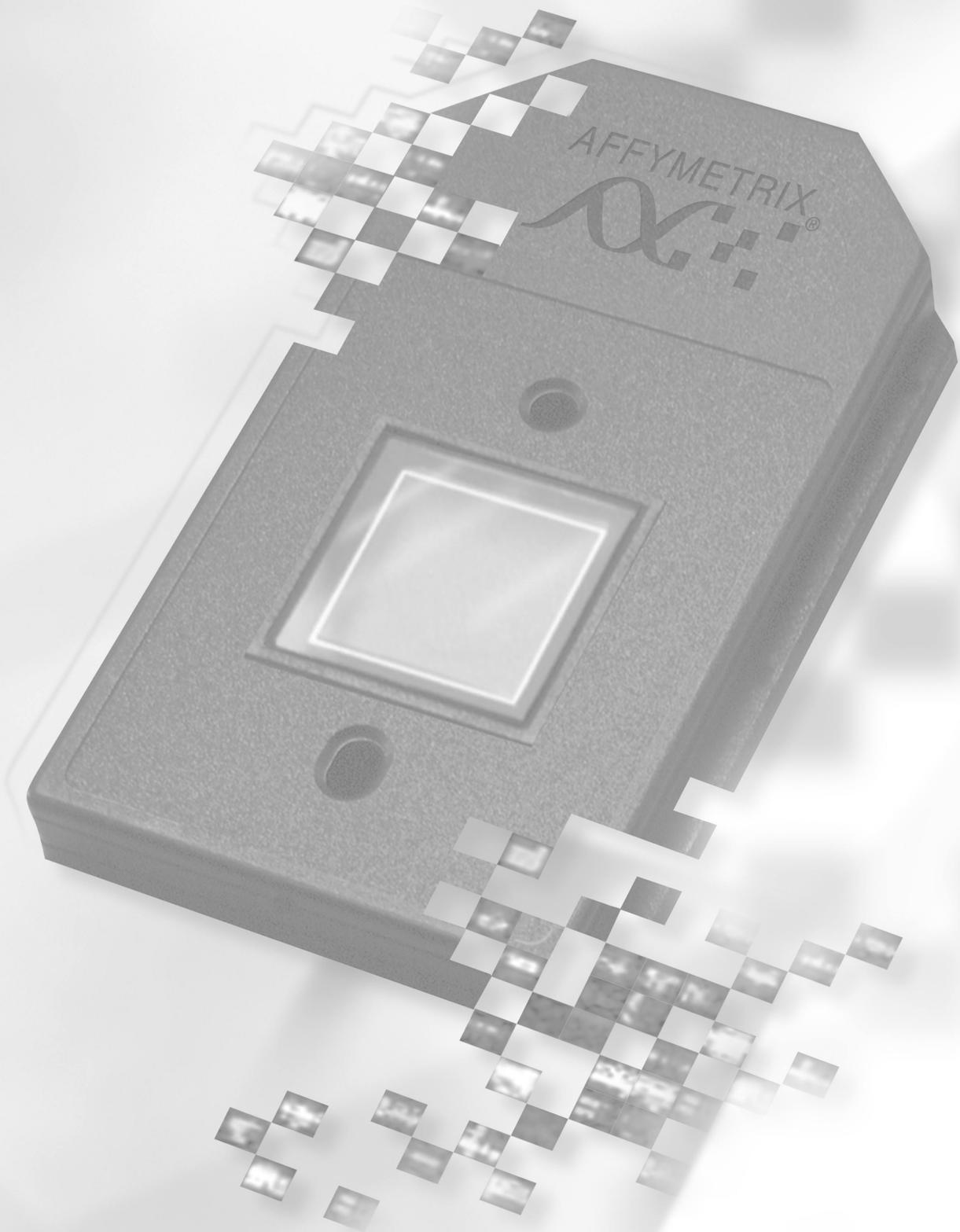


Section 5, Appendix B

Section 5, Appendix B





FAQs & Troubleshooting

FAQs

This section contains frequently asked questions related to GeneChip® expression analysis.

Sample Preparation

What is the minimum amount of total RNA I can use for each microarray experiment?

We currently recommend 5 ug of total RNA for our standard eukaryotic expression arrays. Reducing the amount of starting material used in the standard assay may result in a subsequent decrease in sensitivity. Additionally, a *Technical Note* is available on www.affymetrix.com describing a research small-sample target-labeling protocol and results from Affymetrix using reduced starting material. This protocol has not been fully validated but may be a helpful starting point for customers with limited total RNA.

What is the least amount of labeled eukaryotic cRNA target I can put on an array?

You should always use the recommended quantity of cRNA described in this manual. Please refer to **Table 2.3.1** for detailed instructions on the amount of cRNA needed for different array formats. Although there is a tolerance for some variation in quantity, we have found that hybridization of significantly less cRNA results in reduced sensitivity, particularly for low-copy transcripts.

How long can I store my eukaryotic cRNA target after its first hybridization?

Assuming no RNase contamination, cRNA targets can be stored for at least one year at -80°C without significant loss of signal intensity. The fact that the cRNA is fragmented prior to hybridization reduces the risk of subsequent degradation.

What parameters should I use to QC my GeneChip probe array data?

Quality assessments are critical in obtaining highly reproducible GeneChip probe array results. QC procedures should be performed at various key checkpoints:

- 1.** RNA sample quality: As described in this manual, the quality of starting RNA is very important. Ratio of 260/280 absorbance values, as well as appearance of samples by gel electrophoresis, are suggest methods to detect any degradation of your RNA samples.
- 2.** Target labeling: Various QC protocols described in this manual can be employed at different stopping points of the assay. For example, gel electrophoresis after cDNA synthesis (if using poly-A mRNA as starting material), after cRNA synthesis, and after fragmentation is helpful in estimating quantity and size distribution. Spectrophotometric measurements are also important after cRNA synthesis. Low cRNA yield can be a sensitive indicator of problematic labeling procedures and/or

starting material. You may also want to experiment with using real-time PCR analysis on house-keeping genes after each of these reactions to monitor the efficiency of each step.

3. GeneChip array image and basic data analysis. Routine QC parameters to monitor include visual array inspection, background, scaling factor, noise, 3'/5' GAPDH and Actin ratios, and % Present calls.

Can I hybridize samples to an array from a species other than the organism for which the array was designed?

Affymetrix has not validated the use of GeneChip expression arrays with alternate species. Although there may be high homology between different species, the sequence differences may be sufficient to interfere with hybridization, and more importantly, data interpretation. However, some customers have explored this approach. The following publication is an example of this type of study. Please note that this reference is listed for the convenience of our customers and is not endorsed or supported by Affymetrix.

Kayo, T., Allison, D.B., Weindruch, R., Prolla, T.A. Influences of aging and caloric restriction on the transcriptional profile of skeletal muscles from rhesus monkeys. *Proceedings of the National Academy of Sciences of the USA* **98**:5093-5098 (2001).

When I follow your recommended protocol of isolating total RNA from mammalian tissues, first using Trizol reagents, then with RNeasy columns, I sometimes see a reduced recovery off the RNeasy columns.

Trizol reagents and RNeasy columns are based on very different principles for nucleic acids purification. RNeasy columns exclude certain contaminants that may give rise to a falsely higher spectrophotometric reading, including carried-over phenol and transcripts shorter than 200 nucleotides in length. These shorter transcripts include the 5S rRNA and tRNA molecules that may account for 10% or more of the total RNA isolated.

To verify that the RNA of interest has been cleaned up efficiently during column purification, it may be helpful to run aliquots of your samples on a gel or perform some gene-specific real-time PCR quantitation. In addition, you can estimate how much total RNA you anticipate to recover since the yield is highly dependent on tissue type. These reference numbers can be obtained through your own experience or can be found in published literature, for example, the RNeasy Mini HandBook (www.qiagen.com/literature/handbooks/rna/rnamini/1016272HBRNY_062001WW.pdf).

If you continue to observe significant loss of material on RNeasy columns, please contact QIAGEN Technical Support directly.

Does the GeneChip Sample Cleanup Module generate comparable results relative to the previously recommended phenol/chloroform extraction for cDNA purification?

Highly concordant results have been obtained during our product development process by comparing global array hybridization results obtained from samples cleaned up with both protocols. The concordance was determined based on the overall signal intensity, as well as the qualitative calls. However, due to the different mechanisms associated with each cleanup procedure, there will be minor differences in the data obtained. For example, cDNA cleanup column reduces the recovery of fragments of 100 nucleotides or less, whereas these fragments are retained in the phenol/chloroform method. However, we do believe these differences are minor in magnitude. Customers are encouraged to perform their own comparisons and analysis to determine when to adopt the Sample Cleanup Module into their laboratories.

Hybridization, Washing, and Staining**What happens if the hybridization time is extended beyond 16 hours?**

The standard gene expression hybridization time is 14-16 hours at 45°C. At high temperatures and longer incubation times the sample will evaporate. Loss of sample is undesirable for several reasons:

1. Low volume of hybridization solution in the probe array can lead to dry spots that will show up as uneven hybridization and thus, compromise data.
2. Sample loss compromises the possibility of repeating the experiment with the identical sample.
3. Sample evaporation can lead to changes in the salt concentration of the solution which can affect the stringency conditions for hybridization.

How long can I keep my arrays in non-stringent wash A buffer before scanning?

The arrays may be stored in the dark for up to 16 hours, at 4° C (or 4 hours if stored at room temperature) prior to scanning with no noticeable loss of signal intensity. To avoid condensation while scanning, equilibrate the arrays to room temperature prior to the scan.

How many times can I scan an array before the data is affected?

It is always best to capture the data on the initial scan. Scanning bleaches the fluorophore and will result in reduction in signal intensity of 10-20% with each scan. Therefore, subsequent scans will not give signals equivalent to the initial scan.

How often do I need to do maintenance on the fluidics station?

With normal use (e.g., 20 arrays/module/week), we recommend the following schedule: Every week, the needle bleaching protocol (i.e., “Bleach” fluidics protocol) should be performed; on a monthly basis, the full-fluidics bleaching protocol (i.e., “Monthly Decontamination” protocol) should be performed and the peristaltic-pump tubing replaced. Please refer to Section 4, *Fluidics Station Maintenance Procedures*, for more detail.

What fluidics script do I use?

The appropriate fluidics script is specific to the array format (standard, midi, mini, or micro) and the organism (eukaryotic or prokaryotic). Information on the array format and appropriate script is contained in the package insert that comes with each array package. Please refer to the hybridization protocols in the respective sections of this manual for more detail.

Is there a possibility of contaminating the fluidics station with RNase when gene expression, genotyping, and health management applications are being performed on a shared station?

It is extremely important to change the vials each time a sample is removed or loaded onto a probe array. This prevents cross-contamination as well as sample loss. RNase contamination is not an issue with gene expression applications due to the fact that the cRNA sample is fragmented prior to hybridization and is removed prior to array processing on the fluidics station.

I have a bubble in the array. How do I get rid of it?

After the final wash on the fluidics station, if the door is still open, place the array in the probe array holder and close the door. The fluidics module will automatically run a drain and fill protocol with buffer A. If one cycle does not remove the bubble, repeat the process and try again. If this doesn't work or the door has already been closed, manually drain the array and refill with buffer A.

What are the safe stopping points in the assay?

It is safe to stop work after each of the major steps in the sample preparation process: first strand cDNA synthesis, second strand synthesis, IVT, fragmentation, or after preparing the hybridization cocktail. If possible, work with extracted RNA samples immediately rather than freezing them. Although it is common practice to use stored, frozen RNA samples in the process, eliminating a freeze-thaw will most likely yield higher-quality cRNA.

Data Analysis

I have observed on occasion that multiple _at probe sets are mapped to the same gene but give different expression results. How do I reconcile the difference?

There are various reasons why this happens. With increasing knowledge of the genome, the unique probe sets (_at probe sets) that were initially designed may turn out to represent subclusters that have collapsed into a single cluster in a later design. Therefore, it may seem that multiple “unique” _at probe sets now correspond to a single gene.

Different results from the probe sets could be observed due to the following reasons:

1. They represent splice variants or may cross-hybridize to different members that belong to a highly similar gene family or transcripts with different poly-A sites
2. One probe set is more 5' than the other
3. One probe set is better designed than the other

In these cases, it is important to use the resources available on the NetAffx™ Analysis Center (www.affymetrix.com) to understand if any of the above scenarios apply. Other expression analysis techniques may also be used to confirm which probe set reflects the transcript level more accurately.

What 3'/5' ratio for control genes, for example GAPDH and Actin, should I anticipate to obtain on GeneChip probe arrays?

In addition to the conventional probe sets designed to be within the most 3' 600 bp of a transcript, additional probe sets in the 5' region and middle portion (M) of the transcript have also been selected for certain housekeeping genes, including GAPDH and Actin. Signal intensity ratio of the 3' probe set over the 5' probe set is often referred to as the 3'/5' ratio. This ratio gives an indication of the integrity of your starting RNA, efficiency of first strand cDNA synthesis, and/or *in vitro* transcription of cRNA. The signal of each probe set reflects the sequence of the probes and their hybridization properties. A 1:1 molar ratio of the 3' to 5' transcript regions will not necessarily give a signal ratio of 1.

There is no single threshold cutoff to assess sample quality for all of the diverse organisms and tissues. This is due to the presence of different isoforms of these house-keeping genes and their different expression patterns in various tissues and organisms. Although we routinely refer to a threshold ratio of less than 3 for the most common tissues, such as mammalian liver and brain, this may not be applicable to all situations. It may be more appropriate to document the 3'/5' ratios within a particular study and flag the results that deviate, therefore representing an unusual sample that deserves further investigation.

Can results from different laboratories and different times be compared with each other directly and how do you control the variables in this type of experiment?

Array results can potentially be compared directly. However, it is important to check the following important elements before doing so:

1. Experimental design strategy should be the same at various sites.
2. Identical target labeling protocols should be followed, and yields from cDNA and IVT reactions should be within the same range as specified for that study.
3. Scanners are adjusted to the same PMT setting.
4. Same algorithm parameters are used.
5. Similar results from 3'/5' ratios, background, noise, and scaling factors. Check arrays for scratches and even hybridization/staining.
6. Comparability of results obtained from different operators should be evaluated before including their results in the same study.

Affymetrix Microarray Suite (MAS) is on the C: drive which is low on space. How can I create more room on the hard drive?

The library and data files can be moved to another drive, then deleted from the C: drive. After moving the files, remember to change your library file default settings in MAS to the appropriate directory by clicking on the **Tools** tab and then select **Defaults** in the drop-down menu, then **File locations** tab in the **Defaults** window.

What is the difference between scaling and normalization when I scale or normalize my data to all genes on the array?

With scaling, you select an arbitrary target intensity and scale the average intensity of all genes (minus the highest 2% and lowest 2% Signal values) on each array within a data set to that number. This enables you to compare multiple arrays within a data set. The scaling factor remains the same for a particular array as long as you use the same arbitrary target intensity for scaling. Scaling can be performed independent of the comparison analysis.

On the other hand, normalization can only be done when performing a comparison analysis. It compares an experimental array with a baseline array and normalizes the average intensity of all genes (minus the highest 2% and lowest 2% Signal values) of the experimental array to the corresponding average intensity of the baseline array when running a comparison analysis in MAS. The normalization factor for a particular array changes when you change the comparison baseline array.

How important is it to evaluate the value of the Scaling Factor between different arrays?

Scaling Factor is the multiplication factor applied to each Signal value on an array. A Scaling Factor of 1.0 indicates that the average array intensity is equal to the Target Intensity. Scaling Factors will vary across different samples and there are no set guidelines for any particular sample type. However, if they differ by too much within a set of experiments, approximately 3-fold or more, this indicates wide variation in the .dat files. Therefore, the analyzed data (in the .chp file) should be treated with caution.

Should I always anticipate the hybridization controls, *bioB*, *bioC*, *bioD*, and *cre*, to be called as Present?

The four transcripts are added to the hybridization cocktail at staggered concentrations. At 1.5 pM, *bioB* is at the detection limit for most expression arrays and is anticipated to be called Present at least 70% of the time. In contrast, the other controls should be called Present all of the time, with increasing Signal values (*bioC*, *bioD*, and *cre*, respectively). Absent calls, or relatively low Signal values, indicate a potential problem with the hybridization reaction or subsequent washing and staining steps. Check to see if the hybridization cocktail was prepared correctly, if the recommended hybridization temperature and Fluidics Protocol were used, and make sure the SAPE staining solution did not deteriorate.

Other than qualitative calls and Signal values, the 3'/5' ratio data for these controls are not as informative since they do not relate to the quality of the samples and data.

What does high background mean?

A high background implies that impurities, such as cell debris and salts, are binding to the probe array in a nonspecific manner and that these substances are fluorescing at 570 nm (the scanning wavelength). This nonspecific binding causes a low signal to noise ratio (SNR), meaning that genes for transcripts present at very low levels in the sample may incorrectly be called Absent. High background creates an overall loss of sensitivity in the experiment.

What are masks?

Masks are rarely used features in MAS. There are three types of mask files:

Image mask files: You may want to use an image mask if there is a large visible aberration on an image. You define the image mask based on the physical location of the image. Probe pairs included in the mask are excluded from the analysis. Image masks are associated with a given .dat/.cel file and cannot be used on other images.

Probe mask files: Probe masks are defined by the probe set and probe pair number. Probe pairs included in this type of probe mask are excluded from the analysis when the probe mask is used. Probe masks can be applied across a data set. For a detailed description, please refer to *Affymetrix Microarray Suite User's Guide* (P/N 701099).

A second type of probe mask defines a select group of probe sets that can be used in normalization or scaling. Please refer to *Affymetrix Microarray Suite User's Guide* where this type of probe set mask file is described.

If I realign the grid, how do I create a new .cel file?

If manual adjustment of the grid is necessary, the corresponding .cel file present at the time of adjustment will no longer be a valid representation of the realigned image data. Microarray Suite automatically detects this situation either on initial reopening of the readjusted .dat file or during the analysis process. Once the readjusted .dat file is opened, the .cel file is automatically created. The user does not need to carry out any overt steps to accomplish this.

How do I add additional probe sets in the .rpt file?

Use the **Report Settings** dialog on the short cut menu in Microarray Suite to open the **Expression Report**. You may add any probe sets desired by simply typing in the probe set name(s) you wish to add (this can also be accomplished by cutting and pasting from a text file). Keep in mind that the probe set name must be entered exactly as it appears in the analysis file, including the suffixes such as “12345_s_at”.

Why can't I analyze data files stored on a CD?

Files in CD-ROM format are copied to the hard drive in read-only mode. MAS requires that this attribute be removed. To do this, open NT Explorer and select the file(s) you copied from the CD. Click the right mouse button and select **Properties**. Clear the **Read-only** check box near the bottom of the **Properties** screen and click **OK**.

How can the mismatch probe cell have a higher intensity than its corresponding perfect match probe cell?

There could be a number of reasons for this. It is possible that this probe sequence has high homology with another unknown sequence resulting in a high mismatch-to-perfect match ratio. Another possibility is a mutation or set of mutations in the sequence of the target transcript which causes specific binding to the Mismatch. Regardless of the cause, the built-in redundancy using multiple probe pairs to represent a single sequence on the probe array mitigates any significant impact on the final interpretation of the data.

There are too many files showing in the file window in Microarray Suite. What can I do?

By placing files for projects in their own directories and changing the default settings for data in Microarray Suite appropriately, you can manage large numbers of files.

In addition, with the Windows NT operating system, users can specify their own directory defaults in Microarray Suite while logging on and create their own directories for data. To do so, each user should have a unique logon name and organize files in subdirectories, for example, by project, user, date, or lab. Each user can then set the data default to a subdirectory of choice.

Experimental Design**Which is greater, sample or assay variability?**

Sample variability, which arises mainly from biological heterogeneity, is certainly higher than assay variability, and has been estimated to be at least 10-fold greater. We recommend that researchers run multiple samples per data point to account for sample-to-sample variability. In addition, carefully design the experiment in order to minimize potential variation associated with the samples.

Troubleshooting

Problem	Likely Cause	Solution
Sample Quality		
High 3'/5' ratio	Most often caused by degradation of the RNA during the isolation process.	Start with a fresh sample and minimize the possibility of RNase activity. Look for the presence of Ribosomal RNA bands on a non-denaturing agarose gel.
Low cRNA yield	Low RNA quality, which interferes with reverse transcription and subsequent labeling.	It sometimes helps to do a Trizol-based isolation followed by cleanup with an RNeasy column. For samples with a high lipid content, such as brain, use procedures to reduce the lipid content prior to the reverse transcription reaction.
Enzo BioArray HighYield RNA Transcript Labeling Kit		
Apparent insufficient volume in reagent tubes	The reagent tubes are opened before centrifugation.	The small volume may be expelled by opening. The tubes should be centrifuged briefly before use to ensure that reagents remain at the bottom of the tube.
Precipitation in the reaction buffer	After many freeze-thaw cycles, a precipitate may form.	Centrifuge briefly to remove precipitate before use. The precipitate formation does not interfere with the reaction.
Low yield	The most likely cause of low yield in a transcription reaction is poor quality template. The presence of excess T7 promoter-containing primers can also decrease yield.	Carry over of phenol will inhibit the reaction. To remove phenol, wash the template twice with 70% or 80% ethanol. Following synthesis of the cDNA template the primers can be removed by precipitating the cDNA with 2.5M ammonium acetate and 2.5 volumes of absolute ethanol. The precipitate should be spun immediately at room temperature for 20 minutes. If other salts are used or if the sample is frozen the primers may also precipitate resulting in their incomplete removal. If interference by excess primers persists, the starting concentration of primers can be reduced. This is recommended when starting with reduced amounts of RNA. Some cDNA synthesis reactions may produce cDNA that has been primed with RNA instead of with the T7 promoter-containing oligo primer. This is more likely to occur when starting with total RNA. The RNA-primed cDNA contains no T7 promoter sequence and thus will not be transcribed.
Image / Array Quality		
Low or absent Oligo B2 hybridization	Addition of control Oligo B2 and hybridization, washing or staining.	Make sure that the Control Oligo B2 has been added to the hybridization cocktail at the correct concentration. Also, check the makeup of the hybridization buffer, the stain solution, and hybridization temperature.
Dim Corners	In need of fluidics maintenance.	Bleach the fluidics as recommended and change the peristaltic pump tubing. If the problem persists, call Affymetrix Technical Support.

Problem	Likely Cause	Solution
Image / Array Quality (continued)		
Dim Arrays	Hybridization problems.	Check the signal from control Oligo B2 to see if the signals are also weak. If it appears to be a hybridization issue, check all hybridization reagents and equipment settings before running another assay. Test arrays can be useful for troubleshooting this issue.
	Sample preparation problems.	Re-check each of the quality control procedures recommended in the manual, such as absorbance measurement and running an aliquot on gel, to ensure that there is no significant loss of sample during target preparation due to manipulation of the sample or RNase contamination. Also see above for "low cRNA yield"
Leaking septa	Leaking septa are most often created during the array filling with a pipette.	Be sure to use pipette tips without a beveled end. When filling the arrays, be careful to push the pipette tip straight through the septum and maintain a constant perpendicular angle during filling and draining of the array.
Software Problems		
In Microarray Suite (MAS), I received the error message, "Could not find the .cif file."	The default path for the library files in MAS is incorrect.	Set the correct path for the library files.
	The library files for those specific arrays are not installed on the computer.	Install the library files for that array, making sure to check the box appropriate for that array during the installation process.
The probe array type is missing from the pull-down menu when creating an .exp file.	The default path for the library files in MAS is incorrect.	Set the correct path for the library files.
	The library files for those specific arrays are not installed on the computer.	Install the library files for that array, making sure to check the box appropriate for that array during the installation process.
The fluidics protocols are missing from the pull-down menu in the Fluidics control window.	The default path for the protocol files in MAS is incorrect.	Check that the location of the fluidics files on the hard drive corresponds to the default protocol path in MAS.
	The library files are not installed on the computer.	Install the library files, making sure the protocols are in the same directory as the default path set in MAS.
After putting the computer on the network, the probe array descriptions are not available and a SQL error message appears.	When networking computers, the name of the computer is often changed to correspond to an organization's standard conventions. This results in a breakdown of the connection between MAS and the Microsoft Data Engine (MSDE).	After the computer is renamed, uninstall MAS and MSDE and reinstall MAS.
Microarray Suite is on the C: drive and it's filling up.		The library and protocol files can be moved (or dragged) to another, larger drive. Remember to change the default path for the library and protocol files in MAS, and modify this path for each log in name. In addition, GeneChip data should always be stored locally on the largest available drive on the workstation.
The gene descriptions show up for some users and not for others.	This is a result of different security settings between users and administrators of the workstation.	Call Affymetrix Technical Support for information on how to change the registry to correct this.