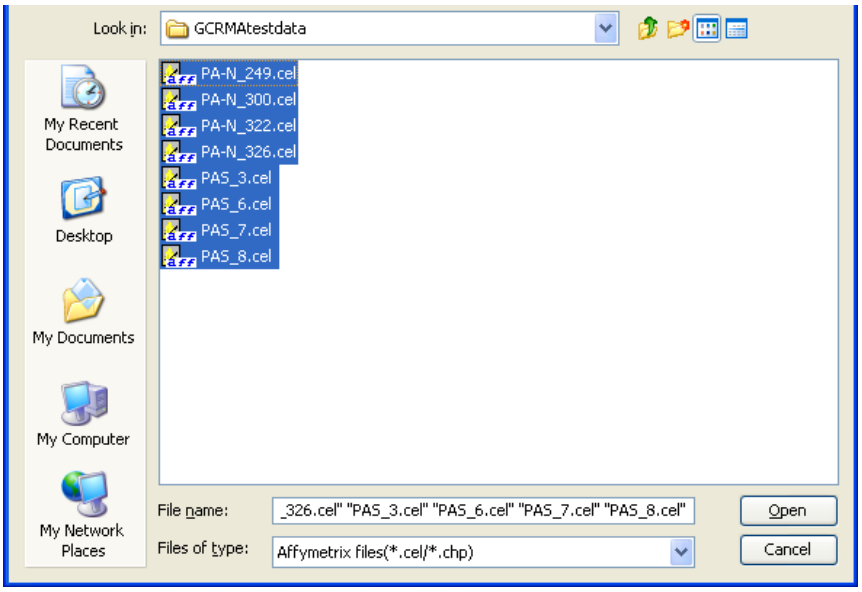
An experiment can be created using either the data files or else using samples. Upon loading data files, GeneSpring GX associates the files with the technology (see below) and creates samples. These samples are stored in the system and can be used to create another experiment via the *Choose Samples*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

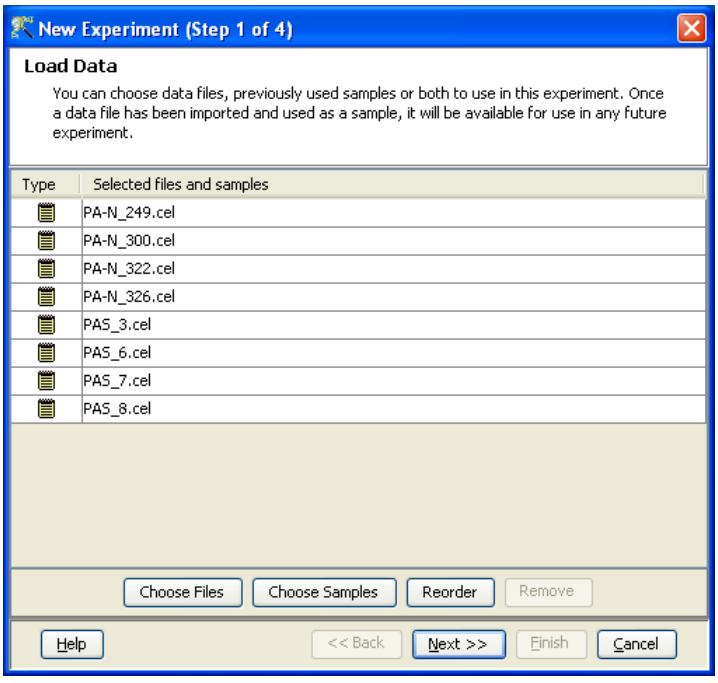


Navigate to the appropriate folder



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

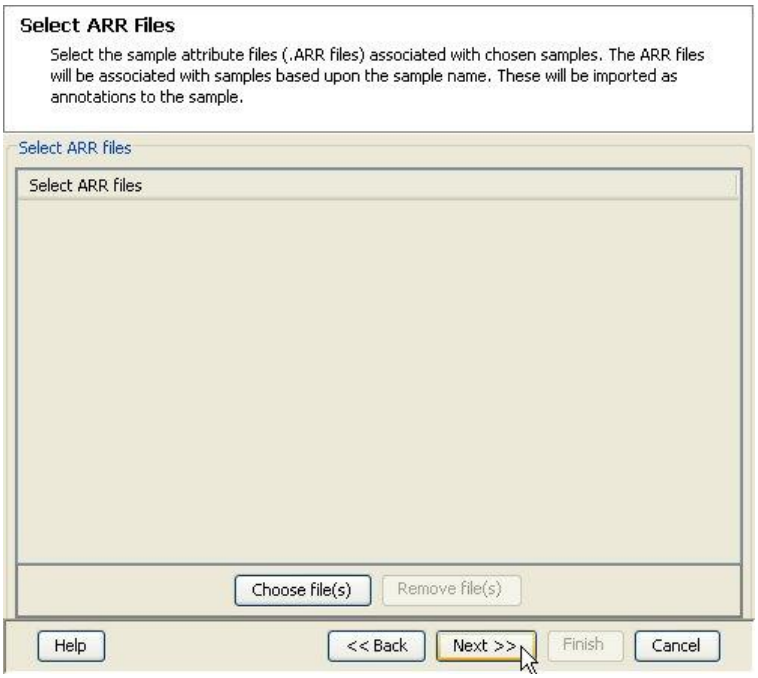




There are two things to be noted here. Upon creating an experiment of a specific chip type for the first time, the tool asks to download the technology from the GeneSpring GX update server. If an experiment has been created previously with the same technology, GeneSpring GX then directly proceeds with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



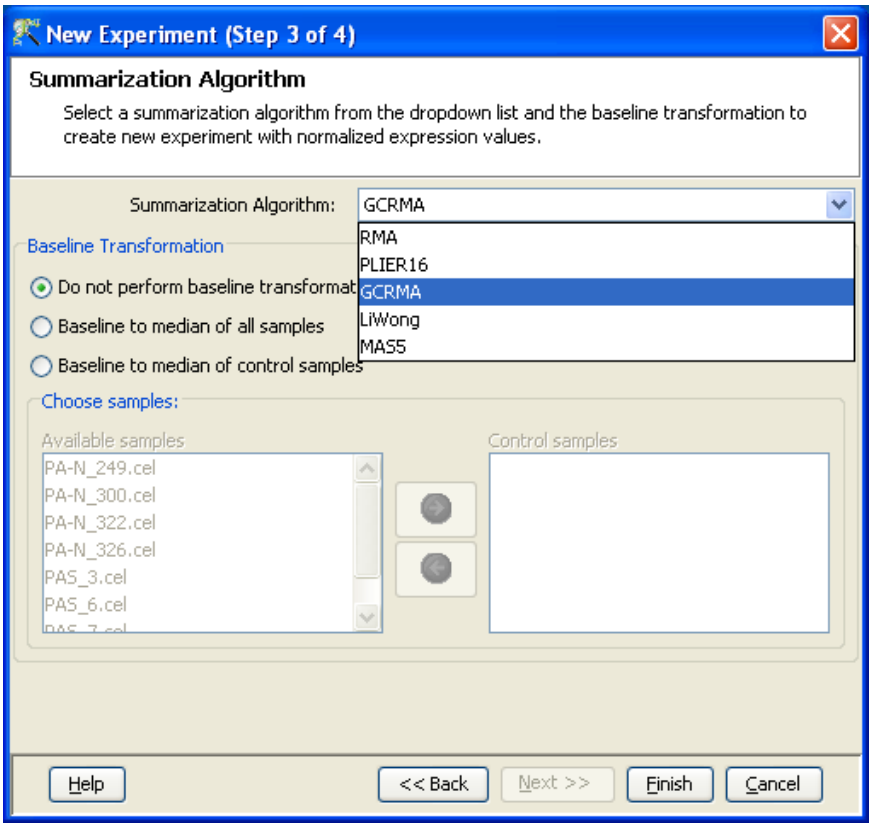
Step 5 : Select Probe Summarization and Normalization options

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

As an optional step, 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.