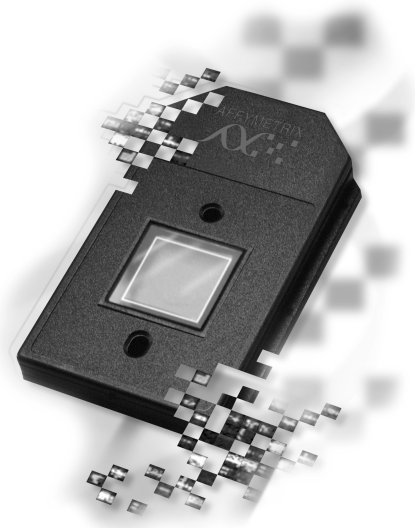
GENE EXPRESSION MONITORING

GeneChip® Expression Analysis



Technical Manual

For Research Use Only.
Not for use in diagnostic procedures.

701021 Rev. 3

AFFYMETRIX®

Trademarks

Affymetrix®, GeneChip®, EASI™,  ,  , , HuSNP™, GenFlex™, Jaguar™, MicroDB™, NetAffx™, CustomExpress™, 417™, 418™, 427™, 428™, Pin-and-Ring™ and Flying Objective™ are trademarks owned or used by Affymetrix, Inc.

GeneArray® is a registered U.S. trademark of Agilent Technologies, Inc.

Limited License

PROBE ARRAYS, INSTRUMENTS, SOFTWARE, AND REAGENTS ARE LICENSED FOR RESEARCH USE ONLY AND NOT FOR USE IN DIAGNOSTIC PROCEDURES. NO RIGHT TO MAKE, HAVE MADE, OFFER TO SELL, SELL, OR IMPORT OLIGONUCLEOTIDE PROBE ARRAYS OR ANY OTHER PRODUCT IN WHICH AFFYMETRIX HAS PATENT RIGHTS IS CONVEYED BY THE SALE OF PROBE ARRAYS, INSTRUMENTS, SOFTWARE, OR REAGENTS HEREUNDER. THIS LIMITED LICENSE PERMITS ONLY THE USE OF THE PARTICULAR PRODUCT(S) THAT THE USER HAS PURCHASED FROM AFFYMETRIX.

Patents

Products may be covered by one or more of the following patents and/or sold under license from Oxford Gene Technology: U.S. Patent Nos. 5,445,934; 5,744,305; 6,261,776; 6,291,183; 5,700,637; 5,945,334, and 6,346,413; and EP 619 321; 373 203 and other U.S. or foreign patents.

Software products may be covered by one or more of the following patents:
U.S. Patent Nos. 5,733,729; 5,795,716; 5,974,164; 6,066,454; 6,090,555; 6,185,561 and 6,188,783; and other U.S. or foreign patents.

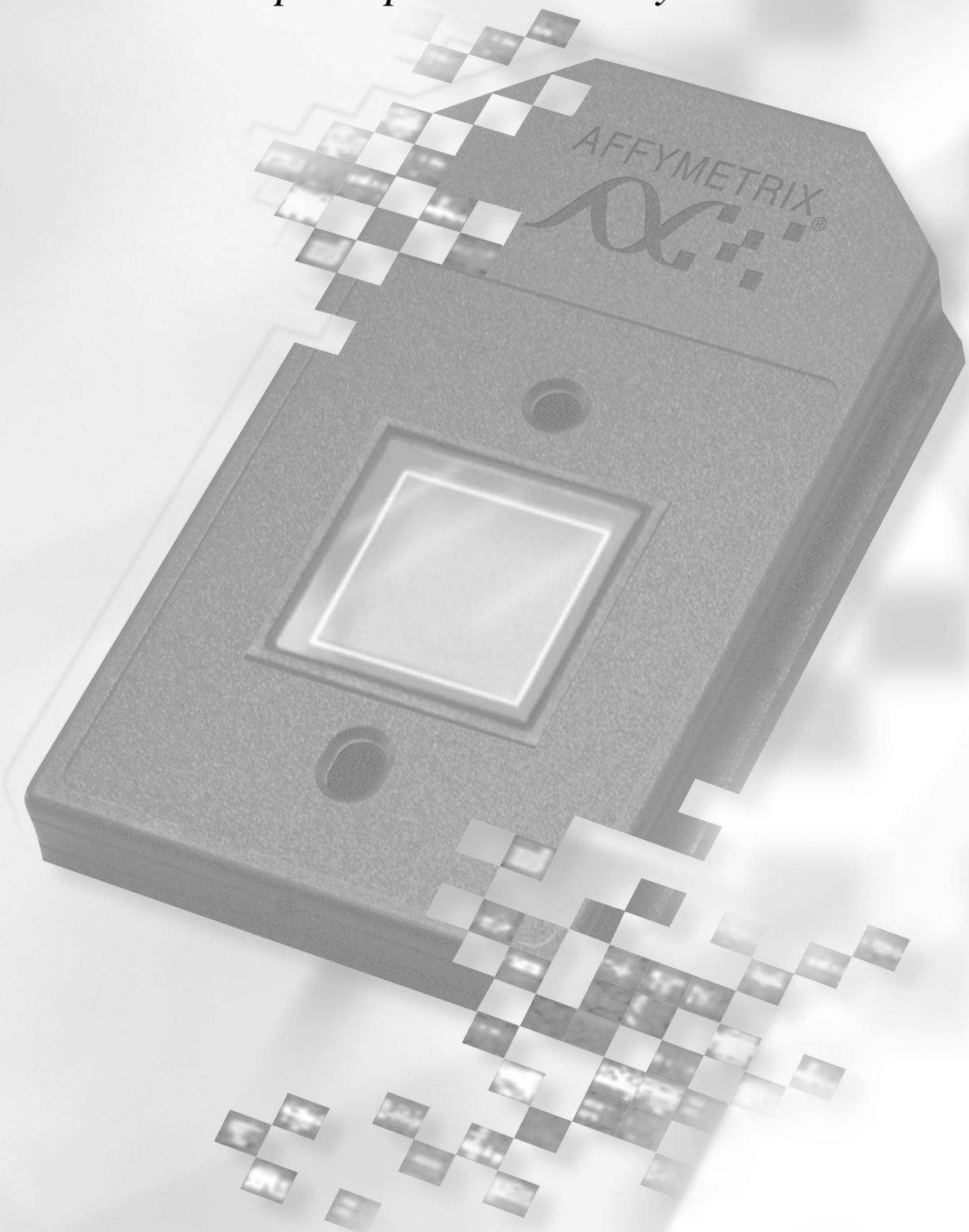
Copyright

Contents

SECTION 5 *Appendices*

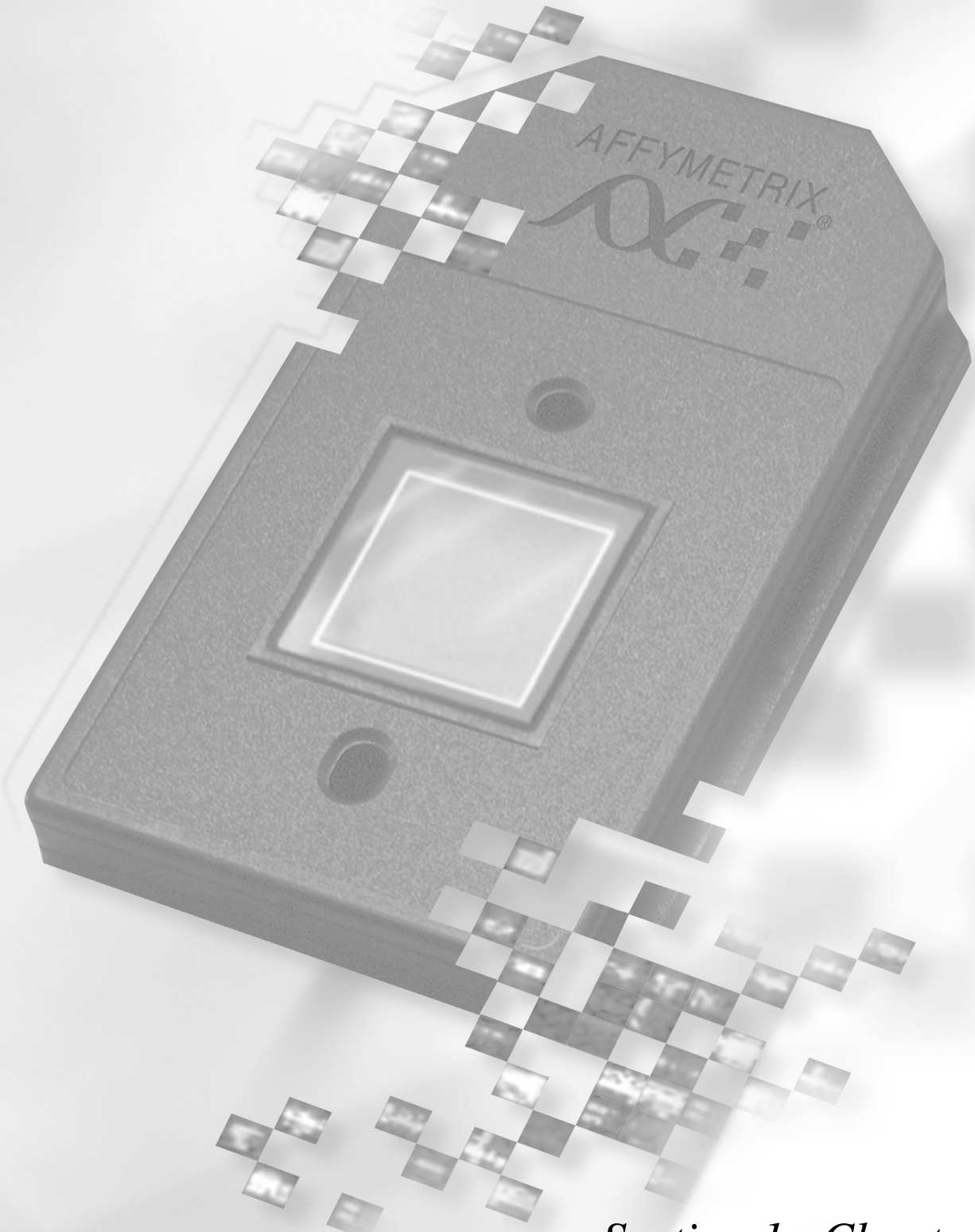
Section 1:

GeneChip® Expression Analysis Overview



Contents

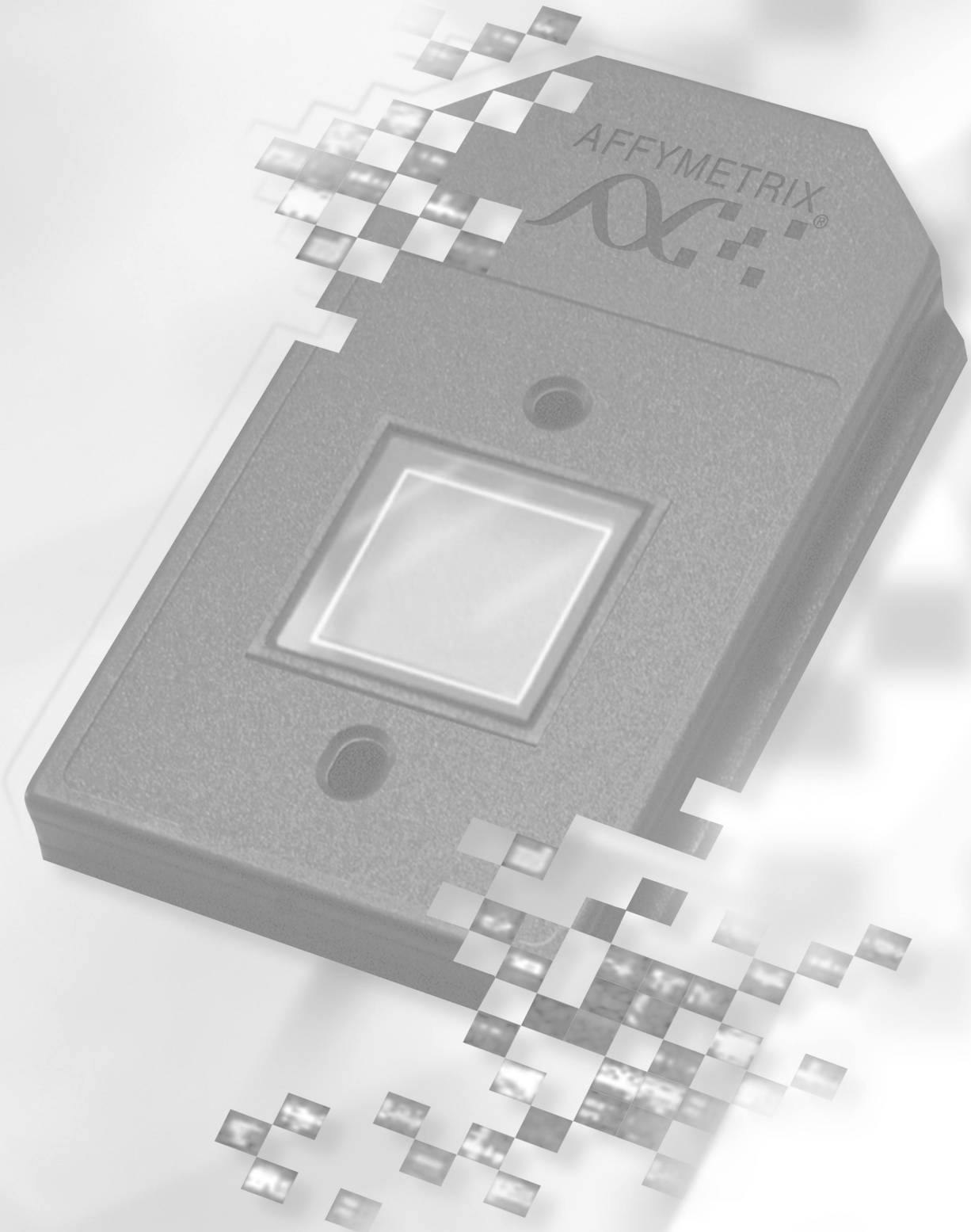
Section 1



Section 1, Chapter 1

701023 Rev. 2

Section 1, Chapter 1



GeneChip® Expression Analysis Overview

This Chapter Contains:

- An overview of GeneChip® Expression Analysis.
- A summary of the procedures covered in the remainder of the manual.

Introduction and Objectives

Welcome to the *Affymetrix® GeneChip® Expression Analysis Technical Manual*. This manual is a technical guide for using GeneChip expression analysis probe arrays. All protocols included in this manual have been used successfully by scientists at Affymetrix, or have been recommended by our collaborators during the development of particular products. The field of mRNA gene expression monitoring is rapidly evolving and periodic technical updates to this manual will reflect the newest protocols and information for using GeneChip probe arrays. This manual applies to all GeneChip expression products.

As an Affymetrix GeneChip user, your feedback is welcome. Please contact our technical support team with any input on how we can improve this resource.

Explanation of GeneChip® Probe Arrays

GeneChip probe arrays are manufactured using technology that combines photolithography and combinatorial chemistry.[1,2] Tens to hundreds of thousands of different oligonucleotide probes are synthesized on each array. Each oligonucleotide is located in a specific area on the array called a probe cell. Each probe cell contains millions of copies of a given oligonucleotide or probe.

Probe arrays are manufactured in a series of cycles. Initially, a glass substrate is coated with linkers containing photolabile protecting groups. Then, a mask is applied that exposes selected portions of the probe array to ultraviolet light. Illumination removes the photolabile protecting groups enabling selective nucleoside phosphoramidite addition only at the previously exposed sites. Next, a different mask is applied and the cycle of illumination and chemical coupling is performed again. By repeating this cycle, a specific set of oligonucleotide probes is synthesized with each probe type in a known location. The completed probe arrays are packaged into cartridges.

During the laboratory procedure described in this manual, biotin-labeled RNA fragments referred to as the “target” are hybridized to the probe array. The hybridized probe array is stained with streptavidin phycoerythrin conjugate and scanned by the GeneArray® Scanner at the excitation wavelength of 488 nm. The amount of light emitted at 570 nm is proportional to the bound target at each location on the probe array.

GeneChip® Expression Analysis Overview

The following major steps outline GeneChip Expression Analysis:

1. Target Preparation
2. Target Hybridization
3. Experiment and Fluidics Station Setup
4. Probe Array Washing and Staining
5. Probe Array Scan
6. Data Analysis

Due to the differences in the RNA species between eukaryotic and prokaryotic organisms, different target labeling protocols have been optimized for eukaryotic and *E. coli* samples. Sections 2 and 3 provide detailed protocols for target preparation, hybridization, array washing and staining for eukaryotic arrays and *E. coli* arrays, respectively. Please refer to the sections in this manual for detailed protocols appropriate for your arrays.

Step 1: Target Preparation

This manual describes procedures for preparing biotinylated target RNA from purified eukaryotic and prokaryotic RNA samples suitable for hybridization to GeneChip expression probe arrays. These procedures are recommendations only. For more information on these procedures, please contact Affymetrix Technical Support at 1-888-DNA-CHIP, or +44 (0)1628 552550 in Europe.

For eukaryotic samples, using protocols in this manual Section 2, double-stranded cDNA is synthesized from total RNA or purified poly-A messenger RNA isolated from tissue or cells. An *in vitro* transcription (IVT) reaction is then done to produce biotin-labeled cRNA from the cDNA. The cRNA is fragmented before hybridization.

For *E. coli* samples, Section 3 describes a detailed protocol to isolate total RNA followed by enrichment of messenger RNA. After fragmentation, the RNA is end-modified and conjugated with biotin.

Step 2: Target Hybridization

A hybridization cocktail is prepared, including the fragmented target, probe array controls, BSA, and herring sperm DNA. It is then hybridized to the probe array during a 16-hour incubation. The hybridization process is described in the respective sections for the different probe array types.

Step 3: Experiment and Fluidics Station Setup

Specific experimental information is defined using Affymetrix Microarray Suite on a PC-compatible workstation with a Windows NT operating system. The probe array type, sample description, and comments are entered in Microarray Suite and saved with a unique experiment name. The fluidics station is then prepared for use by priming with the appropriate buffers. For more information on the fluidics station, refer to the *Fluidics Station 400 User's Guide*.

Step 4: Probe Array Washing and Staining

Immediately following hybridization, the probe array undergoes an automated washing and staining protocol on the fluidics station.

Step 5: Probe Array Scan

Once the probe array has been hybridized, washed, and stained, it is scanned. Each workstation running Affymetrix Microarray Suite can control one scanner. Each probe array is scanned twice, taking up to ten minutes, depending on the array format. The software calculates an average of the two images, defines the probe cells and computes an intensity for each cell. The double scan improves assay sensitivity and reduces background noise.

Each complete probe array image is stored in a separate data file identified by the experiment name and is saved with a data image file (.dat) extension.

Review the scanner user's manual for safety precautions and for more information on using the scanner.

Step 6: Data Analysis

Data is analyzed using the Microarray Suite Expression Analysis window. The .dat image is analyzed for probe intensities; results are reported in tabular and graphical formats. Information on data analysis is provided in the enclosed *GeneChip® Expression Analysis: Data Analysis Fundamentals* booklet (P/N 701190).

Precautions

1. FOR RESEARCH USE ONLY; NOT FOR USE IN DIAGNOSTIC PROCEDURES.
2. Avoid microbial contamination, which may cause erroneous results.

▲ WARNING *All biological specimens and materials with which they come into contact should be handled as if capable of transmitting infection and disposed of with proper precautions in accordance with federal, state, and local regulations. This includes adherence to the OSHA Bloodborne Pathogens Standard (29 CFR 1910.1030) for blood-derived and other samples governed by this act. Never pipet by mouth. Avoid specimen contact with skin and mucous membranes.*

3. Exercise standard precautions when obtaining, handling, and disposing of potentially carcinogenic reagents.
4. Exercise care to avoid cross contamination of samples during all steps of this procedure, as this may lead to erroneous results.
5. Use powder-free gloves whenever possible to minimize introduction of powder particles into sample or probe array cartridges.

Terminology

Probes	The oligonucleotides on the surface of the probe arrays are called probes because they probe, or interrogate, the sample.
Target	The target is the labeled nucleic acid that is being interrogated. It is hybridized to the probes on the array.
Probe Cell	Specific areas on the probe array that contain oligonucleotides of a specific sequence.

Interfering Conditions

! CAUTION *Wear powder-free gloves throughout procedure. Take steps to minimize the introduction of exogenous nucleases. Water used in the protocols below is molecular biology-grade (nuclease-free).*

Proper storage and handling of reagents and samples is essential for robust performance.

All laboratory equipment used to prepare the target during this procedure should be calibrated and carefully maintained to ensure accuracy, as incorrect measurement of reagents may affect the outcome of the procedure.

Instruments

The Affymetrix GeneChip Expression Analysis Technical Manual is designed for use in a system consisting of a Fluidics Station 400, a Hybridization Oven 640, and a GeneArray® Scanner.

References

1. Sambrook, J., Fritsch, E.F., Maniatis, T. *Molecular Cloning: A Laboratory Manual, v.1* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY p 21-52 (1989).
2. See www.affymetrix.com for current GeneChip technology references.

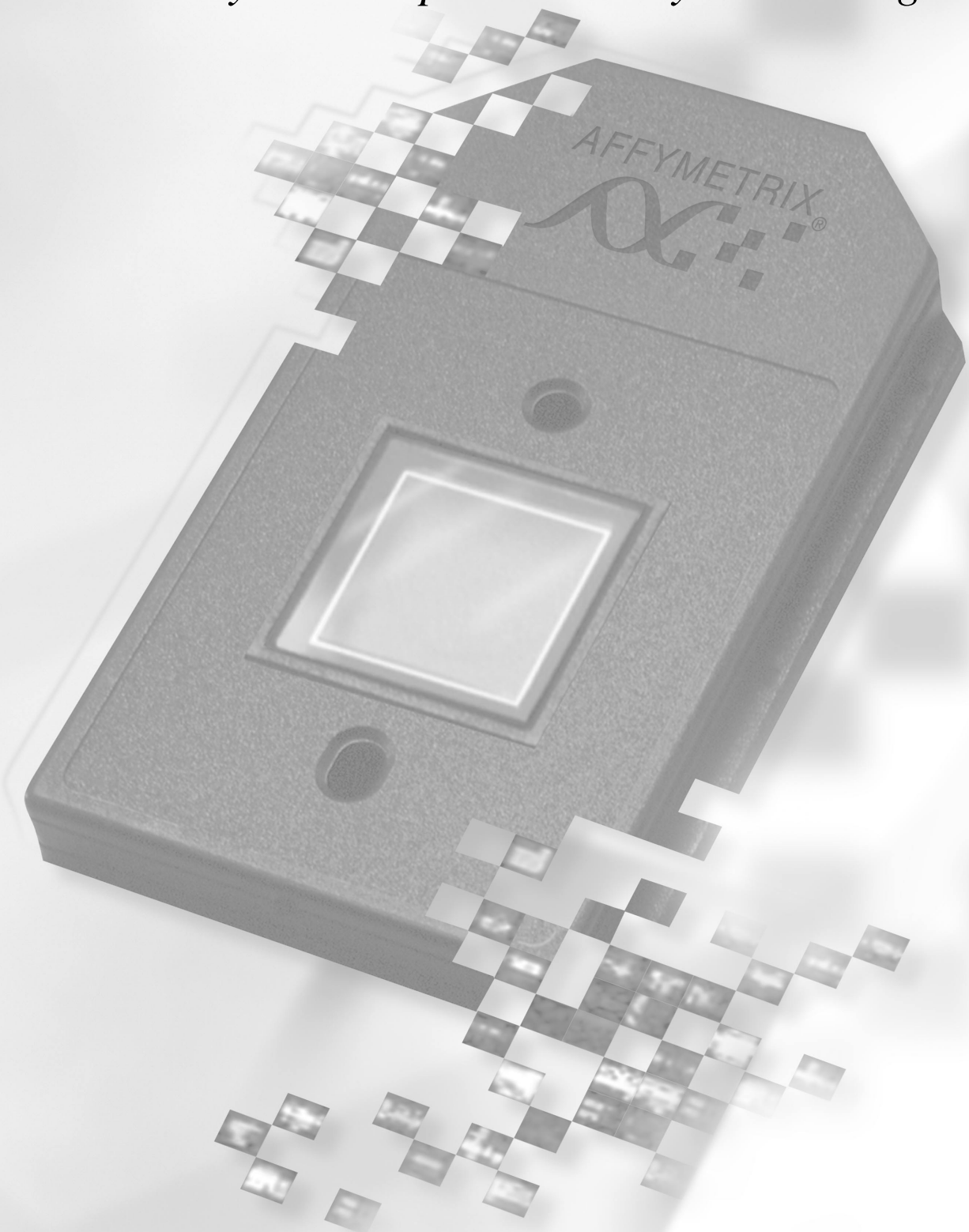
Limitations

- The results of the assay kit are dependent upon the quality of the input RNA, subsequent proper handling of nucleic acids and other reagents.
- The results should be evaluated by a qualified individual.

IMPORTANT *Do not store enzymes in a frost-free freezer.*

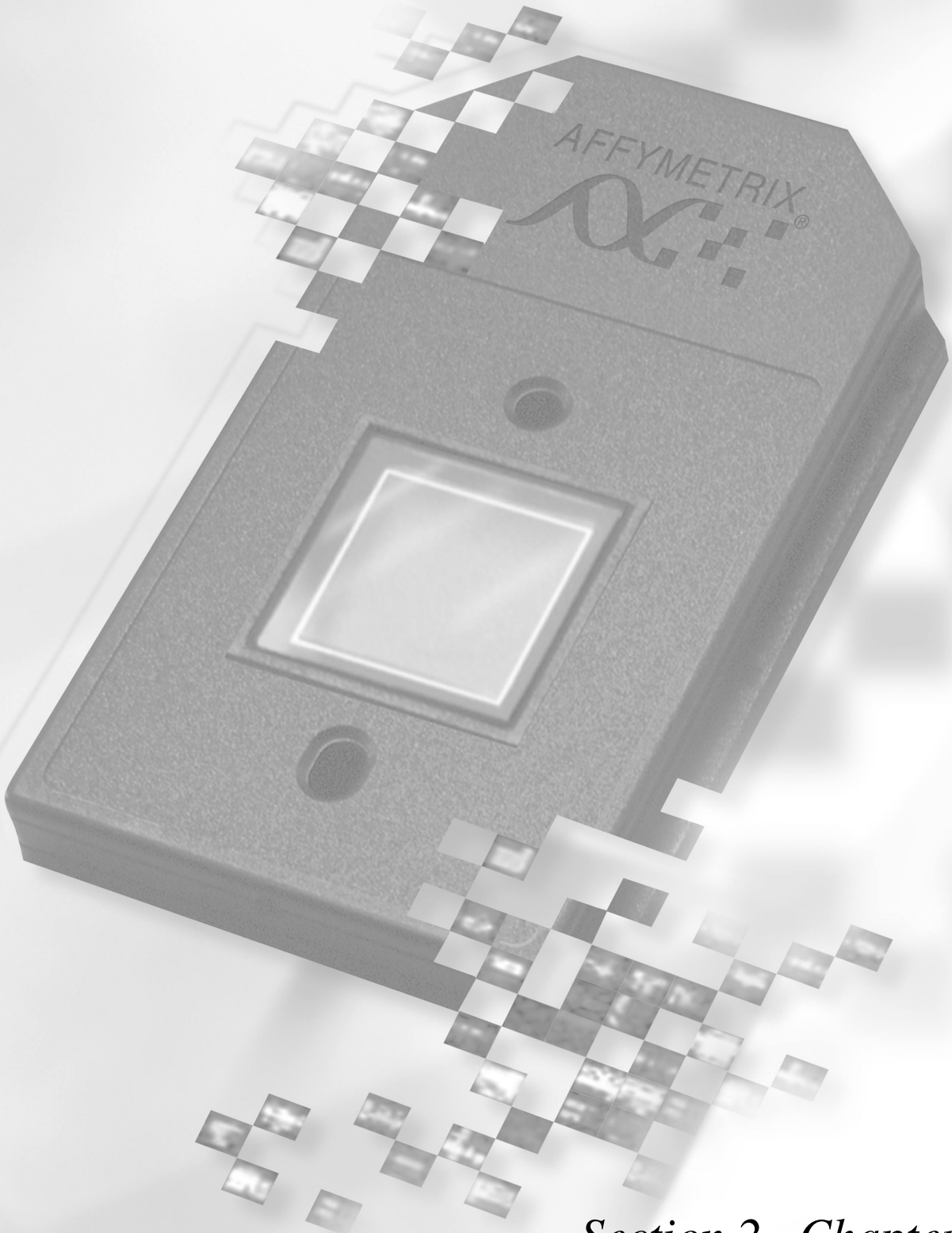
Section 2:

Eukaryotic Sample and Array Processing



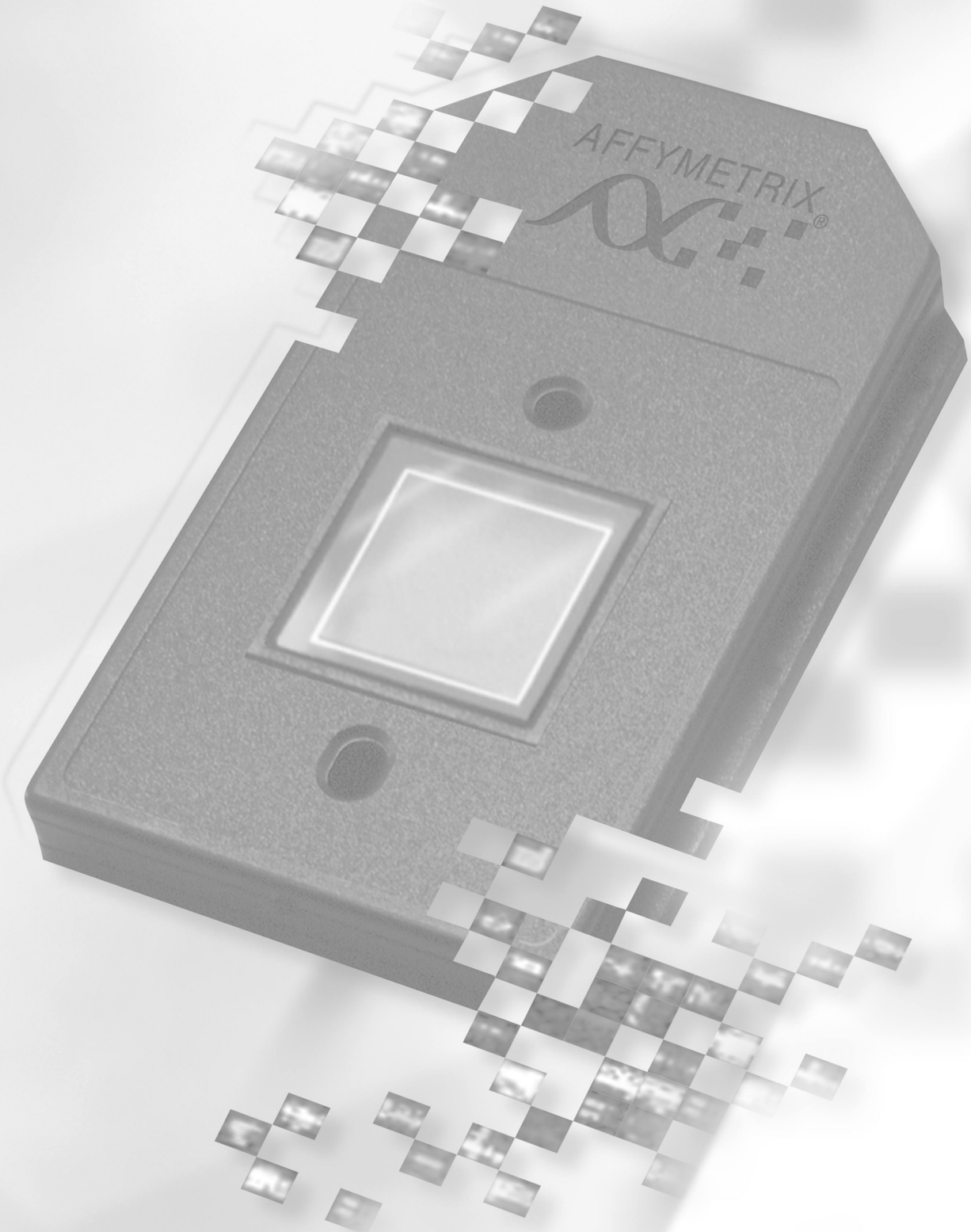
Contents

Section 2 *Eukaryotic Sample and Array Processing*



Section 2, Chapter 1

Section 2, Chapter 1



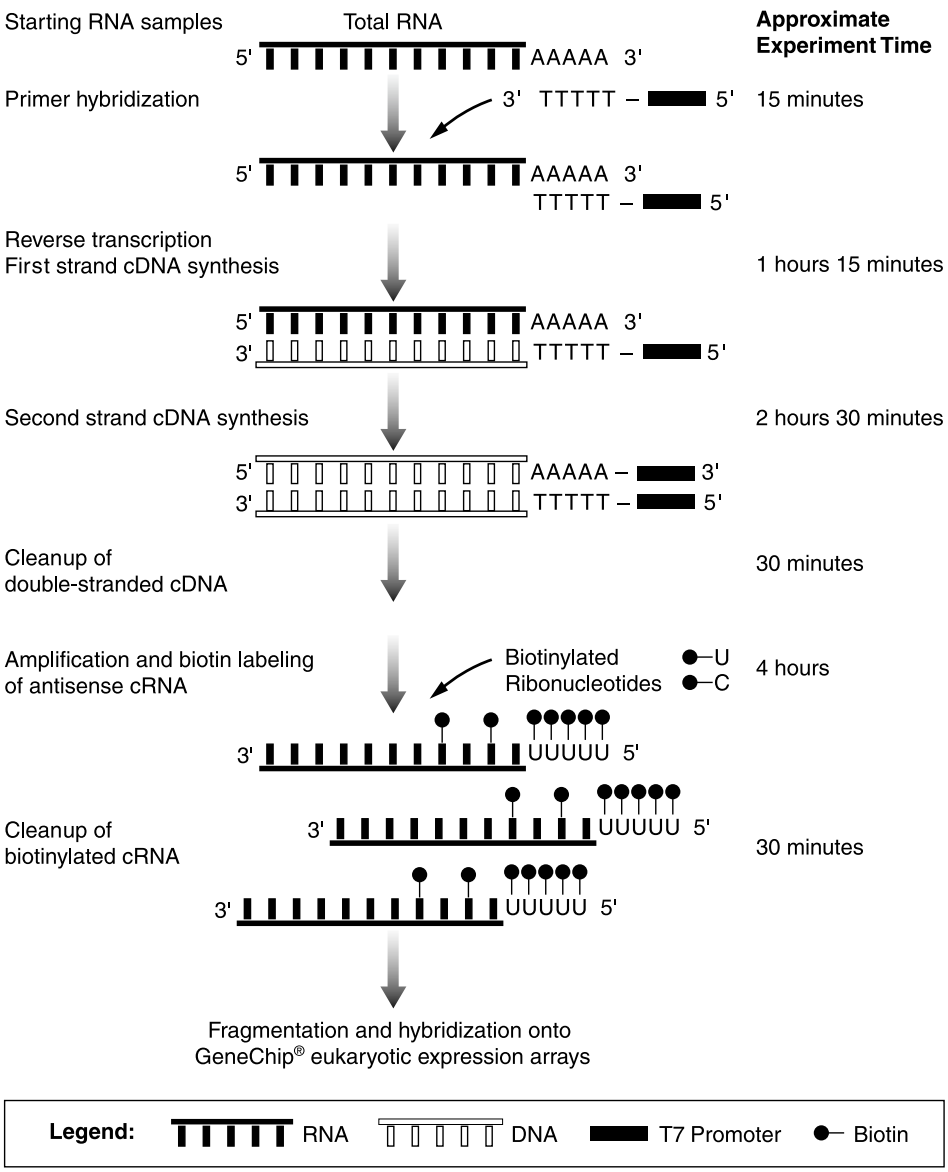
Eukaryotic Target Preparation

This Chapter Contains:

- General Guidelines for extracting RNA from eukaryotic cells or tissues using commercially available reagents and kits
- Detailed steps for making double-stranded cDNA from extracted RNA.
- Guidelines for producing biotin-labeled antisense cRNA (target) using *in vitro* transcription reaction (IVT) and the ENZO® BioArray™ HighYield™ RNA Transcript Labeling Kit
- Instructions for fragmenting the labeled cRNA target

After completing the procedures described in this chapter, the labeled and fragmented cRNA target is hybridized to the probe array as described in Section 2, Chapter 3.

Eukaryotic Target Labeling GeneChip® Probe Arrays



Reagents and Materials Required

The following reagents and materials are recommendations and have been tested and evaluated by Affymetrix scientists. For supplier phone numbers in the U.S. and Europe, please refer to the Supplier and Reagent Reference List, Appendix A, of this manual. Information and part numbers listed are based on U.S. catalog information. Additional reagents needed for the complete analysis are listed in the appropriate chapters. Appendix A contains a master list of all reagents used in this manual.

IMPORTANT *Do not store enzymes in a frost-free freezer.*

Total RNA Isolation

- TRIzol Reagent, Invitrogen Life Technologies, P/N 15596-018
- RNeasy Mini Kit, QIAGEN, P/N 74104

Poly-A mRNA Isolation

- Oligotex Direct mRNA Kit (isolation of mRNA from whole cells), QIAGEN, P/N 72012, 72022, or 72041
- Oligotex mRNA Kit (isolation of mRNA from total RNA), QIAGEN, P/N 70022, 70042, or 70061
- Qiashredder, QIAGEN, P/N 79654 (Required only for use with QIAGEN Oligotex Direct Kit)
- DEPC-Treated Water, Ambion, P/N 9920

cDNA Synthesis

- GeneChip T7-Oligo(dT) Promoter Primer Kit, 5´ - GGCCAGTGAATTGTAATACGACTCACTATAGGGAGGCGG-$(dT)_{24}$ - 3´ 50 µM, HPLC purified, Affymetrix, P/N 900375
- SuperScript II, Invitrogen Life Technologies, P/N 18064-014 or SuperScript Choice System for cDNA Synthesis, Invitrogen Life Technologies, P/N 18090-019

Note *SuperScript Choice System contains, in addition to SuperScript II Reverse Transcriptase, other reagents for cDNA synthesis. However, not all components provided in the Choice System are used in the GeneChip cDNA synthesis protocol.*

- *E. coli* DNA Ligase, Invitrogen Life Technologies, P/N 18052-019
- *E. coli* DNA Polymerase I, Invitrogen Life Technologies, P/N 18010-025
- *E. coli* RNaseH, Invitrogen Life Technologies, P/N 18021-071
- T4 DNA Polymerase, Invitrogen Life Technologies, P/N 18005-025
- 5X Second strand buffer, Invitrogen Life Technologies, P/N 10812-014
- 10 mM dNTP, Invitrogen Life Technologies, P/N 18427-013
- GeneChip Sample Cleanup Module, Affymetrix, P/N 900371

Synthesis of Biotin-Labeled cRNA

Enzo BioArray™ HighYield™ RNA Transcript Labeling Kit, Affymetrix, P/N 900182

IVT cRNA Cleanup and Quantification

- GeneChip Sample Cleanup Module, Affymetrix, P/N 900371
- 10X TBE, Cambrex, P/N 50843

cRNA Fragmentation

- GeneChip Sample Cleanup Module, Affymetrix, P/N 900371

Alternative Protocol for Cleanup of Double-Stranded cDNA

- Phase Lock Gel, Brinkmann Instruments, P/N 955 15 415
- Phenol/chloroform/isoamyl alcohol, Ambion, P/N 9732
- 7.5 M Ammonium Acetate (NH_4OAc), Sigma-Aldrich, P/N A2706
- Water, Molecular Biology Grade, BioWhittaker Molecular Applications / Cambrex, P/N 51200

Miscellaneous Reagents

- Absolute ethanol (stored at -20°C for RNA precipitation; store ethanol at room temperature for use with the GeneChip Sample Cleanup Module)
- 80% ethanol (stored at -20°C for RNA precipitation; store ethanol at room temperature for use with the GeneChip Sample Cleanup Module)
- SYBR Green II, Cambrex, P/N 50523; or Molecular Probes, P/N S7586 (optional)
- Pellet Paint, Novagen, P/N 69049-3 (optional)
- Glycogen, Ambion, P/N 9510 (optional)
- 3 M Sodium Acetate (NaOAc), Sigma-Aldrich, P/N S7899
- Ethidium Bromide, Sigma-Aldrich, P/N E8751
- 1 N NaOH
- 1 N HCl
- 50 mM $MgCl_2$
- 0.5 M EDTA

Miscellaneous Supplies

- Sterile, RNase-free, microcentrifuge tubes, 1.5 mL, USA Scientific, P/N 1415-2600 (or equivalent)
- Micropipettors, (P-2, P-20, P-200, P-1000), Rainin Pipetman or equivalent
- Sterile-barrier, RNase-free pipette tips (Tips must be pointed, not rounded, for efficient use with the probe arrays.) Beveled pipette tips may cause damage to the array septa and cause leakage.
- Mini agarose gel electrophoresis unit with appropriate buffers
- Vacuum filter units (1 liter capacity, 0.22 µm or 0.45 µm), VWR Scientific Products, P/N 28199-730
- UV spectrophotometer
- Cooling waterbath

Isolation of RNA

Protocols are provided for preparing labeled cRNA from either total RNA or purified poly-A mRNA. We have found that results obtained from samples prepared by both of these methods are similar, but not identical. Therefore, to get the best results we suggest only comparing samples prepared using the same type of RNA material.

Please review precautions and interfering conditions in Section 1.

IMPORTANT

The quality of the RNA is essential to the overall success of the analysis. Since the most appropriate protocol for the isolation of RNA can be source dependent, we recommend using a protocol that has been established for the tissues or cells being used. In the absence of an established protocol, we suggest using one of the commercially available kits designed for RNA isolation.

When using a commercial kit, follow the manufacturer's instructions for RNA isolation.

Isolation of RNA from Yeast

Total RNA

We have successfully isolated good quality total RNA from yeast cells using a hot phenol protocol described by Schmitt, *et al. Nucl Acids Res*, **18**:3091-3092 (1990).

Poly-A mRNA

Affymetrix recommends first purifying total RNA from yeast cells before isolating poly-A mRNA from total RNA. Good quality mRNA has been successfully isolated from total RNA using QIAGEN's Oligotex mRNA Kit. A single round of poly-A mRNA selection provides mRNA of sufficient purity and yield to use as a template for cDNA synthesis. Two rounds of poly-A mRNA selection will result in significantly reduced yields of yeast mRNA and is not generally recommended.

Isolation of RNA from Arabidopsis

Total RNA

We have been using TRIzol Reagent from Invitrogen Life Technologies to isolate total RNA from Arabidopsis. Please follow the instructions provided by the supplier and, when necessary, use the steps outlined specifically for samples with high starch and/or high lipid content.

Poly-A mRNA

We have successfully isolated Arabidopsis poly-A RNA using QIAGEN Oligotex products. However, other standard isolation products are likely to be adequate.

Isolation of RNA from Mammalian Cells or Tissues

Total RNA

We have successfully isolated high-quality total RNA from mammalian **cells** (such as cultured cells and lymphocytes) using the RNeasy Mini Kit from QIAGEN.

If mammalian **tissue** is used as the source of RNA, we recommend isolating total RNA with a commercial reagent such as TRIzol.

IMPORTANT

If going directly from TRIzol-isolated total RNA to cDNA synthesis, it may be beneficial to perform a second cleanup on the total RNA before starting. After the ethanol precipitation step in the TRIzol extraction procedure, perform a cleanup using QIAGEN RNeasy Mini Kit. Much better yields of labeled cRNA are obtained from the in vitro transcription-labeling reaction when this second cleanup is performed.

Poly-A mRNA

Good quality mRNA has been successfully isolated from mammalian **cells** (such as cultured cells and lymphocytes) using QIAGEN's Oligotex Direct mRNA kit and from total RNA using the Oligotex mRNA kit. If mammalian **tissue** is used as the source of mRNA, total RNA should be first purified using a commercial reagent such as TRIzol and then using a poly-A mRNA isolation procedure or a commercial kit.

Precipitation of RNA

Total RNA

It is not necessary to precipitate total RNA following isolation or cleanup with RNeasy Mini Kit. Please adjust elution volumes from the RNeasy column to prepare for cDNA synthesis based upon expected RNA yields from your experiment. Ethanol precipitation is required following TRIzol isolation and hot phenol extraction methods; see methods on page 2.1.9.

✔ Note

Affymetrix recommends starting the cDNA synthesis protocol with a minimum of 0.2 µg poly-A mRNA at a minimum concentration of 0.02 µg/µL, or 5 µg of total RNA at a minimum concentration of 0.5 µg/µL, in order to obtain sufficient quantity of labeled cRNA for target assessment and hybridization to GeneChip expression probe arrays. There are two major advantages to starting with at least the recommended amount of material:

1. *Enough material to check sample yield and quality at the various steps of this protocol.*
2. *Production of enough cRNA for hybridization of the target to multiple probe arrays.*

For smaller amounts of starting material, please refer to the alternative research protocol for target labeling described in *GeneChip Eukaryotic Small Sample Target Labeling Technical Note*, available at www.affymetrix.com.

Poly-A mRNA

Most poly-A mRNA isolation procedures will result in dilution of RNA. It is necessary to concentrate mRNA prior to the cDNA synthesis.

Precipitation Procedure

1. Add 1/10 volume 3 M NaOAc, pH 5.2, and 2.5 volumes ethanol.*
2. Mix and incubate at -20°C for at least 1 hour.
3. Centrifuge at ≥ 12,000 x g in a microcentrifuge for 20 minutes at 4°C.
4. Wash pellet twice with 80% ethanol.
5. Air dry pellet. Check for dryness before proceeding.
6. Resuspend pellet in DEPC-treated H_2O. The appropriate volume for resuspension depends on the expected yield and the amount of RNA required for the cDNA synthesis. Please read ahead to the cDNA synthesis protocol in order to determine the appropriate resuspension volume at this step.

*Addition of Carrier to Ethanol Precipitations

Adding carrier material has been shown to improve the RNA yield of precipitation reactions.

- **Pellet Paint**

  Affymetrix has found that adding 0.5 µL of Pellet Paint per tube to nucleic acid precipitations makes the nucleic acid pellet easier to visualize and helps reduce the chance of losing the pellet during washing steps. The pellet paint does not appear to affect the outcome of subsequent steps in this protocol; however, it can contribute to the absorbance at 260 nm when quantifying the mRNA.

- **Glycogen**

  Addition of 0.5 to 1 µL of glycogen (5 mg/mL) to nucleic acid precipitations aids in visualization of the pellet and may increase recovery. The glycogen does not appear to affect the outcome of subsequent steps in this protocol.

Quantification of RNA

Quantify RNA yield by spectrophotometric analysis using the convention that 1 absorbance unit at 260 nm equals 40 µg RNA per mL.

- The absorbance should be checked at 260 and 280 nm for determination of sample concentration and purity.
- The A_{260}/A_{280} ratio should be close to 2.0 for pure RNA (ratios between 1.9 and 2.1 are acceptable).

Synthesis of Double-Stranded cDNA From Total RNA

This protocol is a supplement to instructions provided in the Invitrogen Life Technologies SuperScript Choice system. Please note the following before proceeding:

- Read all information and instructions that come with reagents and kits.
- Use the GeneChip T7-Oligo(dT) Promoter Primer Kit[1] for priming first strand cDNA synthesis in place of the oligo(dT) or random primers provided with the SuperScript Choice kit. The GeneChip T7-Oligo(dT) Promoter Primer Kit provides high-quality HPLC-purified T7-oligo(dT) primer which is essential for this reaction.
- It is recommended that each step of this protocol be checked by gel electrophoresis.

T7-oligo(dT) primer

5´ - GGCCAGTGAATTGTAATACGACTCACTATAGGGAGGCGG-$(dT)_{24}$ - 3´

Step 1: First Strand cDNA Synthesis

Starting material: High-quality total RNA (5.0 µg - 20.0 µg)

Note

When using the GeneChip Sample Cleanup Module for the cDNA and IVT cRNA cleanup steps, there is a potential risk of overloading the columns if greater than the recommended amount of starting material is used.

After purification the RNA concentration is determined by absorbance at 260 nm on a spectrophotometer (one absorbance unit = 40 µg/mL RNA). The A_{260}/A_{280} ratio should be approximately 2.0, with ranges between 1.9 to 2.1 considered acceptable. We recommend checking the quality of the RNA by running it on an agarose gel prior to starting the assay. The rRNA bands should be clear without any obvious smearing patterns.

Before starting cDNA synthesis, the correct volumes of DEPC-treated H_2O and Reverse Transcriptase (RT) must be determined. These volumes will depend on both the concentration and total volume of RNA that is being added to the reaction.

IMPORTANT

Use Table 2.1.1 and Table 2.1.2 for variable component calculations. Determine the volumes of RNA and SuperScript II RT required in Table 2.1.1, then calculate the amount of DEPC-treated H_2O needed in Step 1 Table 2.1.2 to bring the final volume in Step 3 Table 2.1.2 to 20 µL.

1. Users who do not purchase the GeneChip T7-Oligo(dT) Promoter Primer Kit may be required to obtain a license under U.S. Patent Nos. 5,569,584, 5,716,785, 5,891,636, 6,291,170 and 5,545,522 or to purchase another licensed kit.

Table 2.1.1
Reverse Transcriptase Volumes for First Strand cDNA Synthesis Reaction

Total RNA (µg)	SuperScript II RT (µL), 200U/µL
5.0 to 8.0	1.0
8.1 to 16.0	2.0
16.1 to 20.0	3.0

✔ Note

The combined volume of RNA DEPC-treated H_2O and SuperScript II RT should not exceed 11 µL as indicated in Table 2.1.2.

Table 2.1.2
First Strand cDNA Synthesis Components

	Reagents in reaction	Volume	Final concentration or amount in reaction
1: *Primer Hybridization* Incubate at 70°C for 10 minutes Quick spin and put on ice	DEPC-treated H_2O (variable) T7-oligo(dT) primer, 50 µM RNA (variable)	for final reaction volume of 20 µL 2 µL 5.0 to 20 µg	 100 pmol 5.0 to 20 µg
2: *Temperature Adjustment* Add to the above tube and mix well Incubate at **42**°C for 2 minutes	5X First strand cDNA buffer 0.1 M DTT 10 mM dNTP mix	4 µL 2 µL 1 µL	1X 10 mM DTT 500 µM each
3: *First Strand Synthesis* Add to the above tube and mix well Incubate at **42**°C for 1 hour	SuperScript II RT (variable) (200 U/µL)	See Table 2.1.1	200 U to 1000 U
Total Volume		20 µL	

✔ Note

The above incubations have been changed from the SuperScript protocols and are done at 42°C.

Step 2: Second Strand cDNA Synthesis

1. Place First Strand reactions on ice. Centrifuge briefly to bring down condensation on sides of tube.
2. Add to the First Strand synthesis tube the reagents listed in the following Second Strand Final Reaction Composition Table **(Table 2.1.3).**

Table 2.1.3
Second Strand Final Reaction Composition

Component	Volume	Final Concentration or Amount in Reaction
DEPC-treated water	91 µL	
5X Second Strand Reaction Buffer	30 µL	1X
10 mM dNTP mix	3 µL	200 µM each
10 U/µL *E. coli* DNA Ligase	1 µL	10 U
10 U/µL *E. coli* DNA Polymerase I	4 µL	40 U
2 U/µL *E. coli* RNase H	1 µL	2 U
Final Volume	150 µL	

3. Gently tap tube to mix. Then, briefly spin in a microcentrifuge to remove condensation and incubate at 16°C for 2 hours in a cooling waterbath.
4. Add 2 µL [10 U] T4 DNA Polymerase.
5. Return to 16°C for 5 minutes.
6. Add 10 µL 0.5 M EDTA.
7. Proceed to cleanup procedure for cDNA, *Cleanup of Double-Stranded cDNA* on page 2.1.15, or store at -20°C for later use.

Synthesis of Double-Stranded cDNA From Purified Poly-A mRNA

This protocol is a supplement to instructions provided in the Invitrogen Life Technologies SuperScript Choice system. Please note the following before proceeding:

- Read all information and instructions that come with reagents and kits.
- Use the GeneChip T7-Oligo(dT) Promoter Primer Kit[2] for priming first strand cDNA synthesis in place of the oligo(dT) or random primers provided with the SuperScript Choice kit. The GeneChip T7-Oligo(dT) Promoter Primer Kit provides high-quality HPLC-purified T7-oligo(dT) primer which is essential for this reaction.
- It is recommended that each step of this protocol be checked by gel electrophoresis.

T7-oligo(dT) primer

5´ - GGCCAGTGAATTGTAATACGACTCACTATAGGGAGGCGG-$(dT)_{24}$ - 3´

Step 1: First Strand cDNA Synthesis

Starting material: High-quality poly-A mRNA (0.2 µg to 2.0 µg).

✔ Note *When using the GeneChip Sample Cleanup Module for the cDNA and IVT cRNA cleanup steps, there is a potential risk of overloading the columns if greater than the recommended amount of starting material is used.*

Before starting cDNA synthesis, the correct volumes of DEPC-treated H_2O and Reverse Transcriptase (RT) must be determined. These volumes will depend on both the concentration and total volume of mRNA that is being added to the reaction. For every µg of mRNA, you will need to add 1 µL of SuperScript II RT (200 U/µL). For mRNA quantity ≤ 1 µg, use 1 µL of SuperScript II RT. Synthesis reactions should be done in a polypropylene tube (RNase-free).

IMPORTANT *Use Table 2.1.4 for variable component calculations. Determine volumes of mRNA and SuperScript II RT required, and then calculate the amount of DEPC-treated H_2O needed in the **Primer Hybridization Mix** step to bring the final First Strand Synthesis reaction volume to 20 µL.*

2. Users who do not purchase the GeneChip T7-Oligo(dT) Promoter Primer Kit may be required to obtain a license under U.S. Patent Nos. 5,569,584, 5,716,785, 5,891,636, 6,291,170 and 5,545,522 or to purchase another licensed kit.

Table 2.1.4
First Strand cDNA Synthesis Components

	Reagents in Reaction	Volume	Final Concentration or Amount in Reaction
1: *Primer Hybridization* Incubate at 70°C for 10 minutes Quick spin and put on ice	DEPC-treated H_2O (variable) T7-oligo(dT) primer, 50 µM mRNA (variable)	for final reaction volume of 20 µL 2 µL 0.2 to 2 µg	 100 pmol 0.2 to 2 µg
2: *Temperature Adjustment* Add to the above tube and mix well Incubate at 37°C for 2 minutes	5X First Strand cDNA buffer 0.1 M DTT 10 mM dNTP mix	4 µL 2 µL 1 µL	1X 10 mM 500 µM each
3: *First Strand Synthesis* Add to the above tube and mix well Incubate at 37°C for 1 hour	SuperScript II RT (variable) (200 U/µL)	1 µL per µg mRNA	200 U to 1000 U
Total Volume		20 µL	

Step 2: Second Strand cDNA Synthesis

1. Place First Strand reactions on ice. Centrifuge briefly to bring down condensation on sides of tube.
2. Add to the First Strand synthesis tube the reagents listed in the following Second Strand Final Reaction Composition Table **(Table 2.1.5).**

Table 2.1.5
Second Strand Final Reaction Composition

Component	Volume	Final Concentration or Amount in Reaction
DEPC-treated water	91 µL	
5X Second Strand Reaction Buffer	30 µL	1X
10 mM dNTP mix	3 µL	200 µM each
10 U/µL *E. coli* DNA Ligase	1 µL	10 U
10 U/µL *E. coli* DNA Polymerase I	4 µL	40 U
2 U/µL *E. coli* RNase H	1 µL	2 U
Final Volume	**150 µL**	

3. Gently tap tube to mix. Then, briefly spin in a microcentrifuge to remove condensation and incubate at 16°C for 2 hours in a cooling waterbath.
4. Add 2 µL [10 U] T4 DNA Polymerase.
5. Return to 16°C for 5 minutes.
6. Add 10 µL 0.5 M EDTA.
7. Proceed to cleanup procedure for cDNA, *Cleanup of Double-Stranded cDNA* on page 2.1.15, or store at -20°C for later use.

Cleanup of Double-Stranded cDNA

Reagents to be Supplied by User

- Ethanol, 96-100% (v/v)

All other components needed for cleanup of double-stranded cDNA are supplied with the GeneChip Sample Cleanup Module.

IMPORTANT

***BEFORE STARTING**, please note:*

- *cDNA Wash Buffer is supplied as a concentrate. Before using for the first time, add 24 mL of ethanol (96-100%), as indicated on the bottle, to obtain a working solution, and checkmark the box on the left-hand side of the bottle label to avoid confusion.*
- *All steps of the protocol should be performed at room temperature. During the procedure, work without interruption.*
- *If cDNA synthesis was performed in a reaction tube smaller than 1.5 mL, transfer the reaction mixture into a 1.5 or 2 mL microfuge tube (not supplied) prior to addition of cDNA Binding Buffer.*

1. Add 600 µL cDNA Binding Buffer to the 162 µL final double-stranded cDNA synthesis preparation (page 2.1.10 or 2.1.13). Mix by vortexing for 3 seconds.
2. Check that the color of the mixture is yellow (similar to cDNA Binding Buffer without the cDNA synthesis reaction).

✔ Note

If the color of the mixture is orange or violet, add 10 µL of 3 M sodium acetate, pH 5.0, and mix. The color of the mixture will turn to yellow.

3. Apply 500 µL of the sample to the cDNA Cleanup Spin Column sitting in a 2 mL Collection Tube, and centrifuge for 1 minute at ≥ 8,000 x g (≥ 10,000 rpm).

   Discard flow-through.
4. Reload the spin column with the remaining mixture (262 µL) and centrifuge as above.

   Discard flow-through and Collection Tube.
5. Transfer spin column into a new 2 mL Collection Tube (supplied). Pipet 750 µL cDNA Wash Buffer onto the spin column. Centrifuge for 1 minute at ≥ 8,000 x g (≥ 10,000 rpm).

   Discard flow-through.

✔ Note

*cDNA Wash Buffer is supplied as a concentrate. Ensure that ethanol is added to the cDNA Wash Buffer before use (see **IMPORTANT** note above before starting).*

6. Open the cap of the spin column and centrifuge for 5 minutes at maximum speed (≤ 25,000 x g). Discard flow-through and Collection Tube.

   Place columns into the centrifuge using every second bucket. Position caps over the adjoining bucket so that they are oriented in the opposite direction to the rotation (i.e., if the microcentrifuge rotates in a clockwise direction, orient the caps in a counter-clockwise direction). This avoids damage of the caps.

   Centrifugation with open caps allows complete drying of the membrane.

7. Transfer spin column into a 1.5 mL Collection Tube, and pipet 14 µL of cDNA Elution Buffer directly onto the spin column membrane. Incubate for 1 minute at room temperature and centrifuge 1 minute at maximum speed (≤ 25,000 x g) to elute.

   Ensure that the cDNA Elution Buffer is dispensed directly onto the membrane. The average volume of eluate is 12 µL from 14 µL Elution Buffer.

✔ Note

We do not recommend RNase treatment of the cDNA prior to the in vitro transcription and labeling reaction; the carry-over ribosomal RNA does not seem to inhibit the reaction.

8. An aliquot of the cDNA prepared from isolated poly-A RNA can be analyzed for size distribution and yield on a 1% agarose gel. One microliter of double-stranded cDNA should be sufficient to detect on an agarose gel stained with ethidium bromide. A representative gel is shown in **Figure 2.1.1** on page 2.1.22. We do not recommend gel analysis for cDNA prepared from total RNA.

✔ Note

Quantifying the amount of double-stranded cDNA by absorbance at 260 nm is not recommended. The primer can contribute significantly to the absorbance. Subtracting the theoretical contribution of the primer based on the amount added to the reaction is not practical.

9. After cleanup, please proceed to *Synthesis of Biotin-Labeled cRNA* on page 2.1.17.

Synthesis of Biotin-Labeled cRNA

✔ Note

The purity and quality of template cDNA is important for high yields of biotin-labeled RNA.

Use only RNase-free water, buffers, and pipette tips.

IMPORTANT

Store all reagents at -20°C, in a freezer that is not self-defrosting.

Prior to use, centrifuge all reagents briefly to ensure that the components remain at the bottom of the tube.

The product should be used only until the expiration date stated on the label.

1. Enzo BioArray HighYield RNA Transcript Labeling Kit[3] (Affymetrix, P/N 900182) is used for generating labeled cRNA target. Use the following tables to determine the amount of cDNA used for each IVT reaction. Done properly, each reaction should produce sufficient biotin-labeled target to hybridize to at least three standard format GeneChip expression probe arrays in parallel.

Table 2.1.6
cDNA in IVT (Total RNA)

Total RNA (µg)	Volume of cDNA to use in IVT*
5.0 to 8.0	10 µL
8.1 to 16.0	5 µL
16.1 to 20.0	3.3 µL

* assuming 12 µL was eluted from the column, as previously described.

Table 2.1.7
cDNA in IVT (Poly-A RNA)

Poly-A RNA (µg)	Volume cDNA*
0.2 - 0.5	10 µL
0.6 - 1.0	8 µL
1 - 2	5 µL

* assuming 12 µL was eluted from the column, as previously described.

3. For Research Use Only. This product is manufactured by ENZO LIFE SCIENCES, INC. for distribution by Affymetrix, Inc. for research purposes only by the end-user and is not intended for diagnostic or therapeutic use. Purchase does not include a license or the right to utilize this product except for research purposes. Purchase does not include the right to distribute or sell this product commercially. As distributed by Affymetrix. Inc., this product may be used only in conjunction with and is permitted for use only with Affymetrix® GeneChip® probe arrays.

   Enzo is a registered trademark of Enzo Biochem, Inc. and BioArray is a trademark of Enzo Biochem, Inc.

   This product or the use of this product is covered by one or more claims of Enzo patents including, but not limited to, the following: U.S. Patent Nos. 5,328,824; 5,449,767; 5,476,928; 4,711,955 and 4,994,373; EP 0 063 879 B1; EP 0 329 198 B1; DK 171 822 B; Canadian Patent Nos. 1,219,824 and 1,309,672; Japanese Patent Nos. 2,131,266; 1,416,584 and other patents pending.

IMPORTANT *Each GeneChip® Sample Cleanup Module contains 30 cDNA cleanup columns and 30 IVT cRNA cleanup columns. If more than one IVT is carried out from a single cDNA sample and is purified on separate IVT cRNA cleanup columns, there will not be sufficient IVT cRNA columns in each kit for 30 samples.*

2. Add to RNase-free microfuge tubes template cDNA and additions of other reaction components in the order indicated in the following table. Keep reactions at room temperature while additions are made to avoid precipitation of DTT.

Table 2.1.8
IVT cRNA Labeling

Reagent	Volume
Template cDNA	Variable. Refer to Table 2.1.6 and Table 2.1.7.
Distilled or deionized water	Variable (to give a final reaction volume of 40 µL).
10X HY Reaction Buffer (Vial 1)	4 µL
10X Biotin-Labeled Ribonucleotides (Vial 2)	4 µL
10X DTT (Vial 3)	4 µL
10X RNase Inhibitor Mix (Vial 4)	4 µL
20X T7 RNA Polymerase (Vial 5)	2 µL
Total Volume	**40 µL**

3. Carefully mix the reagents and collect the mixture in the bottom of the tube by brief (5 second) microcentrifugation.
4. Immediately place the tube in a 37°C water bath. Incubate for 4 to 5 hours, gently mixing the contents of the tube every 30-45 minutes during the incubation.

Note *Overnight incubation may produce shorter products which is less desirable.*

5. Store labeled cRNA at -70°C, or -20°C if not purifying immediately.

Cleanup and Quantification of Biotin-Labeled cRNA

Reagents to be Supplied by User

- Ethanol, 96-100% (v/v)
- Ethanol, 80% (v/v)

All other components needed for cleanup of biotin-labeled cRNA are supplied with the GeneChip Sample Cleanup Module.

Step 1: Cleanup of Biotin-Labeled cRNA

IMPORTANT

BEFORE STARTING *please note:*

- *It is essential to remove unincorporated NTPs, so that the concentration and purity of cRNA can be accurately determined by 260 nm absorbance.*
- *DO NOT extract biotin-labeled RNA with phenol-chloroform. The biotin will cause some of the RNA to partition into the organic phase. This will result in low yields.*
- *Save an aliquot of the unpurified IVT product for analysis by gel electrophoresis.*
- *IVT cRNA Wash Buffer is supplied as a concentrate. Before using for the first time, add 20 mL of ethanol (96-100%), as indicated on the bottle, to obtain a working solution, and checkmark the box on the left-hand side of the bottle label to avoid confusion.*
- *IVT cRNA Binding Buffer may form a precipitate upon storage. If necessary, redissolve by warming in a water bath at 30°C, and then place the buffer at room temperature.*
- *All steps of the protocol should be performed at room temperature. During the procedure, work without interruption.*

1. Add 60 µL of RNase-free water to the *in vitro* transcription reaction and mix by vortexing for 3 seconds.
2. Add 350 µL IVT cRNA Binding Buffer to the sample and mix by vortexing for 3 seconds.
3. Add 250 µL ethanol (96-100%) to the lysate, and mix well by pipetting. Do not centrifuge.
4. Apply sample (700 µL) to the IVT cRNA Cleanup Spin Column sitting in a 2 mL Collection Tube. Centrifuge for 15 seconds at ≥ 8,000 x g (≥ 10,000 rpm). Discard flow-through and Collection Tube.
5. Transfer the spin column into a new 2 mL Collection Tube (supplied). Pipet 500 µL IVT cRNA Wash Buffer onto the spin column. Centrifuge for 15 seconds at ≥ 8,000 x g (≥ 10,000 rpm) to wash. Discard flow-through.

✔ Note

IVT cRNA Wash Buffer is supplied as a concentrate. Ensure that ethanol is added to the IVT cRNA Wash Buffer before use (see ***IMPORTANT*** *note above before starting).*

6. Pipet 500 µL 80% (v/v) ethanol onto the spin column and centrifuge for 15 seconds at ≥ 8,000 x g (≥ 10,000 rpm). Discard flow-through.

7. Open the cap of the spin column and centrifuge for 5 minutes at maximum speed (≤ 25,000 x g). Discard flow-through and Collection Tube.

   Place columns into the centrifuge using every second bucket. Position caps over the adjoining bucket so that they are oriented in the opposite direction to the rotation (i.e., if the microcentrifuge rotates in a clockwise direction, orient the caps in a counter-clockwise direction). This avoids damage of the caps. Centrifugation with open caps allows complete drying of the membrane.

8. Transfer spin column into a new 1.5 mL Collection Tube (supplied), and pipet 11 µL of RNase-free Water directly onto the spin column membrane. Ensure that the water is dispensed directly onto the membrane. Centrifuge 1 minute at maximum speed (≤ 25,000 x g) to elute.

9. Pipet 10 µL of RNase-free Water directly onto the spin column membrane. Ensure that the water is dispensed directly onto the membrane. Centrifuge 1 minute at maximum speed (≤ 25,000 x g) to elute.

   For subsequent photometric quantification of the purified cRNA, we recommend dilution of the eluate between 1:100 fold and 1:200 fold.

IMPORTANT *The minimum concentration for purified cRNA is 0.6 µg/µL before starting the following fragmentation reaction in "Fragmenting the cRNA for Target Preparation" on page 2.1.21.*

Step 2: Quantification of the cRNA

Use spectrophotometric analysis to determine the cRNA yield. Apply the convention that 1 absorbance unit at 260 nm equals 40 µg/mL RNA.

- Check the absorbance at 260 nm and 280 nm to determine sample concentration and purity.
- Maintain the A_{260}/A_{280} ratio close to 2.0 for pure RNA (ratios between 1.9 and 2.1 are acceptable).

For quantification of cRNA when using total RNA as starting material, an adjusted cRNA yield must be calculated to reflect carryover of unlabeled total RNA. Using an estimate of 100% carryover, use the formula below to determine adjusted cRNA yield:

$$\text{adjusted cRNA yield} = RNA_m - (\text{total } RNA_i)(y)$$

RNA_m = amount of cRNA measured after IVT (µg)
total RNA_i = starting amount of total RNA (µg)
y = fraction of cDNA reaction used in IVT

Example: Starting with 10 µg total RNA, 50% of the cDNA reaction is added to the IVT, giving a yield of 50 µg cRNA. Therefore, adjusted cRNA yield = 50 µg cRNA - (10 µg total RNA) (0.5 cDNA reaction) = 45.0 µg.

Use adjusted yield in *Eukaryotic Target Hybridization* on page 2.3.3.

Note *Please refer to Table 2.3.1 on page 2.3.7 for the amount of cRNA required for one array hybridization experiment. The amount varies depending on the array format. Please refer to a specific probe array package insert for information on the array format.*

Step 3: Checking Unfragmented Samples by Gel Electrophoresis

Gel electrophoresis of the IVT product is done to estimate the yield and size distribution of labeled transcripts. Parallel gel runs of unpurified and purified IVT product can help determine the extent of a loss of sample during the cleanup process.

- Run 1% of each sample on a 1% agarose gel.
- Mix RNA (samples or markers) with loading dye and heat to 65°C for 5 minutes before loading on the gel.
- Ethidium bromide can be used to visualize the RNA in the gel. Alternatively, gels can be stained with SYBR Green II at a 1:10,000 dilution in 1X TBE buffer. SYBR Green II stains single-stranded RNA with greater sensitivity than ethidium bromide, but it requires a special photographic filter available from Molecular Probes to photograph stained bands.
- As an option, run a denaturing gel to obtain a more accurate estimation of the RNA size distribution. Please refer to Figure 2.1.1 for the typical size distribution of unfragmented cRNA.

Fragmenting the cRNA for Target Preparation

5X Fragmentation Buffer is supplied with the GeneChip Sample Cleanup Module.

Fragmentation of cRNA target before hybridization onto GeneChip probe arrays has been shown to be critical in obtaining optimal assay sensitivity.

Affymetrix recommends that the cRNA used in the fragmentation procedure be sufficiently concentrated to maintain a small volume during the procedure. This will minimize the amount of magnesium in the final hybridization cocktail. The cRNA must be at a minimum concentration of 0.6 µg/µL. Fragment an appropriate amount of cRNA for hybridization cocktail and gel analysis (see Section 2, Chapter 3, Table 2.3.1).

1. Add 2 µL of 5X Fragmentation Buffer for every 8 µL of RNA plus H_2O. The fragmentation buffer has been optimized to break down full-length cRNA to 35 to 200 base fragments by metal-induced hydrolysis.

   The final concentration of RNA in the fragmentation mix can range from 0.5 µg/µL to 2 µg/µL. The following table shows an example of a fragmentation mix for a 20 µg cRNA sample at a final concentration of 0.5 µg/µL.

   For fragmentation, use **adjusted** cRNA concentration, as described in *Step 2: Quantification of the cRNA* on page 2.1.20.

   Example for 0.5 µg/µL final concentration:

Table 2.1.9
Example of Fragmentation Reaction

Component	Volume
20 µg cRNA	1 to 21 µL
5X Fragmentation Buffer	8 µL
RNase-free water	to 40 µL

2. Incubate at 94°C for 35 minutes. Put on ice following the incubation.

3. Save an aliquot for gel analysis.

   At least 1 µg fragmented cRNA is needed if using ethidium bromide for staining the gel. Less RNA can be used with SYBR Green II staining. See *Step 3: Checking Unfragmented Samples by Gel Electrophoresis* on page 2.1.21, for information regarding gel electrophoresis. The standard fragmentation procedure should produce a distribution of RNA fragment sizes from approximately 35 to 200 bases. An example of a gel with cRNA samples before and after fragmentation is shown below.

4. Store undiluted, fragmented sample RNA at -20°C until ready to perform the hybridization, as described in Section 2, Chapter 3.

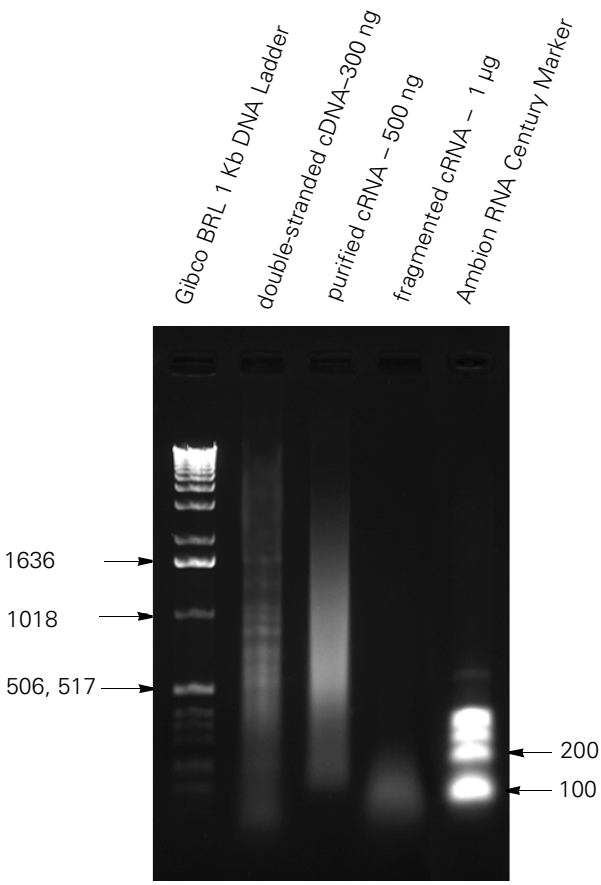


Figure 2.1.1
Monitoring of target preparation by agarose gel electrophoresis

Alternative Protocol for Cleanup of Double-Stranded cDNA

✔ Note *The following procedure may be followed as an alternative to using the cDNA cleanup columns in the GeneChip Sample Cleanup Module. Global concordance on array results obtained from using the two protocols has been established. However, Affymetrix encourages customers to evaluate their unique samples carefully before comparing results from the two cleanup protocols directly.*

Step 1: Phase Lock Gels (PLG)-Phenol/Chloroform Extraction

Phase Lock Gels (PLG) form an inert, sealed barrier between the aqueous and organic phases of phenol-chloroform extractions. The solid barrier allows more complete recovery of the sample (aqueous phase) and minimizes interface contamination of the sample. PLG's are sold as premeasured aliquots in 1.5 mL tubes to which sample and phenol chloroform are directly added.

✔ Note *A standard phenol/chloroform extraction can be performed as an alternative to the PLG procedure.*

1. Pellet the Phase Lock Gel (1.5 mL tube with PLG I - heavy) in a microcentrifuge at 12,000 x g for 20 to 30 seconds.
2. Add 162 µL (equal volume) of (25:24:1) Phenol:chloroform:isoamyl alcohol (saturated with 10 mM Tris-HCl pH 8.0, 1 mM EDTA) to the final cDNA synthesis preparation (162 µL) to a final volume of 324 µL. Vortex briefly.

   (See Reagents and Materials Required on page 2.1.5 for ordering information for phenol:chloroform:isoamyl alcohol.)

✔ Note *Store phenol:chloroform:isoamyl alcohol at 4°C. Dispose of solution when it turns pink.*

3. Transfer the entire cDNA-phenol/chloroform mixture to the PLG tube.
4. **DO NOT VORTEX**. PLG will now become part of the suspension. Microcentrifuge at full speed (12,000 x g) for 2 minutes.
5. Transfer the aqueous upper phase to a fresh 1.5 mL tube.

Step 2: Ethanol Precipitation

Please refer to Precipitation of RNA on page 2.1.9 for information on the use of carriers during ethanol precipitation.

1. Add 0.5 volumes of 7.5 M NH4OAc and 2.5 volumes of absolute ethanol (stored at -20°C) to the sample and vortex.
2. Immediately centrifuge at 12,000 x g in a microcentrifuge at room temperature for 20 minutes.
3. Remove supernatant. Wash pellet with 0.5 mL of 80% ethanol (stored at -20°C).
4. Centrifuge at 12,000 x g at room temperature for 5 minutes.
5. Remove the 80% ethanol very carefully; the pellet may be loose. Repeat the 80% ethanol wash one additional time.

6. Air dry the pellet. Check pellet for dryness before proceeding.
7. Resuspend dried pellet in a small volume of RNase-free water. We recommend resuspending in 12 µL.

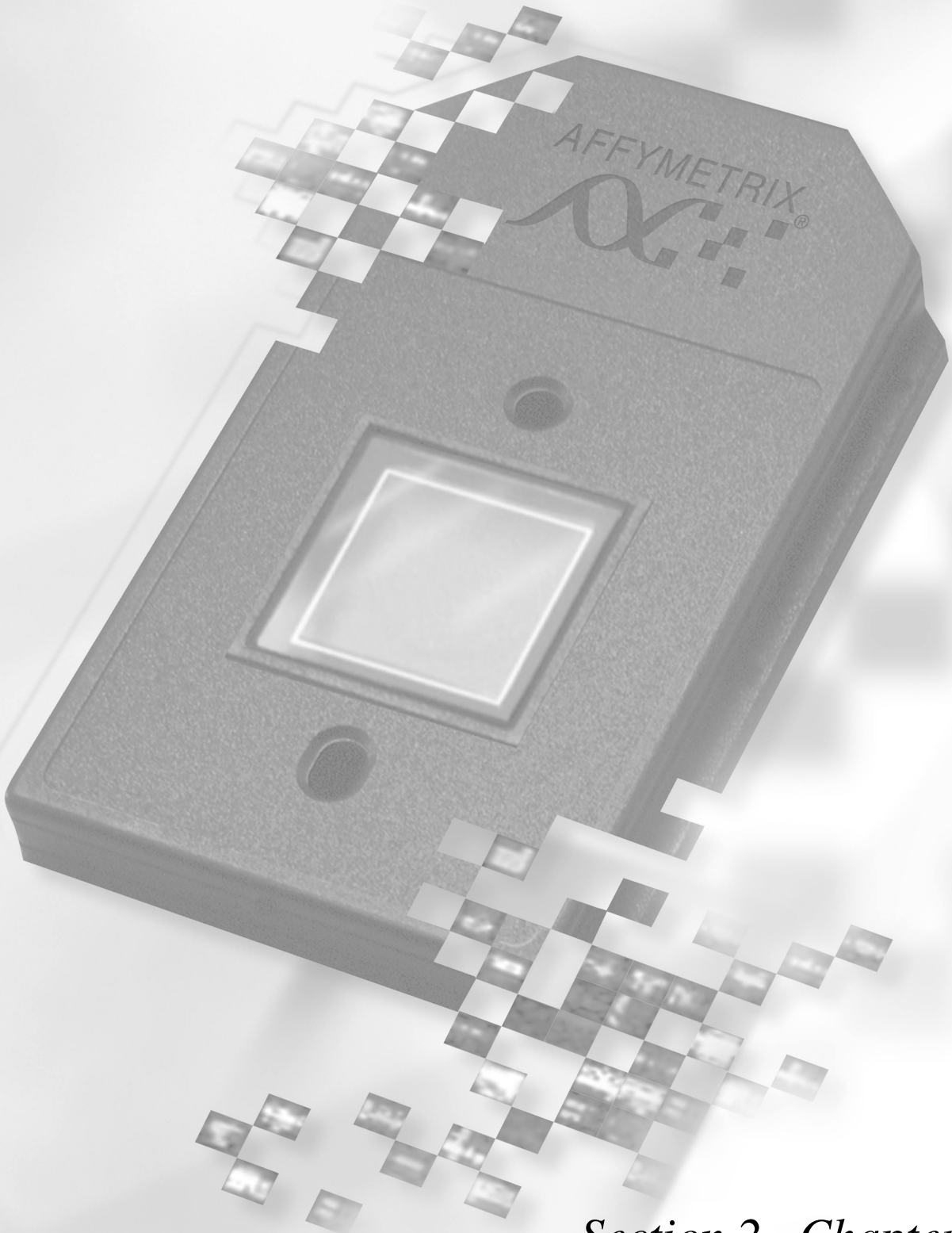
Alternative Protocol for Preparing 5X Fragmentation Buffer

5X RNA Fragmentation Buffer
(200 mM Tris-acetate, pH 8.2, 500 mM KOAc, 150 mM MgOAc)

The fragmentation buffer is made with RNase-free reagents. Tris-containing solutions should not be treated with DEPC; however, once H_2O has been DEPC-treated and autoclaved it can be used for making the Tris solution.

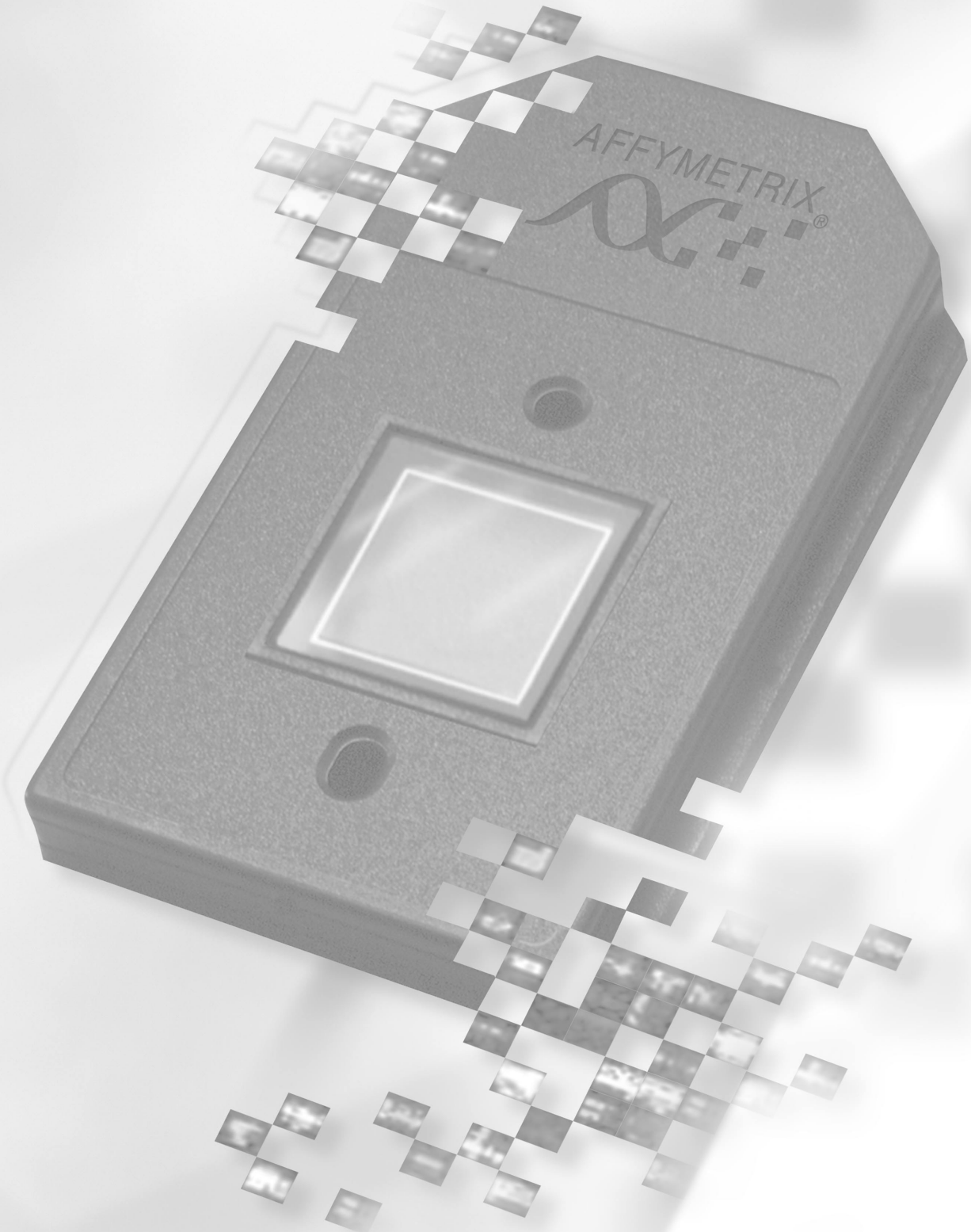
1. Combine the following components to a total volume of 20 mL.
   - 4.0 mL 1 M Tris acetate pH 8.1 (Trizma Base, pH adjusted with glacial acetic acid)
   - 0.64 g MgOAc
   - 0.98 g KOAc
   - DEPC-treated H_2O to 20 mL

   Final pH without adjustment should be 8.2.
2. Mix thoroughly and filter through a 0.2 µm vacuum filter unit. This reagent should be aliquotted and stored at room temperature.



Section 2, Chapter 2

Section 2, Chapter 2



Controls for Eukaryotic Arrays

This Chapter Contains:

- General guidelines for producing controls for eukaryotic arrays.

After completing the procedures described in this chapter, the control transcripts are combined in variable concentrations before adding to the target hybridization mix as explained in Section 2, Chapter 3.

Reagents and Materials Required

The following reagents and materials are recommendations and have been tested and evaluated by Affymetrix scientists. For supplier phone numbers in the U.S. and Europe, please refer to the Supplier and Reagent Reference List, Appendix A, of this manual. Information and part numbers listed are based on U.S. catalog information. Additional reagents needed for the complete analysis are listed in the appropriate chapters. Appendix A contains a master list of all reagents used in this manual.

GeneChip Eukaryotic Hybridization Control Kit (Complete Kit)

- Affymetrix, P/N 900299 (30 reactions) or P/N 900362 (150 reactions)

Poly-A Spike Controls

- pGIBS-lys ATCC 87482
- pGIBS-phe ATCC 87483
- pGIBS-thr ATCC 87484
- pGIBS-trp ATCC 87485
- pGIBS-dap ATCC 87486

Hybridization Control Kit

Each commercially available eukaryotic probe array contains probe sets for several prokaryotic genes as controls. These probe sets are readily identified by the AFFX prefix in the probe set name. The .chp data for these control probe sets can be examined in the Summary Report File (.rpt).

Control Oligo B2

Control Oligo B2 hybridizes to features along the outer edge of all expression arrays and to the checkerboard pattern in each corner. These predefined patterns provide signals for the Microarray Suite software to perform automatic grid alignment during image analysis. They can also be used to align the grid manually. The fluorescence intensities for Control Oligo B2 are not used for analyzing data.

A 60X stock of the B2 oligo is provided as part of the GeneChip Eukaryotic Hybridization Control Kit (P/N 900299 or 900362, for 30 or 150 reactions, respectively), or can be purchased alone (P/N 900301). Please refer to the instructions in Section 2, Chapter 3 for detailed information on including the B2 oligo in preparing the hybridization cocktail.

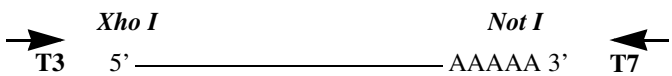
Biotinylated Hybridization Controls: *bioB, bioC, bioD,* and *cre*

BioB, bioC, and *bioD* are genes of the biotin synthesis pathway from the bacteria *E. coli,* and *cre* is the recombinase gene from P1 bacteriophage. A ready-prepared mixture of these biotinylated controls at staggered concentrations can be added with labeled eukaryotic cRNA samples to hybridize onto GeneChip probe arrays. Signal intensities obtained on these genes provide information on how well the hybridization, washing and staining procedures have performed.

Affymetrix provides a kit that contains a 20X pre-mixed control reagents (P/N 900299 or 900362) and the final concentrations in the hybridization cocktail are 1.5 pM, 5 pM, 25 pM and 100 pM for the four transcripts *bioB*, *bioC*, *bioD,* and *cre*, respectively.

Poly-A Spike Controls

Five poly-A-tailed control clones encoding *B. subtilis* genes (*dap, thr, trp, phe, lys*) are cloned into pBluescript as an *Xho I* to *Not I* insert, 5´ to 3´, respectively.

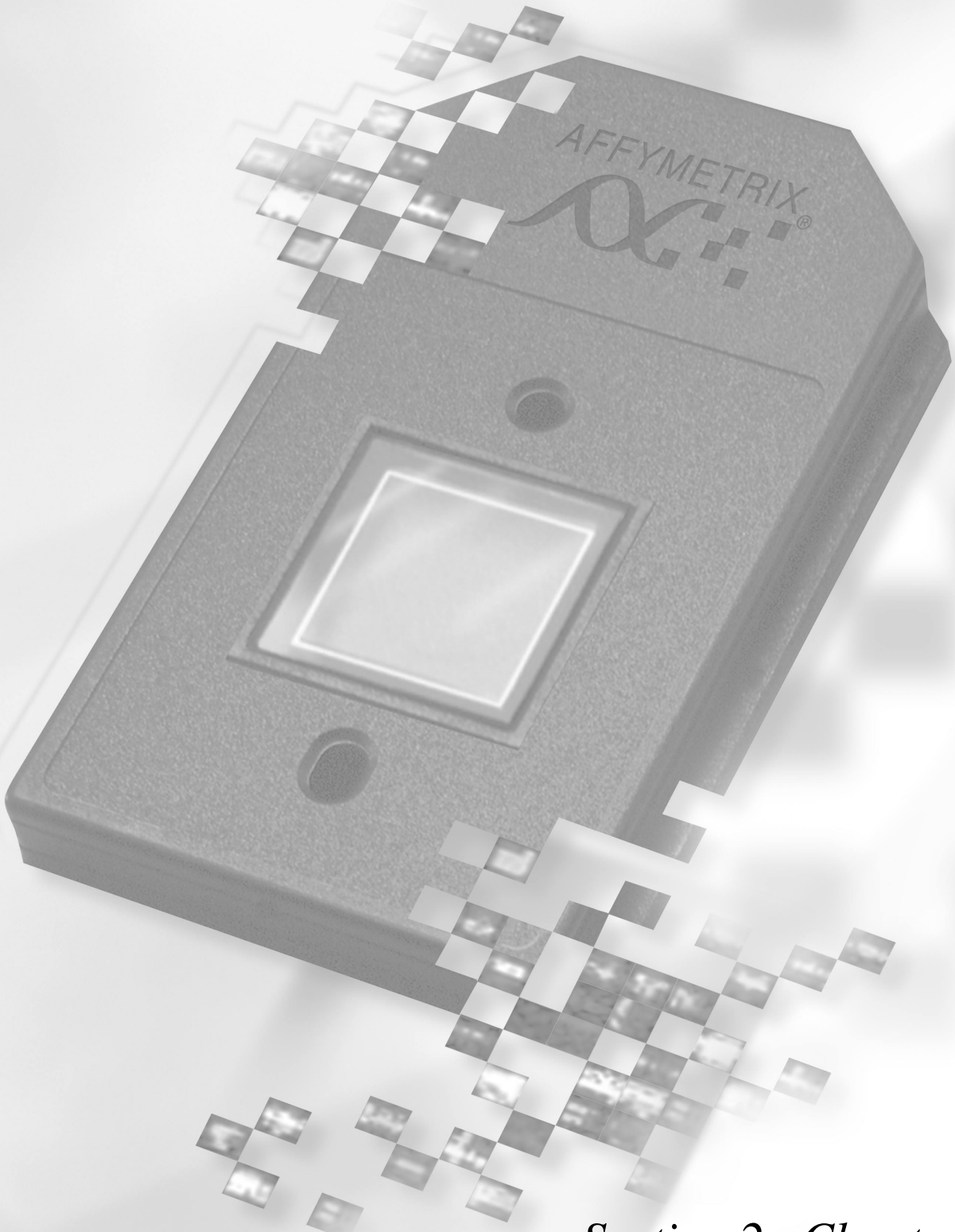


Poly-A-tailed constructs (*dap, thr, trp, phe, lys*)

These clones can be cut with different restriction enzymes to produce template DNA for either sense strand RNA synthesis or antisense RNA synthesis. The antisense control RNA for each *B. subtilis* gene is synthesized from linearized plasmid using T7 RNA polymerase with biotinylated nucleotides. The sense RNA for each *B. subtilis* gene is synthesized from linearized plasmid using T3 RNA polymerase with unlabeled nucleotides. For detailed preparation of sense RNA controls, please refer to Section 3, Chapter 2.

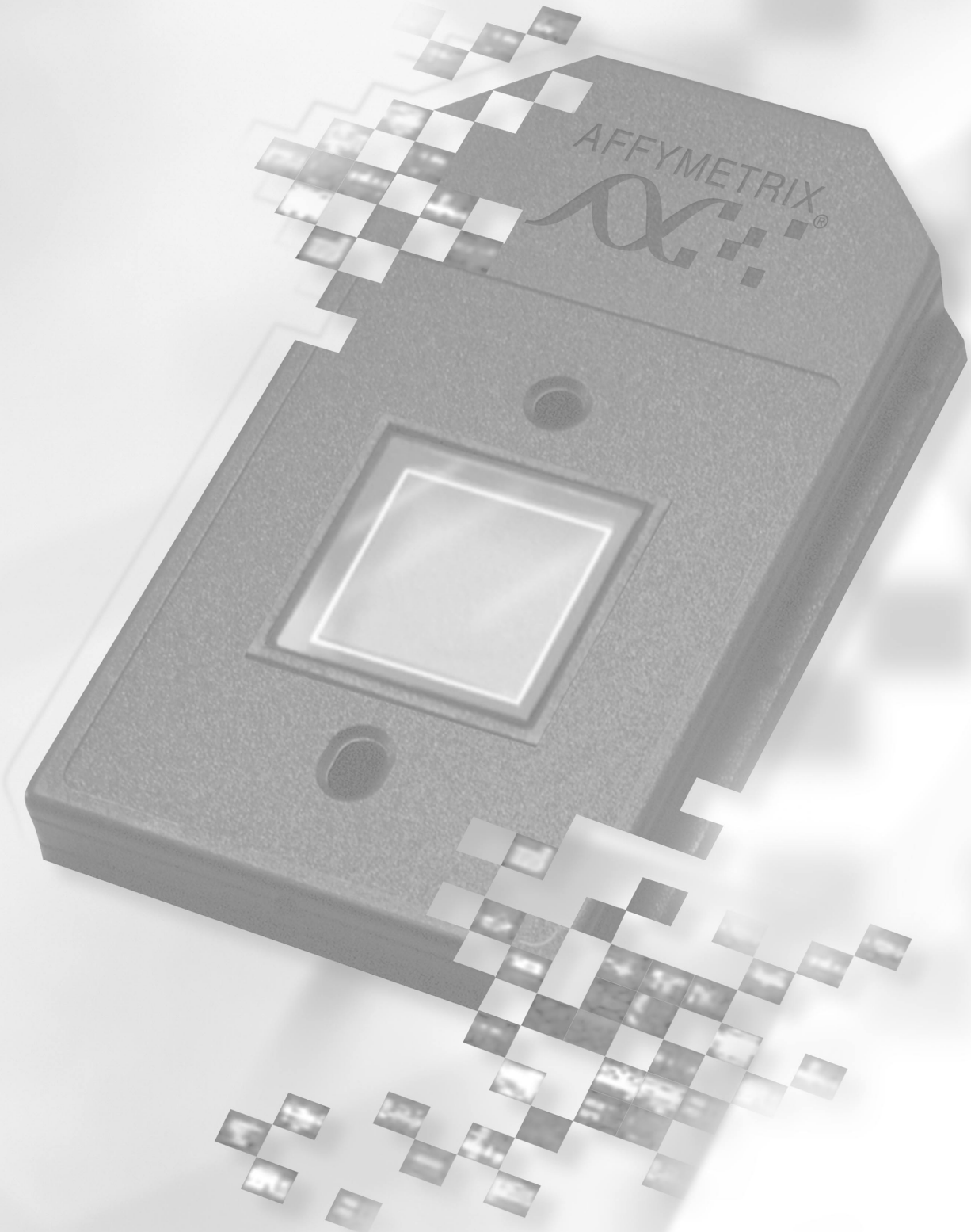
The antisense strand *B. subtilis* RNA controls are used as described above for *bioB*, *bioC,* and *bioD* genes. The sense strand RNA controls can be spiked into samples during mRNA preparation to monitor the efficiency of target preparation, hybridization, wash, and stain.

Bacteria containing these recombinant plasmids can be obtained from the American Type Culture Collection (ATCC). See *Reagents and Materials Required* on page 2.2.5 for details.



Section 2, Chapter 3

Section 2, Chapter 3



Eukaryotic Target Hybridization

This Chapter Contains:

- Detailed steps for preparing the eukaryotic hybridization mix containing labeled target and control cRNA.
- Instructions for hybridizing the target mix to a eukaryotic GeneChip probe array.

After completing the procedures described in this chapter, the hybridized probe array is ready for washing, staining, and scanning as detailed in Section 2, Chapter 4.

Reagents and Materials Required

The following reagents and materials are recommendations and have been tested and evaluated by Affymetrix scientists. For supplier phone numbers in the U.S. and Europe, please refer to the Supplier and Reagent Reference List, Appendix A, of this manual. Information and part numbers listed are based on U.S. catalog information. Additional reagents needed for the complete analysis are listed in the appropriate chapters. Appendix A contains a master list of all reagents used in this manual.

- Water, Molecular Biology Grade, BioWhittaker Molecular Applications / Cambrex, P/N 51200
- Acetylated Bovine Serum Albumin (BSA) solution (50 mg/mL), Invitrogen Life Technologies, P/N 15561-020
- Herring Sperm DNA, Promega Corporation, P/N D1811
- Micropure Separator, Millipore, P/N 42512 (optional)
- GeneChip Eukaryotic Hybridization Control Kit, Affymetrix, P/N 900299 (contains Control cRNA and Control Oligo B2)
- Control Oligo B2, 3 nM, Affymetrix, P/N 900301 (can be ordered separately)
- 5 M NaCl, RNase-free, DNase-free, Ambion, P/N 9760G
- MES Free Acid Monohydrate SigmaUltra, Sigma-Aldrich, P/N M5287
- MES Sodium Salt, Sigma-Aldrich, P/N M5057
- EDTA Disodium Salt, 0.5 M solution (100 mL), Sigma-Aldrich, P/N E7889

Miscellaneous Reagents

- Tough Spots, Label Dots, USA Scientific, P/N 9185 (optional)
- Surfact-Amps 20 (Tween-20), 10%, Pierce Chemical, P/N 28320

Miscellaneous Supplies

- Hybridization Oven 640, Affymetrix, P/N 800139
- Sterile, RNase-free, microcentrifuge tubes, 1.5 mL, USA Scientific, P/N 1415-2600 (or equivalent)
- Micropipettors, (P-2, P-20, P-200, P-1000), Rainin Pipetman or equivalent
- Sterile-barrier pipette tips and non-barrier pipette tips
- Heatblock

Reagent Preparation

12X MES Stock

(1.22 M MES, 0.89 M [Na^+])

For 1000 mL:

70.4 g MES-free acid monohydrate
193.3 g MES Sodium Salt
800 mL of Molecular Biology Grade water
Mix and adjust volume to 1000 mL.
The pH should be between 6.5 and 6.7. Filter through a 0.2 µm filter.

IMPORTANT *Do not autoclave. Store at 2°C to 8°C, and shield from light. Discard solution if yellow.*

2X Hybridization Buffer

(Final 1X concentration is 100 mM MES, 1 M [Na^+], 20 mM EDTA, 0.01% Tween 20)

For 50 mL:

8.3 mL of 12X MES Stock
17.7 mL of 5 M NaCl
4.0 mL of 0.5 M EDTA
0.1 mL of 10% Tween 20
19.9 mL of water
Store at 2°C to 8°C, and shield from light

Eukaryotic Target Hybridization

Please refer to the table below for the necessary amount of cRNA for appropriate probe array format. These recipes take into account that it is necessary to make extra hybridization cocktail due to a small loss of volume (10-20 µL) during each hybridization.

1. Mix the following for each target, scaling up volumes for hybridization to multiple probe arrays.

Table 2.3.1
Hybridization Cocktail for Single Probe Array*

Component	Micro/Mini Array	Midi Array	Standard Array	Final Concentration
Fragmented cRNA **	5 µg	10 µg	15 µg	0.05 µg/µL
Control Oligonucleotide B2 (3 nM)	1.7 µL	3.3 µL	5 µL	50 pM
20X Eukaryotic Hybridization Controls (*bioB*, *bioC*, *bioD*, *cre*)	5 µL	10 µL	15 µL	1.5, 5, 25 and 100 pM respectively
Herring Sperm DNA (10 mg/mL)	1 µL	2µL	3 µL	0.1 mg/mL
Acetylated BSA (50 mg/mL)	1 µL	2 µL	3 µL	0.5 mg/mL
2X Hybridization Buffer	50 µL	100 µL	150 µL	1X
H_2O	to final volume of 100 µL	to final volume of 200 µL	to final volume of 300 µL	
Final volume	**100 µL**	**200 µL**	**300 µL**	

*Please refer to specific probe array package insert for information on array format.

**Please see Section 2, Chapter 1, page 2.1.20 for amount of adjusted fragmented cRNA to use when starting from total RNA.

IMPORTANT *It is imperative that frozen stocks of 20X GeneChip Eukaryotic Hybridization Control be heated to 65°C for 5 minutes to completely resuspend the cRNA before aliquotting.*

2. Equilibrate probe array to room temperature immediately before use.

Note *It is important to allow the arrays to equilibrate to room temperature completely. Specifically, if the rubber septa are not equilibrated to room temperature, they may be prone to cracking which can lead to leaks.*

3. Heat the hybridization cocktail to 99°C for 5 minutes in a heat block.
4. Meanwhile, wet the array by filling it through one of the septa (see **Figure 2.3.1** for location of the probe array septa) with appropriate volume 1X Hybridization Buffer using a micropipettor and appropriate tips **(Table 2.3.2)**.

Note *It is necessary to use two pipette tips when filling the probe array cartridge: one for filling and the second to allow venting of air from the hybridization chamber. After the addition of hybridization cocktails to the array, the septa may be covered with Tough Spots to prevent evaporation.*

5. Incubate the probe array filled with 1X Hybridization Buffer at 45°C for 10 minutes with rotation.

Table 2.3.2
Probe Array Cartridge Volumes

Array	Hybridization Volume	Total Fill Volume
Standard	200 µL	250 µL
Midi	130 µL	160 µL
Mini	80 µL	100 µL
Micro	80 µL	100 µL

6. Transfer the hybridization cocktail that has been heated at 99°C, in step 3, to a 45°C heat block for 5 minutes.
7. Spin hybridization cocktail(s) at maximum speed in a microcentrifuge for 5 minutes to remove any insoluble material from the hybridization mixture.
8. Remove the buffer solution from the probe array cartridge and fill with appropriate volume (**Table 2.3.2** on page 2.3.8) of the clarified hybridization cocktail avoiding any insoluble matter in the volume at the bottom of the tube.
9. Place probe array in rotisserie box in 45°C oven.

   Avoid stress to rotisserie motor; load probe arrays in a balanced configuration around rotisserie axis. Rotate at 60 rpm.
10. Hybridize for 16 hours.

    During the latter part of the 16-hour hybridization, proceed to Section 2, Chapter 4 to prepare reagents required immediately after completion of hybridization.

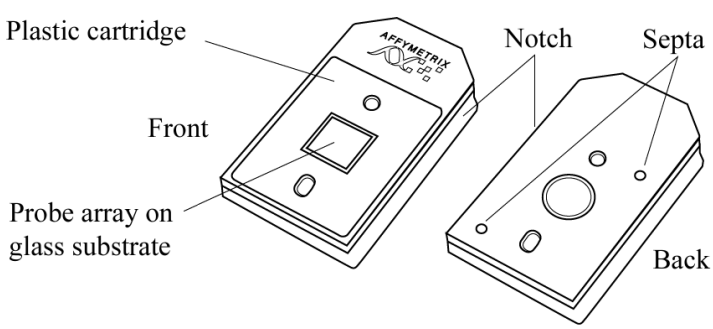
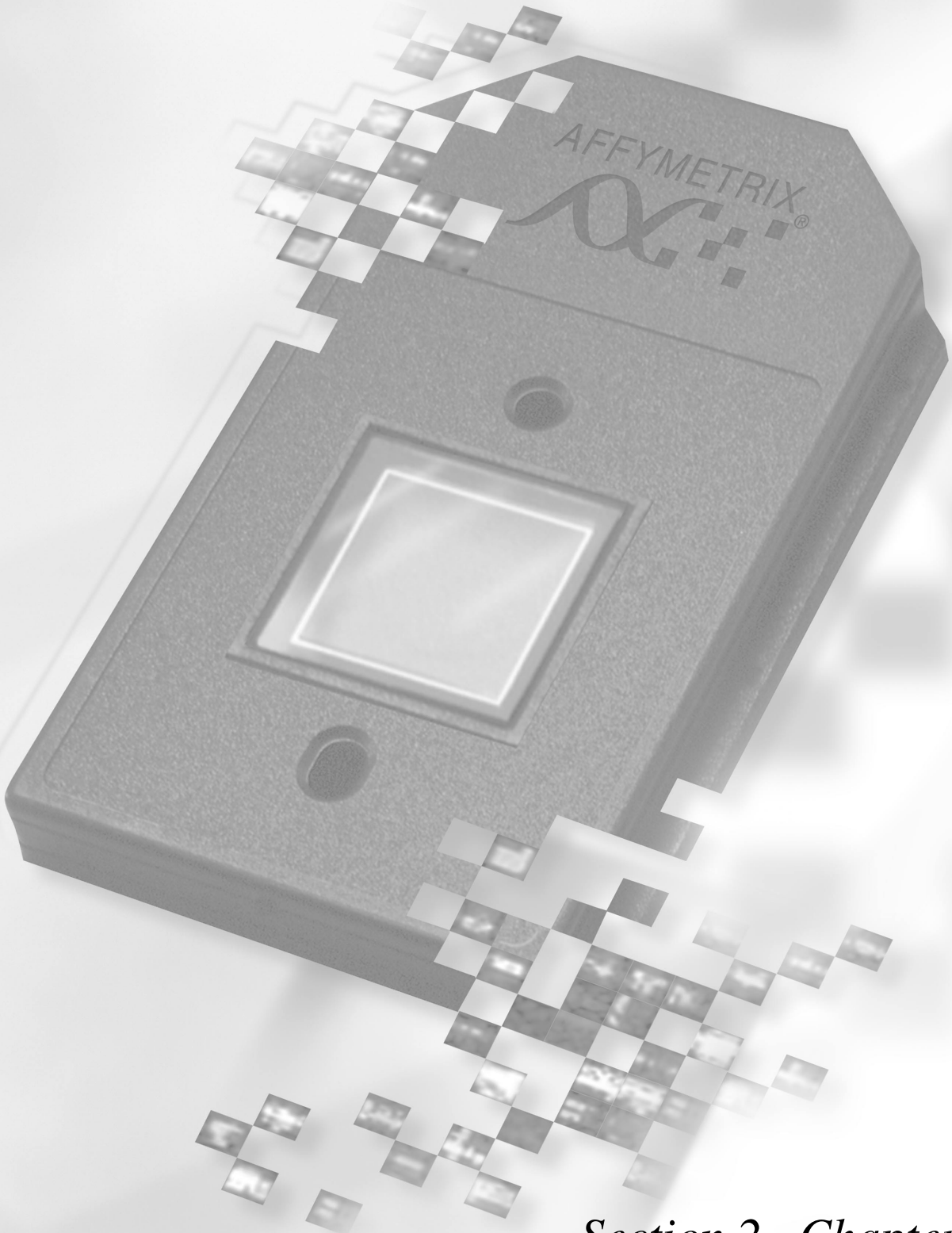
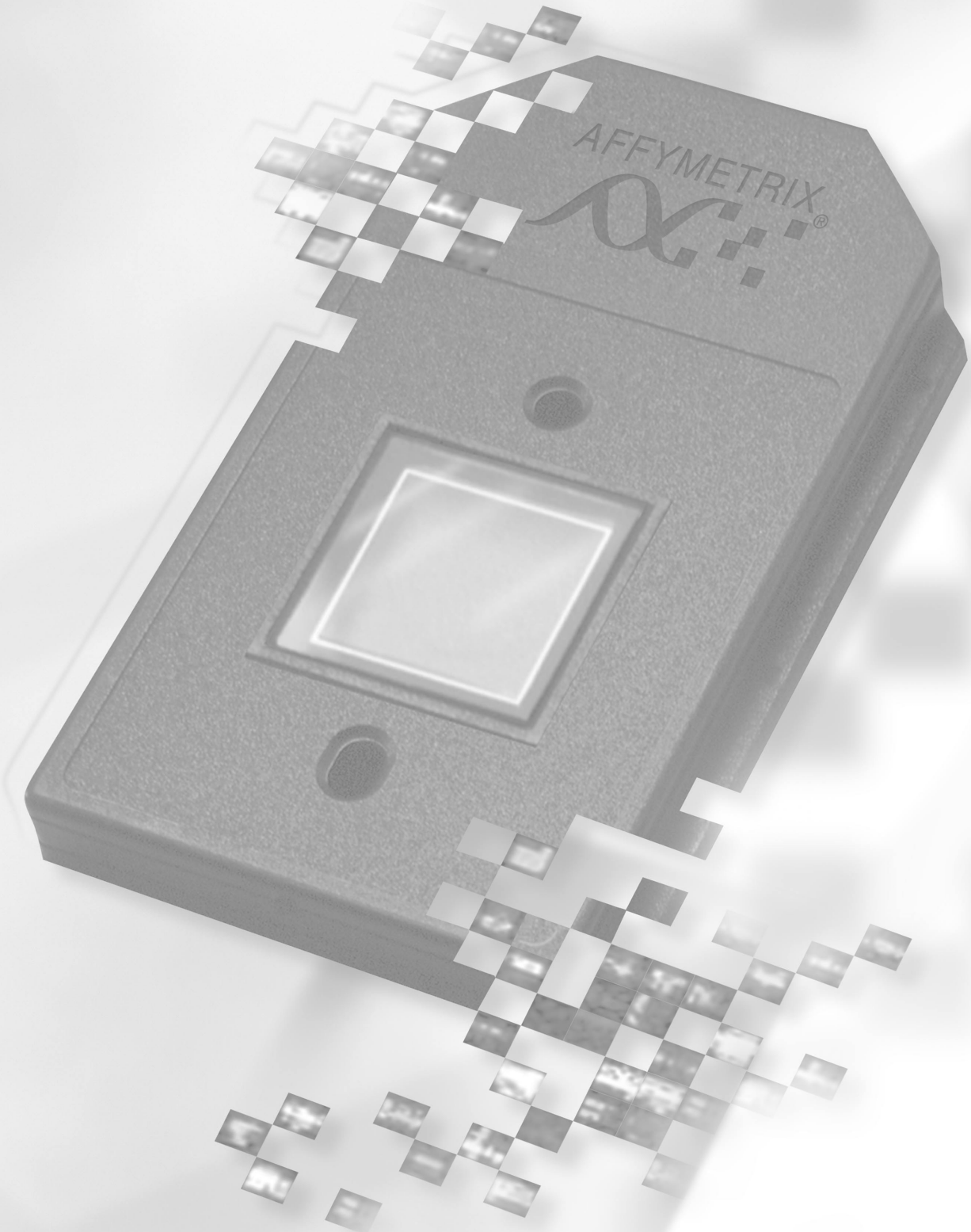


Figure 2.3.1
GeneChip® Probe Array



Section 2, Chapter 4

Section 2, Chapter 4



Eukaryotic Arrays: Washing, Staining, and Scanning

This Chapter Contains:

- Instructions for using the Fluidics Station 400 to automate the washing and staining of eukaryotic GeneChip® expression probe arrays.
- Instructions for scanning probe arrays using the GeneArray® Scanner.

After completing the procedures described in this chapter, the scanned probe array image (.dat file) is ready for analysis, as explained in the enclosed *GeneChip Expression Analysis: Data Analysis Fundamentals* booklet (P/N 701190).

Reagents and Materials Required

The following reagents and materials are recommendations and have been tested and evaluated by Affymetrix scientists. For supplier phone numbers in the U.S. and Europe, please refer to the Supplier and Reagent Reference List, Appendix A, of this manual. Information and part numbers listed are based on U.S. catalog information. Additional reagents needed for the complete analysis are listed in the appropriate chapters. Appendix A contains a master list of all reagents used in this manual.

- Water, Molecular Biology Grade, BioWhittaker Molecular Applications / Cambrex, P/N 51200
- Distilled water, Invitrogen Life Technologies, P/N 15230147
- Acetylated Bovine Serum Albumin (BSA) solution (50 mg/mL), Invitrogen Life Technologies, P/N 15561-020
- R-Phycoerythrin Streptavidin, Molecular Probes, P/N S-866
- 5 M NaCl, RNase-free, DNase-free, Ambion, P/N 9760G
- PBS, pH 7.2, Invitrogen Life Technologies, P/N 20012-027
- 20X SSPE (3 M NaCl, 0.2 M NaH_2PO_4, 0.02 M EDTA), BioWhittaker Molecular Applications / Cambrex, P/N 51214
- Goat IgG, Reagent Grade, Sigma-Aldrich, P/N I 5256
- Anti-streptavidin antibody (goat), biotinylated, Vector Laboratories, P/N BA-0500
- 10% surfact-Amps20 (Tween-20), Pierce Chemical, P/N 28320
- Bleach (5.25% Sodium Hypochlorite), VWR Scientific, P/N 21899-504 (or equivalent)

Miscellaneous Supplies

- Sterile, RNase-free, microcentrifuge tubes, 1.5 mL, USA Scientific, P/N 1415-2600 (or equivalent)
- Micropipettors, (P-2, P-20, P-200, P-1000), Rainin Pipetman (or equivalent)
- Sterile-barrier pipette tips and non-barrier pipette tips
- Tygon Tubing, 0.04″ inner diameter, Cole-Palmer, P/N H-06418-04
- Water, Molecular Biology Grade, BioWhittaker Molecular Applications / Cambrex, P/N 51200

Reagent Preparation

Wash A: Non-Stringent Wash Buffer

(6X SSPE, 0.01% Tween 20)

For 1000 mL:
300 mL of 20X SSPE
1.0 mL of 10% Tween-20
699 mL of water
Filter through a 0.2 µm filter

Wash B: Stringent Wash Buffer

(100 mM MES, 0.1 M [Na^+], 0.01% Tween 20)

For 1000 mL:
83.3 mL of 12X MES Stock Buffer (see Section 2, Chapter 3 for reagent preparation)
5.2 mL of 5 M NaCl
1.0 mL of 10% Tween 20
910.5 mL of water
Filter through a 0.2 µm filter
Store at 2°C to 8°C and shield from light

2X Stain Buffer

(Final 1X concentration: 100 mM MES, 1 M [Na^+], 0.05% Tween 20)

For 250 mL:
41.7 mL 12X MES Stock Buffer (see Section 2, Chapter 3 for reagent preparation)
92.5 mL 5 M NaCl
2.5 mL 10% Tween 20
113.3 mL water
Filter through a 0.2 µm filter
Store at 2°C to 8°C and shield from light

10 mg/mL Goat IgG Stock

Resuspend 50 mg in 5 mL 150 mM NaCl

Store at 4°C

✔ Note

If a larger volume of the 10 mg/mL IgG stock is prepared, aliquot and store at -20°C until use. After the solution has been thawed it should be stored at 4°C. Avoid additional freezing and thawing.

Experiment and Fluidics Station Setup

Step 1: Defining File Locations

Before working with Microarray Suite it is important to define where the program stores and looks for files.

1. Launch Microarray Suite from the workstation and select **Tools** → **Defaults** → **File Locations** from the menu bar.
2. The File Locations window displays the locations of the following files:
   - Probe Information (library files, mask files)
   - Fluidics Protocols (fluidics station scripts)
   - Experiment Data (.exp, .dat, .cel, and .chp files are all saved to location selected here)
3. Verify that all three file locations are set correctly and click **OK**.

   Contact Affymetrix Technical Support if you have any questions regarding this procedure.

Step 2: Entering Experiment Information

To wash, stain and scan a probe array, an experiment must first be defined in Microarray Suite.

1. Select **Run** → **Experiment Info** from the menu bar. Alternatively, click the New Experiment icon on the tool bar.

   ⇒ The Experiment Information dialog box appears allowing the experiment name to be defined along with several other parameters, such as probe array type, sample description, and comments.
2. Type in the **Experiment Name**.
3. In the **Probe Array Type** box, click the arrow and select the probe array type from the drop-down list.

   Experiment name and probe array type are required. Complete as much of the other information as desired. The protocol information at the bottom of the dialog box is exported to the experiment information dialog box after the hybridization and scan are completed.
4. Save the experiment by selecting **Save**.

   The name of the experiment is used by Microarray Suite to access the probe array type and data for the sample while it is being processed. Data files generated for the sample are automatically labeled to correspond to the experiment name. Microarray Suite automatically fills in the **Protocol** section of this dialog box with information on array processing from the fluidics station.
5. Close the Experiment Information dialog box.

Step 3: Preparing the Fluidics Station

The Fluidics Station 400 is used to wash and stain the probe arrays. It is operated using Microarray Suite.

Setting Up the Fluidics Station

1. Turn on the Fluidics Station using the toggle switch on the lower left side of the machine.
2. Select **Run** → **Fluidics** from the menu bar.
   ⇒ The Fluidics Station dialog box appears with a drop-down list for selecting the experiment name for each of the fluidics station modules. A second drop-down list is accessed for choosing the Protocol for each of the four fluidics station modules.

✔ Note

Refer to the Fluidics Station 400 User's Guide for instructions on connecting and addressing multiple fluidics stations.

Priming the Fluidics Station

Priming ensures that the lines of the fluidics station are filled with the appropriate buffers and the fluidics station is ready for running fluidics station protocols.

Priming should be done:

- when the fluidics station is first started.
- when wash solutions are changed.
- before washing, if a shutdown has been performed.
- if the LCD window instructs the user to prime.

1. To prime the fluidics station, select **Protocol** in the Fluidics Station dialog box.
2. Choose **Prime** for the respective modules in the Protocol drop-down list.
3. Change the intake buffer reservoir A to **Non-Stringent Wash Buffer** and intake buffer reservoir B to **Stringent Wash Buffer**.
4. Click **Run** for each module to begin priming.

Probe Array Wash and Stain

Affymetrix offers two staining protocols: 1) the single stain protocol for eukaryotic targets (page 4.9), and 2) a signal amplification protocol for eukaryotic targets (page 4.12). Please use the *Antibody Amplification Washing and Staining Protocol* for all arrays with probe cells of 24 µm or smaller.

1. After 16 hours of hybridization, remove the hybridization cocktail from the probe array and set it aside in a microcentrifuge tube. Store on ice during the procedure or at -80°C for long-term storage.
2. Fill the probe array completely with the appropriate volume of Non-Stringent Wash Buffer, as given in Table 2.3.2 on page 2.3.8.

✔ Note

If necessary, at this point, the probe array can be stored at 4°C for up to 3 hours before proceeding with washing and staining. Equilibrate the probe array to room temperature before washing and staining.

Washing and Staining Procedure 1: Single Stain for Eukaryotic Targets

✔ Note

Volumes needed will be the same for all fluidics protocols. This procedure takes approximately 75 minutes to complete.

Preparing the SAPE Stain Solution

Streptavidin Phycoerythrin (SAPE) should be stored in the dark at 4°C, either foil-wrapped or kept in an amber tube. Remove SAPE from refrigerator and tap the tube to mix well before preparing stain solution. Do not freeze SAPE. Always prepare the SAPE stain solution immediately before use.

For each probe array to be stained, combine the following components in a microcentrifuge tube:

Table 2.4.1
SAPE Solution Mix

Components	Volume	Final Concentration
2X MES Stain Buffer	300.0 µL	1X
50 mg/mL acetylated BSA	24.0 µL	2 mg/mL
1 mg/mL Streptavidin Phycoerythrin (SAPE)	6.0 µL	10 µg/mL
DI H_20	270.0 µL	—
Total	**600 µL**	

Table 2.4.2
Fluidics Protocols - Single Stain for Eukaryotic Targets

	Standard Format EukGE-WS1	**Mini Format** Mini_euk1
Post Hyb Wash #1	10 cycles of 2 mixes/cycle with Wash Buffer A at 25°C	10 cycles of 2 mixes/cycle with Wash Buffer A at 25°C
Post Hyb Wash #2	4 cycles of 15 mixes/cycle with Wash Buffer B at 50°C	8 cycles of 15 mixes/cycle with Wash Buffer B at 50°C
Stain	Stain the probe array for 30 minutes in SAPE solution at 25°C	Stain the probe array for 10 minutes in SAPE solution at 25°C
Final Wash	10 cycles of 4 mixes/cycle with Wash Buffer A at 25°C. The holding temperature is 25°C	10 cycles of 4 mixes/cycle with Wash Buffer A at 30°C. The holding temperature is 25°C

- Wash Buffer A = non-stringent wash buffer
- Wash Buffer B = stringent wash buffer

Washing and Staining the Probe Array

1. In the Fluidics Station dialog box on the workstation, select the correct experiment name in the drop-down **Experiment** list. The probe array type will appear automatically.
2. In the **Protocol** drop-down list, select the specific single stain protocol to control the washing and staining of the probe array format being used: **Table 2.4.2**.
3. Choose **Run** in the Fluidics Station dialog box to begin the washing and staining. Follow the instructions on the LCD window on the fluidics station.

   If you are unfamiliar with inserting and removing probe arrays from the fluidics station modules, please refer to the *Fluidics Station 400 User's Guide*, *Fluidics Station 400 Video In-Service CD* (P/N 900374), or *Quick Reference Card* (P/N 08-0072).
4. Insert the appropriate probe array into the designated module of the fluidics station while the probe array lever is in the **EJECT** position. When finished, verify that the probe array lever is returned to the **ENGAGE** position.
5. Remove any microcentrifuge tube remaining in the sample holder of the fluidics station module(s) being used.
6. Place the microcentrifuge tube containing the SAPE stain solution into the sample holder, verifying that the metal sampling needle is in the tube with its tip near the bottom.

   ⇒ The Fluidics Station dialog box and the LCD window display the status of the washing and staining as they progress. When the wash is complete, the LCD window displays the message **EJECT CARTRIDGE**.
7. Remove microcentrifuge tube containing stain and replace with an empty microcentrifuge tube.
8. Remove the probe arrays from the fluidics station modules by first moving the probe array holder lever to the **EJECT** position.
9. Check the probe array window for large bubbles or air pockets.

- If bubbles are present, proceed to **Table 2.4.3**.
- If the probe array has no large bubbles, it is ready to scan on the GeneArray Scanner. **ENGAGE** wash block and proceed to *Probe Array Scan* on page 2.4.15.

If you do not scan the arrays right away, keep the probe arrays at 4°C and in the dark until ready for scanning.

If there are no more samples to hybridize, shut down the fluidics station following the procedure outlined in the section, *Shutting Down the Fluidics Station* on page 2.4.16.

✔ Note

For proper cleaning and maintenance of the fluidics station including the bleach protocol, refer to Section 4, Fluidics Station Maintenance Procedures.

Table 2.4.3
If Bubbles are Present

Return the probe array to the probe array holder. Latch the probe array holder by gently pushing it up until a light click is heard. Engage the wash block by firmly pushing up on the probe array lever to the **ENGAGE** position. The fluidics station will drain the probe array and then fill it with a fresh volume of the last wash buffer used. When it is finished, if the LCD window displays **EJECT CARTRIDGE** again, remove the probe array and inspect it again for bubbles. If no bubbles are present, it is ready to scan. Proceed to *Probe Array Scan* on page 2.4.15. If several attempts to fill the probe array without bubbles are unsuccessful, the array should be filled with **Wash A (non-stringent buffer)** manually, using a micropipette. Excessive washing will result in a loss of signal intensity.

Washing and Staining Procedure 2: Antibody Amplification for Eukaryotic Targets

This protocol is recommended for use with probe arrays with probe cells of 24 µm or smaller. This procedure takes approximately 90 minutes to complete.

Preparing the Staining Reagents

Prepare the following reagents. Volumes given are sufficient for one probe array.

SAPE Stain Solution

Streptavidin Phycoerythrin (SAPE) should be stored in the dark at 4°C, either foil-wrapped or kept in an amber tube. Remove SAPE from refrigerator and tap the tube to mix well before preparing stain solution. Do not freeze concentrated SAPE or diluted SAPE stain solution. Always prepare the SAPE stain solution immediately before use.

Table 2.4.4
SAPE Solution Mix

Components	Volume	Final Concentration
2X MES Stain Buffer	600.0 µL	1X
50 mg/mL acetylated BSA	48.0 µL	2 mg/mL
1 mg/mL Streptavidin Phycoerythrin (SAPE)	12.0 µL	10 µg/mL
DI H_20	540.0 µL	—
Total	**1200 µL**	

Mix well and divide into two aliquots of 600 µL each to be used for stains 1 and 3, respectively.

Antibody Solution

Table 2.4.5
Antibody Solution Mix

Components	Volume	Final Concentration
2X MES Stain Buffer	300.0 µL	1X
50 mg/mL acetylated BSA	24.0 µL	2 mg/mL
10 mg/mL Normal Goat IgG	6.0 µL	0.1 mg/mL
0.5 mg/mL biotinylated antibody	3.6 µL	3 µg/mL
DI H_20	266.4 µL	—
Total	**600 µL**	

Table 2.4.6
Fluidics Protocols - Antibody Amplification for Eukaryotic Targets

	Standard Format EukGE-WS2	**Midi Format** Midi_euk2	**Micro / Mini Format** Micro_1v1 / Mini_euk2
Post Hyb Wash #1	10 cycles of 2 mixes/cycle with Wash Buffer A at 25°C	10 cycles of 2 mixes/cycle with Wash Buffer A at 30°C	10 cycles of 2 mixes/cycle with Wash Buffer A at 25°C
Post Hyb Wash #2	4 cycles of 15 mixes/cycle with Wash Buffer B at 50°C	6 cycles of 15 mixes/cycle with Wash Buffer B at 50°C	8 cycles of 15 mixes/cycle with Wash Buffer B at 50°C
Stain	Stain the probe array for 10 minutes in SAPE solution at 25°C	Stain the probe array for 5 minutes in SAPE solution at 35°C	Stain the probe array for 10 minutes in SAPE solution at 25°C
Post Stain Wash	10 cycles of 4 mixes/cycle with Wash Buffer A at 25°C	10 cycles of 4 mixes/cycle with Wash Buffer A at 30°C	10 cycles of 4 mixes/cycle with Wash Buffer A at 30°C
2nd Stain	Stain the probe array for 10 minutes in antibody solution at 25°C	Stain the probe array for 5 minutes in antibody solution at 35°C	Stain the probe array for 10 minutes in antibody solution at 25°C
3rd Stain	Stain the probe array for 10 minutes in SAPE solution at 25°C	Stain the probe array for 5 minutes in SAPE solution at 35°C	Stain the probe array for 10 minutes in SAPE solution at 25°C
Final Wash	15 cycles of 4 mixes/cycle with Wash Buffer A at 30°C The holding temperature is 25°C	15 cycles of 4 mixes/cycle with Wash Buffer A at 35°C The holding temperature is 25°C	15 cycles of 4 mixes/cycle with Wash Buffer A at 35°C The holding temperature is 25°C

- Wash Buffer A = non-stringent wash buffer
- Wash Buffer B = stringent wash buffer

Washing and Staining the Probe Array

1. In the Fluidics Station dialog box on the workstation, select the correct experiment name from the drop-down **Experiment** list.

   ⇒ The **Probe Array Type** appears automatically.

2. In the **Protocol** drop-down list, select the appropriate antibody amplification protocol to control the washing and staining of the probe array format being used.

✔ Note

Three-stain protocols require the user to replace stain solutions as directed by the LCD window during staining steps.

3. Choose **Run** in the Fluidics Station dialog box to begin the washing and staining. Follow the instructions in the LCD window on the fluidics station.

   If you are unfamiliar with inserting and removing probe arrays from the fluidics station modules, please refer to the *Fluidics Station 400 User's Guide*, *Fluidics Station 400 Video In-Service CD (P/N 900374),* or *Quick Reference Card (P/N 08-0072).*

4. Insert the appropriate probe array into the designated module of the fluidics station while the probe array lever is in the **EJECT** position. When finished, verify that the probe array lever is returned to the **ENGAGE** position.

5. Remove any microcentrifuge tube remaining in the sample holder of the fluidics station module(s) being used.

6. Place the microcentrifuge tube containing the SAPE stain solution into the sample holder, verifying that the metal sampling needle is in the tube with its tip near the bottom.

   ⇒ The Fluidics Station dialog box and the LCD window display the status of the washing and staining as they progress. When the wash is complete, the LCD window displays the message **EJECT CARTRIDGE**.

7. Remove microcentrifuge tube containing stain and replace with an empty microcentrifuge tube.

8. Remove the probe arrays from the fluidics station modules by first moving the probe array holder lever to the **EJECT** position.

9. Check the probe array window for large bubbles or air pockets.

   - If bubbles are present, proceed to **Table 2.4.3**.
   - If the probe array has no large bubbles, it is ready to scan on the GeneArray Scanner. **ENGAGE** wash block and proceed to *Probe Array Scan* on page 2.4.15.

   If you do not scan the arrays right away, keep the probe arrays at 4°C and in the dark until ready for scanning.

   If there are no more samples to hybridize, shut down the fluidics station following the procedure outlined in the section, *Shutting Down the Fluidics Station* on page 2.4.16.

✔ Note

For proper cleaning and maintenance of the fluidics station, including the bleach protocol, refer to Section 4, Fluidics Station Maintenance Procedures.

Table 2.4.7
If Bubbles are Present

Return the probe array to the probe array holder. Latch the probe array holder by gently pushing it up until a light click is heard. Engage the washblock by firmly pushing up on the probe array lever to the **ENGAGE** position. The fluidics station will drain the probe array and then fill it with a fresh volume of the last wash buffer used. When it is finished, if the LCD window displays **EJECT CARTRIDGE** again, remove the probe array and inspect it again for bubbles. If no bubbles are present, it is ready to scan. Proceed to *Probe Array Scan* on page 2.4.15. If several attempts to fill the probe array without bubbles are unsuccessful, the array should be filled with **Wash A (non-stringent buffer)** manually, using a micropipette. Excessive washing will result in a loss of signal intensity.

Probe Array Scan

The scanner is also controlled by Affymetrix Microarray Suite. The probe array is scanned after the wash protocols are complete. Make sure laser is warmed up prior to scanning by turning the laser on at least 15 minutes before use. If probe array was stored at 4°C, warm to room temperature before scanning. Refer to the Microarray Suite online help and the appropriate scanner user's manual for more information on scanning.

If necessary, clean the glass surface of probe array with a non-abrasive towel or tissue before scanning. **Do not use alcohol to clean glass**.

✔ Note

The scanner uses an argon-ion laser and is equipped with a safety interlock system. Defeating the interlock system may result in exposure to hazardous laser light.

1. Click **Run** → **Scanner** from the menu bar. Alternatively, click the Start Scan icon in the tool bar.
   ⇒ The Scanner dialog box appears with a drop-down list of experiments that have not been run.
2. Select the experiment name that corresponds to the probe array to be scanned.
   A previously run experiment can also be selected by using the **Include Scanned Experiments** option box. After selecting this option, previously scanned experiments appear in the drop-down list.
3. By default, after selecting the experiment the number [**2**] is displayed in the **Number of Scans** box to perform the recommended 2X image scan.
4. Once the experiment has been selected, click the **Start** button.
   ⇒ A dialog box prompts you to load a sample into the scanner.
5. Click the **Options** button to check for the correct pixel value and wavelength of the laser beam.
   For Probe Arrays with Probe Cells 24 µm or Less
   - Pixel value = 3 µm
   - Wavelength = 570 nm

   For a 50 µm Probe Array with a Phycoerythrin Stain
   - Pixel value = 6 µm
   - Wavelength = 570 nm
6. Open the sample door on the scanner and insert the probe array into the holder. Do not force the probe array into the holder. Close the sample door of the scanner.
7. Click **OK** in the Start Scanner dialog box.
   ⇒ The scanner begins scanning the probe array and acquiring data. When **Scan in Progress** is selected from the **View** menu, the probe array image appears on the screen as the scan progresses.

Shutting Down the Fluidics Station

1. After removing a probe array from the probe array holder, the LCD window displays the message **ENGAGE WASHBLOCK**.
2. Engage the washblock by firmly pushing up on the probe array lever to the **ENGAGE** position.
   ⇒ The fluidics station automatically performs a Cleanout procedure. The LCD window indicates the progress of the Cleanout procedure**.**
3. When the fluidics station LCD window indicates **REMOVE VIAL**, the Cleanout procedure is complete.
4. Remove the sample microcentrifuge tube from the sample holder.
5. If no other hybridizations are to be performed, place wash lines into a bottle filled with deionized water.
6. Choose **Shutdown** for all modules from the drop-down **Protocol** list in the Fluidics Station dialog box. Click the **Run** button for all modules.

   The Shutdown protocol is critical to instrument reliability. Refer to the *Fluidics Station 400 User's Guide* for more information.
7. After Shutdown protocol is complete, flip the ON/OFF switch of the fluidics station to the OFF position.

IMPORTANT

To maintain the cleanliness of the fluidics station and obtain the highest quality image and data possible, a weekly bleach protocol and a monthly decontamination protocol are highly recommended. Please refer to Section 4, Fluidics Station Maintenance Procedures for further detail.

Customizing the Protocol

There may be times when the fluidics protocols need to be modified. Modification of protocols must be done before downloading the protocol to the fluidics station. Protocol changes will not affect runs in progress. For more specific instructions, refer to the Microarray Suite online help.

1. Select **Tools** → **Edit Protocol** from the menu bar.
   ⇒ The Edit Protocol dialog box appears.
2. Select the protocol to be changed from the **Protocol Name** drop-down list.
   ⇒ The name of the protocol is displayed in the Protocol Name box. The conditions for that protocol are displayed on the right side of the Edit Protocol dialog box.
3. Select the item to be changed and input the new parameters as needed, keeping the parameters within the ranges shown below in **Table 2.4.8**.

Table 2.4.8
Valid Ranges for Wash/Stain Parameters

Parameter	Valid Range
Wash Temperature for A1, B, A2, or A3 (°C)	15 to 50
Number of Wash Cycles for A1, B, A2, or A3	0 to 99
Mixes / Wash cycle for A1, B, A2, or A3	15 to 50
Stain Time (seconds)	0 to 86399
Stain Temperature (°C)	15 to 50
Holding Temperature (°C)	15 to 50

- Wash A1 corresponds to Post Hyb wash #1 in Tables 2.4.2 and 2.4.6.
- Wash B corresponds to Post Hyb wash #2 in Tables 2.4.2 and 2.4.6.
- Wash A2 corresponds to Post Stain Wash in Tables 2.4.2 and 2.4.6.
- Wash A3 corresponds to Final Wash in Tables 2.4.2 and 2.4.6.

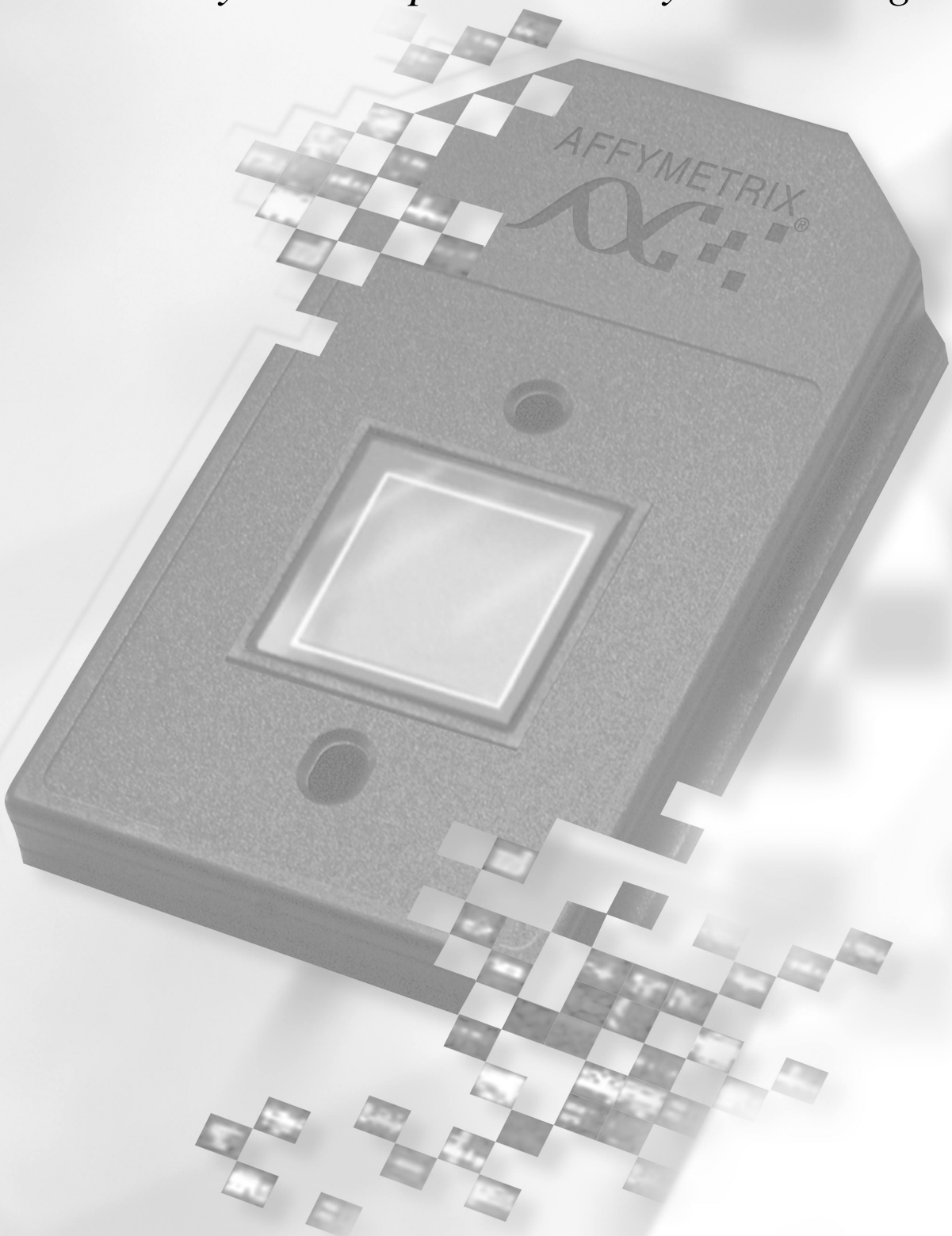
4. To return to the default values for the protocol selected, click the **Defaults** button.
5. After all the protocol conditions are modified as desired, change the name of the edited protocol in the **Protocol Name** box.

! CAUTION *If the protocol is saved without entering a new **Protocol Name**, the original protocol parameters will be overwritten.*

6. Click **Save**, then close the dialog box.
   Enter **0** (zero) for hybridization time if hybridization step is not required. Likewise, enter **0** (zero) for the stain time if staining is not required. Enter **0** (zero) for the number of wash cycles if a wash solution is not required.

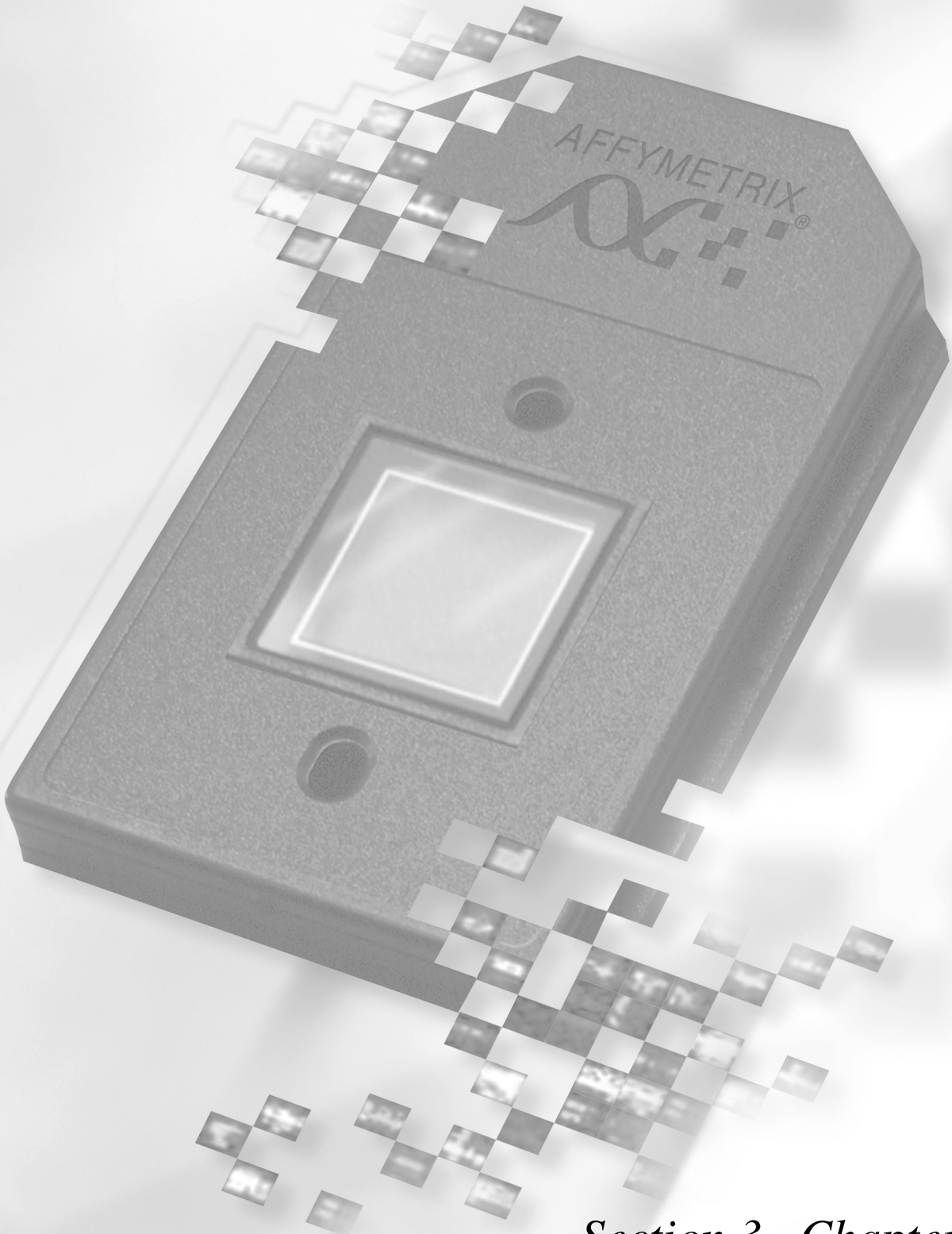
Section 3:

Prokaryotic Sample and Array Processing



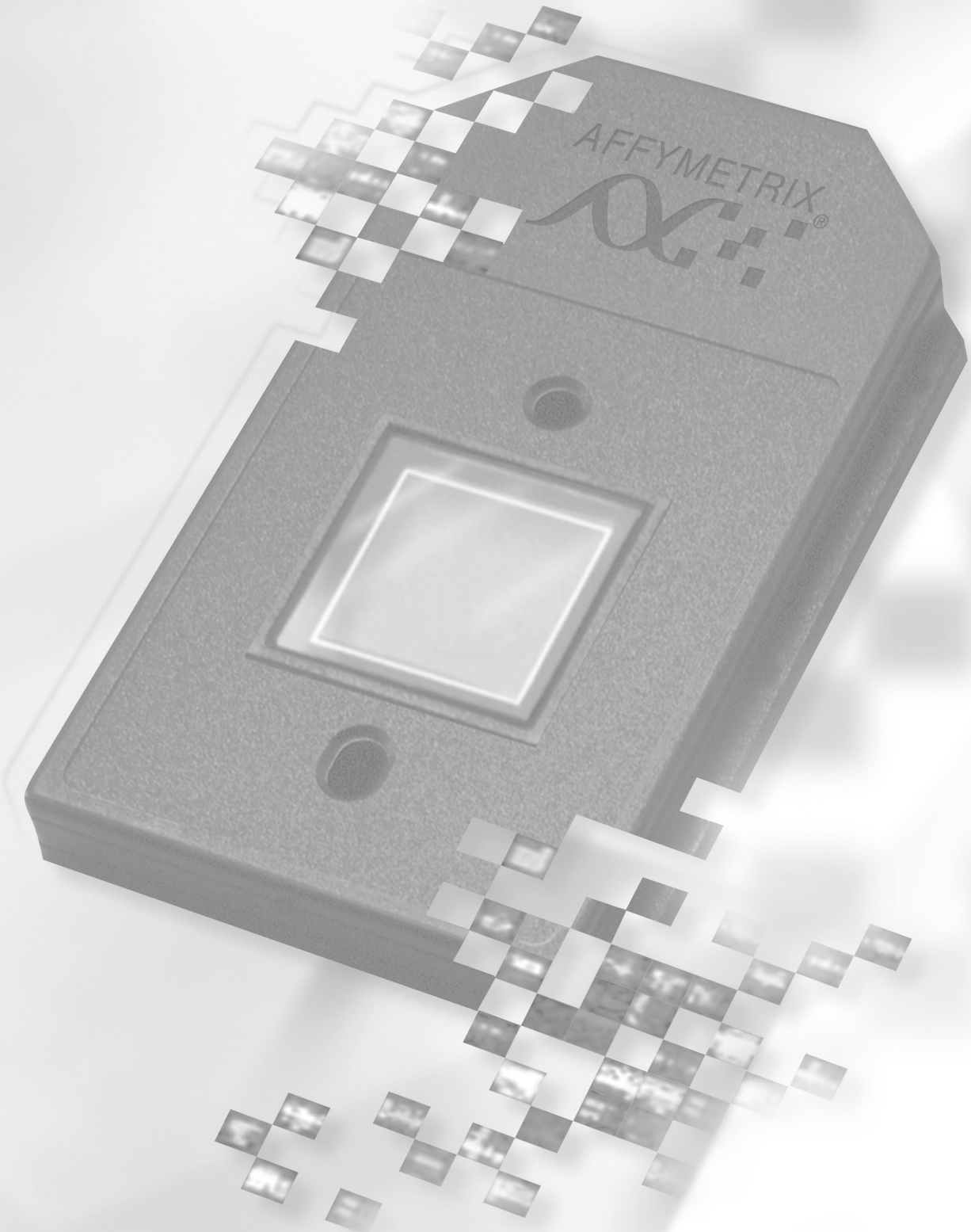
Contents

Section 3 *Prokaryotic Sample and Array Processing*



Section 3, Chapter 1

Section 3, Chapter 1



Direct Labeling of Enriched E. coli mRNA

This Chapter Contains:

- Instructions for enriching, fragmenting, and labeling RNA target from *E. coli* cells.

After completing the procedures described in this chapter, the labeled and fragmented target is hybridized to the GeneChip *E. coli* Genome Array, as described in Section 3, Chapter 3.

Introduction

This chapter describes the assay procedures recommended for use with GeneChip *E. coli* Genome Array. These procedures differ from those described under the previous section primarily because, unlike their eukaryotic counterpart, the majority of *E. coli* messenger RNAs do not contain poly-A tails. To enrich for mRNA species, a procedure designed to remove 16S and 23S ribosomal RNAs, which constitute approximately 90% of the total RNA population, is utilized. The resulting RNA population is then directly labeled with biotin.

mRNA Enrichment Procedure

The enrichment procedure is a series of enzymatic steps that specifically eliminate the 16S and 23S rRNA species in the total *E. coli* RNA. Reverse transcriptase and primers specific to 16S and 23S rRNA are used to synthesize complementary DNAs. Then rRNA is removed enzymatically by treatment with RNase H, which specifically digests RNA within an RNA:DNA hybrid. The cDNA molecules are then removed by DNase I digestion and the enriched mRNA is purified on QIAGEN RNeasy columns.

RNA Fragmentation and Labeling Procedure

The direct labeling of RNA consists of the following steps:

1. The RNA is fragmented by heat and ion-mediated hydrolysis.
2. The 5´-end RNA termini are enzymatically modified by T4 polynucleotide kinase with γ-S-ATP.
3. Biotin is conjugated to 5´-ends of the RNA. After purification of the product, the efficiency of the labeling procedure can be assessed using a gel-shift assay described in *Gel-Shift Assay* on page 3.1.14.

Reagents and Materials Required

The following reagents and materials are recommendations and have been tested and evaluated by Affymetrix scientists. For supplier phone numbers in the U.S. and Europe, please refer to the Supplier and Reagent Reference List, Appendix A, of this manual. Information and part numbers listed are based on U.S. catalog information. Additional reagents needed for the complete analysis are listed in the appropriate chapters. Appendix A contains a master list of all reagents used in this manual.

Total RNA Isolation

- MasterPure™ RNA Purification Kit, Epicentre Technologies, P/N MCR85102
- Isopropanol
- 75% Ethanol

mRNA Enrichment Procedure

- MMLV Reverse Transcriptase, New England BioLabs, P/N M0253L
- 10X MMLV Reverse Transcriptase Buffer, New England BioLabs, P/N M0253L (contains DTT and is also supplied with MMLV Reverse Transcriptase)
- 16S rRNA Primers

Name	Sequence
16S1514	5′-CCTAC GGTTA CCTTG TT-3′
16S889	5′TTAAC CTTGC GGCCG TACTC-3′
16S541	5′-TCCGA TTAAC GCTTG CACCC-3′

- 23s rRNA Primers

Name	Sequence
23S2878	5′-CCTCA CGGTT CATTA GT-3′
23SEco2064	5′-CTATA GTAAA GGTTC ACGGG-3′
23SEco1595	5′-CCTGT GTCGG TTTGG GGT-3′
23S1022	5′-TCCCA CATCG TTTCC CAC-3′
23S539	5′-CCATT ATACA AAAGG TAC-3′

✔ Note *All primers should be purchased in the 1 µmole amount and purified by HPLC. Oligonucleotides purchased from Operon have been used successfully in the procedure described in this chapter.*

- dATP, dCTP, dGTP, dTTP, Amersham Pharmacia Biotech, P/N 27-2035-01
- SUPERase•In™, Ambion, P/N 2696
- Ribonuclease H (RNase H), *E. coli*, Epicentre Technologies, P/N R0601K
- Deoxyribonuclease I (DNase I), Amersham Pharmacia Biotech, P/N 27-0514-01
- 0.5 M EDTA, pH 8.0, Invitrogen Life Technologies, P/N 15575-038
- RNeasy Mini Kit, QIAGEN, P/N 74104
- β-Mercaptoethanol
- 96% to 100% Ethanol
- Nuclease-free Water, Ambion, P/N 9930

RNA Fragmentation and Labeling Reaction

- T4 Polynucleotide Kinase and 10X NEBuffer, New England BioLabs, P/N 201L
- γ-S-ATP (20 µmoles), Roche Molecular Biochemical, P/N 1162306
- MOPS, Sigma-Aldrich, P/N M3183
- PEO-Iodoacetyl-Biotin (50 mg), Pierce Chemical, P/N 21334ZZ
- 3 M Sodium Acetate (NaOAc), pH 5.2, Sigma-Aldrich, P/N S 7899
- 96% to 100% Ethanol
- RNA/DNA Mini Column Kit, QIAGEN, P/N 14123
- β-Mercaptoethanol
- Isopropanol
- 70% Ethanol
- Glycogen (20 mg/mL), Roche Molecular Biochemical, P/N 901393

Gel-Shift Assay

- Novex XCell SureLock™ Mini-Cell, Invitrogen, P/N EI9001
- 4-20% TBE Gel, 1.0 mm, 12 well, Invitrogen, P/N EC62252
- Sucrose Gel Loading Dye, 5X, Amresco, P/N E-274
- 10X TBE Running Buffer
- SYBR Gold, Molecular Probes, P/N S-11494
- 10 bp and 100 bp DNA ladder, Invitrogen Life Technologies, P/N 10821-015 and 15628-019, respectively
- ImmunoPure NeutrAvidin, Pierce Chemical, P/N 31000ZZ
- 1M Tris pH 7.0, Ambion, P/N 9850G

Miscellaneous Supplies

- Hybridization Oven 640, Affymetrix, P/N 800139
- Sterile, RNase-free, microcentrifuge tubes, 1.5 mL, USA Scientific, P/N 1415-2600 (or equivalent)
- Micropipettors, (P-2, P-20, P-200, P-1000), Rainin Pipetman or equivalent
- Sterile-barrier pipette tips and non-barrier pipette tips
- Thermocycler
- Refrigerated microcentrifuge
- Spectrophotometer and quartz micro cuvettes
- Gel imaging system with appropriate filter for SYBR Green II or Gold.
- PCR tubes
- 0.2 µm filters for liquid sterilization
- Tough Spots, Label Dots, USA Scientific, P/N 9902 (optional)

Reagent Preparation

mRNA Enrichment Procedure

rRNA Removal Primer Stock

Individual primers are maintained at a stock concentration of 100 µM. Each primer is diluted 10-fold in the final primer mix to a concentration of 10 µM. Individual stocks and the primer mix are stored at -20°C.

25 mM dNTP Mix

Add 50 µL of each dNTP stock solution (100 mM each) to a sterile microcentrifuge tube. Mix thoroughly. Store at -20°C.

RNA Fragmentation and Labeling Reaction

500 mM MOPS, pH 7.5

For 100 mL:

10.5 g MOPS

80 mL Distilled water

Mix and adjust pH to 7.5 with 1 M NaOH. Add Distilled water to 100 mL. Filter through 0.2 µM filter. Store at room temperature.

Gel-Shift Assay

2 mg/mL NeutrAvidin

Resuspend 10 mg NeutrAvidin in 5 mL solution containing 50 mM Tris-Cl, pH 7.6, 200 mM NaCl. Store at 4°C.

Total RNA Isolation

Prior to the enrichment process, total RNA is isolated. For *E. coli*, Affymetrix has successfully used the Epicentre MasterPure RNA Purification Kit. Typical yields from a 10 mL log-phase culture range from 80 to 400 µg of total RNA. After purification, the RNA concentration is determined by absorbance at 260 nm on a spectrophotometer (1 absorbance unit = 40 µg/mL RNA). The A_{260}/A_{280} ratio should be approximately 2.0, with ranges between 1.8 to 2.1 considered acceptable. We recommend checking the quality of the RNA by running it on an agarose gel prior to starting the assay. The figure below shows 1 µg samples from three acceptable RNA preparations. The 23S and 16S rRNA bands should be clear without any obvious smearing patterns.

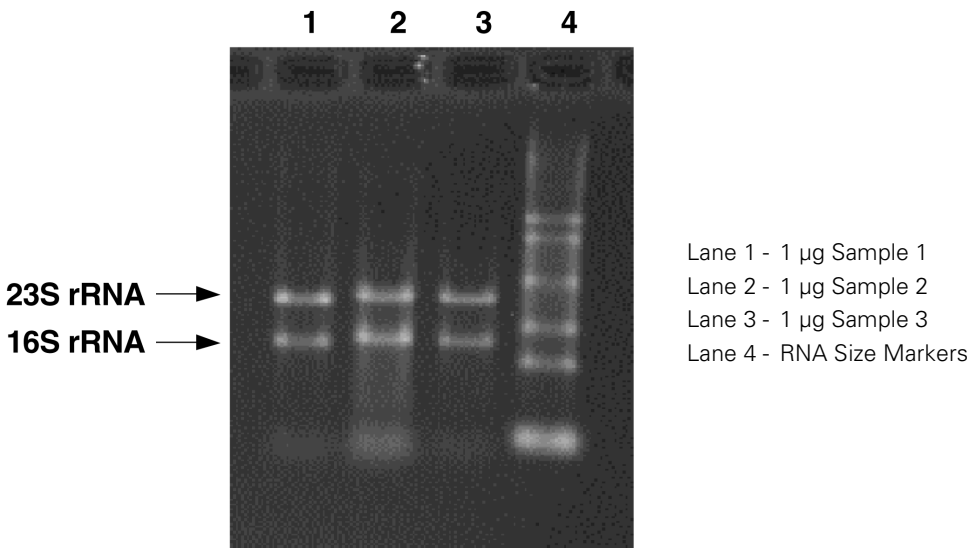


Figure 3.1.1
Typical RNA preparations from *E. coli*

mRNA Enrichment Procedure

The following protocol starts with 100 µg of total RNA. Incubations are performed in a thermocycler. Each sample is equally split between four 200-µL PCR tubes. It takes approximately 5 hours to obtain enriched mRNA.

✔ Note

The enrichment procedure involves three enzymatic steps, and the amount of enzymes used and incubation time are critical to the success of the assay. Please exercise precautions and follow standard laboratory procedures when handling RNA samples.

Step 1: cDNA Synthesis

1. Prepare the following mixture in each of the four PCR tubes.

✔ Note

Optionally, 1 µL of control transcript mix (lys, phe, dap, thr, and trp from B. subtilis) can be added to total RNA prior to the enrichment procedure for a final concentration of 10 pM of each control transcript to be applied to the arrays. Please refer to Section 3, Chapter 2 for the detailed protocol to prepare control sense RNA.

The detection limit of the assay is estimated to be around 5 pM. Assuming complete recovery of spike transcripts, the final hybridization mix contains each transcript at 10 pM concentration that is slightly above detection limit. Detection of these controls on the GeneChip E. coli Genome Array serves as indicators of the labeling efficiency. Alternatively, various control transcripts can also be spiked in at variable concentrations to demonstrate the dynamic range of the assay.

Table 3.1.1
Primer Mixture for mRNA Enrichment

Components	Volume or Amount	Final Concentration
Total RNA	25.0 µg	0.83 µg/µL
10 µM rRNA Removal Primer Stock	7 µL	1.75 µM
500 pM Control Transcript (optional)	1 µL	
Nuclease-free DI H_2O	Up to 40.0 µL	—
Total Volume Added	**40 µL**	

2. Heat RNA and primer mixture to 70°C for 5 minutes and then cool to 4°C.
3. Add the following components to each of the four tubes.

Table 3.1.2
Reverse Transcription Components

Components	Volume	Final Concentration
10X MMLV RT Buffer	10.0 µL	1X
25.0 mM dNTP Mix	2.0 µL	0.5 mM
20 U/µL SUPERase•In	3.0 µL	0.6 U/µL
200 U/µL MMLV RT	2.5 µL	5 U/µL
Nuclease-free H_2O	42.5 µL	—
Total Volume Added	**60 µL**	

4. Incubate the reaction at 42°C for 25 minutes, then at 45°C for 20 minutes. Cool to 4°C.
5. Inactivate the enzyme at 65°C for 5 minutes, then hold at 4°C. Proceed immediately to *Step 2: rRNA Digestion*, below.

Step 2: rRNA Digestion

1. Add the following to each of the four tubes.

Table 3.1.3
RNase Digestion of rRNAs

Components	Volume	Final Concentration
10.0 U/µL RNase H	4.0 µL	0.4 U/µL
20 U/µL SUPERase•In	3.0 µL	0.6 U/µL
Total Volume Added	**7 µL**	

2. Incubate the reaction at 37°C for 25 minutes, then hold at 4°C. Proceed immediately to *Step 3: cDNA Digestion* below.

Step 3: cDNA Digestion

1. Add the following to each of the four tubes.

Table 3.1.4
Digestion of cDNA

Components	Volume	Final Concentration
5.0 U/µL DNase I*	4.0 µL	0.2 U/µL

*Dilute 10 U/µL DNase I to 5 U/µL with nuclease-free H_2O.

Note *The concentration of DNase I is critical. Excessive digestion may cause degradation of mRNA. Lot-to-lot variation of enzyme activity may occur. Therefore, titrate DNase concentrations when necessary.*

2. Incubate the reaction at 37°C for 20 minutes.
3. Inactivate the enzyme by adding 3 µL of 500 mM EDTA to each tube for a final concentration of 10.0 mM.

IMPORTANT *The concentration and incubation time for DNase I are critical. Under-digestion is preferable to over-digestion.*

4. Pool reaction product from all four tubes and clean up the enriched mRNA with QIAGEN RNeasy Mini column.
5. Quantify the enriched mRNA preparation by 260 nm absorbance. Typical yields for the procedure are 20 to 40 µg of RNA (1.0 A_{260} unit = 40 µg/mL single strand RNA).

Note *The enriched mRNA is stored at -20°C until ready for use in the subsequent fragmentation and labeling reaction.*

RNA Fragmentation and Labeling Reaction

As in the previous procedure, the fragmentation and labeling reactions are done in PCR tubes in a thermocycler. A maximum of 20 µg of RNA per tube is used in the fragmentation step. Incomplete fragmentation may occur if excess RNA is used. Split samples into multiple tubes if the yield of RNA from the enrichment step is greater than 20 µg. Following the 5´ thiolation and biotin addition reaction, the target is purified with ethanol precipitation or RNA/DNA Mini Columns. The reactions take about 6 hours to complete.

✔ Note

The reaction of PEO-Iodoacetyl-Biotin is highly pH-dependent. MOPS is used as the buffer because of its inability to react with the iodoacetyl group under the reaction conditions. Other commonly used laboratory buffers containing amino groups such as Tris should not be used because of their reactivity with the iodoacetyl moiety.

Step 1: RNA Fragmentation

1. Prepare the following mixture.

Table 3.1.5
RNA Fragmentation Mix

Components	Volume or Amount	Final Concentration
10X NEBuffer for T4 Polynucleotide Kinase	10.0 µL	1.1X
enriched mRNA	up to 20.0 µg	—
DI H_2O	up to 88.0 µL total volume	—
Final Volume	**88 µL**	

2. Incubate the reaction at 95°C for 30 minutes.
3. Cool to 4°C.

Step 2: RNA 5´-Thiolation

1. Prepare the following mixture.

Table 3.1.6
RNA Thiolation Mix

Components	Volume	Final Concentration
Fragmented RNA (from Step 1)	88.0 µL	—
5 mM γ-S-ATP	2.0 µL	0.1 mM
10 U/µL T4 Polynucleotide Kinase	10.0 µL	1 U/µL
Final Volume	**100.0 µL**	

2. Incubate the reaction at 37°C for 50 minutes.
3. Inactivate the reaction by heating at 65°C for 10 minutes and then cool to 4°C.

4. Remove excess γ-S-ATP by ethanol precipitation. If you have started with multiple tubes, combine all samples in one sterile microcentrifuge tube. Add 1/10 volume of 3 M sodium acetate, pH 5.2, and 2.5 volumes of ethanol. Leave on ice for 15 minutes.
5. Spin at 14,000 rpm at 4°C for 30 minutes to pellet the RNA.
6. Resuspend the RNA pellet in 90 µL of DI H_2O.

Step 3: Biotin Addition

1. Prepare the following mixture.

Table 3.1.7
RNA Labeling Mix

Components	Volume	Final Concentration
500 mM MOPS, pH 7.5	6.0 µL	30 mM
Fragmented thiolated RNA (from Step 2)	90.0 µL	—
50 mM PEO-Iodoacetyl-Biotin	4.0 µL	2 mM
Final Volume	**100.0 µL**	

2. Incubate the reaction at 37°C for one hour.
3. Cool to 4°C.
4. Remove unincorporated biotin label using the QIAGEN RNA/DNA Mini Columns.

✔ Note

For increased RNA recovery, use one RNA/DNA column and 5.4 mL Buffer QRV2 for every 10.0 µg RNA. It is recommended to add 50 µg of glycogen as carrier during the precipitation step.

Alternatively, clean up the labeling product by ethanol precipitation instead of Qiagen columns (add 50 µg of glycogen as carrier, 1/10 volume of 3 M sodium acetate and 2.5 volume of ethanol to samples), followed by twice washing the pellets with 750 µL of 70% ethanol.

5. Dissolve the pelleted RNA in 20 to 30 µL of nuclease-free water.
6. Quantify product by 260 nm absorbance. Typical yields for the procedure are 2 to 4 µg of RNA. The labeled RNA should be stored at -20°C until ready for gel analysis or hybridization, as described in Section 3, Chapter 3.

✔ Note

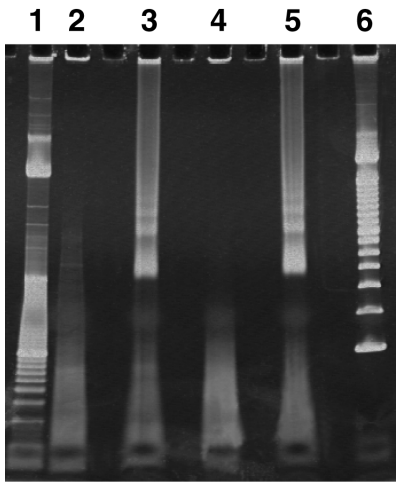
The process may be stopped during ethanol precipitation following the RNA/DNA column purification.

Gel-Shift Assay

After purification of the target, the efficiency of the labeling procedure can be assessed using the following procedure. This quality control protocol prevents hybridizing poorly labeled target onto the probe array. The addition of biotin residues is monitored in a gel-shift assay where the fragments are incubated with avidin prior to electrophoresis. Biotin-containing residues are retarded or shifted during the electrophoresis due to avidin binding. The nucleic acids are then detected by staining. Affymetrix routinely obtains approximately 70% labeling efficiency as shown in the gel photograph **(Figure 3.1.2)**. The procedure takes approximately 90 minutes to complete.

✔ Note

The absence of a shift pattern indicates poor biotin labeling. The problem should be addressed before proceeding to the hybridization step.



Lane 1 - 10 bp DNA Ladder
Lane 2 - Fragmented and labeled enriched *E. coli* RNA
Lane 3 - Fragmented and labeled enriched *E. coli* RNA with avidin
Lane 4 - Fragmented and labeled total *E. coli* RNA
Lane 5 - Fragmented and labeled total *E. coli* RNA with avidin
Lane 6 - 100 bp DNA Ladder

Figure 3.1.2
Gel-shift assay for monitoring E. coli target labeling efficiency

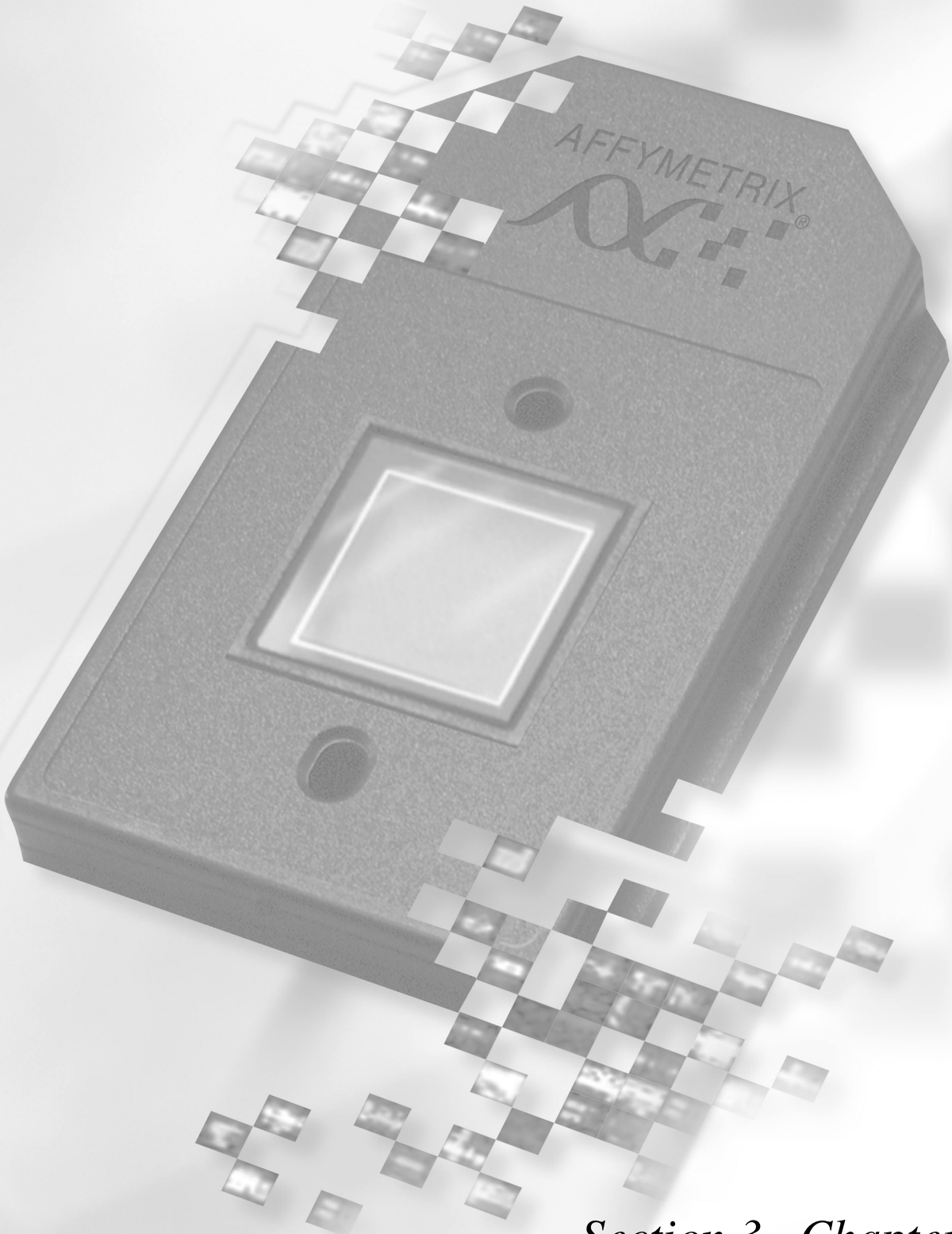
1. Prepare a NeutrAvidin solution of 2 mg/mL or higher. 50 mM Tris, pH 7.0 can be used to dilute the NeutrAvidin solution.
2. Place a 4-20% TBE gel into the gel holder and load system with 1X TBE Buffer.
3. For each sample to be tested, remove two 150 to 200 ng aliquots of fragmented and biotinylated sample to fresh tubes.
4. Add 5 µL of 2 mg/mL NeutrAvidin to each tube.
5. Mix and incubate at room temperature for 5 minutes.
6. Add loading dye to all samples to a final concentration of 1X loading dye.
7. Prepare 10 bp and 100 bp DNA ladders (1µL ladder + 7 µL water + 2 µL loading dye for each lane).
8. Carefully load samples and two ladders on gel. (Each gel well can hold a maximum of 20 µL.)

9. Run the gel at 150 volts until the front dye (red) almost reaches the bottom. The electrophoresis takes approximately 1 hour.
10. While gel is running, prepare at least 100 mL of a 1X solution of SYBR Green II or Gold for staining.

✔ Note

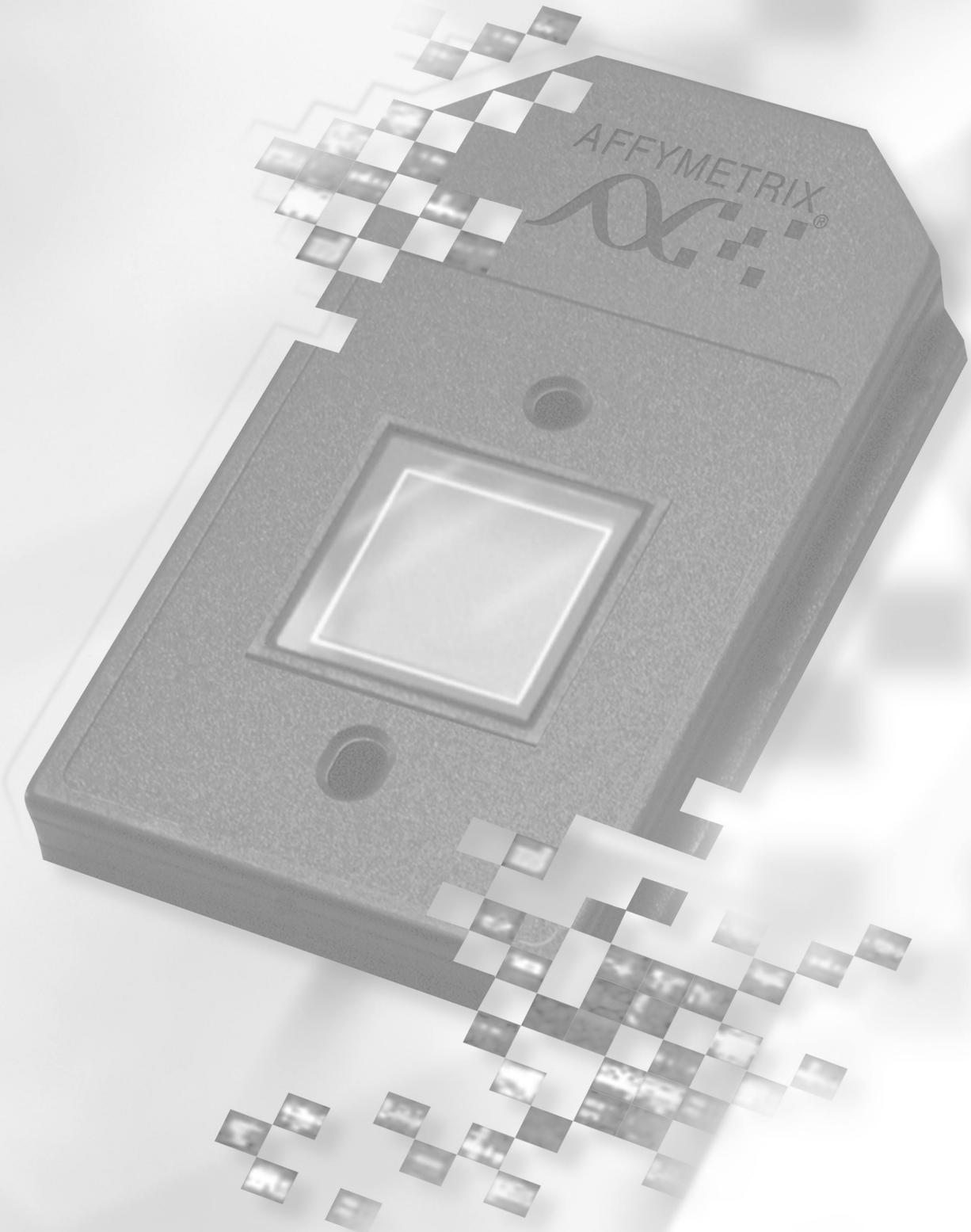
SYBR Green II and Gold are light sensitive. Therefore use caution and shield the staining solution from light. Prepare a new batch of stain at least once a week.

11. After the gel is complete, break open cartridge and stain the gel in 1X SYBR Green II or Gold for 10 minutes.
12. Place the gel on the UV light box and produce an image following standard procedure. Be sure to use the appropriate filter for SYBR Green II or Gold.



Section 3, Chapter 2

Section 3, Chapter 2



Preparation of Control Spike Transcripts for GeneChip® E. coli Genome Array

This Chapter Contains:

- Detailed steps for producing full-length control spike sense RNA.

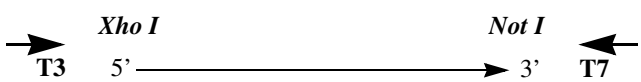
After completing the procedures described in this chapter, the control sense transcripts can be added to purified *E. coli* RNA samples prior to enrichment and labeling procedure as described in Section 3, Chapter 1.

Overview

This chapter describes protocols used to generate sense RNA controls from *B. subtilis* genes. These control transcripts can be spiked into *E. coli* total RNA used for target preparation at a predetermined concentration to monitor labeling, hybridization, and staining efficiency.

To be used as control for assay performance, the GeneChip® *E. coli* Genome Array contains probe sets with sequences of *dap, thr, phe, lys,* and *trp* genes from *B. subtilis*. These genes have been cloned into Stratagene pBluescript as an *Xho I* to *Not I* insert, 5´ to 3´, respectively (see Section 2, Chapter 2, *Controls for Eukaryotic Arrays*).

pGIBS-lys	ATCC 87482
pGIBS-phe	ATCC 87483
pGIBS-thr	ATCC 87484
pGIBS-trp	ATCC 87485
pGIBS-dap	ATCC 87486



These clones can be digested with the *Not I* restriction enzyme to produce linear template DNA for the subsequent *in vitro* transcription (IVT) to produce sense strand RNA by T3 RNA polymerase as control molecules.

Bacteria containing these recombinant plasmids can be obtained from the American Type Culture Collection (ATCC).

Reagents and Materials Required

The following reagents and materials are recommendations and have been tested and evaluated by Affymetrix scientists. For supplier phone numbers in the U.S. and Europe, please refer to the Supplier and Reagent Reference List, Appendix A, of this manual. Information and part numbers listed are based on U.S. catalog information. Additional reagents needed for the complete analysis are listed in the appropriate chapters. Appendix A contains a master list of all reagents used in this manual.

- Expression Control Clones, American Type Culture Collection (ATCC)
  - pGIBS-lys ATCC 87482
  - pGIBS-phe ATCC 87483
  - pGIBS-thr ATCC 87484
  - pGIBS-trp ATCC 87485
  - pGIBS-dap ATCC 87486
- *Not I* restriction Endonuclease, New England BioLabs, P/N R0189S
- Phase Lock Gel, Brinkmann Instruments, P/N 955 15 415
- Phenol/chloroform/isoamyl alcohol, Ambion, P/N 9732
- MEGAscript™ T3 Kit, Ambion, P/N 1338

Miscellaneous Reagents

- 3 M NaAcetate (NaOAc)
- Absolute Ethanol
- 80% Ethanol
- RNeasy Mini Kit, QIAGEN, P/N 74104

Bacterial Plasmid DNA Preparation

1. Grow *E. coli* bacterial cultures containing recombinant plasmids according to established protocols (a minimum 50 mL of culture volume is recommended).
2. Prepare plasmid DNA from overnight cultures using standard procedures or commercial kits.

   Affymetrix has obtained reliable results using QIAGEN Plasmid Kits for plasmid DNA isolation.

Linearization of Plasmid DNA Preparation

1. In a 50 µL reaction volume, digest 10 µg of plasmid with the restriction enzyme, *NotI*, according to the enzyme manufacturer's recommendations.
2. Analyze 50 ng of the uncut and linearized plasmid by gel electrophoresis on a 1% agarose gel. Complete digestion of the plasmid is required for IVT. Repeat restriction enzyme digestion, if necessary.

Purification of Linearized Plasmid DNA

Purify the linearized plasmid from restriction enzymes and potential RNase contaminants before proceeding to IVT using a Phase Lock Gel-phenol/chloroform extraction procedure.

Phase Lock Gels (PLG) form an inert, sealed barrier between the aqueous and organic phases of phenol-chloroform extractions. The solid barrier allows more complete recovery of the sample (aqueous phase) and minimizes interface contamination of the sample. PLG's are sold as premeasured aliquots in 1.5 mL tubes to which sample and phenol chloroform are directly added.

1. Pellet the Phase Lock Gel (1.5 mL tube with PLG I-heavy) in a microcentrifuge at ≥ 12,000 x g for 20 seconds.
2. Dilute the linearized plasmid to final volume of 150 µL with TE and add equal volume of (25:24:1) Phenol:chlororform:isoamyl alcohol (saturated with 10 mM Tris-HCl pH8.0/1 mM EDTA). Vortex.
3. Transfer the mix to the PLG tube and microcentrifuge at ≥ 12,000 x g for 2 minutes.
4. Transfer the top aqueous phase to a new 1.5 mL tube.
5. Add 0.1 volumes (15 µL) of 3 M NaOAc and 2.5 volumes (375 µL) of absolute ethanol to the samples. Vortex.
6. Immediately centrifuge at ≥ 12,000 x g in a microcentrifuge at room temperature for 20 minutes.
7. Carefully remove supernatent.
8. Wash pellet with 0.5 mL of 80% ethanol, then centrifuge at ≥ 12,000 x g at room temperature for 5 minutes.
9. Remove the supernatent very carefully and air dry the pellet.
10. Resuspend DNA pellet in 15 µL of RNase-free water.
11. Quantify the DNA by absorbance at 260 nm (50 µg/mL of DNA for 1 absorbance unit at 260 nm).

✔ Note

The quality of DNA template can be monitored by the A_{260}/A_{280} ratio, which should be between 1.8 and 2.0 for pure DNA.

In Vitro Transcription (IVT) to Produce Control Sense Transcripts

Use MEGAscript™ T3 High Yield Transcription Kit for the IVT reaction.

1. To make up the reaction mix, follow the procedures in the instruction manual provided by Ambion.

✔ Note — *No tracer is involved in this assay.*

2. Incubate the reaction for 4 hours at 37°C.
3. Cleanup the reaction product with RNeasy Mini column.
4. Quantify the transcript by absorbance at 260 nm (40 µg/mL RNA = 1 absorbance unit at 260 nm).

✔ Note — *It is recommended to examine the quality and integrity of the IVT product on an agarose gel.*

➲ IMPORTANT — *Aliquot and freeze the IVT transcripts at -80°C. Avoid repeated freeze / thaw cycles.*

Preparing the Control Transcript Mix

1. Prepare stock solutions for each of the five transcripts separately at 2.5 nM for each transcript.

   Use the following table to calculate the amount of transcript needed to prepare 2.5 nM stock.

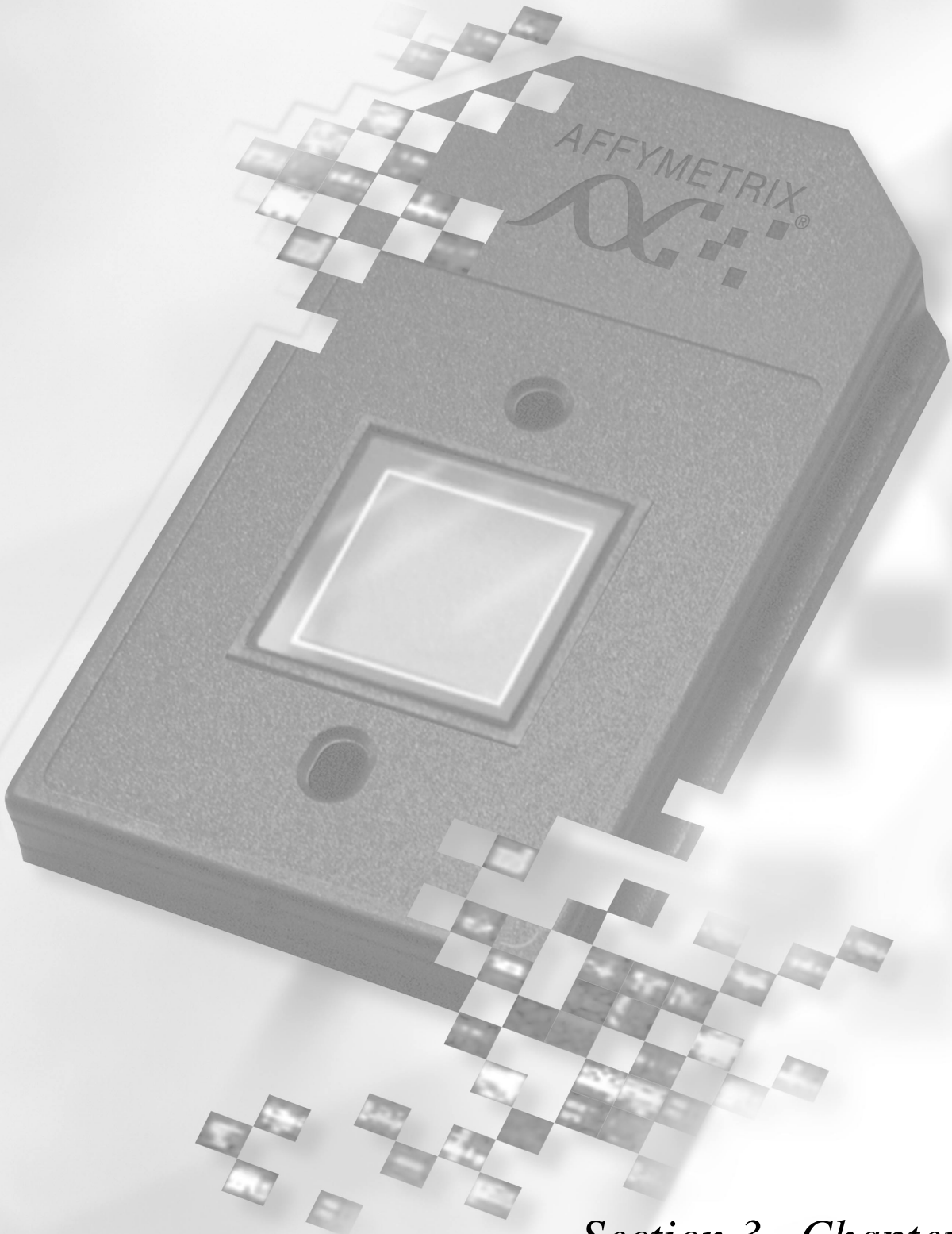
Table 3.2.1
Conversions for Preparing 2.5 nM Control Transcript Mix

Control RNA	Size (kb)	Molecular Weight	pMoles / µg
Lys	1	330,000	3.03
Phe	1.32	435,600	2.30
Dap	1.82	607,200	1.65
Thr	1.98	653,400	1.53
Trp	2.5 kb	825,000	1.21

2. Mix equal volume of all five transcripts so that final concentration for each is 500 pM.
3. Apply 4 µL of the transcript mix with each 100 µg of total RNA prior to the enrichment procedure as described in Chapter 1 of this section. Final concentration applied on the array for the control transcripts would be 10 pM, assuming 100% recovery.

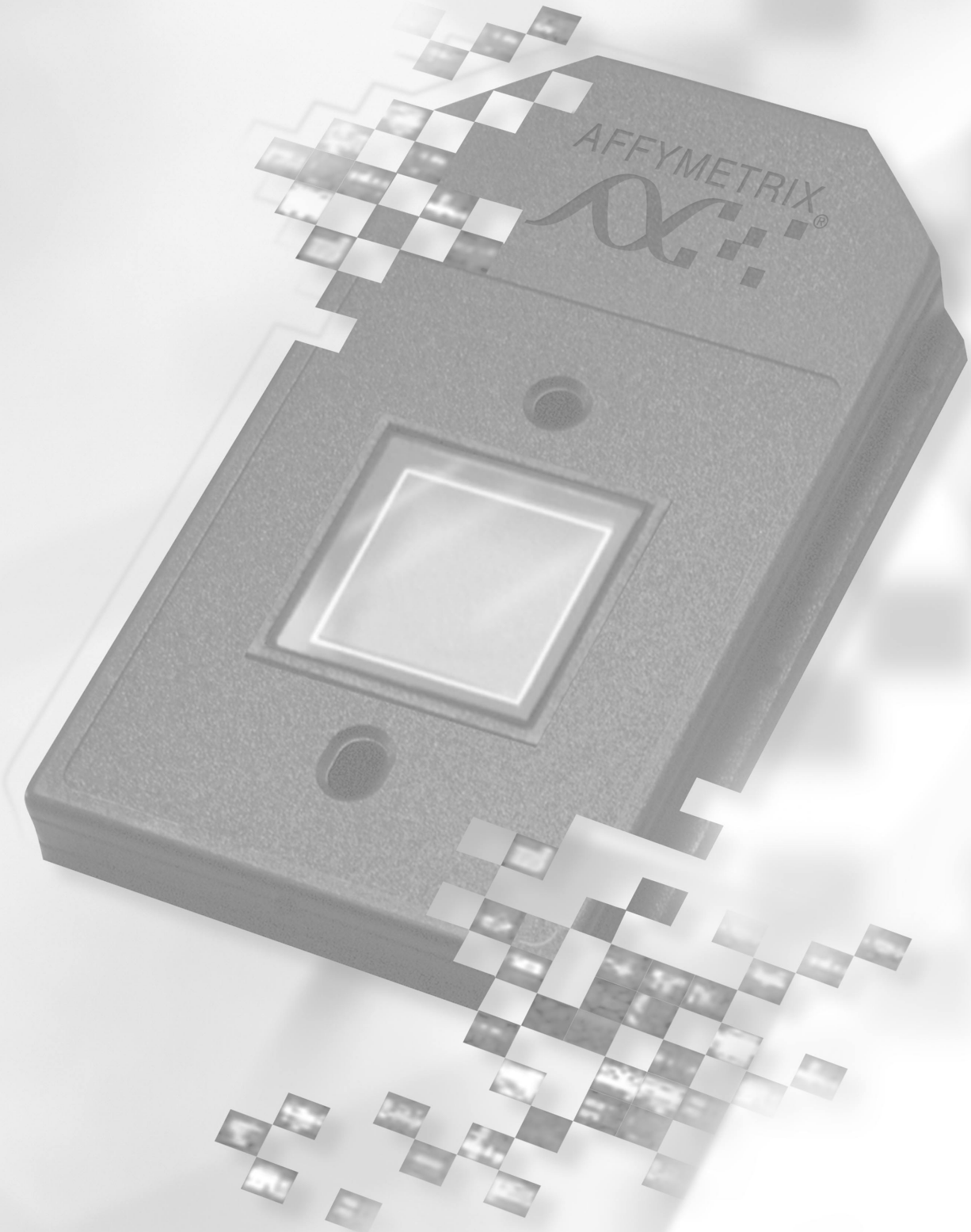
✔ Note — *Since the hybridization mix has a volume of 200 µL, 10 pM would be the final concentration for the transcripts assuming 100% recovery throughout the enrichment and labeling procedure. Alternatively, different concentrations of transcript stock can be prepared to generate "variable" concentrations for different transcripts to monitor the dynamic range of the assay.*

➲ IMPORTANT — *Aliquot and freeze the IVT transcripts at -80°C. Avoid repeated freeze / thaw cycles.*



Section 3, Chapter 3

Section 3, Chapter 3



E. coli Target Hybridization

This Chapter Contains:

- Detailed steps for preparing the hybridization mix containing labeled target.
- Instructions for hybridizing the target mix to a GeneChip *E.coli* Genome Array.

After completing the procedures described in this chapter, the hybridized probe array is then ready for washing, staining, and scanning as detailed in Section 3, Chapter 4.

Reagents and Materials Required

The following reagents and materials are recommendations and have been tested and evaluated by Affymetrix scientists. For supplier phone numbers in the U.S. and Europe, please refer to the Supplier and Reagent Reference List, Appendix A, of this manual. Information and part numbers listed are based on U.S. catalog information. Additional reagents needed for the complete analysis are listed in the appropriate chapters. Appendix A contains a master list of all reagents used in this manual.

- Water, Molecular Biology Grade, BioWhittaker Molecular Applications / Cambrex, P/N 51200
- Acetylated Bovine Serum Albumin (BSA) solution (50 mg/mL), Invitrogen Life Technologies, P/N 15561-020
- Herring Sperm DNA, Promega Corporation, P/N D1811
- Micropure Separator, Millipore, P/N 42512 (optional)
- Control Oligo B2, 3 nM, Affymetrix, P/N 900301 (can be ordered separately)
- 5 M NaCl, RNase-free, DNase-free, Ambion, P/N 9760G
- MES Free Acid Monohydrate SigmaUltra, Sigma-Aldrich, P/N M5287
- MES Sodium Salt, Sigma-Aldrich, P/N M5057
- EDTA Disodium Salt, 0.5 M solution (100 mL), Sigma-Aldrich, P/N E7889

Miscellaneous Reagents

- Tough Spots, Label Dots, USA Scientific, P/N 9185 (optional)
- Tween-20, 10%, Pierce Chemical, P/N 28320

Miscellaneous Supplies

- Hybridization Oven 640, Affymetrix, P/N 800139
- Sterile, RNase-free, microcentrifuge tubes, 1.5 mL, USA Scientific, P/N 1415-2600 (or equivalent)
- Micropipettors, (P-2, P-20, P-200, P-1000), Rainin Pipetman or equivalent
- Sterile-barrier pipette tips and non-barrier pipette tips

Reagent Preparation

12X MES Stock

(1.22 M MES, 0.89 M [Na^+])

For 1000 mL:

70.4 g MES-free acid monohydrate
193.3 g MES Sodium Salt
800 mL of Molecular Biology Grade water
Mix and adjust volume to 1000 mL.
The pH should be between 6.5 and 6.7. Filter through a 0.2 µm filter.

IMPORTANT *Do not autoclave, store at 2°C to 8°C, and shield from light. Discard solution if yellow.*

2X Hybridization Buffer

(Final 1X concentration is 100 mM MES, 1 M [Na^+], 20 mM EDTA, 0.01% Tween 20)

For 50 mL:

8.3 mL of 12X MES Stock
17.7 mL of 5 M NaCl
4.0 mL of 0.5 M EDTA
0.1 mL of 10% Tween 20
19.9 mL of water
Store at 2°C to 8°C, and shield from light

E. coli Target Hybridization

After determining that the fragmented RNA is labeled with biotin, prepare the hybridization solution mix. The recommended amount of RNA per assay is 1.5 to 4.0 µg. The solution is stable for approximately 6 to 8 hours at 4°C. The following protocol can be used for freshly prepared or frozen hybridization cocktail. Re-use of prokaryotic sample is not recommended, since the samples are end-labeled rather than internally labeled.

1. Prepare the following hybridization solution mix.

Table 3.3.1
Hybridization Solution Mix

Components	Volume	Final Concentration
2X MES Hybridization Buffer	100.0 µL	1X
3 nM Control Oligo B2	3.3 µL	50 pM
10 mg/mL Herring Sperm DNA	2.0 µL	0.1 mg/mL
50 mg/mL BSA	2.0 µL	0.5 mg/mL
Fragmented Labeled RNA	1.5 to 4.0 µg	—
Molecular Biology Grade Water	to a final volume of 200 µL	—
Final Volume	**200 µL**	

2. Equilibrate probe array to room temperature immediately before use.

✔ Note

It is important to allow the arrays to normalize to room temperature completely. Specifically, if the rubber septa are not equilibrated to room temperature, they may be prone to cracking which can lead to leaks.

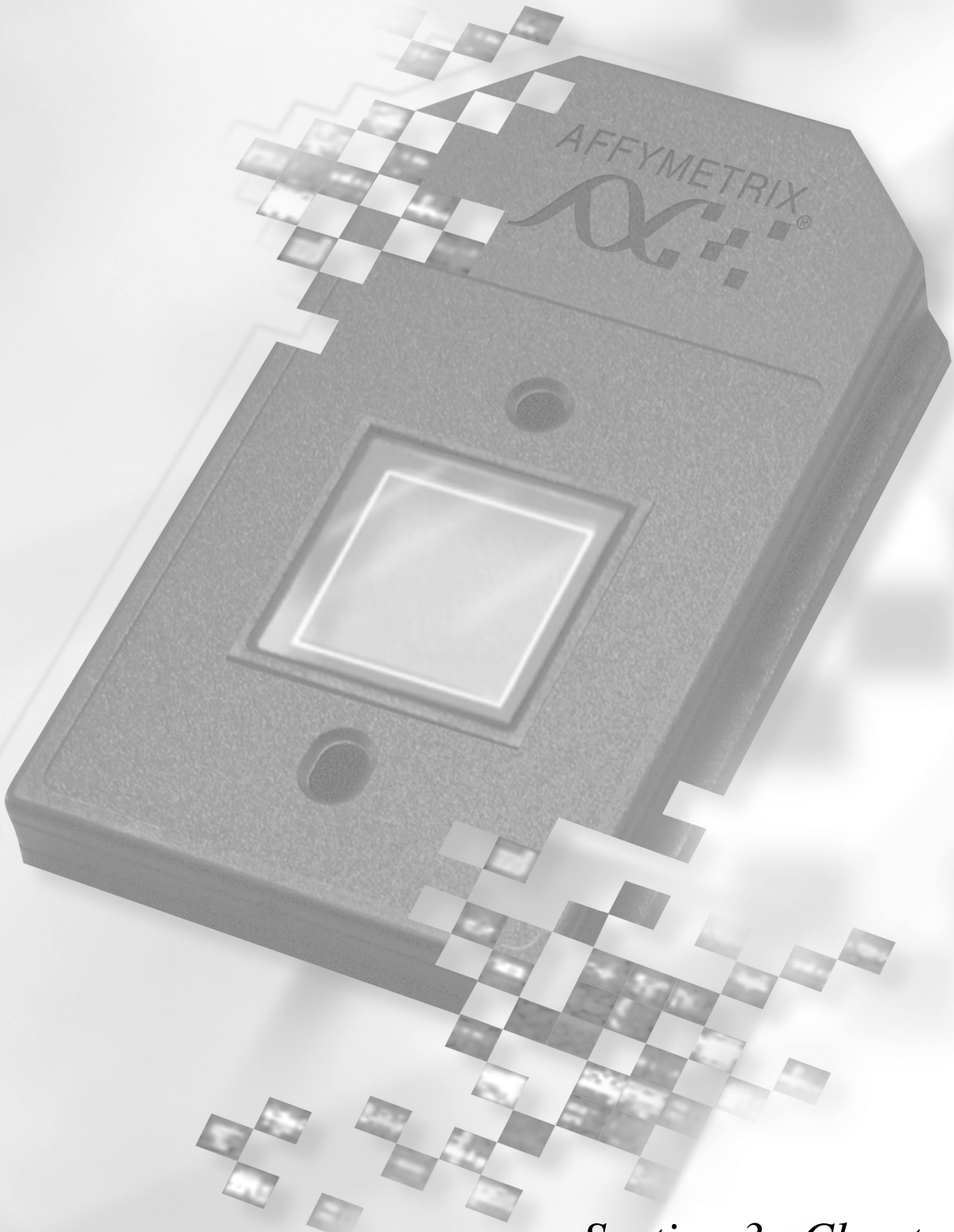
3. Add the hybridization solution mix to the probe array.

✔ Note

It is necessary to use two pipette tips when filling the probe array cartridge: one for filling, and the second to allow venting of air from the hybridization chamber. After the addition of hybridization cocktail, the septa may be covered with Tough Spots to prevent evaporation.

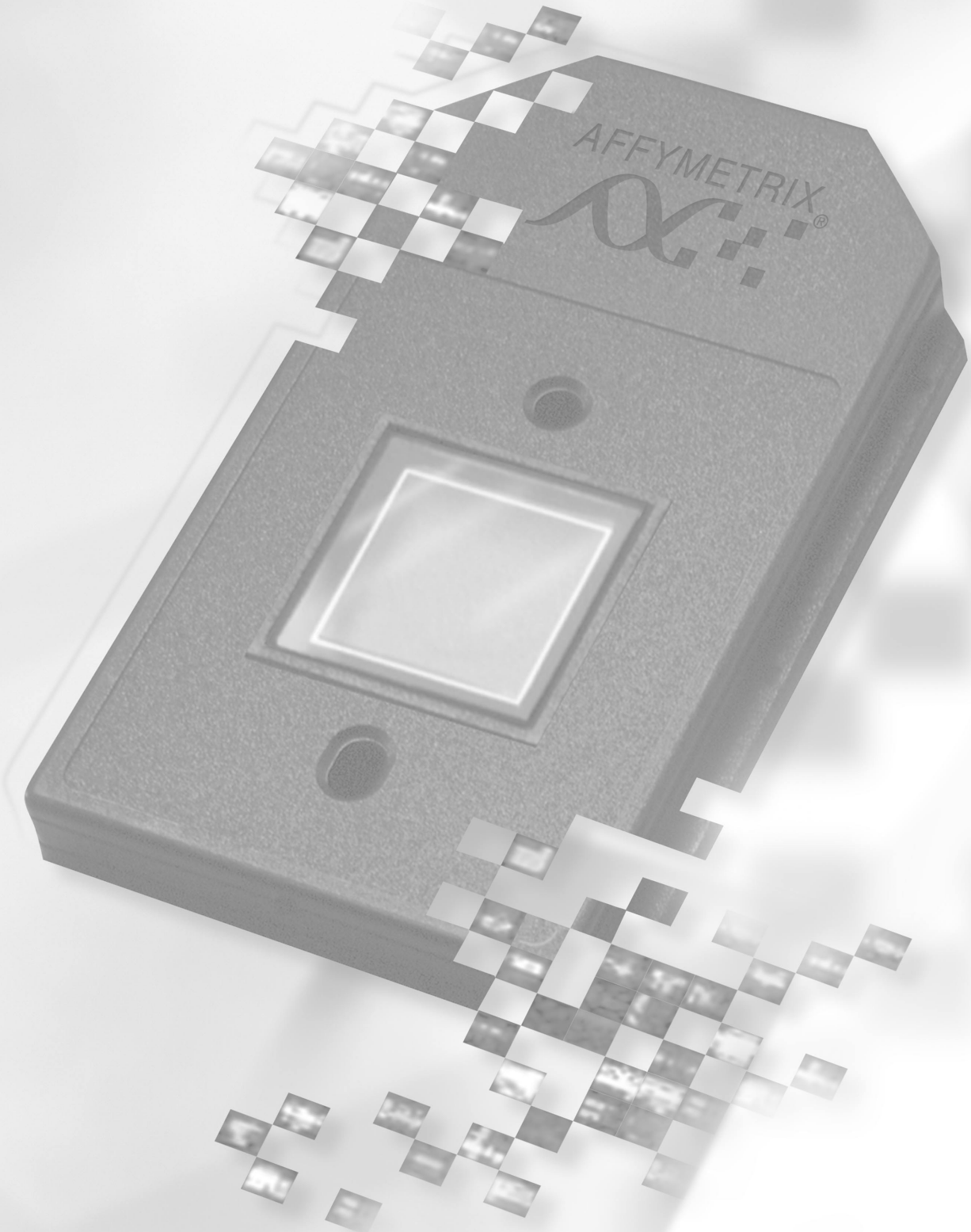
4. Place probe array in the hybridization oven set at 45°C.
5. Avoid stress to the motor; load probe arrays in a balanced configuration around rotisserie axis. Rotate at 60 rpm.
6. Hybridize for 16 hours.

   During the latter part of the 16-hour hybridization, proceed to Section 3, Chapter 4 to prepare reagents required immediately after completion of hybridization.



Section 3, Chapter 4

Section 3, Chapter 4



GeneChip E. coli Genome Array: Washing, Staining, and Scanning

This Chapter Contains:

- Instructions for using the Fluidics Station 400 to automate the washing and staining of GeneChip *E.coli* Genome Arrays.
- Instructions for scanning probe arrays using the GeneArray Scanner.

After completing the procedures described in this chapter, the scanned probe array image (.dat file) is ready for analysis, as explained in the enclosed *GeneChip Expression Analysis: Data Analysis Fundamentals* booklet (P/N 701190).

Reagents and Materials Required

The following reagents and materials are recommendations and have been tested and evaluated by Affymetrix scientists. For supplier phone numbers in the U.S. and Europe, please refer to the Supplier and Reagent Reference List, Appendix A, of this manual. Information and part numbers listed are based on U.S. catalog information. Additional reagents needed for the complete analysis are listed in the appropriate chapters. Appendix A contains a master list of all reagents used in this manual.

- Water, Molecular Biology Grade, BioWhittaker Molecular Applications / Cambrex, P/N 51200
- Distilled water, Invitrogen Life Technologies, P/N 15230-147
- Acetylated Bovine Serum Albumin (BSA) solution (50 mg/mL), Invitrogen Life Technologies, P/N 15561-020
- R-Phycoerythrin Streptavidin, Molecular Probes, P/N S-866
- 5 M NaCl RNase-free, DNase-free, Ambion, P/N 9760G
- PBS, pH 7.2, Invitrogen Life Technologies, P/N 20012-027
- 20X SSPE (3 M NaCl, 0.2 M NaH_2PO_4, 0.02 M EDTA), BioWhittaker Molecular Applications / Cambrex, P/N 51214
- Goat IgG, Reagent Grade, Sigma-Aldrich, P/N I 5256
- Anti-streptavidin antibody (goat), biotinylated, Vector Laboratories, P/N BA-0500
- 10% surfact-Amps20 (Tween-20), Pierce Chemical, P/N 28320
- Bleach (5.25% Sodium Hypochlorite), VWR Scientific, P/N 21899-504 (or equivalent)
- ImmunoPure Streptavidin, Pierce Chemical, P/N 21125

Miscellaneous Supplies

- Sterile, RNase-free, microcentrifuge tubes, 1.5 mL, USA Scientific, P/N 1415-2600 (or equivalent)
- Micropipettors, (P-2, P-20, P-200, P-1000), Rainin Pipetman (or equivalent)
- Sterile-barrier pipette tips and non-barrier pipette tips
- Tygon Tubing, 0.04″ inner diameter, Cole-Palmer, P/N H-06418-04

Reagent Preparation

Wash A: Non-Stringent Wash Buffer

(6X SSPE, 0.01% Tween 20)

For 1000 mL:
300 mL of 20X SSPE
1.0 mL of 10% Tween-20
699 mL of water
Filter through a 0.2 µm filter

Wash B: Stringent Wash Buffer

(100 mM MES, 0.1 M [Na^+], 0.01% Tween 20)

For 1000 mL:
83.3 mL of 12X MES Stock Buffer (see Section 3, Chapter 3 for reagent preparation)
5.2 mL of 5 M NaCl
1.0 mL of 10% Tween 20
910.5 mL of water
Filter through a 0.2 µm filter
Store at 2°C to 8°C and shield from light

2X Stain Buffer

(Final 1X concentration: 100 mM MES, 1 M [Na^+], 0.05% Tween 20)

For 250 mL:
41.7 mL 12X MES Stock Buffer (see Section 3, Chapter 3 for reagent preparation)
92.5 mL 5 M NaCl
2.5 mL 10% Tween 20
113.3 mL water
Filter through a 0.2 µm filter
Store at 2°C to 8°C and shield from light

10 mg/mL Goat IgG Stock

Resuspend 50 mg in 5 mL PBS

Store at 4°C

1 mg/mL Streptavidin Stock

Resuspend 5 mg in 5 mL of PBS

Store at 4°C

Experiment and Fluidics Station Setup

Step 1: Defining File Locations

Before working with Affymetrix Microarray Suite it is important to define where the program stores and looks for files.

1. Launch Microarray Suite from the workstation and select **Tools** → **Defaults** → **File Locations** from the menu bar.

   The File Locations window displays the locations of the following files:

   - Probe Information (library files, mask files)
   - Fluidics Protocols (fluidics station scripts)
   - Experiment Data (.exp, .dat, .cel, and .chp files are all saved to location selected here)

2. Verify that all three file locations are set correctly and click **OK**.

   Contact Affymetrix Technical Support if you have any questions regarding this procedure.

Step 2: Entering Experiment Information

To wash, stain, and scan a probe array, an experiment must first be defined in Microarray Suite.

1. Select **Run** → **Experiment Info** from the menu bar. Alternatively, click the New Experiment icon on the tool bar.

   ⇒ The Experiment Information dialog box appears allowing the experiment name to be defined along with several other parameters, such as probe array type, sample description, and comments.

2. Type in the **Experiment Name**.

3. In the **Probe Array Type** box, click the arrow and select **Ecoli** from the drop-down list.

   Experiment name and probe array type are required. Complete as much of the other information as desired. The protocol information at the bottom of the dialog box is exported to the experiment information dialog box after the hybridization and scan are complete.

4. Save the experiment by selecting **Save**.

   The name of the experiment is used by Microarray Suite to access the probe array type and data for the sample while it is being processed. Data files generated for the sample are automatically labeled to correspond to the experiment name. Microarray Suite automatically fills in the **Protocol** section of this dialog box with information on array processing from the fluidics station.

5. Close the Experiment Information dialog box.

Step 3: Preparing the Fluidics Station

The Fluidics Station 400 is used to wash and stain the probe arrays. It is operated using Microarray Suite.

Setting Up the Fluidics Station

1. Turn on the Fluidics Station using the toggle switch on the lower left side of the machine.
2. Select **Run** → **Fluidics** from the menu bar.
   ⇒ The Fluidics Station dialog box appears with a drop-down list for selecting the experiment name for each of the fluidics station modules. A second drop-down list is accessed for choosing the Protocol for each of the four fluidics station modules.

✔ Note

Refer to the Fluidics Station 400 User's Guide for instructions on connecting and addressing multiple fluidics stations.

Priming the Fluidics Station

Priming ensures that the lines of the fluidics station are filled with the appropriate buffers and the fluidics station is ready for running fluidics station protocols.

Priming should be done:

- when the fluidics station is first started.
- when wash solutions are changed.
- before washing, if a shutdown has been performed.
- if the LCD window instructs the user to prime.

1. To prime the fluidics station, select **Protocol** in the Fluidics Station dialog box.
2. Choose **Prime** for the respective modules in the Protocol drop-down list.
3. Change the intake buffer reservoir A to **Non-Stringent Wash Buffer** and intake buffer reservoir B to **Stringent Wash Buffer**.
4. Click **Run** for each module to begin priming.

Probe Array Wash and Stain

Following hybridization, the wash and stain procedures are carried out by the fluidics station using the **ProkGE-WS2** fluidics script. The procedure takes approximately 90 minutes to complete. The use of streptavidin in the first part of the stain procedure enhances the overall signal.

Preparing the Staining Reagents

1. Prepare the following stain and wash solutions the day of the procedure. The solutions are stable for approximately 6 to 8 hours at 4°C. Volumes given are sufficient for one probe array.

   Streptavidin Phycoerythrin (SAPE) should be stored in the dark at 4°C, either foil-wrapped or kept in an amber tube. Remove SAPE from refrigerator and tap the tube to mix well before preparing stain solution. Do not freeze SAPE. Always prepare the SAPE stain solution immediately before use.

Table 3.4.1
Streptavidin Solution Mix

Components	Volume	Final Concentration
2X MES Stain Buffer	300.0 µL	1X
50 mg/mL BSA	24.0 µL	2 mg/mL
1 mg/mL Streptavidin	6.0 µL	10 µg/mL
DI H_20	270.0 µL	—
Total Volume	**600 µL**	

Table 3.4.2
Antibody Solution Mix

Components	Volume	Final Concentration
2X MES Stain Buffer	300.0 µL	1X
50 mg/mL BSA	24.0 µL	2 mg/mL
10 mg/mL Normal Goat IgG	6.0 µL	0.1 mg/mL
0.5 mg/mL Biotin Anti-streptavidin	6.0 µL	5 µg/mL
DI H_20	264.0 µL	—
Total Volume	**600 µL**	

Table 3.4.3
SAPE Solution Mix

Components	Volume	Final Concentration
2X MES Stain Buffer	300.0 µL	1X
50 mg/mL BSA	24.0 µL	2 mg/mL
1 mg/mL Streptavidin Phycoerythrin	6.0 µL	10 µg/mL
DI H_20	270.0 µL	—
Total Volume	**600 µL**	

Table 3.4.4
Fluidics Protocols - Antibody Amplification for *E. coli* Targets

	Standard Format ProkGE-WS2
Post Hyb Wash #1	10 cycles of 2 mixes/cycle with Wash Buffer A at 25°C
Post Hyb Wash #2	4 cycles of 15 mixes/cycle with Wash Buffer B at 45°C
Stain	Stain the probe array for 10 minutes in Streptavidin Solution. Mix at 25°C
Post Stain Wash	10 cycles of 4 mixes/cycle with Wash Buffer A at 30°C
2nd Stain	Stain the probe array for 10 minutes in antibody solution. Mix at 25°C
3rd Stain	Stain the probe array for 10 minutes in SAPE Solution at 25°C
Final Wash	15 cycles of 4 mixes/cycle with Wash Buffer A at 30°C. The holding temperature is 25°C

- Wash Buffer A = non-stringent wash buffer
- Wash Buffer B = stringent wash buffer

2. In the Fluidics Station dialog box on the workstation, select the correct experiment name from the drop-down **Experiment** list. The probe array type will appear automatically.
3. In the **Protocol** drop-down list, select the **ProkGE-WS2** protocol to control the washing and staining of the probe array format being used.

✔ Note

Three-stain protocols require the user to replace stain solutions as directed by the LCD window during staining steps.

4. Choose **Run** in the Fluidics Station dialog box to begin the washing and staining. Follow the instructions in the LCD window on the fluidics station.

   If you are unfamiliar with inserting and removing probe arrays from the fluidics station modules, please refer to the *Fluidics Station 400 User's Guide*, *Fluidics Station 400 Video In-Service CD* (P/N 900374), or *Quick Reference Card* (P/N 08-0072).
5. Insert the appropriate probe array into the designated module of the fluidics station while the probe array lever is in the **EJECT** position. When finished, verify that the probe array lever is returned to the **ENGAGE** position.
6. Remove any microcentrifuge tube remaining in the sample holder of the fluidics station module(s) being used.
7. Place a microcentrifuge tube containing 600 µL streptavidin solution into the sample holder, making sure that the metal sampling needle is in the tube with its tip near the bottom.

   ⇒ The Fluidics Station dialog box and the LCD window display the status of the washing and staining as they progress.
8. When the LCD window indicates, replace the microcentrifuge tube containing the streptavidin stain with a microcentrifuge tube containing 600 µL antibody stain solution into the sample holder, making sure that the metal sampling needle is in the tube with its tip near the bottom.
9. When the LCD window indicates, replace the microcentrifuge tube containing antibody solution with the microcentrifuge tube containing the 600 µL of SAPE solution.
10. When the protocol is complete, the LCD window displays the message **EJECT CARTRIDGE**.
11. Remove microcentrifuge tube containing stain and replace with an empty microcentrifuge tube.
12. Remove the probe arrays from the fluidics station modules by first moving the probe array holder lever to the **EJECT** position.
13. Check the probe array window for large bubbles or air pockets.
    - If bubbles are present, refer to **Table 3.4.5** on page 3.4.12.
    - If the probe array has no large bubbles, it is ready to scan on the GeneArray Scanner. **ENGAGE** wash block and proceed to *Probe Array Scan* on page 3.4.13.

    If you do not scan the arrays right away, keep the probe arrays at 4°C and in the dark until ready for scanning.

    If there are no more samples to hybridize, shut down the fluidics station following the procedure outlined in the section, *Shutting Down the Fluidics Station* on page 3.4.14.

✔ Note

For proper cleaning and maintenance of the fluidics station, including the bleach protocol, refer to Section 4, Fluidics Station Maintenance Procedures.

Table 3.4.5
If bubbles are present

Return the probe array to the probe array holder. Latch the probe array holder by gently pushing it up until a light click is heard. Engage the washblock by firmly pushing up on the probe array lever to the **ENGAGE** position. The fluidics station will drain the probe array and then fill it with a fresh volume of the last wash buffer used. When it is finished, if the LCD window displays **EJECT CARTRIDGE** again, remove the probe array and inspect it again for bubbles. If no bubbles are present, it is ready to scan. Proceed to *Probe Array Scan* on page 3.4.13. If several attempts to fill the probe array without bubbles are unsuccessful, the array should be filled with **Wash A (non-stringent buffer)** manually, using a micropipette. Excessive washing will result in a loss of signal intensity.

Probe Array Scan

The scanner is also controlled by Affymetrix Microarray Suite. The probe array is scanned after the wash protocols are complete. Make sure laser is warmed up prior to scanning by turning the laser on at least 15 minutes before use. If probe array was stored at 4°C, warm to room temperature before scanning. Refer to the Microarray Suite online help and the appropriate scanner user's manual for more information on scanning.

If necessary, clean the glass surface of probe array with a non-abrasive towel or tissue before scanning. **Do not use alcohol to clean glass**.

✔ Note

The scanner uses an argon-ion laser and is equipped with a safety interlock system. Defeating the interlock system may result in exposure to hazardous laser light.

1. Select **Run** → **Scanner** from the menu bar. Alternatively, click the Start Scan icon in the tool bar.
   ⇒ The Scanner dialog box appears with a drop-down list of experiments that have not been run.
2. Select the experiment name that corresponds to the probe array to be scanned.
   A previously run experiment can also be selected by using the **Include Scanned Experiments** option box. After selecting this option, previously scanned experiments appear in the drop-down list.
3. By default, after selecting the experiment the number [**2**] is displayed in the **Number of Scans** box to perform the recommended 2X image scan.
4. Once the experiment has been selected, click the **Start** button.
   ⇒ A dialog box prompts you to load a sample into the scanner.
5. Click the **Options** button to check for the correct pixel value and wavelength of the laser beam.
   - Pixel value = 3 µm
   - Wavelength = 570 nm
6. Open the sample door on the scanner and insert the probe array into the holder. Do not force the probe array into the holder. Close the sample door of the scanner.
7. Click **OK** in the Start Scanner dialog box.
   ⇒ The scanner begins scanning the probe array and acquiring data. When **Scan in Progress** is Selected from the **View** menu, the probe array image appears on the screen as the scan progresses.

Shutting Down the Fluidics Station

1. After removing a probe array from the probe array holder, the LCD window displays the message **ENGAGE WASHBLOCK**.
2. Engage the washblock by firmly pushing up on the probe array lever to the **ENGAGE** position.

   The fluidics station will automatically perform a Cleanout procedure. The LCD window will indicate the progress of the Cleanout procedure.
3. When the fluidics station LCD window indicates **REMOVE VIAL**, the Cleanout procedure is complete.
4. Remove the sample microcentrifuge tube from the sample holder.
5. If no other hybridizations are to be performed, place wash lines into a bottle filled with deionized water.
6. Select **Shutdown** for all modules from the drop-down **Protocol** list in the Fluidics Station dialog box. Click the **Run** button for all modules.

   The Shutdown protocol is critical to instrument reliability. Refer to the *Fluidics Station 400 User's Guide* for more information.
7. After Shutdown protocol is complete, flip the ON/OFF switch of the fluidics station to the OFF position.

IMPORTANT

To maintain the cleanliness of the fluidics station and obtain the highest quality image and data possible, a weekly bleach protocol and a monthly decontamination protocol are highly recommended. Please refer to Section 4, Fluidics Station Maintenance Procedures for further detail.

Customizing the Protocol

There may be times when the fluidics protocols need to be modified. Modification of protocols must be done before downloading the protocol to the fluidics station. Protocol changes will not affect runs in progress. For more specific instructions, refer to the Microarray Suite online help.

1. Select **Tools** → **Edit Protocol** from the menu bar.
2. In the Edit Protocol dialog box under Protocol Name, click the arrow to open a list of protocols. Click the protocol to be changed.
   ⇒ The name of the protocol is displayed in the **Protocol Name** text box. The conditions for that protocol are displayed on the right side of the Edit Protocol dialog box.
3. Select the item to be changed and input the new parameters as needed, keeping parameters within the ranges shown below in **Table 3.4.6**.

Table 3.4.6
Valid Ranges for Wash/Stain Parameters

Parameter	Valid Range
Wash Temperature for A1, B, A2, or A3 (°C)	15 to 50
Number of Wash Cycles for A1, B, A2, or A3	0 to 99
Mixes / Wash cycle for A1, B, A2, or A3	15 to 50
Stain Time (seconds)	0 to 86,399
Stain Temperature (°C)	15 to 50
Holding Temperature (°C)	15 to 50

- Wash A1 corresponds to Post Hyb wash #1 in Table 3.4.4.
- Wash B corresponds to Post Hyb wash #2 in Table 3.4.4.
- Wash A2 corresponds to Post Stain Wash in Table 3.4.4.
- Wash A3 corresponds to Final Wash in Table 3.4.4.

4. To return to the default values for the protocol selected, click the **Defaults** button.
5. Once all the protocol conditions are modified as desired, change the name of the edited protocol in the **Protocol Name** box.

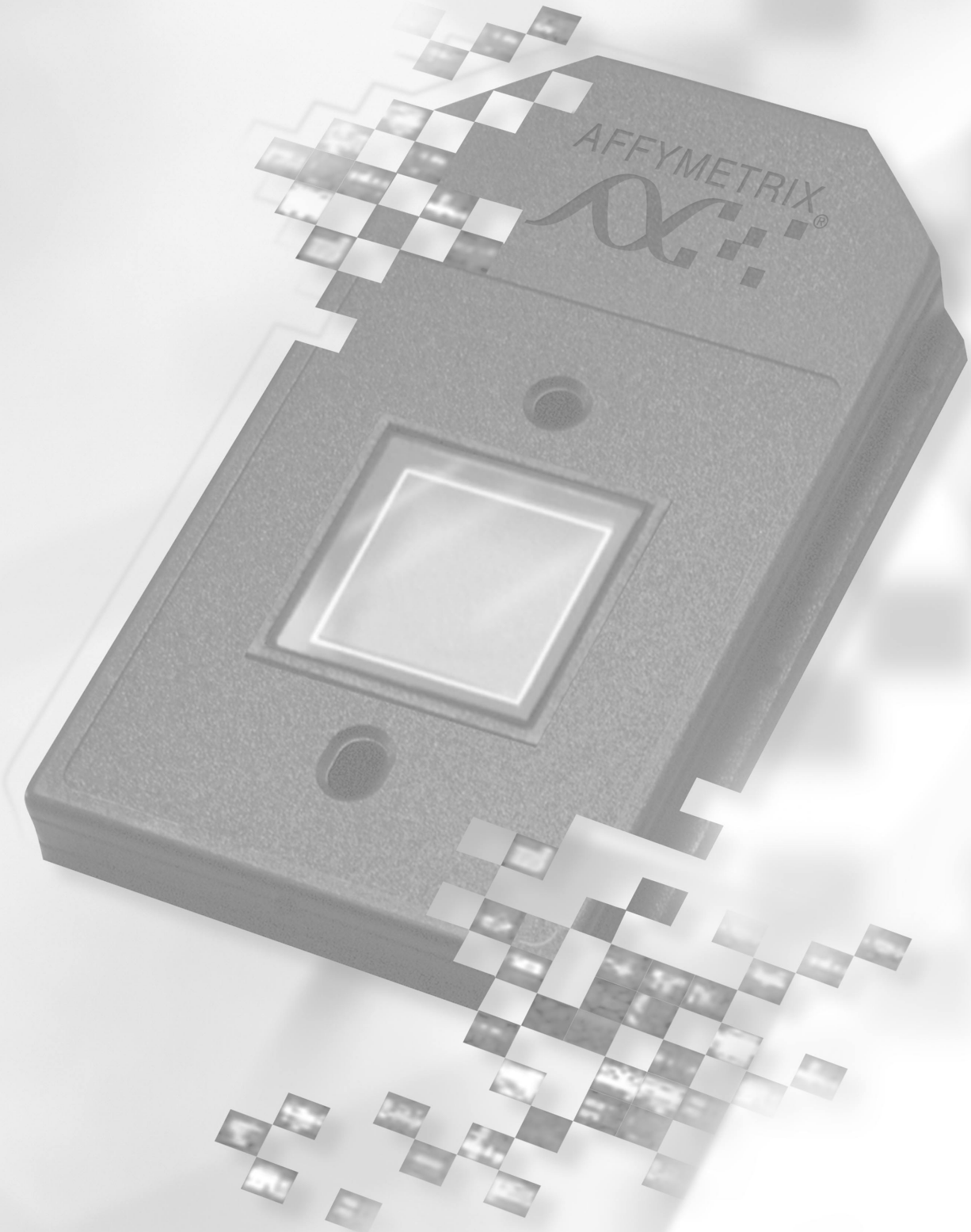
! CAUTION *If the protocol is saved without entering a new **Protocol Name**, the original protocol parameters will be overwritten.*

6. Click **Save**, then close the dialog box.

   Enter **0** (zero) for hybridization time if hybridization step is not required. Likewise, enter **0** (zero) for the stain time if staining is not required. Enter **0** (zero) for the number of wash cycles if a wash solution is not required.

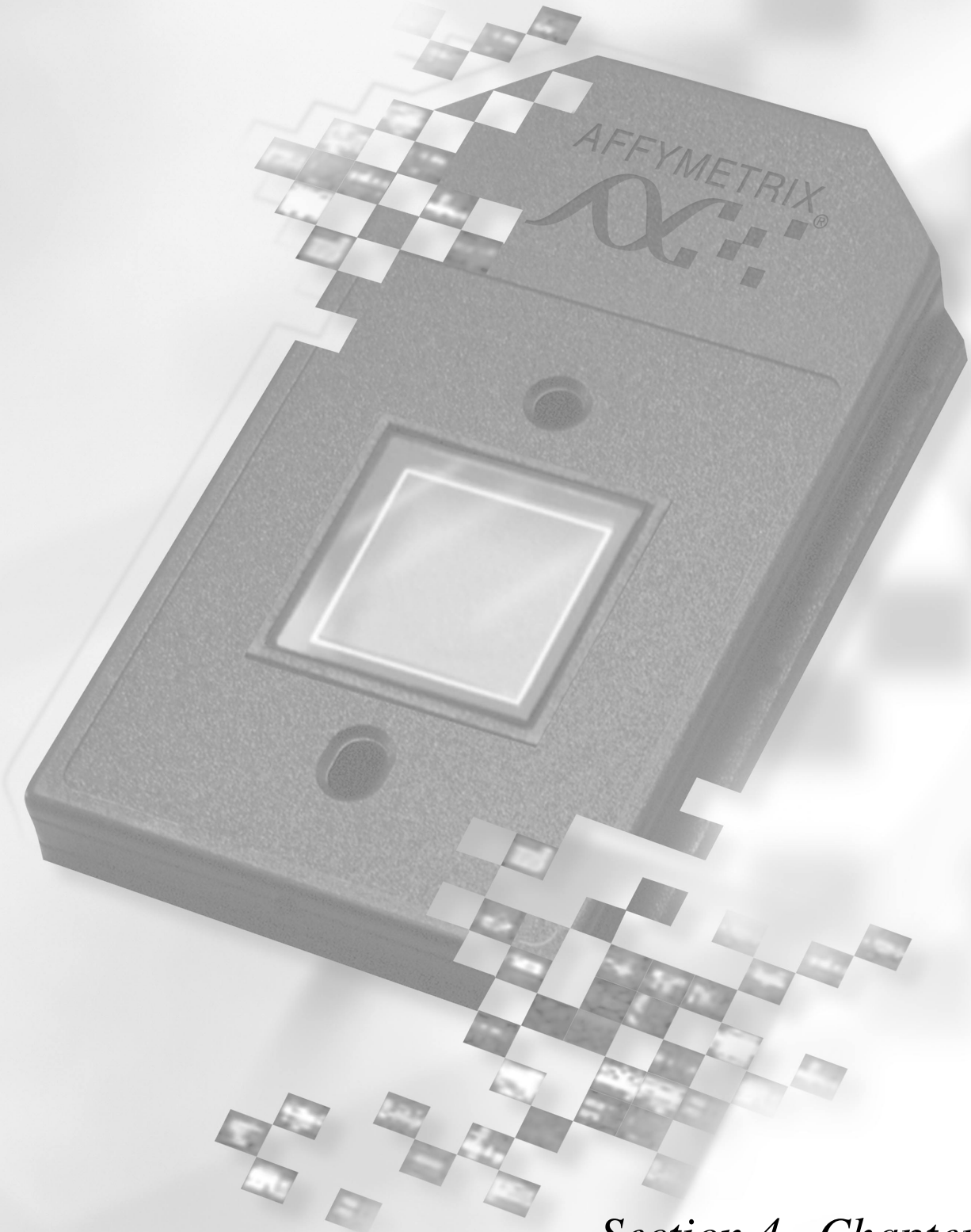
Section 4:

Fluidics Station Maintenance Procedures



Contents

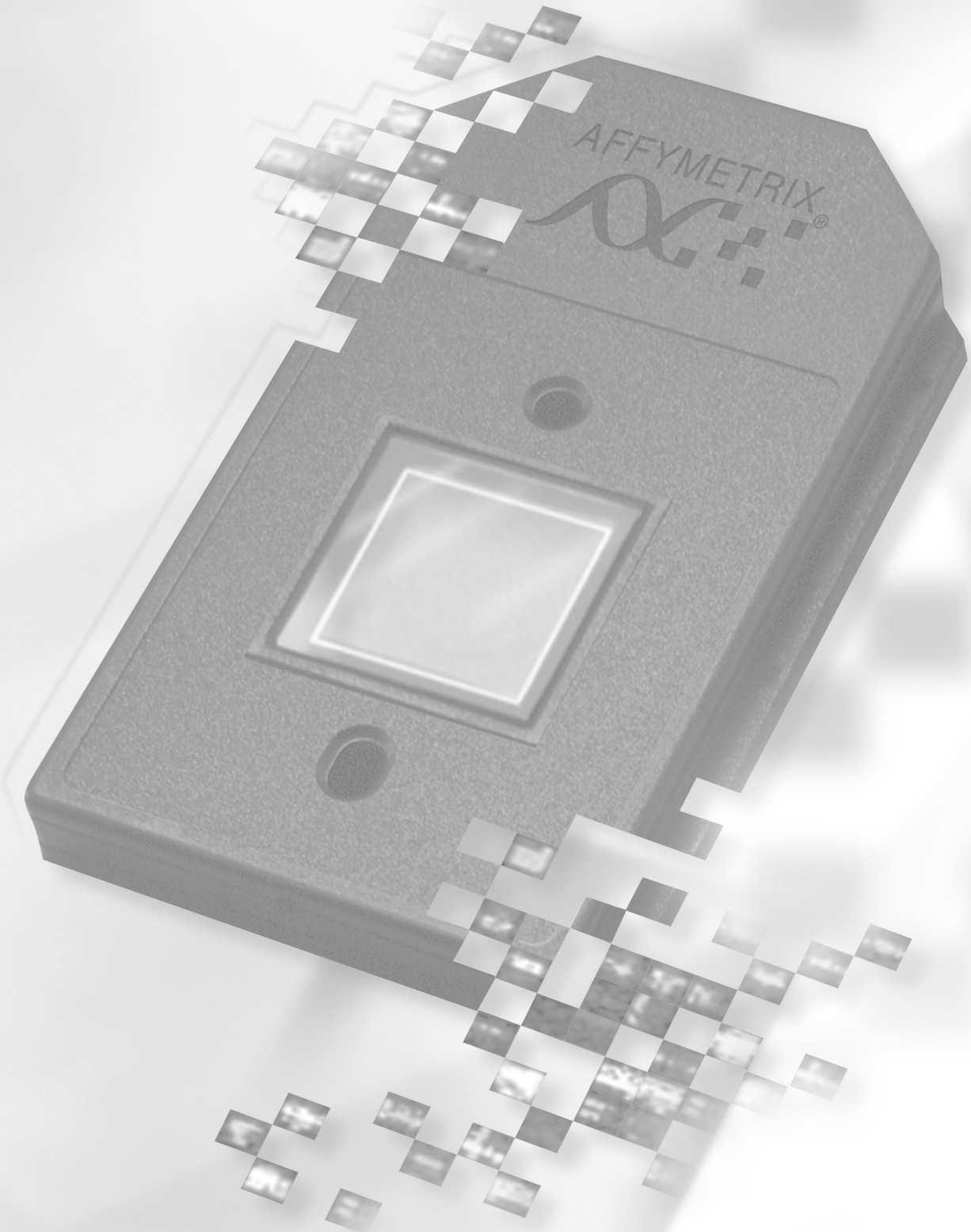
Section 4



F.S. Maintenance

Section 4, Chapter 1

Section 4, Chapter 1



Fluidics Station Maintenance Procedures

This Section Contains:

- A weekly fluidics station bleach protocol.
- A monthly fluidics station decontamination protocol.

Weekly Fluidics Station Cleanout

A cleaning protocol is recommended for fluidics station maintenance if the antibody staining procedure is used. Choose **Bleach** for all modules from the drop-down list in the Fluidics Station dialog box. Click the **Run** button for all modules and follow LCD instructions.

Bleach Protocol

This protocol is designed to eliminate any residual SAPE-antibody complex that may be present in the fluidics station tubing and needles. We recommend running this protocol at least once a week.

1. Prepare 1 liter of 0.525% sodium hypochlorite solution using distilled water. Shake well.

✔ Note *Each fluidics station with four modules requires at least 200 mL of the 0.525% sodium hypochlorite solution.*

2. Cut at least two feet of tubing (Tygon tubing, 0.04") for each module of each fluidics station. These can be reused for subsequent BLEACH runs.
3. Place all three wash lines of each fluidics station in 1 liter of distilled water.

✔ Note *The BLEACH protocol requires at least 550 mL of distilled water.*

4. Choose **Fluidics** from the **Run** menu. Alternatively, click the down arrow Protocol list on the toolbar.
5. Choose **Bleach** for the respective modules in the **Protocol** drop-down list.
6. Disengage washblock for each module.

✔ Note *Temperature will ramp up to 50°C.*

7. Connect one end of the plastic tubing to the needle at the bottom of each module and insert the other end into 0.525% sodium hypochlorite solution (at least 200 mL for all four modules).
8. **ENGAGE** washblock. The fluidics station will empty the lines and perform three cleaning cycles of 10 rinses each using bleach solution.
9. When the fluidics station LCD window displays **Remove Tube from Needle**, carefully remove tubing from each module needle by pushing the tubing down with one hand while holding the needle with the other.

➲ IMPORTANT *Do not pull the tube as this may damage the needle in the process.*

10. Load empty microcentrifuge tube onto each module. The fluidics station will empty the lines and run three cycles with three rinses each. In addition, the fluidics station will rinse the needle 20 times, twice using distilled water, then bring the temperature back to 25°C and drain the lines with air.
11. The LCD display will read **CLEANING DONE**.

Monthly Fluidics Station Decontamination Protocol

To maintain your Fluidics Station in the best possible working condition, we recommend that the following decontamination protocol be performed on your fluidics station **at least once a month**, in addition to the weekly cleaning described above. The protocol requires approximately 2 hours to run.

This protocol ensures that all of the tubing associated with the station is kept thoroughly clean. Keeping this tubing as clean as possible ensures that array images will be optimized and high-quality results will be obtained.

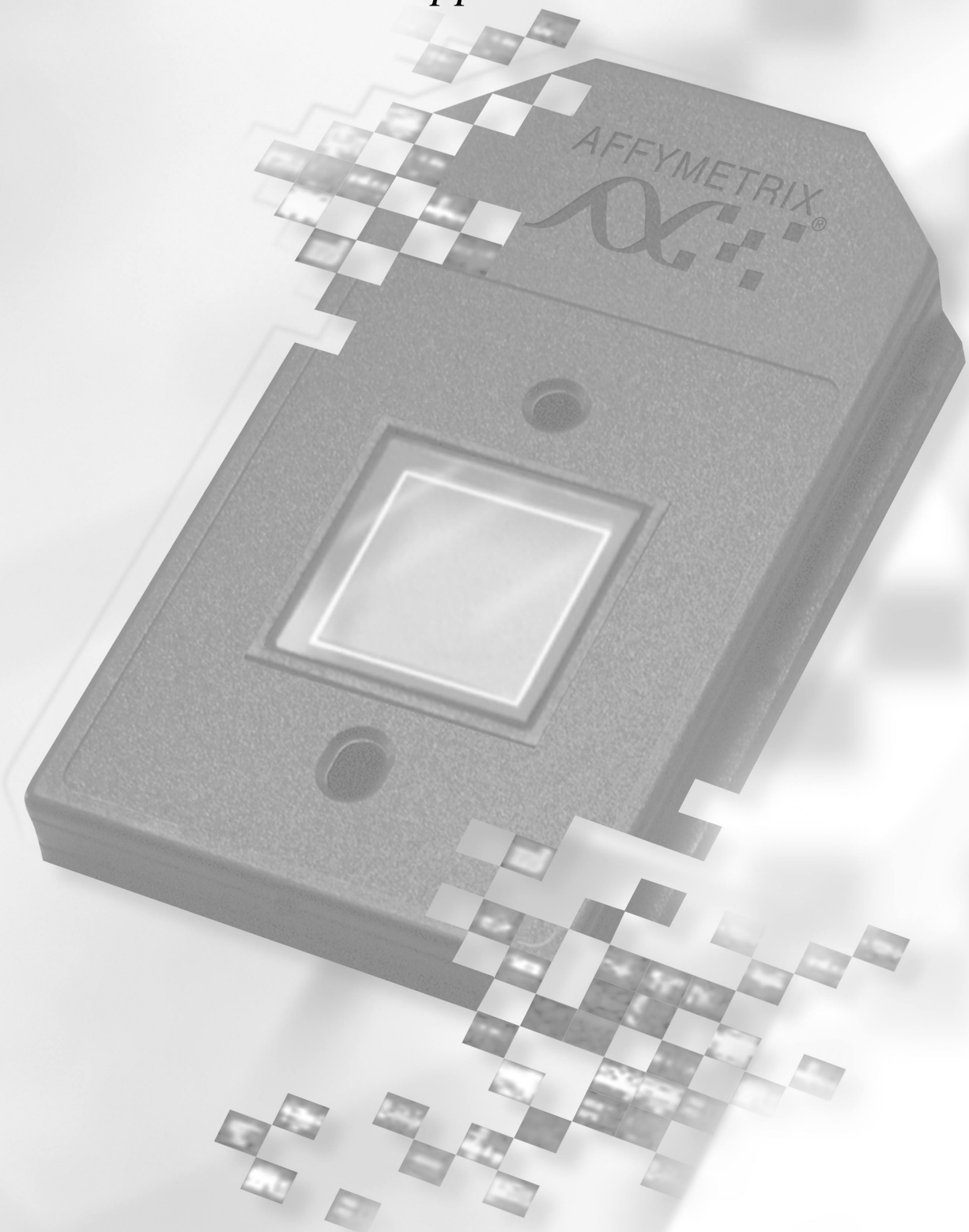
1. Prepare 2 liters of 0.525% sodium hypochlorite solution using distilled water. Mix well.
2. Place all three wash lines of the fluidics station in 1 liter of 0.525% sodium hypochlorite solution.
3. Run the Prime protocol (page 2.4.8) on all four modules with wash lines in 0.525% sodium hypochlorite solution (instead of wash buffers A and B).
4. Run the Shutdown protocol (page 2.4.16) on all four modules with wash lines in 0.525% sodium hypochlorite solution (instead of distilled water).
5. Follow Bleach Protocol (as described on page 4.1.4) with the following change in Step 3: place the three wash lines of the fluidics station in 1 liter of 0.525% sodium hypochlorite solution instead of distilled water.
6. Change intake tubing and peristaltic tubing, if required (as described in the *Fluidics Station 400 User's Guide*).
7. Run the **Bleach** protocol with three wash lines of the fluidics station in distilled water.
8. Run the **Prime** protocol with wash lines in distilled water (instead of wash buffers A and B).
9. Run the **Shutdown** protocol with wash lines in distilled water.
10. Run the **Prime** protocol with wash lines in distilled water (instead of wash buffers A and B).
11. Run the **Shutdown** protocol with distilled water.

✔ Note

At the end of each step, the fluidics station will indicate a 'ready' status. The fluidics station should not be used until this entire procedure (steps 1-11) is complete.

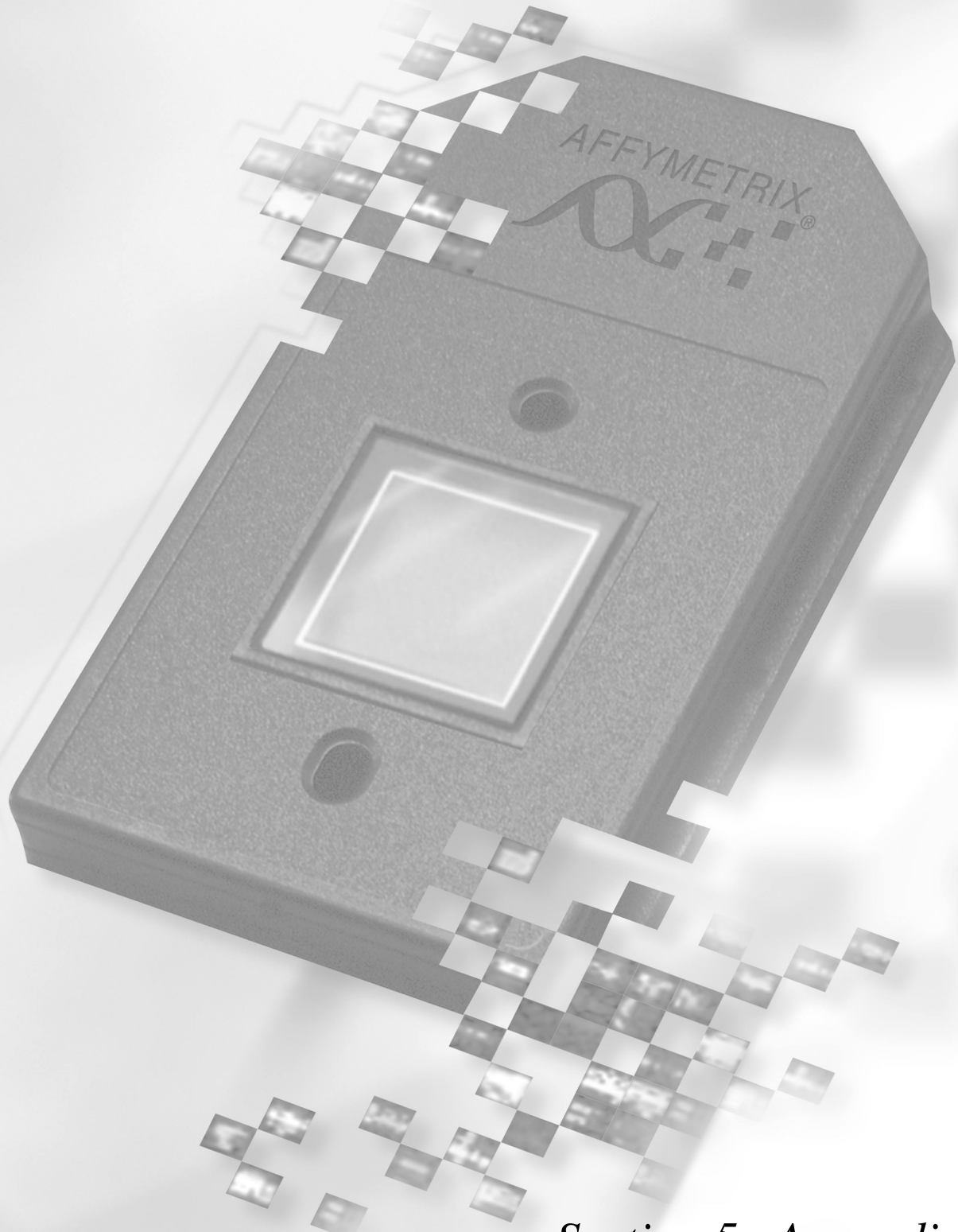
Section 5:

Appendices



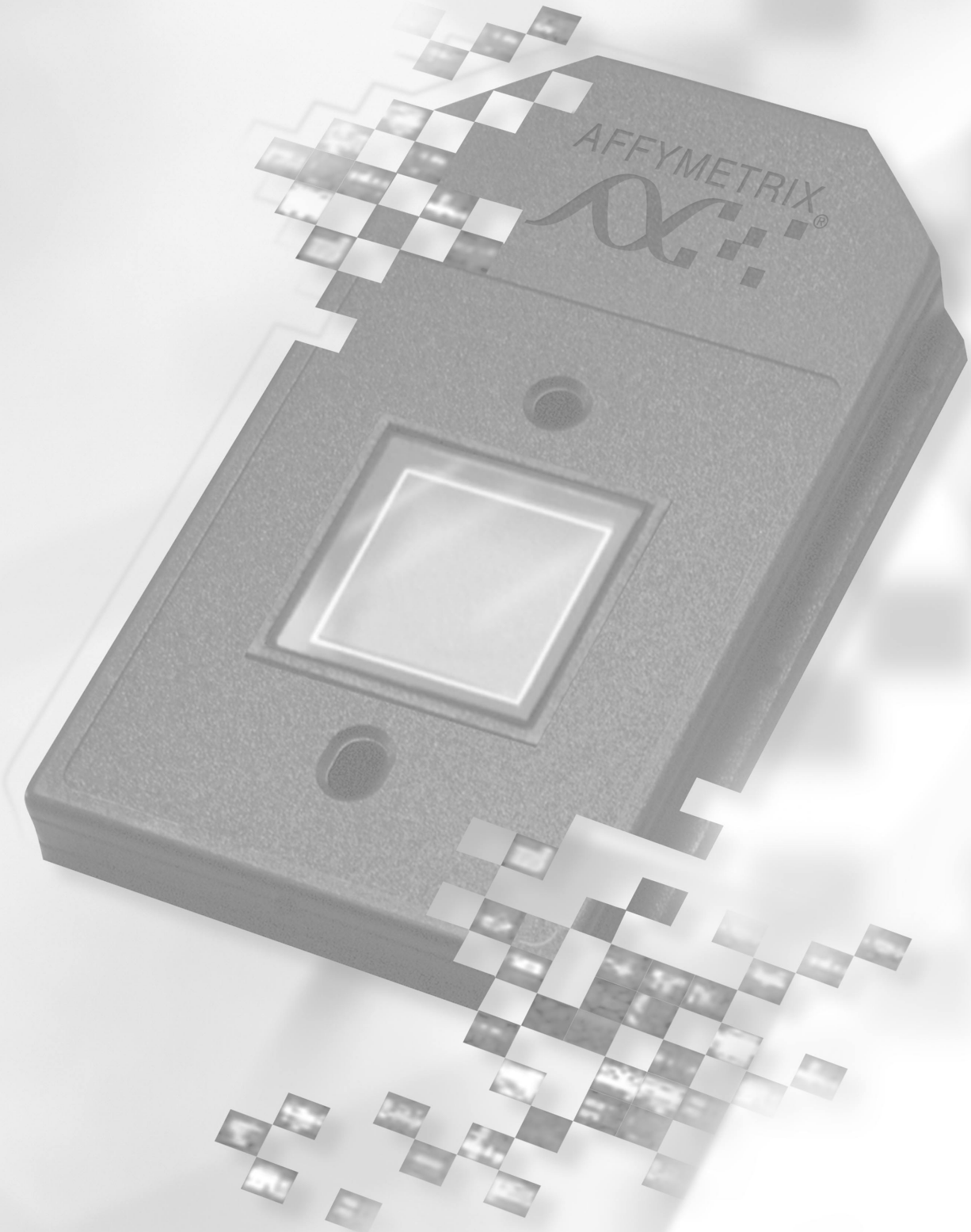
Contents

Section 5 *Appendices*



Section 5, Appendix A

Section 5, Appendix A



Supplier and Reagent Reference List

Affymetrix Technical Support

Affymetrix provides technical support via phone or e-mail. To contact Affymetrix Technical Support:

Affymetrix Inc.
3380 Central Expressway
Santa Clara, CA 95051
USA

Tel: 1-888-362-2447 (1-888-DNA-CHIP)
Fax: 1-408-731-5441
E-mail: support@affymetrix.com

Affymetrix UK, Ltd
Voyager, Mercury Park,
Wycombe Lane, Wooburn Green,
High Wycombe HP10 0HH
United Kingdom

Tel: +44 (0)1628 552550
Fax: +44 (0)1628 552585
E-mail: supporteurope@affymetrix.com

www.affymetrix.com

Supplier Contact Information

Supplier	US	United Kingdom	France	Germany
Ambion www.ambion.com	(800) 888-8804	+44 (0)1993 706 500	+33 (0)1 42 53 14 53	+49 (0)692 88082
Amersham/Pharmacia Biotech www.apbiotech.com	(800) 323-9750	+44 (0)800 515 313	+33 (0)1 69 35 67 00	+49 (0)761 49030
Amresco www.amresco-inc.com	(800) 448-4442	+44 (0)1582 745 000	+33 (0)4 70 03 88 55	+49 (0) 551 506860
ATCC www.atcc.org	(800) 638-6597		US (703) 365-2700	
BioWhittaker Molecular Applications / Cambrex www.cambrex.com	(800) 341-1574	+44 (0)1189 795 234	(Belgium) +32 8-732-1611	
Brinkmann Instruments www.brinkmann.com	(800) 421-9988	see web site for nearest distributor		
Cambrex www.cambrex.com	(800) 341-1574	see web site for nearest distributor		
Cole-Parmer www.coleparmer.com	(800) 323-4340	+44 (0)1815 747 556	+33 (0)3 88 67 14 14	+49 (0)785 17069
CLONTECH www.clontech.com	(800) 662-CLON	+44 (0)1256 476 500	+33 (0)1 34 60 24 24	+49 (0)622 134170
Epicentre Technologies www.epicentre.com	(800) 284-8474	+44 (0)1223 366 500	+33 (0)1 30 46 39 00	+49 (0)515 29020
Eppendorf - 5 Prime www.5prime.com	(800) 533-5703		US (303) 440-3705	
FMC Bioproducts www.fmc.com		US (215) 299-6000		
GENSET Corp www.genset.com	(800) 995-0308		+33 (0)1 43 56 59 00	
Invitrogen Life Technologies www.invitrogen.com	(800) 955-6288	00 800 5345 5345 (Toll-free for Europe)		
Millipore Corp www.millipore.com	(800) 645-5476	+44 (0)1923 816 375	+33 (0)1 30 12 70 00	+49 (0)619 64940
Molecular Probes www.probes.com	(541) 465-8300	+44 (0)1223 316 855	+33 (0)4 70 03 88 55	+49 (0)551 371062
New England Biolabs www.neb.com	(800) 632-5227	+44 (0)800 318 846	+33 (0)1 34 60 24 24	+49 (0)800 2465 227
Novagen www.novagen.com	(800) 207-0144	+44 (0)800 622 935	+33 (0)1 30 46 39 00	+49 (0)800 6931 000
Operon Technologies www.operon.com	(800) 688-2248	+44 (0)1914 100 323	see web site for nearest distributor	
Pierce Chemical www.piercenet.com	(800) 874-3723	+44 (0)1244 382 525	+33 (0)4 70 03 88 55	+49 2241 96850
Promega Corporation www.promega.com	(800) 356-9526	+44 (0)800 378 994	+33 (0)8 00 48 79 99	+49 (0)621 85010

Supplier	US	United Kingdom	France	Germany
QIAGEN www.qiagen.com	(800) 426-8157	+44 (0)1293 422 911	+33 (0)1 60 92 09 20	+49 (0)210 3892 230
Rainin www.rainin.com	(800) 472-4646	+44 (0)1582 456 666	see web site for nearest distributor	
Roche Molecular Biochemical biochem.roche.com	(800) 262-1640	+44 (0)1273 480 444	+33 (0)4 76 76 30 87	+49 (0)621 75985 68
Sigma-Aldrich www.sigma-aldrich.com	(800) 325-3010	+44 (0)1202 733 114	+33 (0)4 74 82 28 88	+49 (0)896 5131130
USA Scientific www.usascientific.com	(800) 522-8477	US (352)-237-6288		
Vector Laboratories www.vectorlabs.com	(800) 227-6666	+44 (0)1733 237 999	44 86 22 75	+49 (0)9342 39499 or 0800 253 9472
VWR Scientific Products www.vwrsp.com	(800) 932-5000		US (908) 757-4045	

Reagent List

A

Acetic Acid, Glacial, Sigma-Aldrich, P/N A 6283
Acetylated Bovine Serum Albumin (BSA) solution, 50 mg/mL, Invitrogen Life Technologies, P/N 15561020
Ammonium Acetate, 7.5 M, Sigma-Aldrich, P/N A 2706
Anti-streptavidin antibody (goat), biotinylated, Vector Laboratories, P/N BA-0500
Antibody (goat), Anti-streptavidin, biotinylated, 0.5 mg, Vector Laboratories, P/N BA-0500
Antibody, IgG, Goat, Reagent Grade, 50 mg, Sigma-Aldrich, P/N I5256
γ-S-ATP, 20 μmoles, Roche Molecular Biochemical, P/N 1162306

B

Bleach (5.25% Sodium Hypochlorite), VWR Scientific Products, P/N 21899-504
Bovine Serum Albumin (BSA) solution, 50 mg/mL, Acetylated, Invitrogen Life Technologies, P/N 15561020

C

CHROMA SPIN-100 Columns in Swing Bucket Format, CLONTECH, P/N K1302-1
Control Oligo B2, 30 nM, Affymetrix, P/N 900301

D

dATP, dCTP, dGTP, dTTP, Amersham Pharmacia Biotech, P/N 27-2035-01
Deoxyribonuclease I (DNase I), Amersham Pharmacia Biotech, P/N 27-0514-01
DMSO, Hybrid-Max®, Sigma-Aldrich, P/N D2650
DNA, Herring Sperm, Promega Corporation, P/N D1811
DNA Ligase, *E. coli*, Invitrogen Life Technologies, P/N 18052-019
DNA Polymerase, *E. coli*, Invitrogen Life Technologies, P/N 18010-025
DNA Polymerase, T4, Invitrogen Life Technologies, P/N 18005-025
dNTP, 10 mM, Invitrogen Life Technologies, P/N 18427-013
DTT, 100 mM, Epicentre Technologies, P/N M4410K (supplied with MMLV Reverse Transcriptase)

E

EDTA Disodium Salt, 0.5 M solution, 100 mL, Sigma-Aldrich, P/N E7889
EDTA, 0.05 M, pH 8.0, Invitrogen Life Technologies, P/N 15575-038
Enzo BioArray HighYield RNA Transcript Labeling Kit, Affymetrix, P/N 900182
Ethidium Bromide, Sigma-Aldrich, P/N E 8751
Expression Control Clones, American Type Culture Collection (ATCC)

pGIKS-bioB	ATCC 87487
pGIKS-bioC	ATCC 87488
pGIKS-bioD	ATCC 87489
pGIKS-cre	ATCC 87490
pGIBS-lys	ATCC 87482
pGIBS-phe	ATCC 87483

pGIBS-thr	ATCC 87484
pGIBS-trp	ATCC 87485
pGIBS-dap	ATCC 87486

G

GeneChip® Eukaryotic Hybridization Control Kit, Affymetrix, P/N 900299
Glycogen, Ambion, P/N 9510
Glycogen, 20 mg/mL, Roche Molecular Biochemical, P/N 901393
Goat IgG, Reagent Grade, Sigma-Aldrich, P/N I5256

H

HCl, 1N solution, VWR Scientific Products, P/N MK638860
Herring Sperm DNA, Promega Corporation, P/N D1811
Hybridization Oven 640, Affymetrix, P/N 800139

I

ImmunoPure® NeutrAvidin, Pierce Chemical, P/N 31000
ImmunoPure® Streptavidin, Pierce Chemical, P/N 21125

L

10 bp and 100 bp ladder, Invitrogen Life Technologies, P/N 10821-015 and15628-019, respectively

M

Magnesium acetate (MgOAc), Sigma-Aldrich, P/N M2545
MasterPure™ RNA Purification Kit, Epicentre Technologies, P/N MCR85102
MEGAscript T3 Kit, Ambion, P/N 1338
MEGAscript T7 Kit, Ambion, P/N 1334
MES Free Acid Monohydrate SigmaUltra, Sigma-Aldrich, P/N M5287
MES Sodium Salt, Sigma-Aldrich, P/N 5057
MMLV Reverse Transcriptase, New England BioLabs, P/N M0253L
MMLV Reverse Transcriptase Buffer, 10X, New England BioLabs, P/N M0253L
MOPS, Sigma-Aldrich, P/N M3183

N

NaCl, 5 M, RNase-free, DNase-free, Ambion, P/N 9760G
NaOH, 1N Solution, VWR Scientific Products, P/N MK469360
Novex XCell SureLock™ Mini-Cell, Invitrogen Life Technologies, P/N EI9001
Nuclease-free Water, Ambion, P/N 9930
Nucleotides, labeled, Biotin-11-CTP and Biotin-16-UTP, Enzo, P/N 42818 (CTP) and P/N 42814 (UTP)
Nucleotides, Biotin-11-CTP, Sigma-Aldrich, P/N B7048
Nucleotides, Biotin-16-UTP, Roche Molecular Biochemicals, P/N 1388908
Nucleotides, Biotin-16-UTP, Sigma-Aldrich, P/N B6923

O

Oligo B2, Control, Control Oligo for the antisense probe array, HPLC purified
5´ -bio GTCGTCAAGATGCTACCGTTCAGGA- 3 ´
Oligotex Direct mRNA Kit, QIAGEN, P/N 72012, 72022, or 72041
Oligotex mRNA Kit, QIAGEN, P/N 70022, 70042, or 70061

P

PBS, pH 7.2, Invitrogen Life Technologies, P/N 20012-027
Pellet Paint, Novagen, P/N 69049-3
PEO-Iodoacetyl-Biotin, 50 mg, Pierce Chemical, P/N 21334ZZ
Phase Lock Gel, Brinkmann Instruments, P/N 955 15 415
Phenol/chloroform/isoamyl alcohol, Ambion, P/N 9732
Phycoerythrin-Streptavidin, Molecular Probes, P/N S-866
Polynucleotide Kinase, T4, New England BioLabs, P/N 201L
Potassium acetate (KOAc), Sigma-Aldrich, P/N P5708
Primer, T7– $(dT)_{24}$, (Genset Corp), HPLC purified
5´ - GGCCAGTGAATTGTAATACGACTCACTATAGGGAGGCGG-(T)24 - 3´

Q

Qiashredder, QIAGEN, P/N 79654

R

R-Phycoerythrin Streptavidin, Molecular Probes, P/N S-866
Ribonuclease H (RNase H), *E. coli*, Epicentre Technologies, P/N R0601K
RNA/DNA Mini Column Kit, QIAGEN, P/N 14123
RNase H, *E. coli*, Invitrogen Life Technologies, P/N 18021-021, or Epicentre Technologies, P/N R0601K
RNeasy Mini Kit, QIAGEN, P/N 74104, 74106

S

Second strand buffer, 5X, Invitrogen Life Technologies, P/N 10812-014
Separator, Micropure, Millipore, P/N 42512
Sodium Acetate, 3 M, pH 5.2, Sigma-Aldrich, P/N S7899
Sodium Hypochlorite, Sigma-Aldrich, P/N 7681-52-9
SSPE, 20X, BioWhittaker Molecular Applications / Cambrex, P/N 51214
Streptavidin, ImmunoPure®, Pierce Chemical, P/N 21125
Sucrose Gel Loading Dye, 5X, Amresco, P/N E-274
SUPERase•In™, Ambion, P/N 2696
SuperScript II RT, Invitrogen Life Technologies, P/N 18064-071
SuperScript Choice system, Invitrogen Life Technologies, P/N 18090-019
SYBR Gold, Molecular Probes, P/N S-11494
SYBR Green II, Cambrex, P/N 50523, FMC Bioproducts, P/N 50523, or Molecular Probes, P/N S7586

T

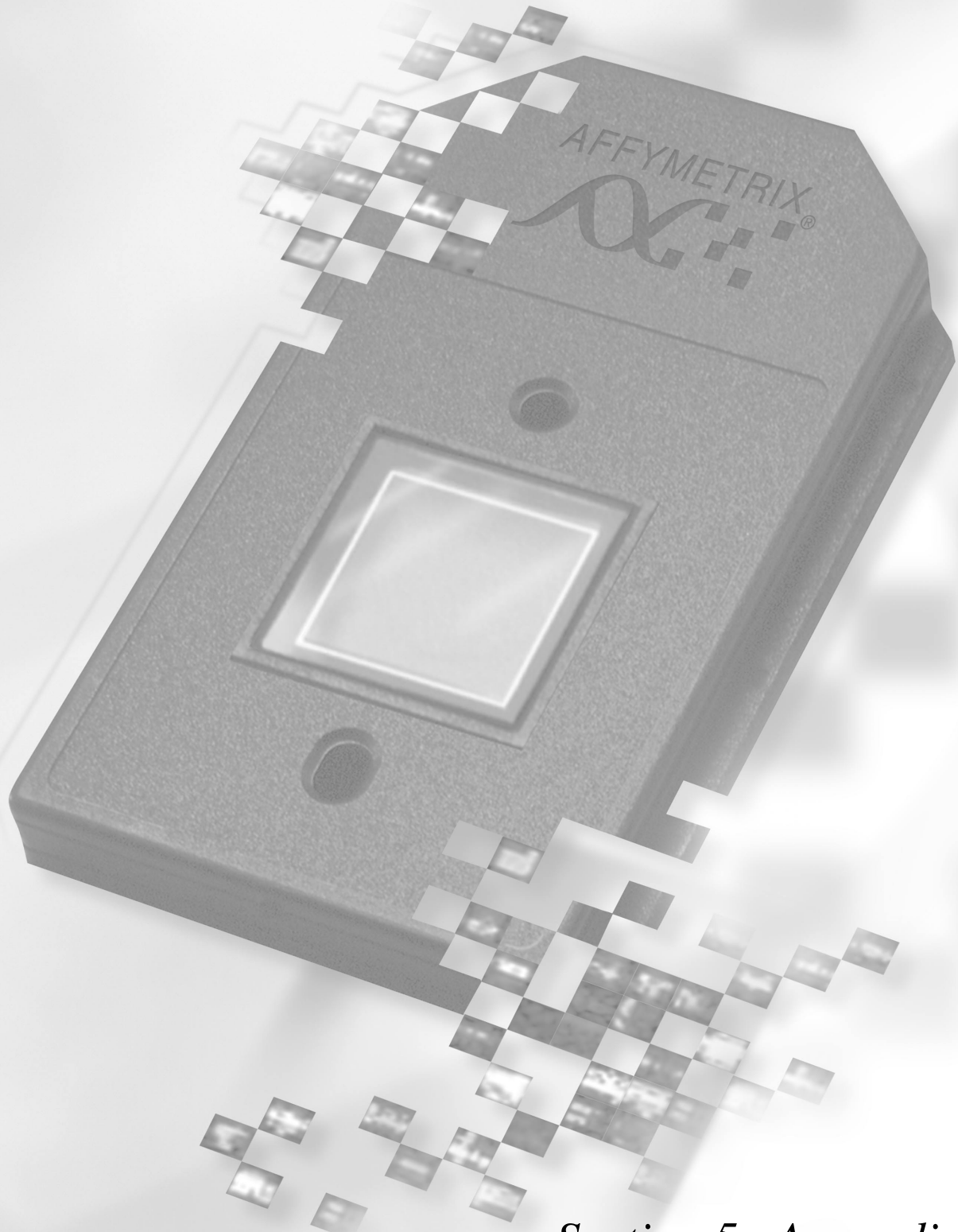
TBE, 10X, Cambrex, P/N 50843
TBE Gel (4-20%), 1.0 mm, 12 well, Invitrogen Life Technologies, P/N EC62252
TE, 1X, BioWhittaker Molecular Applications / Cambrex, P/N 51235
Tough Spots, Label Dots, USA Scientific, P/N 9185
Tris pH 7.0, 1M, Ambion, P/N 9850G
Trizma Base, Sigma-Aldrich, P/N T 1503
TRIzol Reagent, Invitrogen Life Technologies, P/N 15596-018
Tubes, Sterile, RNase-free, microcentrifuge, 1.5 mL, USA Scientific, P/N 1415-2600
Tubing, Tygon, 0.04" inner diameter, Cole Palmer, P/N H-06418-04
Tween-20, 10%, Pierce Chemical, P/N 28320

V

Vacuum filter units 1 liter capacity, 0.20 µm or 0.45 µm, Corning, P/N 25988-1L

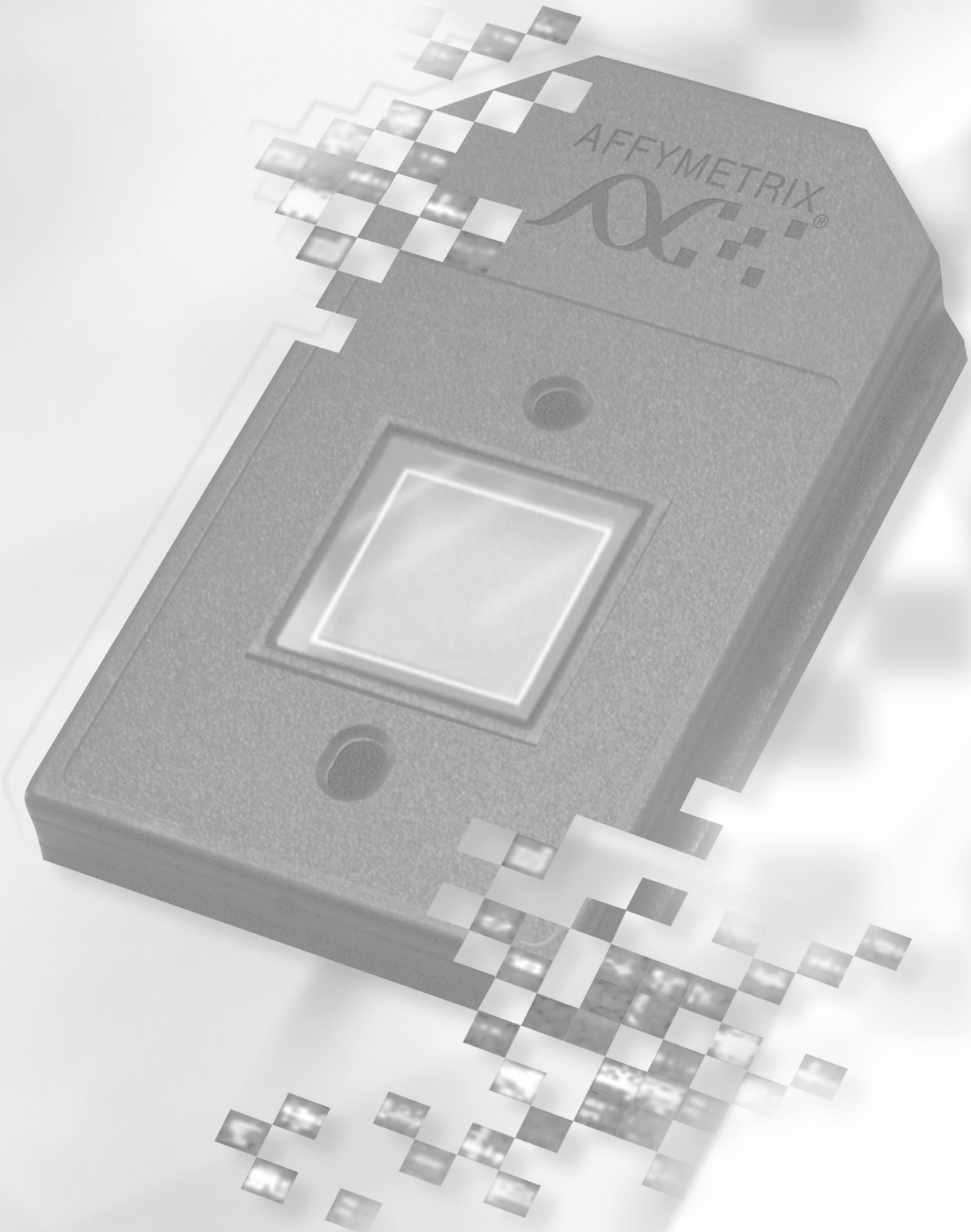
W

Water, DEPC-Treated, Ambion, P/N 9902
Water, Distilled, Invitrogen Life Technologies, P/N 15230147
Water, Molecular Biology Grade, BioWhittaker Molecular Applications / Cambrex, P/N 51200



Section 5, Appendix B

Section 5, Appendix B



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 4º 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

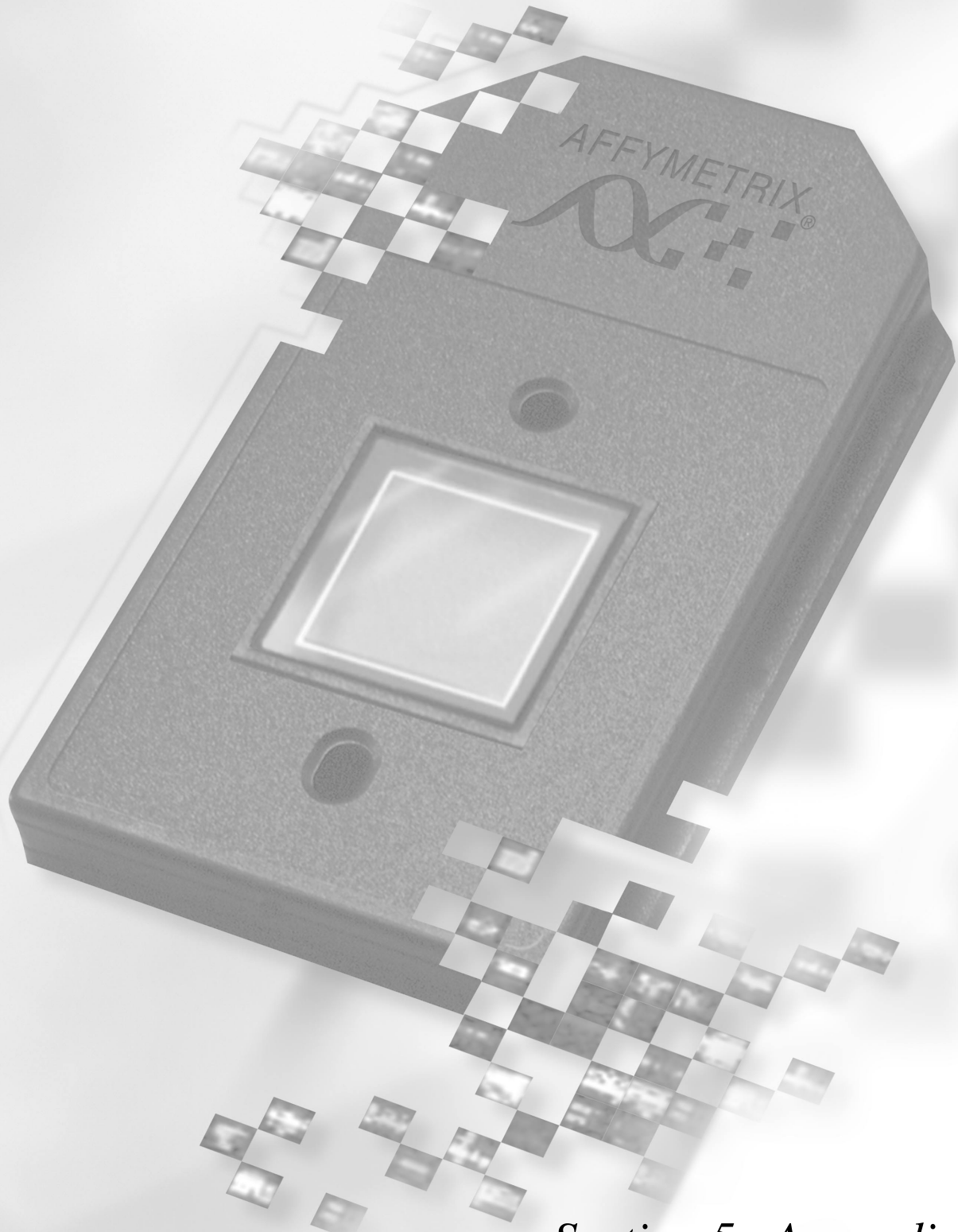
Which is greater, sample or assay variability?

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

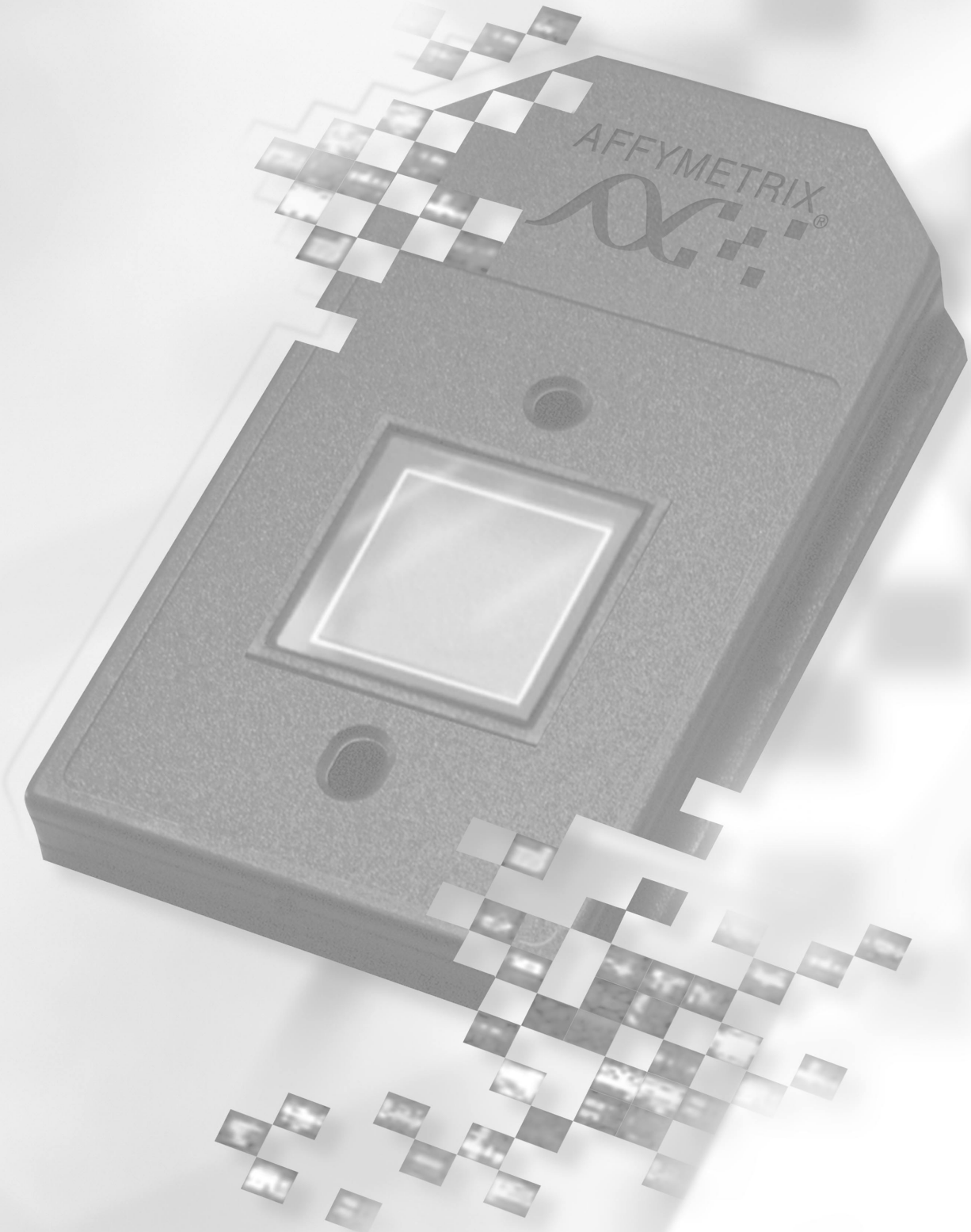
Problem	Likely Cause	Solution
Sample Quality		
High 3′/5′ ratio	Most often caused by degradation of the RNA during the isolation process.	Start with a fresh sample and minimize the possibility of RNase activity. Look for the presence of Ribosomal RNA bands on a non-denaturing agarose gel.
Low cRNA yield	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.	Carry over of phenol will inhibit the reaction. To remove phenol, wash the template twice with 70% or 80% ethanol.
	The presence of excess T7 promoter-containing primers can also decrease yield.	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.



Section 5, Appendix C

Section 5, Appendix C



List of Controls on GeneChip Probe Arrays

Table 5.C.1
Control Genes on GeneChip® probe arrays

Array Type	Origin of Organism	Control Gene Name	Utility for GeneChip® Experiments	Associated Affymetrix Products
Eukaryotic Arrays	synthetic	B2 Oligo	Grid alignment.	Control Oligo B2, P/N 900301 Section 2, Chapter 2 and Chapter 3
	E. coli	*bioB* *bioC* *bioD*	Antisense biotinylated cRNA probes are used as hybridization controls.	GeneChip Eukaryotic Hybridization Control Kit, P/N 900299 Section 2, Chapter 3
	P1 Bacteriophage	*cre*	Antisense biotinylated cRNA probes are used as hybridization controls.	
	B. subtilis	*dap* *thr* *trp* *phe* *lys*	Poly-A-tailed sense RNA can be produced by IVT and spiked into isolated RNA samples as controls for the labeling and hybridization process. The spikes can also be used to estimate assay sensitivity.	N/A Section 2, Chapter 2
GeneChip *E. coli* Genome Array	synthetic	B2 Oligo	Grid alignment.	Control Oligo B2, P/N 900301 Section 3, Chapter 3
	B. subtilis	*dap* *thr* *trp* *phe* *lys*	Sense RNA can be produced by IVT and spiked into purified sample RNA as control for the labeling and hybridization process. The spikes can also be used to estimate assay sensitivity.	N/A Section 3, Chapter 2 and Chapter 3

Registration

To automatically receive updates to the Expression Analysis Technical Manual, please register on our web site at:
www.affymetrix.com/support/technical/expression_registration.affx

GeneChip® Expression Analysis

Data Analysis Fundamentals

AFFYMETRIX®

Table of Contents

Introduction

The purpose of this manual is to provide users with a comprehensive description of different terms used in GeneChip® expression analysis, to present users with information on assessing sample and array quality, and to supply instructions on how to use the Affymetrix® Microarray Suite (MAS) software to analyze expression data. This handbook is a supplement to Affymetrix manuals and does not replace them. Brief descriptions of the different sections covered in this manual are as follows:

Guidelines for Assessing Sample and Array Quality

This section provides guidelines to assess array and sample quality.

Statistical Algorithms Reference Guide

This chapter focuses on the new Affymetrix Statistical Algorithms used in the expression analysis of GeneChip probe arrays. It provides a basic description of the mathematical concepts behind expression measurements for both single array and comparison analysis.

Single Array and Comparison Analyses

These sections provide step-by-step instructions for both single array and comparison analysis using Microarray Suite.

Basic Data Interpretation and Change Calculation Worksheet

These two sections cover step-by-step instructions for sorting data and calculating false change from data generated in Microarray Suite.

The NetAffx™ Analysis Center Summary

This chapter includes background information and functionality of the NetAffx site.

Relevant Publications

This chapter provides additional information and relevant publications that users might find helpful in gathering further information.

Appendices

Appendix A: Glossary

This appendix defines terminology used in expression analysis using GeneChip probe arrays.

Appendix B: GeneChip Probe Array Probe Set Name Designations

Background information of databases from which gene sequences are derived and the definitions of the probe set extensions are also available in this section.

Appendix C: Microarray Suite Expression Defaults

This appendix covers the different defaults that can be used in Microarray Suite.

Appendix D: File Types

This appendix describes all the file types associated with the GeneChip microarray platform.

To learn more about Affymetrix products or technology, please visit www.affymetrix.com.

Chapter 1 Guidelines for Assessing Sample and Array Quality

The purpose of this chapter is to help researchers establish quality control processes for gene expression analyses. To achieve this, Affymetrix has developed several controls which allow researchers to monitor assay performance and evaluate sample quality.

The following are a series of quality control parameters associated with assay and hybridization performance. Affymetrix highly encourages new users to create a **running log** of these parameters in order to monitor quality and potentially flag outlier samples. Evaluation of a particular sample should be based on the examination of all sample and array performance metrics.

RNA Sample QC

All RNA samples should meet assay quality standards to ensure the highest quality RNA is hybridized to the gene expression arrays. Researchers should run the initial total RNA on an agarose gel and examine the ribosomal RNA bands. Non-distinct ribosomal RNA bands indicate degradation.

260/280 absorbance readings should be measured for both total RNA and biotinylated cRNA. Acceptable 260/280 ratios fall in the range of 1.8 to 2.1. Ratios below 1.8 indicate possible protein contamination. Ratios above 2.1 indicate presence of degraded RNA, truncated cRNA transcripts, and/or excess free nucleotides.

For optimal results, please follow the protocols described in the Affymetrix® GeneChip® Expression Analysis Technical Manual.

Probe Array Image (.dat) Inspection

Inspect for the presence of image artifacts (i.e., high/low intensity spots, scratches, high regional, or overall background, etc.) on the array. Depending on the nature of the artifact, you may wish to apply an image mask (use the mouse to click and drag on the desired area, then select "Mask Cells" from the Edit menu) in order to eliminate affected probe cells from data analysis. Please contact your Field Applications Specialist (FAS) for further advice regarding image artifacts.

After scanning the probe array, the resulting image data created is stored on the hard drive of the GeneChip Analysis Suite/Microarray Suite workstation as a .dat file with the name of the scanned experiment. In the first step of the analysis, a grid is automatically placed over the .dat file demarcating each probe cell. One of the probe array library files, the .cif file, is used by Microarray Suite to determine the appropriate grid size used. Confirm the alignment of the grid by zooming in on each of the four corners and on the center of the image.

If the grid is not aligned correctly, adjust the alignment by placing the cursor on an outside edge or corner of the grid. The cursor image will change to a small double-headed arrow. The grid can then be adjusted using the arrow keys on the keyboard or by clicking and dragging the borders with the mouse.

Average Background and Noise Values

The Average Background and raw Noise (Q) values can be found either in the Analysis Info tab of the Data Analysis (.chp) file, or in the Expression Report (.rpt) file. Although there are no official guidelines regarding background, Affymetrix has found that typical Average Background values range from 20 to 100 for arrays scanned with GeneArray® Scanners calibrated to the new PMT setting (10% of maximum). For arrays scanned with GeneArray Scanners under the old PMT setting (100%), values range from 200 to 1,000. Arrays being compared should ideally have comparable background values.

Noise (Q) is a measure of the pixel-to-pixel variation of probe cells on a GeneChip array. The two main factors that contribute to noise are:

1. Electrical noise of the GeneArray Scanner.

2. Sample quality.

Each GeneArray Scanner has a unique inherent electrical noise associated with its operation. Since a significant portion of Q is based on electrical noise, absolute Q values among scanners will vary. Arrays being compared that were scanned on the same scanner should ideally have comparable Q values.

B2 Oligo Performance

The boundaries of the probe area (viewed upon opening the .dat/.cel file) are easily identified by the hybridization of B2 oligo, which is spiked into each hybridization cocktail. Hybridization of B2 is highlighted on the image by the following:

- The alternating pattern of intensities on the border
- The checkerboard pattern at each corner (Refer to Figure 1)
- The array name, located in the upper-left or upper-middle of the array (Refer to Figure 2)

B2 Oligo serves as a positive hybridization control and is used by the software to place a grid over the image. Variation in B2 hybridization intensities across the array is normal and does not indicate variation in hybridization efficiency. If the B2 intensities at the checkerboard corners are either too low or high or are skewed due to image artifacts, the grid will not align automatically. The user must align the grid manually using the mouse to click and drag each grid corner to its appropriate checkerboard corner.

The B2 oligonucleotide is available as part of the GeneChip Eukaryotic Hybridization Control Kit (P/N 900299 and 900362).

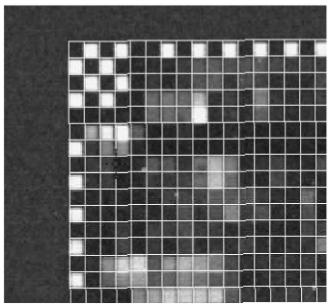


Figure 1. An example of B2 illuminating the corner and edges of the array.

Figure 2. The array name.

Poly-A Controls: *dap, lys, phe, thr, trp*

Dap, *lys*, *phe*, *thr,* and *trp* are *B. subtilis* genes that have been modified by the addition of poly-A tails, and then cloned into pBluescript vectors, which contain both T3 and T7 promoter sequences. Amplifying these poly-A controls with T3 RNA polymerase will yield sense RNAs, which can be spiked into a complex RNA sample, carried through the sample preparation process, and evaluated like internal control genes. Amplifying these controls with T7 RNA polymerase and biotinylated ribonucleotides will yield antisense cRNAs, which can be spiked into a hybridization cocktail and evaluated like the 20x Eukaryotic Hybridization Controls (*bioB*, *bioC*, *bioD*, and *cre*).

Details on poly-A preparation are described in the GeneChip Expression Analysis Technical Manual (Section 2 and Section 3, Chapter 2)

Hybridization Controls: *bioB, bioC, bioD,* and *cre*

BioB, bioC, and *bioD* represent genes in the biotin synthesis pathway of *E. coli*. *Cre* is the recombinase gene from P1 bacteriophage. The GeneChip Eukaryotic Hybridization Control Kit (P/N 900299 and 900362) contains 20x Eukaryotic Hybridization Controls that are composed of a mixture of biotin-labeled cRNA transcripts of *bioB*, *bioC*, *bioD,* and *cre*, prepared in staggered concentrations (1.5 pM, 5 pM, 25 pM, and 100 pM for *bioB*, *bioC*, *bioD,* and *cre*, respectively).

The 20x Eukaryotic Hybridization Controls are spiked into the hybridization cocktail, independent of RNA sample preparation, and are thus used to evaluate sample hybridization efficiency to gene expression arrays. *BioB* is at the level of assay sensitivity (1:100,000 complexity ratio) and should be called "Present" at least 50% of the time. *BioC*, *bioD,* and *cre* should always be called "Present" with increasing Signal values, reflecting their relative concentrations.

The 20x Eukaryotic Hybridization Controls can be used to indirectly assess RNA sample quality among replicates. When global scaling is performed, the overall intensity for each array is determined and is compared to a Target Intensity value in order to calculate the appropriate scaling factor. The overall intensity for a degraded RNA sample,

or a sample that has not been properly amplified and labeled, will have a lower overall intensity when compared to a normal replicate sample. Thus, when the two arrays are globally scaled to the same Target Intensity, the scaling factor for the "bad" sample will be much higher than the "good" sample. However, since the 20x Eukaryotic Hybridization Controls are added to each replicate sample equally (and are independent of RNA sample quality), the intensities of the *bioB*, *bioC*, *bioD,* and *cre* probe sets will be approximately equal. As a result, the Signal values (adjusted by scaling factor) for these control probe sets on the "bad" array will be adjusted higher relative to the Signal values for the control probe sets on the "good" array.

Internal Control Genes

For the majority of GeneChip expression arrays, actin and GAPDH are used to assess RNA sample and assay quality. Specifically, the Signal values of the 3' probe sets for actin and GAPDH are compared to the Signal values of the corresponding 5' probe sets. The ratio of the 3' probe set to the 5' probe set is generally no more than 3. Since the gene expression assay has an inherent 3' bias (i.e., antisense cRNA is transcribed from the sense strand of the synthesized ds cDNA, via the incorporated T7 promoter), a high 3' to 5' ratio may indicate degraded RNA or inefficient transcription of ds cDNA or biotinylated cRNA. 3' to 5' ratios for internal controls are displayed in the Expression Report (.rpt) file.

There are occasions when the 3' to 5' ratio of one internal control gene is normal, but the 3' to 5' ratio of another internal control gene is high. Since the gene expression assay is not biased in terms of the transcripts being amplified, this discrepancy in 3' to 5' ratios is most likely due to a specific transcript-related or image artifact issue and is not an indication of overall sample and assay quality.

Percent Genes Present

The number of probe sets called "Present" relative to the total number of probe sets on the array is displayed as a percentage in the Expression Report (.rpt) file. Percent Present (%P) values depend on multiple factors including cell/tissue type, biological or environmental stimuli, probe array type, and overall quality of RNA. Replicate samples should have similar %P values. Extremely low %P values are a possible indication of poor sample quality. However, the use of this metric must be evaluated carefully and in conjunction with the other sample and assay quality metrics described in this document.

Scaling and Normalization Factors

Details regarding Scaling and Normalization are listed in the Affymetrix Microarray Suite User Guide Version 5.0, Appendix D. Scaling and normalization factors can be found either in the Analysis Info tab of the .chp file output or in the Expression Report (.rpt) file.

For the majority of experiments where a relatively small subset of transcripts is changing, the global method of scaling/normalization is recommended. In this case, since the majority of transcripts are not changing among samples, the overall intensities of the arrays should be similar. Differences in overall intensity are most likely due to assay variables including pipetting error, hybridization, washing, and staining efficiencies, which are all independent of relative transcript concentration. Applying the global method corrects for these variables. For global scaling, it is imperative that the same Target Intensity value is applied to all arrays being compared.

For some experiments, where a relatively large subset of transcripts is affected, the "Selected Probe Sets" method of scaling/normalization is recommended. The global approach does not make sense in this situation since the overall intensities among arrays are no longer comparable. Differences in overall intensity are due to biological and/or environmental conditions. Applying the global method skews the relative transcript concentrations. One option for users of the HG-U133 Set is to apply the "Selected Probe Sets" method using the 100 Normalization Control probe sets.

For replicates and comparisons involving a relatively small number of changes, the scaling/normalization factors (calculated by the global method) should be comparable among arrays. Larger discrepancies among scaling/normalization factors (e.g., three-fold or greater) may indicate significant assay variability or sample degradation leading to noisier data.

Scaling/normalization factors calculated by the "Selected Probe Sets" method should also be equivalent for arrays being compared. Larger discrepancies between scaling/normalization factors may indicate significant assay or biological variability or degradation of the transcripts used for scaling/normalization, which leads to noisier data.

Chapter 2 Statistical Algorithms Reference Guide

This chapter is a reference for the Affymetrix Statistical Algorithms used in the expression analysis of GeneChip probe arrays. It provides the user with a basic description of the mathematical concepts behind expression measurements for either single array or comparison analysis.

The Statistical Algorithms were implemented in Affymetrix Microarray Suite Version 5.0. Previous versions of the GeneChip Analysis Suite and Affymetrix Microarray Suite used the Empirical Algorithms.

The Statistical Algorithms were developed using standard statistical techniques. The performance was validated using an experimental design called the Latin Square. In this experimental design, transcripts, naturally absent in the complex background, were spiked in at known concentrations.

Single Array Analysis

Single array analysis can be used to build databases of gene expression profiles, facilitate sample classification and transcript clustering, and monitor gross expression characteristics. In addition, the analyses provide the initial data required to perform comparisons between experiment and baseline arrays.

This analysis generates a **Detection *p*-value** which is evaluated against user-definable cut-offs to determine the **Detection** call. This call indicates whether a transcript is reliably detected (Present) or not detected (Absent). Additionally, a **Signal** value is calculated which assigns a relative measure of abundance to the transcript.

Figure 1 illustrates the output of Single Array Analysis in Microarray Suite 5.0.

	Stat Pairs	Stat Pairs Used	Signal	Detection	Detection p-value
37984_s_at	16	16	92.2	P	0.000218
32102_at	16	16	59.5	P	0.000218
37900_at	16	16	72.6	P	0.000219
31697_s_at	16	16	664.2	P	0.000219
40567_at	16	16	502.3	P	0.000219
35808_at	16	16	212.6	P	0.000219
34819_at	16	16	143.0	P	0.000219
35787_at	16	16	295.7	P	0.000219
35758_at	16	16	301.0	P	0.000219
34817_s_at	16	16	339.6	P	0.000219
34644_at	16	16	723.9	P	0.000219
34608_at	16	16	3313.0	P	0.000219

Figure 1. Data analysis output (.chp file) for a Single Array Analysis includes Stat Pairs, Stat Pairs Used, Signal, Detection, and the Detection *p*-value.

Detection Algorithm

The Detection algorithm uses probe pair intensities to generate a Detection *p*-value and assign a Present, Marginal, or Absent call. Each probe pair in a probe set is considered as having a potential vote in determining whether the measured transcript is detected (Present) or not detected (Absent). The vote is described by a value called the Discrimination score [R]. The score is calculated for each probe pair and is compared to a predefined threshold Tau. Probe pairs with scores *higher* than Tau vote for the *presence* of the transcript. Probe pairs with scores *lower* than Tau vote for the *absence* of the transcript. The voting result is summarized as a *p*-value. The higher the discrimination scores are above Tau, the smaller the *p*-value and the more likely the transcript will be Present. The lower the discrimination scores are below Tau, the larger the *p*-value and the more likely the transcript will be Absent. The *p*-value associated with this test reflects the confidence of the Detection call.

Detection *p*-value

A two-step procedure determines the Detection *p*-value for a given probe set.

1. Calculate the Discrimination score [R] for each probe pair.

2. Test the Discrimination scores against the user-definable threshold Tau.

The Discrimination score is a basic property of a probe pair that describes its ability to detect its intended target. It measures the target-specific intensity difference of the probe pair (PM-MM) relative to its overall hybridization intensity (PM+MM):

$$R = (PM - MM) / (PM + MM)$$

For example, if the PM is much larger than the MM, the Discrimination score for that probe pair will be close to 1.0 (e.g., probe pair 1 in Figure 2). If the Discrimination scores are close to 1.0 for the majority of the probe pairs, the calculated Detection *p*-value will be lower (more significant). A lower *p*-value is a reliable indicator that the result is valid and that the probability of error in the calculation is small. Conversely, if the MM is larger than or equal to the PM, then the Discrimination score for that probe pair will be negative or zero (e.g., probe pairs 8, 9, and 10 in Figure 2). If the Discrimination scores are low for the majority of the probe pairs, the calculated Detection *p*-value will be higher (less significant).

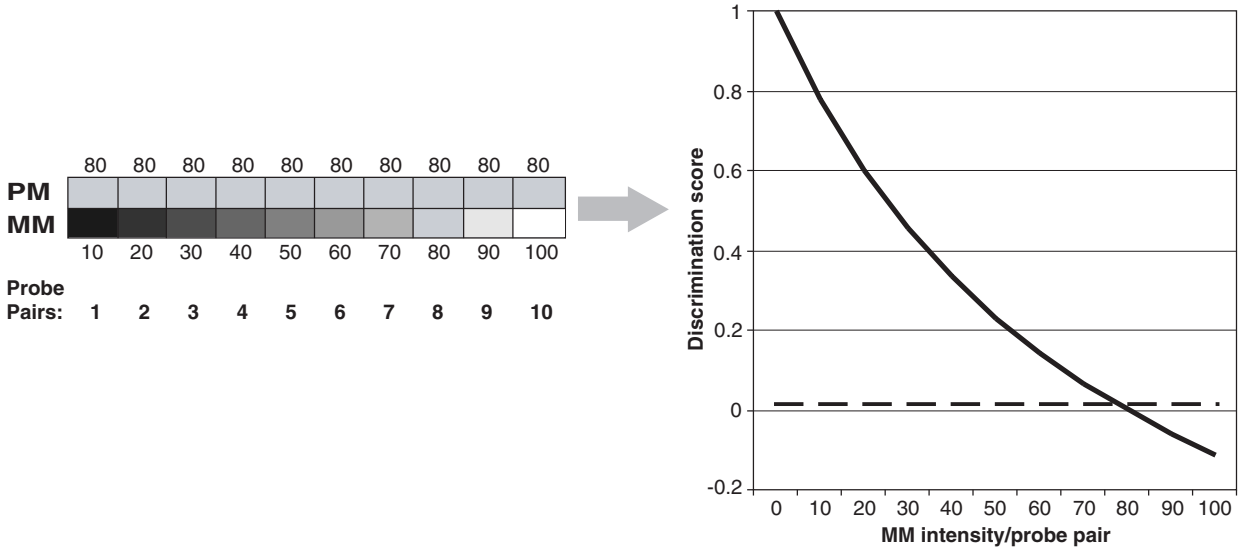


Figure 2. In this hypothetical probe set, the Perfect Match (PM) intensity is 80 and the Mismatch (MM) intensity for each probe pair increases from 10 to 100. The probe pairs are numbered from 1 to 10. As the Mismatch (MM) probe cell intensity, plotted on the x-axis, increases and becomes equal to or greater than the Perfect Match (PM) intensity, the Discrimination score decreases as plotted on the y-axis. More specifically, as the intensity of the Mismatch (MM) increases, our ability to discriminate between the PM and MM decreases. The dashed line is the user-definable parameter Tau (default = 0.015).

The next step toward the calculation of a Detection *p*-value is the comparison of each Discrimination score to the user-definable threshold Tau. Tau is a small positive number that can be adjusted to increase or decrease sensitivity and/or specificity of the analysis (default value = 0.015). The One-Sided Wilcoxon's Signed Rank test is the statistical method employed to generate the Detection *p*-value. It assigns each probe pair a rank based on how far the probe pair Discrimination score is from Tau.

Tunable Parameter Tip:

Increasing the threshold Tau can reduce the number of false Present calls, but may also reduce the number of true Present calls. Note: Changing Tau directly influences the calculation of the Detection *p*-value.

Detection Call

The user-modifiable Detection *p*-value cut-offs, Alpha 1 (α_1) and Alpha 2 (α_2) (See Figure 3), provide boundaries for defining Present, Marginal, or Absent calls. At the default settings, determined for probe sets with 16–20 probe pairs (defaults α_1 = 0.04 and α_2 = 0.06), any *p*-value that falls below α_1 is assigned a Present call, and above α_2 is assigned an Absent call. Marginal calls are given to probe sets which have *p*-values between α_1 and α_2 (see Figure 3).
The *p*-value cut-offs can be adjusted to increase or decrease sensitivity and specificity.

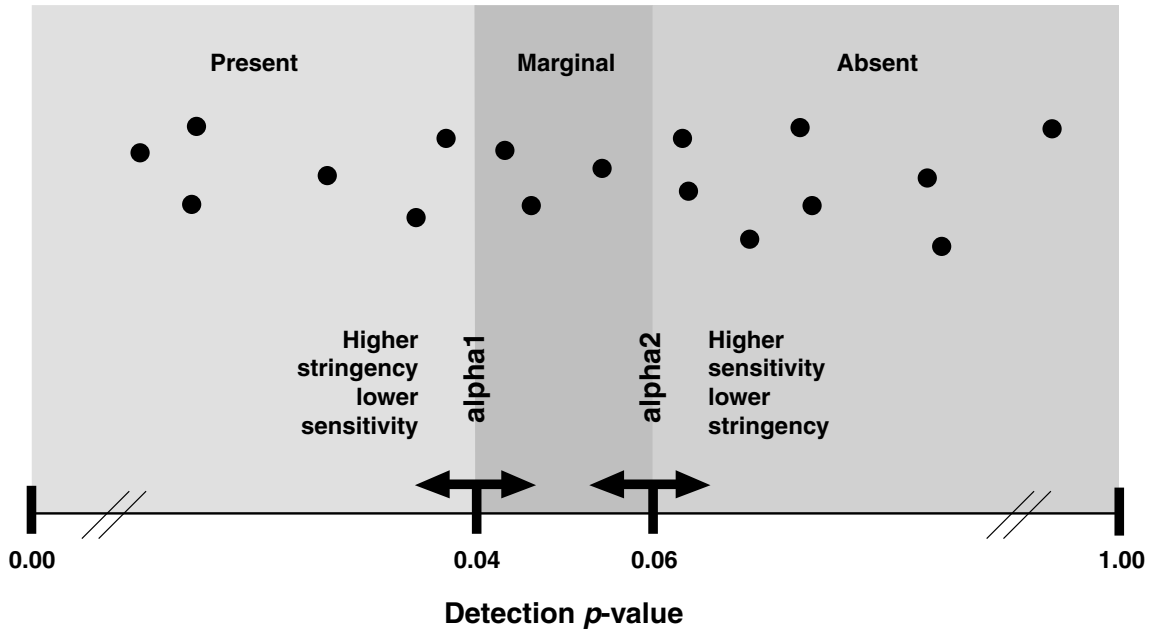


Figure 3. Significance levels α_1 and α_2 define cut-offs of *p*-values for Detection calls. Please note that these cut-offs are for probe sets with 16–20 probe pairs.

It is important to note that prior to the two-step Detection *p*-value calculation, the level of photomultiplier saturation for each probe pair is evaluated. If all probe pairs in a probe set are saturated, the probe set is immediately given a Present call. Note that a probe pair is rejected from further analysis when a Mismatch (MM) probe cell is saturated (MM = 46,000 for the 2500 GeneArray Scanner).

In summary, the Detection Algorithm assesses probe pair saturation, calculates a Detection *p*-value and assigns a Present, Marginal, or Absent call.

Signal Algorithm

Signal is a quantitative metric calculated for each probe set, which represents the relative level of expression of a transcript. Signal is calculated using the One-Step Tukey's Biweight Estimate which yields a robust weighted mean that is relatively insensitive to outliers, even when extreme.

Similar to the Detection algorithm, each probe pair in a probe set is considered as having a potential vote in determining the Signal value. The vote, in this case, is defined as an estimate of the real signal due to hybridization of the target. The mismatch intensity is used to estimate stray signal. The real signal is estimated by taking the log of the Perfect Match intensity after subtracting the stray signal estimate. The probe pair vote is weighted more strongly if this probe pair Signal value is closer to the median value for a probe set. Once the weight of each probe pair is determined, the mean of the weighted intensity values for a probe set is identified. This mean value is corrected back to linear scale and is output as Signal.

When the Mismatch intensity is lower than the Perfect Match intensity, then the Mismatch is informative and provides an estimate of the stray signal. Rules are employed in the Signal algorithm to ensure that negative Signal values are not calculated. Negative values do not make physiological sense and make further data processing, such as log transformations, difficult. Mismatch values can be higher than Perfect Match values for a number of reasons, such as cross hybridization. If the Mismatch is higher than the Perfect Match, the Mismatch provides no additional information about the estimate of stray signal. Therefore, an imputed value called Idealized Mismatch (IM) is used instead of the uninformative Mismatch (see Figure 4).

The following rules are applied:

Rule 1: If the Mismatch value is less than the Perfect Match value, then the Mismatch value is considered informative and the intensity value is used directly as an estimate of stray signal.

Rule 2: If the Mismatch probe cells are generally informative across the probe set except for a few Mismatches, an adjusted Mismatch value is used for uninformative Mismatches based on the biweight mean of the Perfect Match and Mismatch ratio.

Rule 3: If the Mismatch probe cells are generally uninformative, the uninformative Mismatches are replaced with a value that is slightly smaller than the Perfect Match. These probe sets are generally called Absent by the Detection algorithm.

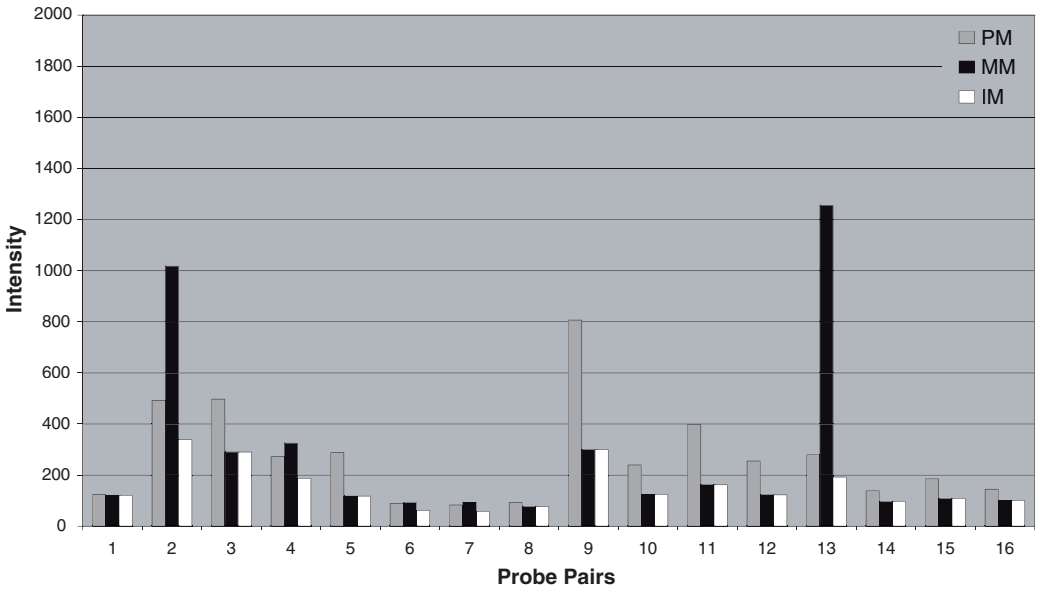


Figure 4. The grey bars illustrate the Perfect Match (PM) intensities and black bars the Mismatch (MM) intensities across a 16-probe pair probe set. The white bars, Idealized Mismatch (IM), are the intensities of the Mismatch based on the Signal rules. In this example, most of the Perfect Match intensities are higher than the Mismatch intensities and therefore Mismatch values can be used directly (e.g., probe pair 9). When the Mismatch is larger than the Perfect Match (e.g., probe pairs 2, 4, and 13) the IM value is used instead of the Mismatch.

Comparison Analysis (Experiment versus Baseline arrays)

In a Comparison Analysis, two samples, hybridized to two GeneChip probe arrays of the same type, are compared against each other in order to detect and quantify changes in gene expression. One array is designated as the baseline and the other as an experiment. The analysis compares the difference values (PM-MM) of each probe pair in the baseline array to its matching probe pair on the experiment array. Two sets of algorithms are used to generate change significance and change quantity metrics for every probe set. A change algorithm generates a **Change *p*-value** and an associated **Change**. A second algorithm produces a quantitative estimate of the change in gene expression in the form of **Signal Log Ratio**.

Figure 5 illustrates the output of Comparison Analysis in Microarray Suite 5.0.

	Stat Common Pairs	Signal Log Ratio	Signal Log Ratio Low	Signal Log Ratio High	Change	Change p-value
35839_at	16	0.3	0.2	0.4	I	0.000014
1799_at	16	0.9	0.5	1.3	I	0.000015
35985_at	16	0.4	0.3	0.5	I	0.000015
34696_at	16	0.4	-0.1	0.9	I	0.000023
31356_at	16	1.8	0.8	2.8	I	0.000025
35202_at	16	0.4	0.2	0.6	I	0.000027
39651_at	16	0.4	0.3	0.5	I	0.000029
39777_at	16	0.4	0.1	0.6	I	0.000031
37610_at	16	0.4	0.2	0.5	I	0.000034
32070_at	16	0.3	0.2	0.4	I	0.000034
1581_s_at	16	0.7	0.1	1.3	I	0.000037
35283_at	16	0.5	0.3	0.6	I	0.000037

Figure 5. Data analysis output (.chp file) for a Comparison Analysis includes Stat Common Pairs, Signal Log Ratio, Signal Log Ratio Low, Signal Log Ratio High, Change, and the Change *p*-value.

Before comparing two arrays, scaling or normalization methods must be applied. Scaling and normalization correct for variations between two arrays. Two primary sources of variation in array experiments are biological and technical differences. Biological differences may arise from many sources such as genetic background, growth conditions, dissection, time, weight, sex, age, and replication. Technical variation can be due to experimental variables such as quality and quantity of target hybridized, reagents, stain, and handling error. The minimization of variation is essential, but scaling and normalization techniques provide a means to remove differences and facilitate comparison analysis.

Normalization and scaling techniques can be applied by using data from a selected user-defined group of probe sets, or from all probe sets. When normalization is applied, the intensity of the probe sets (or selected probe sets) from the experiment array are normalized to the intensity of the probe sets (or selected probe sets) on the baseline array. When scaling is applied, the intensity of the probe sets (or selected probe sets) from the experimental array and that from the baseline array are scaled to a user-defined target intensity. In general, global scaling (scaling to all probe sets) is the preferred method when comparing two arrays.

An additional normalization factor is defined in the Robust Normalization section described below. This 'robust normalization,' which is not user-modifiable, accounts for unique probe set characteristics due to sequence-dependent factors, such as affinity of the target to the probe and linearity of hybridization of each probe pair in the probe set.

Change Algorithm

As in the Single Array Analysis, the Wilcoxon's Signed Rank test is used in Comparison Analysis to derive biologically meaningful results from the raw probe cell intensities on expression arrays. During a Comparison Analysis, each probe set on the experiment array is compared to its counterpart on the baseline array, and a Change *p*-value is calculated indicating an increase, decrease, or no change in gene expression. User-defined cut-offs (gammas) are applied to generate discrete Change calls (Increase, Marginal Increase, No Change, Marginal Decrease, or Decrease).

Robust Normalization

After scaling or normalization of the array (discussed in the Comparison Analysis overview), a further robust normalization of the probe set is calculated. Once the initial probe set normalization factor is determined, two additional normalization factors are calculated that are slightly higher and slightly lower than the original. The range by which the normalization factor is adjusted up and down is specified by a user-modified parameter called perturbation. This supplementary normalization accounts for unique probe set characteristics due to sequence dependent factors, such as affinity and linearity. More specifically, this approach addresses the inevitable error of using an average intensity of the majority of probes (or selected probes) on the array as the normalization factor for every probe set on the array. The noise from this error, if unattenuated, would result in many false positives in expression level changes between the two arrays being compared. The perturbation value directly affects the subsequent *p*-value calculation. Of the *p*-values that result from applying the calculated normalization factor and its two perturbed variants, the one that is most conservative is used to estimate whether any change in level is justified by the data. The lowest value for perturbation is 1.00, indicating no perturbation. The highest perturbation value allowed is set at 1.49. Increasing the perturbation value widens the range allowed before a change is called. For example, changes that were called Increase with a smaller perturbation value, may be called No Change with a higher perturbation value. A default was established at 1.1 based on calls made from the Latin Square data set. The perturbation factor and the Latin Square data set are described in more detail in the Affymetrix Technical Notes referenced in the back of this guide.

Change *p*-value

The Wilcoxon's Signed Rank test uses the differences between Perfect Match and Mismatch intensities, as well as the differences between Perfect Match intensities and background to compute each Change *p*-value.

From Wilcoxon's Signed Rank test, a total of three, one-sided *p*-values are computed for each probe set. These are combined to give one final *p*-value which is provided in the data analysis output (.chp file). The *p*-value ranges in scale from 0.0 to 1.0 and provides a measure of the likelihood of change and direction. Values close to 0.0 indicate likelihood for an increase in transcript expression level in the experiment array compared to the baseline, whereas values close to 1.0 indicate likelihood for a decrease in transcript expression level. Values near 0.5 indicate a weak likelihood for change in either direction (see Figure 6). Hence, the *p*-value scale is used to generate discrete change calls using thresholds. These thresholds will be described in the Change Call section.

	Signal Log Ratio	Signal Log Ratio Low	Signal Log Ratio High	Change	Change p-value
l25069_s_at	-0.4	-0.5	-0.2	D	0.999979
m12303_s_at	-0.3	-0.4	-0.2	D	0.999999
m12347_s_at	-0.3	-0.3	-0.2	D	0.999997
m21495_f_at	-0.2	-0.3	-0.2	D	0.999763
m29793_s_at	-0.2	-0.2	-0.1	D	0.998697
m34173_at	-0.3	-0.4	-0.3	D	0.999505
m62867_s_at	-0.4	-0.5	-0.3	D	1.000000
n28127_rc_at	-0.4	-0.5	-0.3	D	0.997989
l07577_s_at	-0.4	-0.5	-0.4	D	1.000000
aa000380_s_at	0.4	0.2	0.6	I	0.002283
aa002704_at	0.7	0.5	0.9	I	0.000150
aa002761_s_at	0.4	0.2	0.6	I	0.000226
aa000148_s_at	0.3	0.2	0.4	I	0.001246
aa009154_s_at	0.4	0.3	0.6	I	0.000028
aa013647_s_at	0.3	0.2	0.4	I	0.002011
AA023300_at	0.2	-0.1	0.6	I	0.001140
aa023407_s_at	0.9	0.6	1.2	I	0.000088
AA408234_rc_g_at	0.5	0.3	0.6	I	0.000001

Figure 6. Data analysis output (.chp file) for a Comparison Analysis illustrating Change *p*-values with the associated Increase (I) or Decrease (D) call. Increase calls have Change *p*-values closer to zero and Decrease calls have Change *p*-values closer to one.

Tunable Parameter Tip:

Increasing the perturbation value can reduce the number of false changes, but may also decrease the detection of true changes. Note: Changing perturbation factor affects the calculation of the *p*-value directly.

Change Call

The final Change *p*-value described above is categorized by cutoff values called gamma1 (γ_1) and gamma2 (γ_2) (see Figure 7). These cut-offs provide boundaries for the Change calls: Increase (I), Marginal Increase (MI), No Change (NC), Marginal Decrease (MD), or Decrease (D).

The user does not directly set γ_1 and γ_2; rather each is derived from two user-adjustable parameters, γL and γH. In the case of γ_1, the two user-adjustable parameters are called γ_1L and γ_1H (defaults for probe sets with 15-20 probe pairs: γ_1L= 0.0025 and γ_1H= 0.0025), which define the lower and upper boundaries for γ_1. Gamma2 (γ_2) is computed as a linear interpolation of γ_2L and γ_2H (defaults for probe sets with 15-20 probe pairs: γ_2L= 0.003 and γ_2H= 0.003) in an analogous fashion.

The ability to adjust the stringency of calls associated with high and low signal ranges independently makes it possible to compensate for effects that influence calls based on low and high signals. This feature, however, is not used by default because the defaults are set as γ_1L = γ_1H and γ_2L = γ_2H

It is important to note that, like in Detection *p*-value calculation, the level of photomultiplier saturation for each probe pair is evaluated. In the computation of Change *p*-value, any saturated probe cell, either in the Perfect Match or Mismatch, is rejected from analysis. The number of discarded cells can be determined from the Stat Common Pairs parameter.

In summary, the Change algorithm assesses probe pair saturation, calculates a Change *p*-value, and assigns an Increase, Marginal Increase, No Change, Marginal Decrease, or Decrease call.

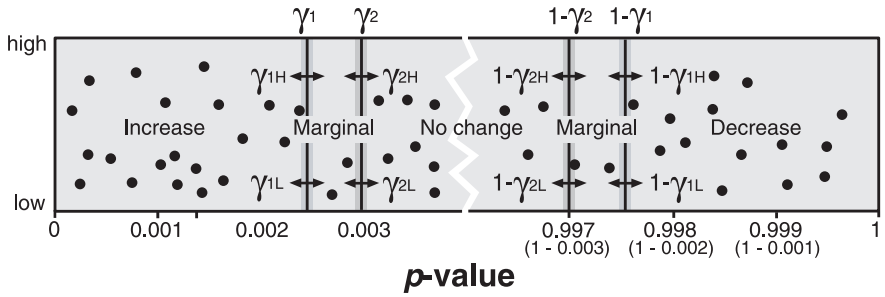


Figure 7. A representation of a range of *p*-values for a data set. The Y-axis is the probe set signal. The arrows on the vertical bars represent the adjustable γ values. The γ_1 value is a linear interpolation of γ_1L and γ_1H. Similarly γ_2 is derived from γ_2L and γ_2H.

Signal Log Ratio Algorithm

The Signal Log Ratio estimates the magnitude and direction of change of a transcript when two arrays are compared (experiment versus baseline). It is calculated by comparing each probe pair on the experiment array to the corresponding probe pair on the baseline array. This strategy cancels out differences due to different probe binding coefficients and is, therefore, more accurate than a single array analysis.

As with Signal, this number is computed using a one-step Tukey's Biweight method by taking a mean of the log ratios of probe pair intensities across the two arrays. This approach helps to cancel out differences in individual probe intensities, since ratios are derived at the probe level, before computing the Signal Log Ratio. The log scale used is base 2, making it intuitive to interpret the Signal Log Ratios in terms of multiples of two. Thus, a Signal Log Ratio of 1.0 indicates an increase of the transcript level by 2 fold and -1.0 indicates a decrease by 2 fold. A Signal Log Ratio of zero would indicate no change.

The Tukey's Biweight method gives an estimate of the amount of variation in the data, exactly as standard deviation measures the amount of variation for an average. From the scale of variation of the data, confidence intervals are generated measuring the amount of variation in the biweight estimate. A 95% confidence interval indicates a range of values, which will contain the true value 95% of the time. Small confidence intervals indicate that the data is more precise while large confidence intervals reflect uncertainty in estimating the true value. For example, the Signal Log Ratio for some transcripts may be measured as 1.0, with a range of 0.5 to 1.5 from low to high. For 95% of transcripts

with such results, the true Signal Log Ratio will lie somewhere in that range. A set of noisy experiments might also report a Signal Log Ratio of 1.0, but with a range of -0.5 to 2.5, indicating that the true effect could easily be zero, since the uncertainty in the data is very large. The confidence intervals associated with Signal Log Ratio are calculated from the variation between probes, which may not reflect the full extent of experimental variation.

Terminology Comparison Table (Statistical Algorithms versus Empirical Algorithms)

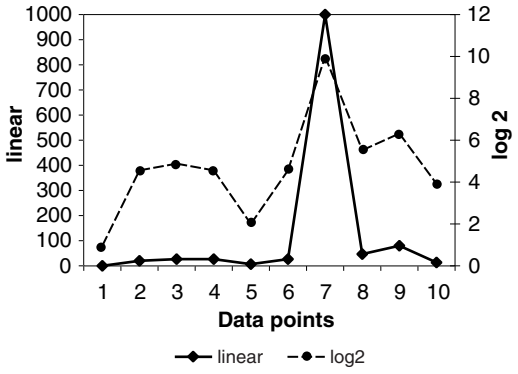
Statistical Algorithms	Empirical Algorithms
Signal	Average Difference
Detection	Absolute Call
Change	Difference Call
Signal Log Ratio	Fold Change

The Logic of Logs

Quantitative changes in gene expression are reported as a Signal Log Ratio in the Statistical Algorithms as opposed to a Fold Change that was reported in the Empirical Algorithms.

The Benefit of Logs:

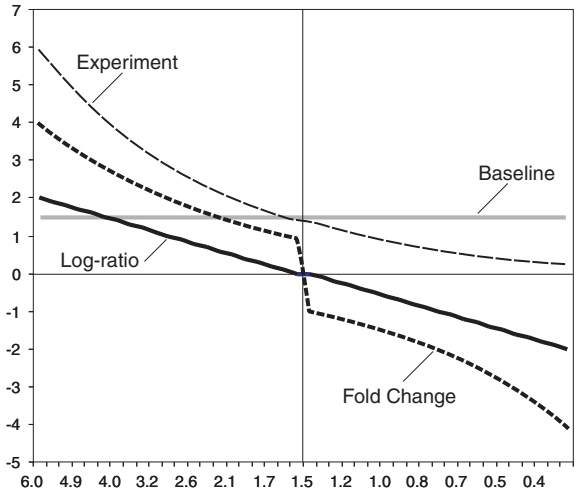
Hybridized probe intensities tend to be distributed over exponential space due to hybridization behavior that is governed by exponential functions of sequence-dependent base-pairing energetics. Thus, log transformation is an appropriate process for analyzing hybridization data. Some of the benefits are apparent in this graph where we show the same data set plotted on two scales. When the data is plotted on a linear scale (solid) the single, high data point (7) overwhelms the graph and obscures information contained in the low values. When the same data is plotted on a Log_2 scale (dashed line), we can see variations in the low values as well as the very high values.



Signal Log Ratio vs. Fold Change

In this graph, Signal Log Ratio is compared to Fold Change in a hypothetical experiment. Baseline values were set to 1.5 and experiment values were reduced progressively from 6 to 0.375. The X-axis illustrates the values that were decreased in the hypothetical experiment. The Y-axis represents units (e.g., signal log ratio, fold change, or signal for baseline and experiment).

There is a discontinuous transition where the experiment and the baseline converge and the fold change approaches 1 or -1. At this point (smaller changes), the fold change is less sensitive. Since we use log_2, a Signal Log Ratio of 1 equals a Fold Change of 2 and a Signal Log Ratio of 2 equals a Fold Change of 4. Alternatively, use the following formula:



$$\text{Fold Change} = \begin{cases} 2^{\text{Signal Log Ratio}} & \text{Signal Log Ratio} \geq 0 \\ (-1) * 2^{-(\text{Signal Log Ratio})} & \text{Signal Log Ratio} < 0 \end{cases}$$

Chapter 3 Single Array Analysis

This section describes a basic GeneChip array analysis procedure that can be applied to many analysis situations. This procedure can be modified to account for specific experimental situations. It is highly recommended that before attempting to modify this procedure, users familiarize themselves with the scaling strategies and settings involved in GeneChip array analysis. More detailed information can be found in the Affymetrix Microarray Suite User Guide Version 5.0 (P/N 700293) or by contacting your Affymetrix Field Applications Specialist.

The following instructions assume that a GeneChip probe array has been hybridized, washed, stained, and scanned according to the directions detailed in the Affymetrix GeneChip Expression Analysis Technical Manual. Upon completion of the scan, the image file (.dat) is displayed in the Microarray Suite 5.0 software. After analysis of arrays, the procedures in the preceding chapters can be used to assess the quality the data generated.

These instructions relate to analyses performed in Microarray Suite 5.0. Notes for using Microarray Suite 4.0 can be found at the end of this chapter.

Quality Assessment of .dat Image

Prior to conducting array analysis, the quality of the array image (.dat file) should be assessed following the guidelines in Chapter 1 of this training manual.

NOTE: Refer to Chapter 1 to aid in quality assessment of the array.

Select a Scaling Strategy

These instructions use a global scaling strategy that sets the average signal intensity of the array to a Target Signal of 500. The key assumption of the global scaling strategy is that there are few changes in gene expression between the arrays being analyzed. This is a common strategy used by many users, however, it should be noted that this strategy is not appropriate for all experiments. Further discussion on scaling strategies and how to implement them can be found in Appendix D of the Microarray Suite User Guide Version 5.0 or by contacting your Affymetrix Field Applications Specialist.

Expression Analysis Set-Up

A single array analysis will create a .chp file from a .cel image file. Microarray Suite automatically generates the .cel image file from the .dat file. To perform a single array analysis, settings relating to file locations and the analysis must first be defined.

Specifying File-Related Settings

1. Select “Defaults” from the “Tools” pull-down menu.

2. Select the “Analysis Settings” tab.

    a) Check “Prompt For Output File” to ensure display of output file name for confirmation or editing. With this option checked, Microarray Suite will prompt for new file names for each analysis preventing unintentional overwrite.

    b) Check “Display Settings When Analyzing Data” to ensure display of expression settings for confirmation or editing.

3. Select “File Locations” tab to verify:

    a) the Location listed to the right of Probe Information is the directory containing the library files.

    b) the Location listed to the right of Fluidics Protocols is the directory containing the fluidics protocols.

    c) the Location listed to the right of Experiment Data is the directory containing the data files to be analyzed.

NOTE: Errors are commonly found in Microarray Suite due to incorrectly set file locations.

4. Select the “Database” tab. Select “Disk Files” mode to direct where file information will be saved.

NOTE: “Disk Files” refers to data storage on the local hard drive. “Affymetrix® LIMS” refers to storage on a dedicated server using the Affymetrix Laboratory Information Management System.

5. Select “OK.”

Expression Analysis Settings

Select “Analysis Settings>Expression” from the “Tools” pull-down menu. The “Expression Analysis Settings” dialogue box opens.

1. Select the “Probe Array Type” to be analyzed from the drop-down menu.

2. Select the “Scaling” tab.

    a) Select “All Probe Sets” and set “Target Signal” to 500 or to desired Target Signal.

3. Select the “Normalization” tab.

    a) Select “User Defined” and place a “1” in the ″Normalization Value” box. This ensures that no normalization procedure is applied to the data. Normalization is not necessary as the data is being scaled. Further information can be found in Appendix D of Microarray Suite User’s Guide Version 5.0.

4. Select the “Probe Mask” tab. This feature is used to mask user-defined probe cells.

    a) Ensure that the “Use Probe Mask File” option is unchecked.

5. Select the “Baseline” tab. For single array analysis no baseline file should be used.

    a) Ensure “Use Baseline File Comparison” is unchecked.

6. Select the “Parameters” tab.

    a) Confirm default settings appropriate to the version of Microarray Suite and the array being analyzed as specified in Appendix C of this training manual.

NOTE: These Settings should not be adjusted unless the user has advanced experience with the Affymetrix GeneChip system.

7. Once all settings have been adjusted or confirmed select “OK” to define settings and close the dialogue box.

Performing Single Array Analysis

1. Open the file you wish to analyze (.dat or .cel) by double clicking it in the data file tree. Alternatively, select “Open” from the “File” pull-down menu and select the image file you wish to analyze.

2. Select “Analysis” from the “Run” pull-down menu.

    a) Verify the .chp file name. The default corresponds to the name of the .exp and .dat file names. Edit the .chp file name, if necessary, and click “OK.”

NOTE: Microarray Suite will overwrite a .chp file if the filename is the same as an existing .chp file in the directory.

    b) Verify “Expression Analysis Settings” in the subsequent pop-up window as previously set in the above Expression Analysis Settings section and select “OK” to begin analysis and generate the analysis results file (.chp).

    c) The Microarray Suite status window will indicate that the analysis has started.

3. Once the analysis is complete, generate an Expression Analysis report file (.rpt) and review the quality control metrics.

    a) To generate the report select “Report” from the “File” pull-down menu.

    b) Select the appropriate analysis results file (.chp).

NOTE: Alternatively, you can highlight the appropriate .chp file in the data file tree, right click on the mouse and select "Report."

c) Review the quality control data.

NOTE: See Chapter 1 for detailed explanations.

- Review *bioB*, *bioC*, *bioD,* and *cre* sensitivity spikes.
- Review percent present determination.
- Review internal control 3'/5' ratios.
- Review noise (Raw Q).
- Review background.

d) Return to the .chp file by closing the Report (.rpt) file.

NOTE: The open .chp file data is displayed in the Expression Analysis Window (EAW) and can be accessed by clicking on the Expression Analysis button in the Microarray Suite shortcuts window.

4. Select the "Pivot" tab at the bottom of the analysis results .chp file. The Pivot table displays analysis output and descriptions for each transcript represented on the probe array. The far-left column contains the Affymetrix unique probe set identifier and the column on the far-right a brief description of the sequence that the probe set represents.

   a) Display additional Pivot table columns in the analysis by selecting "Pivot Data>Absolute Results" from the "View" pull-down menu. Select the columns desired to be displayed. Columns may include "Signal," "Detection Call," "Detection *p*-value," "Stat Pairs," and "Stat Pairs Used."
   (*See Notes for Microarray Suite 4.0 Users.)

NOTE: Values in the "Signal" column reflect intensity. The "Detection Call" column assigns a call of "Present," "Absent," or "Marginal" to each probe set. The "Detection *p*-value" column provides an assessment of statistical significance of each call. The "Descriptions" column provides some summary information about each transcript. Right click on a transcript of interest to link to an external database for more information.

   b) Select the "Metrics" tab at the bottom of the .chp file.

   c) The Metrics table displays data for each distinct probe set in the .chp file. Columns displayed are similar to the Pivot table.

      (1) Organize the tabular data columns by right clicking at the top of the column to "Hide Column."

      (2) Sort by right clicking on the column header and selecting the desired sorting function.

NOTE: Refer to Chapter 5 for recommendations.

   d) Select the "Analysis Info" tab at the bottom of the analysis results or .chp file. The Analysis Information table displays experimental and sample information and algorithm settings information. Information includes Scaling or Normalization factors, Background, Raw Q, and Sample Type information.

Once a single chip analysis has been completed and a .chp file generated, this file can be further utilized in a number of ways. The file can be used as a "baseline" file in a comparison analysis (see Chapter 4 of this training guide). The .chp file can also be published into either the MicroDB™ or LIMS database, becoming accessible for data mining with the Affymetrix Data Mining Tool or other third-party analysis tools. The .chp file data can also be exported from Microarray Suite as a text file allowing the data to be imported into third-party programs (e.g., Microsoft® Excel).

Note for Microarray Suite 4.0 Users

*Step 4. a) in Performing Single Array Analysis.
Display additional Pivot table columns in the analysis by selecting "Pivot Data>Absolute Results" from the "View" pull-down menu. Select the columns desired to be displayed. Columns may include "Average Difference" and "Absolute Call."

Chapter 4 Comparison Analysis

Comparison analysis is used to compare expression profiles from two GeneChip probe arrays of the same type. One array is designated as a baseline and the other is designated as the experimental. The experimental file is analyzed in comparison to the baseline file. While the designations "experimental" and "baseline" are arbitrary, it is important to keep these designations in mind when examining the changes reported. For example, if the baseline file is derived from a treated sample and the experimental from an untreated sample, all genes activated by the treatment will have decrease calls.

As Microarray Suite 4.0 and Microarray Suite 5.0 use different algorithms, the files being compared must be analyzed using the same version of Microarray Suite. These instructions relate to analyses performed in Microarray Suite 5.0. Notes for Microarray Suite 4.0 users can be found at the end of this section.

Quality Assessment of .dat Image

Prior to conducting analysis of an array, the quality of the array image (.dat file) should be assessed following the guidelines in Chapter 1 of this training manual.

NOTE: Refer to Chapter 1 to aid in quality assessment of the array.

Ensuring Consistency of Files to be Compared

Ensure .dat and .cel files corresponding to both the designated experiment and baseline files along with the baseline .chp file are present in the data file tree. If they are not, verify that the files are in the same directory and that the directory is specified correctly, as described on page 13 of this training manual.

NOTE: Single-array (or 'absolute') analyses must be previously completed and .chp files present for all samples that will be used as baseline files.

When conducting a comparison analysis it is important to ensure that the scaling strategy used for the comparison analysis is the same as was used to generate the baseline file. To examine the analysis settings of the baseline file, right click the baseline .chp file in the Data File Tree and select "Information." The following fields are of note:

TGT	Target Signal value used in this protocol should be 500.
SF	Displays the scaling factor calculated. In this protocol this should **NOT** be 1.0000.
NF	Displays the normalization factor applied. In this protocol the value should be 1.0000, as no normalization was used.
SFGene	Displays the Scaling strategy used. In this protocol the value should be 'All,' as the global scaling strategy was used.

Comparison Analysis Set-Up

Like the single array analysis, comparison analysis will create a .chp file from a .cel image file. Microarray Suite automatically generates the .cel image file from the .dat file. To perform a comparison analysis, settings relating to file locations and the analysis must first be defined.

Expression Analysis Set-Up

Close any .chp files that are currently open and Select "Analysis Settings>Expression" from the "Tools" pull-down menu. The "Expression Analysis Settings" dialogue box opens.

1. Select the "Probe Array Type" to be analyzed from drop-down menu.
2. Select the "Scaling" tab.
    a) Select "All Probe Sets" and set "Target Signal" to 500.
3. Select the "Normalization" tab.
    a) Select "User Defined" and place a "1" in the "Normalization Value" box.
4. Select the "Probe Mask" tab. This feature is used to mask user-defined probe cells.
    a) Ensure that the "Use Probe Mask File" option is unchecked.

5. Select the "Baseline" tab.

   a) Check the "Use Baseline File Comparison" option.

   b) Click the "Browse" button.

   c) Select the baseline .chp file.

   d) Click the "Open" button.

6. Select the "Parameters" tab.

   a) Confirm default settings appropriate to the version of Microarray Suite and array being analyzed as specified in Appendix C of this training manual.

NOTE: These Settings should not be adjusted unless the user has advanced experience with the Affymetrix GeneChip system.

7. Once all settings have been adjusted or confirmed select "OK" to define settings and close the dialogue box. One can now perform comparison analyses based upon these settings.

Performing Comparison Analysis

1. Open the designated experimental file (.dat or .cel) by double clicking in the data file tree. Alternatively, select "Open" from the "File" pull-down menu and select the experimental file.

2. Select "Analysis" from the "Run" pull-down menu.

   a) Verify the .chp filename. The default corresponds to the name of the experimental file .exp and .dat filename. Edit the .chp filename, if necessary, and click "OK."

NOTE: Microarray Suite will overwrite a .chp file if the filename is the same as an existing .chp file in the directory.

   b) Verify "Expression Analysis Settings" in the subsequent pop-up window as previously set in the above **Expression Analysis Settings** section and select "OK" to begin analysis and generate the .chp file.

   c) The Microarray Suite status window will indicate that the analysis has started.

3. Once the analysis is complete, generate an Expression Analysis report file (.rpt) and review the quality control metrics.

   a) To generate the report, select "Report" from the "File" pull-down menu.

   b) Select the appropriate analysis results file (.chp).

NOTE: All metrics reported in a comparison file report refer to the designated experimental file, NOT the baseline file.

   c) Review the quality control data.

NOTE: See Chapter 1 for detailed explanations.

- Review *bioB*, *bioC*, *bioD,* and *cre* sensitivity spikes.
- Review percent Present determination.
- Review internal control 3'/5' ratios.
- Review noise (Raw Q).
- Review background.

   d) Return to the .chp file by closing the Report (.rpt) file.

NOTE: The open .chp file data is displayed in the Expression Analysis Window (EAW) and can be accessed by clicking on the Expression Analysis button in the Microarray Suite shortcuts window.

4. Select the "Pivot" tab at the bottom of the .chp file. The Pivot table displays analysis output and descriptions for each transcript represented on the probe array. The far-left column contains the Affymetrix unique probe set identifier and the column on the far-right provides a brief description of the sequence that the probe set represents.

    a) Display additional Pivot table columns in the analysis by selecting "Pivot Data>Comparison Results" from the "View" pull-down menu. Select the columns desired to be displayed. Suggested columns may include "Signal," "Detection," "Detection *p*-value," "Signal Log Ratio," "Change," and "Change *p*-value." (*See notes for Microarray Suite 4.0 Users).

    b) Select the "Metrics" tab at the bottom of the .chp file. The Metrics table displays data for each distinct probe set in the .chp file. Columns displayed are similar to the Pivot table.

    c) Sort data by right clicking the mouse on the column header and selecting the desired sorting function. These useful functions enable you to sort the data in ascending or descending order and to hide or unhide columns. For example, if you are interested in only those genes which are "Present" and have increased at a "Signal Log Ratio" of >1.

NOTE: Refer to Chapter 5 for recommendations.

After the comparison analysis .chp file has been generated, this file can be further utilized in a number of ways. The .chp file can also be published into either the MicroDB or LIMS database, becoming accessible for data mining with the Affymetrix® Data Mining Tool or other third-party analysis tools. The .chp file data can also be exported from Microarray Suite as a text file allowing the data to be imported into third-party programs (e.g., Microsoft Excel).

Using the Batch Analysis Tool

Batch analysis is a way to analyze many .cel files and generate .chp files with unattended operation. Many files can be simultaneously compared to a selected baseline. Files from different experiments may also be simultaneously analyzed. It is important to select a different name for the analysis output (.chp file) otherwise batch analysis will overwrite the previous files. Either the Drag and Drop method or the Toolbar can be used to select files for batch analysis. Further details can be found in Chapter 13 of the Affymetrix Microarray Suite User Guide Version 5.0.

NOTE: Prior to batch analysis, check the Expression Analysis settings and ensure that they are correct (i.e., Select the "Baseline" tab and ensure "Use Baseline File Comparison" is unchecked).

1. Open the Batch Analysis window by selecting "Batch Analysis" from the "Run" menu.

2. Add files to the Batch Analysis window by:

    a) Dragging and Dropping each .cel or .chp file to the Batch Analysis window from the data file tree to the Batch Analysis window.

    OR

    a) Using the Toolbar, click the "Add" Toolbar or select "Edit>Add."

    b) An open dialog of .cel files appears.

    c) Select the .cel or .chp files to be analyzed.

    d) To select all files hold "shift" while you click on the first and last file.

    e) To select files individually, hold "control" while selecting files.

    f) Click open to place the files into the Batch Analysis window.

3. Verify the Output filenames.

    a) The filename for the .chp file is listed in the Output column. If the .chp filename is already present the filename will be red to indicate that a file is going to be overwritten.

    b) To edit the .chp file name, double click on the output file name and add by typing in a new name.

4. To select the baseline file, double click in the Baseline column corresponding to the .cel file being analyzed or click the .cel file and choose “Select Baseline” from the “Edit” pull-down menu.

a) Double click on the baseline .chp file from the dialog box.

b) Right clicking the baseline file and selecting “Clear Baseline” or selecting “Edit>Clear Baseline” can remove a baseline file in the batch analysis window.

5. To start the Batch Analysis, click on the Analyze button which is found immediately above the Batch Analysis window.

Note for Microarray Suite 4.0 Users

*Step 4 a) in Performing Comparison Analysis.

Select “Analysis>Options...>Pivot Tab” and select the comparison analysis metrics you wish to see from the right side of the menu under Comparison Results. Recommendations include “Fold Change” and “Difference Call.”

Chapter 5 Basic Data Interpretation

The use of GeneChip gene expression arrays allows interrogation of several thousands of transcripts simultaneously. One of the formidable challenges of this assay is to manage and interpret large data sets. This chapter provides users with guidelines for determining the most robust changes from a comparison analysis. The guidelines listed below apply to Microarray Suite 5.0. Notes for Microarray Suite 4.0 users are highlighted at the end of this section.

Metrics for Analysis

Which data analysis metrics should be used to determine the most significant transcripts when comparing an experimental sample to a baseline sample? Microarray Suite provides users with both qualitative and quantitative measures of transcript performance. One standardized approach for sorting gene expression data involves the following metrics:

- Detection
- Change
- Signal Log Ratio

Detection is the qualitative measure of presence or absence for a particular transcript. A fundamental criterion for significance is the correlation of the Detection calls for a particular transcript between samples. When looking for robust increases, it is important to select for transcripts that are called "Present" in the experimental sample. When determining robust decreases, it is important to select for "Present" transcripts in the baseline sample. By following these initial guidelines, you will eliminate "Absent" to "Absent" changes, which are uninformative.

Change is the qualitative measure of increase or decrease for a particular transcript. When looking for both significant increases and decreases, it is important to eliminate "No Change" calls.

Signal Log Ratio is the quantitative measure of the relative change in transcript abundance. The Affymetrix Gene Expression Assay has been shown to identify Fold Changes greater than two 98% of the time by Wodicka *et al.* in 1997 (26). Based on these observations, robust changes can be consistently identified by selecting transcripts with a Fold Change of >2 for increases and <2 for decreases. This corresponds to a Signal Log Ratio of 1 and -1, respectively. These value guidelines apply when performing a single comparison analysis.

NOTE: Please refer to "Introduction to Replicates" below in this chapter for exceptions.

Interpretation of Metrics

When sorting through gene expression data in Microarray Suite, you will notice that some transcripts provide conflicting information. Here are some examples:

1. A transcript is called "Increase" but has a Signal Log Ratio of less than 1.0.

2. A transcript is called "No Change" but has a Signal Log Ratio of greater than 1.0.

3. A transcript is called "Absent" in both experimental and baseline files but is also called "Increase."

These contradictions arise due to the fact that Detection, Change, and Signal Log Ratio are calculated separately. The benefit of this approach is that transcripts can be assessed using three independent metrics.

Thus, in order to determine the most robust changes, it is crucial to use all three metrics in conjunction. The following section outlines this process.

Sorting for Robust Changes

NOTE: For detailed sorting instructions, please refer to Chapter 6.

Basic steps for determining robust increases:

1. Eliminate probe sets in the experimental sample called "Absent."

2. Select for probe sets called "Increase."*

3. Eliminate probe sets with a Signal Log Ratio of below 1.0.

Basic steps for determining robust decreases:

1. Eliminate probe sets in the baseline sample called “Absent.”

2. Select for probe sets called “Decrease.”*

3. Eliminate probe sets with a Signal Log Ratio of above -1.0.

* For those who wish to relax the Change criterion, include “Marginal Increase” and “Marginal Decrease” during selection.

“Real” Changes vs. “False” Changes

The procedures listed above can be used to determine both “Real” and “False” changes. The difference between “Real” and “False” changes lies in the relationship between the samples being compared. If the samples are different (e.g., normal vs. diseased, control vs. treated, etc.), the procedures will highlight transcripts that change significantly from the baseline sample to the experimental sample. If the samples are identical (i.e., hybridization replicates), no changes are expected. Thus, any transcripts showing significant change are false changes.

Note on Signal Log Ratio

When applying the sorting functions on Signal Log Ratio in Microarray Suite (i.e. “Sort Ascending” and “Sort Descending”), you will notice that the column sorts on the magnitude of the Signal Log Ratio value, and not on the sign. Keep this in mind when sorting for robust changes.

Differences in MAS 4.0

The metrics used to sort for robust changes in MAS 5.0 are Detection, Change, and Signal Log Ratio. The equivalent metrics in MAS 4.0 are Absolute Call, Difference Call, and Fold Change, respectively.

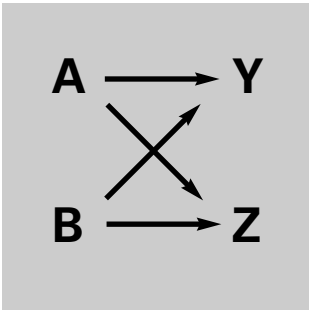
The Signal Log Ratio is essentially the log base 2 of the Fold Change. Thus, when sorting on MAS 4.0 gene expression data for significant increases, probe sets with a Fold Change value below 2.0 should be eliminated. For significant decreases, probe sets with a Fold Change value above -2.0 should be eliminated.

As with Signal Log Ratio, the Fold Change column sorts values on the magnitude and not on the sign. Keep this in mind when sorting for robust changes.

Introduction to Replicates

The guidelines outlined in “Sorting for Robust Changes” above apply to a single comparison analysis. However, when biological replicates are introduced and multiple comparisons are generated, it becomes possible to relax the sorting thresholds based on consensus.

For example, here is an experiment with two sets of replicate samples consisting of two control samples (A and B) and two experimental samples (Y and Z). Performing pair-wise comparisons results in the following matrix:



This set of four analyses (A to Y, B to Y, A to Z, and B to Z) are comparison replicates. Each transcript has essentially been interrogated four times. The following is a hypothetical set of metrics for one transcript to determine whether or not it has increased:

Comparison	Detection in Exp.	Change in Exp.	Signal Log Ratio
A to Y	A	I	1.3
B to Y	P	I	1.2
A to Z	P	I	0.9
B to Z	P	I	1.2

Note: "Exp." refers to the experimental sample.

Following the change guidelines for a single comparison analysis, the "Absent" call in the "A to Y" comparison would throw out this transcript. Likewise, the 0.9 Signal Log Ratio value would throw out the transcript in the "A to Z" comparison.

Overall, the transcript appears to be increasing since two of the four comparisons meet all three conditions for determining robust change and the other two comparisons meet two out of the three conditions. Based on overall consensus, we may choose to accept this transcript as a robust change.

The number of replicates to utilize and the conditions for acceptance of change are variable and up to the discretion of the user. However, the benefit of replicates in gene expression data (as with other assay data) is clear.

More advanced data analysis can be carried out in the Affymetrix Data Mining Tool software.

Chapter 6 Change Calculation Worksheet

This procedure can be used to identify robust changes between two GeneChip probe arrays. These instructions relate to analyses performed in Microarray Suite 5.0. Notes for Microarray Suite 4.0 users can be found at the end of this chapter.

If the samples hybridized to the two arrays are derived from separate samples, this procedure will identify probe sets showing significant change and serves as a useful starting point for further data analysis. If the two samples are derived from the same hybridization cocktail, this procedure will identify false changes. According to the Affymetrix specification, the false change observed should be no more than 2%. This value is based on observations reported by Wodicka *et al.* in 1997 (36).

Data Preparation

1. Choose the two data sets that you wish to analyze.
2. Conduct a single array analysis of the baseline data set as described in Chapter 3 of this manual.
3. Conduct a comparison analysis of the experiment data set using the previous data set as the baseline as described in Chapter 4 of this manual. Ensure that the scaling strategy used in step 2 is also used in step 3.
4. Record the file names of the baseline and experiment in the appropriate spaces on the Change Calculation Worksheet (see page 26).

Calculate Increases

The first step of this procedure is to calculate the number of significant increases.

1. Calculate the number of probe sets that have a Detection call of 'P' in the Experiment file. (*See Notes for Microarray Suite 4.0 Users.)

    a) Open the comparison .chp file in MAS 5.0, with the Pivot table view.

    b) Display additional Pivot table columns in the analysis by selecting "Pivot Data>Absolute Results" from the "View" pull-down menu. Ensure that the Detection, Change and Signal Log Ratio Columns are displayed.

    c) Sort the data on the Detection column in descending order by right-clicking on the Detection column heading and selecting "Sort Descending" from the pop-up menu as shown in Figure 1.

    d) Click on the probe set identifier, contained in the far-left column, at the top of the list.

    e) Use the mouse to scroll down the data list until the last 'P' is visible.

    f) Hold down the 'Shift' key and click on the probe set identifier corresponding to the last 'P' value.

    g) Click the "Hide unselected probe sets" button as shown in Figure 2.

    h) The number of remaining probe sets is displayed in the bottom-right of the window, as shown in Figure 3. Enter this value into the box on Line 1 of the Change Calculation Worksheet.

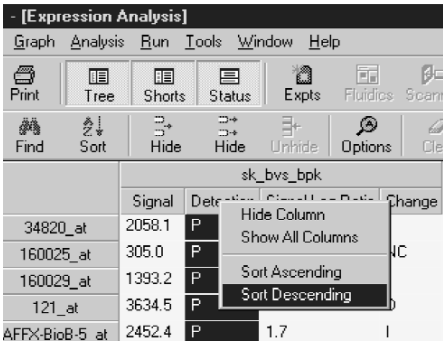


Figure 1

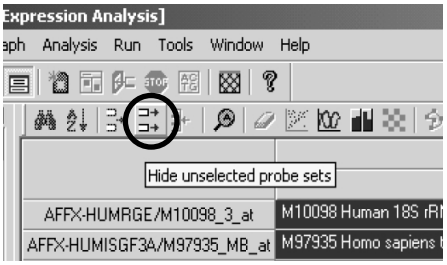


Figure 2

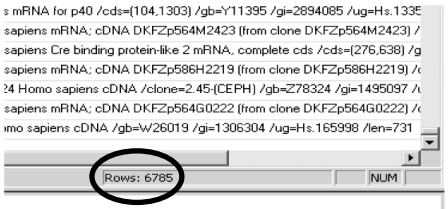


Figure 3

2. Calculate the number of probe sets from above list that also have a Change call of 'I.'
(**See Notes for Microarray Suite 4.0 Users.)

a) After performing step 1 of the Increase calculation, sort the data on the Change column in ascending order, by right-clicking the Change column heading and selecting "Sort Ascending" from the pop-up menu as shown in Figure 1.

b) Scroll down the list of probe sets until the first 'I' call is visible, then click on this probe set identifier.

c) Scroll down the list until the last 'I' call is visible, hold down the 'Shift' key and click on the corresponding probe set identifier.

d) Click the "Hide unselected probe sets" button as shown in Figure 2.

e) The number of remaining probe sets is displayed in the bottom-right of the window as shown in Figure 3. Enter this value into the box on Line 2 of the Change Calculation Worksheet.

3. Calculate the number of probe sets from the above list that also have a Signal Log Ratio of 1.0 or greater.
(***See Notes for Microarray Suite 4.0 Users.)

a) After performing step 2 of the Increase calculation, sort the data on the Signal Log Ratio column in descending order by right-clicking the Signal Log Ratio column heading and selecting "Sort Descending" from the pop-up menu as shown in Figure 1.

b) Click on the probe set identifier at the top of the list.

c) Scroll down the list until the last Signal Log Ratio value (equal to 1.0) is visible, hold down the 'Shift' key and click on the corresponding probe set identifier.

d) Click the "Hide unselected probe sets" button as shown in Figure 2.

e) The number of remaining probe sets is displayed in the bottom-right of the window as shown in Figure 3. Enter this value into the box on Line 3 of the Change Calculation Worksheet.

4. Calculate the number of probe sets that have increased as a percentage of the probe sets detected.

a) Divide the number of probe sets showing significant increase (Line 3) by the number of probe sets detected (Line 1).

b) Multiply the above number by 100 to convert to a percentage.

c) Enter the value in the box on Line 4 of the Change Calculation Worksheet.

Calculate Decreases

The next part of this procedure is to calculate the number of significant decreases.

1. Calculate the number of probe sets that have a Detection call of 'P' in the Baseline file.
(*See Notes for Microarray Suite 4.0 Users.)

a) Open both the comparison .chp and baseline .chp files in MAS 5.0 in the Pivot table view.

b) Display Pivot table columns in the analysis by selecting "Pivot Data>Absolute Results" from the "View" pull-down menu. Ensure that the Detection, Change, and Signal Log Ratio Columns are displayed.

c) Sort the data on the Detection column of the **baseline file** in descending order by right-clicking the Detection column heading and selecting "Sort Descending" from the pop-up menu as shown in Figure 1.

d) Click on the probe set identifier contained in the far-left column at the top of the list.

e) Use the mouse to scroll down the data list until the last 'P' is visible in the baseline file.

f) Hold down the 'Shift' key and click on the probe set identifier corresponding to the last 'P' value.

g) Click the "Hide unselected probe sets" button as shown in Figure 2.

h) The number of remaining probe sets is displayed in the bottom-right of the window as shown in Figure 3. Enter this value into the box on Line 5 of the Change Calculation Worksheet.

2. Calculate the number of probe sets from the above list that also have a Change call of 'D.' (**See Notes for Microarray Suite 4.0 Users.)

    a) After performing step 1 of the Decrease calculation, sort the data on the Change column of the **comparison file** in ascending order by right-clicking the Change column heading and selecting "Sort Ascending" from the pop-up menu as shown in Figure 1.

    b) Click on the probe set identifier contained in the far-left column at the top of the list.

    c) Scroll down the list until the last 'D' call is visible, hold down the 'Shift' key and click on the corresponding probe set identifier.

    d) Click the "Hide unselected probe sets" button as shown in Figure 2.

    e) The number of remaining probe sets is displayed in the bottom-right of the window as shown in Figure 3. Enter this value into the box on Line 6 of the Change Calculation Worksheet.

3. Calculate the number of probe sets from above list that also have a Signal Log Ratio of -1.0 or less. (***See Notes for Microarray Suite 4.0 Users.)

    a) After performing step 2 of the Decrease calculation, sort the data on the Signal Log Ratio column of the **comparison file** in descending order by right-clicking the Signal Log Ratio column heading and selecting "Sort Descending" from the pop-up menu as shown in Figure 1. (Note that Microarray Suite 5.0 sorts the Signal Log Ratio column on the magnitude of the Signal Log Ratio, hence, the sign of the value is ignored.)

    b) Click on the probe set identifier at the top of the list.

    c) Scroll down the list until the last Signal Log Ratio value equal to -1.0 is visible, hold down the 'Shift' key, and click on the corresponding probe set identifier.

    d) Click the "Hide Unselected probe sets" button as shown in Figure 2.

    e) The number of remaining probe sets is displayed in the bottom-right of the window as shown in Figure 3. Enter this value into the box on Line 7 of the Change Calculation Worksheet.

4. Calculate the number of probe sets that have decreased, as a percentage of the probe sets detected.

    a) Divide the number of probe sets showing significant decrease (Line 7) by the number of probe sets detected (Line 5).

    b) Multiply the above number by 100 to convert to a percentage.

    c) Enter the value into the box on Line 8 of the Change Calculation Worksheet.

Calculate Total Percentage Change

Finally, add the Percentage Increase (Line 4) to the Percentage Decrease (Line 8) and place the sum into the box on Line 9 of the Change Calculation Worksheet.

If the two samples being compared are from the same hybridization cocktail, the value in Line 9 should be less than 2.0. If this is not the case, it is likely that the arrays were not analyzed using the same scaling strategy. The data should be re-analyzed paying particular attention to ensure that the scaling strategy is identical for all analyses performed before contacting your Affymetrix Field Applications Specialist for further consultation.

Notes for Microarray Suite 4.0 Users

*Step 1. The equivalent to the Detection call in Microarray Suite 4.0 is the Absolute Call.

**Step 2. The equivalent to the Change call in Microarray Suite 4.0 is the Difference Call.

***Step 3. The equivalent to the Signal Log Ratio in Microarray Suite 4.0 is the Fold Change. To identify the increases, Fold Change values $\geq$2.0 are required. For decreases, fold change values $\leq$-2.0 are required.

Change Calculation Worksheet for Microarray Suite 5.0

Experiment File name: _____

Baseline File name: _____

Increases

Number of probe sets with Detection of 'P' in Experiment:		Line 1
Number of probe sets from Line 1 that have a Change call of 'I' :		Line 2
Number of probe sets from Line 2 that have a Signal Log Ratio of ≥1:		Line 3
% Increase (Line 3 divided by Line 1)*100:		Line 4

Decreases

Number of probe sets with Detection of 'P' in Baseline:		Line 5
Number of probe sets from Line 5 that have a Change call of 'D' :		Line 6
Number of probe sets from Line 6 that have a Signal Log Ratio of ≤-1:		Line 7
% Decrease (Line 7 divided by Line 5) *100:		Line 8

Total Changes

Total % Change (Line 4 + Line 8):		% Line 9

Chapter 7 The NetAffx™ Analysis Center Summary

The NetAffx™ Analysis Center (www.netaffx.com) is an online resource that allows researchers to correlate their GeneChip array results to a catalog of array design and annotation information. The NetAffx Analysis Center uses the SRS (Sequence Retrieval System) data and application integration platform.

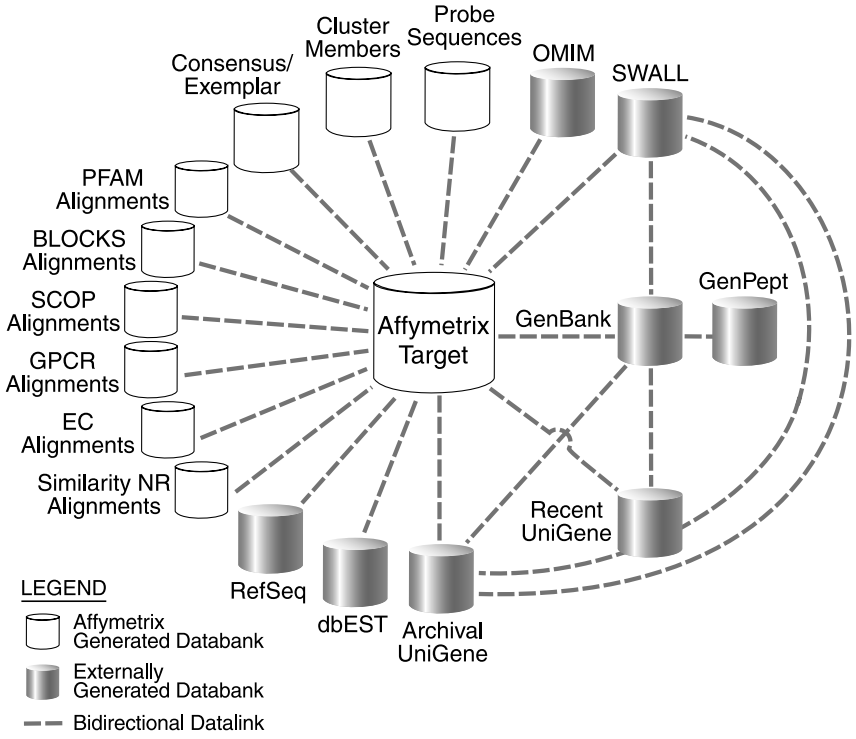


Figure 1

This useful tool enables you to access product-specific biological annotations from both the public domain and Affymetrix (Figure 1). Specifically, you may link from target sequences to the information represented in PFAM, BLOCKS, SCOP, Similarity NR, and EC. These annotations provide further structural and functional information, helping you to draw biologically relevant conclusions about your experimental results.

For instructions on using the NetAffx Analysis Center for downstream analysis, please refer to the NetAffx Interactive Tutorial in the NetAffx Help Center.

Analysis Center

The NetAffx Analysis Center is a comprehensive resource of functional annotations and public database information integrated with the probe sets. Now researchers can access detailed array content, including target and probe sequences. The NetAffx Analysis Center is now available to anyone who completes a short registration form.

Flexible query capabilities are provided to help you retrieve biological information for probe sets from both public and proprietary data. Unless otherwise noted, public data representations are updated on the site once every three months.

A new interactive Probe Set Display tool allows users to visualize information about probe alignments. It is currently available for HG-U133 Set and HG-U95 Set. To explore an interactive display graphic, you will need to download and install an SVG viewer from Adobe (www.adobe.com). For more information about our Probe Set Display tool, please refer to the user's guide.

Download Center

This enables you to efficiently access the data represented on Affymetrix GeneChip catalog arrays. You may download consensus, exemplar, target (SIF), and probe sequences and incorporate this information into your internal data analysis pipelines.

Brief Information on the Databases Available on the NetAffx Analysis Center

dbEST
dbEST is a database for Expressed Sequence Tags (ESTs). More information about dbEST may be found at www.ncbi.nlm.nih.gov/dbEST/.

Domains_PFAM (Affymetrix internal)
Domains_PFAM contains detailed alignment data associated with the computational annotation of protein domains represented in the PFAM database using the HMMer program. PFAM entries are derived from seed alignments largely generated through human curation. More information about PFAM may be found at pfam.wustl.edu.

Domains_BLOCKS (Affymetrix internal)
Domains_BLOCKS contains detailed alignment data associated with the computational annotation of protein domains represented in the BLOCKS database. BLOCKS entries represent domains or motifs from multiply aligned, ungapped segments in the most highly conserved regions of proteins. More information about BLOCKS may be found at www.blocks.fhcrc.org.

Families_GPCR (Affymetrix internal)
Families_GPCR contains alignments to families of G protein coupled receptors as organized by SWISS-PROT. The alignments are generated by scoring against SAM-T99 derived HMM models. The GPCR classification list may be found at www.expasy.ch/cgi-bin/lists?7tmrlist.txt

Families_SCOP (Affymetrix internal)
Families_SCOP contains detailed alignment data associated with the computational prediction of structural classification based on protein sequence similarity to representative sequences from the SCOP database. SCOP is the Structural Classification of Proteins database containing a hierarchical representation of classes, folds, super families, families and individual proteins. Predictions are based on the creation of individual sub-family models using the SAM program and T-99-derived methods for HMM model generation. More information about SCOP may be found at scop.mrc-lmb.cam.ac.uk/scop.

Families_EC (Affymetrix internal)
Families_EC contains detailed alignment data associated with the computational identification of homology to enzymes using the SAM-T99 method for HMM model generation. The Enzyme Commission (EC) classification scheme contains a hierarchical representation based on broad enzymatic classes, sets of substrates and cofactors/reagents. Enzyme commission (EC) numbers and associated pathway data is available via hypertext links to the Kyoto Encyclopedia of Genes and Genomes (KEGG). Information on KEGG can be found at www.genome.ad.jp/kegg/. More information about EC may be found at www.chem.qmw.ac.uk/iupac/jcbn/ or on a page at Rockefeller University at prowl.rockefeller.edu/enzymes/enzymes.htm.

GenBank
GenBank is a public database of genetic sequences and annotations maintained by the National Center for Biotechnology Information (NCBI). You can access the web page at www.ncbi.nlm.nih.gov/.

Summary of available information:

- Brief description of sequence includes information such as source organism, gene name/protein name, or some description of the sequence's function (if the sequence is non-coding).
- Publications by the authors of the sequence that discuss the data reported in the record with links to corresponding MEDLINE records.
- Information about genes and gene products, as well as regions of biological significance reported in the sequence. These can include regions of the sequence that code for proteins and RNA molecules, as well as a number of other features (promoter, 5'UTR, 3'UTR etc).
- Protein and DNA sequence in FASTA format.

GenPept
GenPept is a database of translated protein coding sequence that is copied from the GenBank translation information. It is a duplication of the GenBank protein translation information. GenPept is the sequence format most appropriate to perform similarity searches.

LocusLink
LocusLink provides curated gene sequences and descriptive information about genetic loci. More information about LocusLink can be found at www.ncbi.nlm.nih.gov/LocusLink.

Summary of available information:

- Official gene symbol and link to the Human Genome Nomenclature Database.
- Locus information: alternate gene symbols, links to the corresponding UniGene and OMIM records.
- Map information: chromosomal and cytogenetic location, STS markers associated with the locus, links to the NCBI Map Viewer.
- Links to GenBank, GenPept, PFAM records.
- Gene Ontology categories and links to appropriate PubMed records.

OMIM
OMIM, Online Mendelian Inheritance in Man, is a catalog of human genes and genetic disorders. More information can be found at www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM.

Summary of available information:

- Detailed description of the gene and its protein product.
- Summary of the literature and large set of links to the corresponding MEDLINE records.

Pathways
Pathways contains mappings of signaling, metabolic, and biosynthetic pathways from www.genmapp.org mapped to Affymetrix GeneChip probe sets. A link to the GenMAPP web site enables the user to pull down the GenMAPP software for examining pathways curated by the Conklin group at the Gladstone Institute at UCSF. Future releases of the NetAffx web site will include HTML documents depicting these pathways.

PFAM
PFAM contains a large collection of multiple-sequence alignments and Hidden Markov Models covering many common protein domains. More information on this database can be found at pfam.wustl.edu.

Protein_Summary
Protein_Summary contains the summary results of homology modeling of the translated peptide sequences associated with a probe set. The current databank contains annotations on the publicly annotated protein coding regions (CDS) of known full-length sequences. Sequence similarity is determined through several approaches as follows: Protein domains are identified using the HMMer program to search the PFAM database and by using position-specific weight matrices to search the BLOCKS database. A hidden Markov model is a previously trained statistical model for an ordered sequence of symbols such as bases or amino acids. It functions as a state machine that generates a symbol each time a transition is made from one state to the next. HMMs can function as probabilistic models for multiple sequence alignments where all possible combinations of matches, mismatches, and gaps are used to generate alignment in a series of sequences or may model periodic patterns in a single sequence. HMMs have been found in the exons of a gene or families of similar protein structure. Structural family prediction is based on hidden Markov models representing each SCOP structural sub-family using the SAM program. Enzyme classifications (EC) and associations with pathways are also obtained by using hidden Markov model searching. G protein coupled receptor (GPCR) classifications are obtained using hidden Markov model searching. General sequence similarity is obtained using the BLASTP program to search the non-redundant protein database (nr).

RefSeq (Reference Sequence Project)
RefSeq is a non-redundant set of reference sequences including constructed genomic contigs, mRNAs, and proteins. It is a stable reference point for mutation analysis, gene expression studies, and polymorphism discovery. More information can be found at www.ncbi.nlm.nih.gov/LocusLink/refseq.html.

Records are classified as follows:

(NT_######) constructed genomic contigs

(NM_######) curated mRNAs

(NP_######) curated proteins

(NC_######) chromosomes

(XM_######) model mRNAs corresponding to genomic contig

(XP_######) model proteins corresponding to genomic contig

Similarity_NR
Similarity_NR contains detailed alignment data associated with the annotation of homologous protein sequences determined by sequence similarity searching using the BLASTP program against the non-redundant protein database (nr) from the National Center for Biotechnology Information (NCBI). More information about the BLAST family of programs and the non-redundant protein database can be found at www.ncbi.nlm.nih.gov/.

Swiss_Prot
Swiss_Prot is a curated protein sequence database that provides a high level of annotations, a minimal level of redundancy and a high level of integration with other databases. More information can be found at www.expasy.org/sprot/.

Summary of available information:

- Annotation information: description of protein function, domain structure, post-translational modifications, variants, etc.
- Extensive links to MEDLINE records.

UniGene
UniGene provides a non-redundant set of gene-oriented clusters. More information can be obtained at www.ncbi.nlm.nih.gov/UniGene.

Summary of available information:

- Expression pattern (tissue-specific expression).
- Similarity to proteins in model organisms.
- Links to UniGene, Locus Link, dbEST, HomoloGene records and mapping information.

Chapter 8 Relevant Publications

Affymetrix Technical Note: Fine Tuning Your Data Analysis. (2001).

Affymetrix Technical Note: New Statistical Algorithms for Monitoring Gene Expression on GeneChip® Probe Arrays. (2001).

Alevizos, I. *et al*. Oral cancer *in vivo* gene expression profiling assisted by laser capture microdissection and microarray analysis. *Oncogene* **20**, 6196-204 (2001).

Bumol, T.F., Watanabe, A.M. Genetic information, genomic technologies, and the future of drug discovery. *Journal of the American Medical Associations* **285**, 551-555 (2001).

Cao, S.X., Dhahbi, J.M., Mote, P.L., Spindler, S.R. Genomic profiling of short- and long-term caloric restriction effects in the liver of aging mice. *Proceedings of the National Academy of Sciences of the USA* **98**, 10630-10635 (2001).

Carter, T.A. *et al*. Chipping away at complex behavior: Transcriptome/phenotype correlations in the mouse brain. *Physiology & Behavior* **73**, 849-57 (2001).

Cho, R.J. *et al*. Transcriptional regulation and function during the human cell cycle. *Nature Genetics* **27**, 48-54 (2001).

Cutler, D.J. *et al*. High-throughput variation detection and genotyping using microarrays. *Genome Research* **11**, 1913-25 (2001).

Dong, D. *et al*. Flexible Use of High-Density Oligonucleotide Arrays for Single-Nucleotide Polymorphism Discovery and Validation. *Genome Research* **11**, 1418-1424 (2001).

Electronic Textbook, StatSoft, Inc. (1984-2000). www.statsoft.com/textbook/stathome.

Fodor, S.P.A. Genes, Chips and the Human Genome. *FASEB Journal* **11**, A879 (1997).

Gerhold, D. et al. Monitoring expression of genes involved in drug metabolism and toxicology using DNA microarrays. *Physiological Genomics* **5**, 161-170 (2001).

Golub, T.R. *et al*. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531-537 (1999).

Harrington, C., Rosenow, C., Retief, J. Monitoring gene expression using DNA microarrays. *Current Opinion in Microbiology* **3**, 285-291 (2000).

Hoaglin, D.C., Mosteller, F., Tukey, J.W. Understanding Robust and Exploratory Data Analysis. John Wiley & Sons, New York (2000).

Hollander, M., Wolfe, D.A. Nonparametric Statistical Methods (second edition). John Wiley & Sons, New York (1999).

Hsiao, L.L. *et al*. A Compendium of Gene Expression in Normal Human Tissues Reveals Tissue-Selective Genes and Distinct Expression Patterns of Housekeeping Genes Physiological Genomics. *In press* (2001).

Jin, H. *et al*. Effects of Early Angiotensin-Converting Enzyme Inhibition on Cardiac Gene Expression After Acute Myocardial Infarction. *Circulation* **103**, 736 (2001).

Lindblad-Toh, K. *et al*. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nature Genetics* **24**, 381-386 (2000).

Lander, E.S. *et al*. Initial sequencing and analysis of the human genome. *Nature* **409**, 860 - 921 (2001).

Lee, C.K., Weindruch, R., Prolla, T.A. Gene-expression profile of the aging brain in mice. *Nature Genetics* **25**, 294-297 (2000).

Lee, C.K., Klopp, R.G., Weindruch, R., Prolla, T.A. Gene Expression Profile of Aging and Its Retardation by Caloric Restriction. *Science* **285**, 1390-1393 (1999).

Lipshutz, R.J., Fodor, S.P.A., Gingeras, T.R., Lockhart, D.J. High density synthetic oligonucleotide arrays. *Nature Genetics Chipping Forecast* **21**, 20-24 (1999).

Liu, W.M. *et al*. Analysis of high density expression microarrays with signed-rank call algorithms. *In Preparation* (2001).

Liu, W.M. *et al*. Rank-based algorithms for analysis of microarrays. *Proceedings SPIE* **4266**, 56-67 (2001).

Ly, D.H., Lockhart, D.J., Lerner, R.A., Schultz, P.G. Mitotic misregulation and human aging. *Science* **287**, 2486-2492 (2000).

McDonald, M.J., Rosbash, M. Microarray Analysis and Organization of Circadian Gene Expression in Drosophila. *Cell* **107**, 567-578 (2001).

Motulsky, H. Intuitive Biostatistics. Oxford University Press, New York (1995).

Notterman, D.A., Alon, U., Sierk, A.J., Levine, A.J. Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. *Cancer Research* **61**, 3124-30 (2001).

Ramaswamy, S. *et al*. Multiclass cancer diagnosis using tumor gene expression signatures. *Proceedings of the National Academy of Sciences of the USA* **98**, 15149-15154 (2001).

Little, R. and Rubin, D. Statistical analysis with missing data. Wiley & Sons, New York (1987).

Stamey, T.A. *et al*. Molecular genetic profiling of Gleason grade 4/5 prostate cancers compared to benign prostatic hyperplasia. *Journal of Urology* **166**, 2171-7 (2001).

Warrington, J.A., Nair, A., Mahadevappa, M., Tsyganskaya, M. Comparison of human adult and fetal expression and identification of 535 housekeeping / maintenance genes. *Physiological Genomics* **2**, 143-147 (2000).

Wilcoxon, F. Individual comparisons by ranking methods. *Biometrics* **1**, 80-83 (1945).

Winzeler, E.A. *et al*. Functional Characterization of the S. cerevisiae Genome by Gene deletion and Parallel Analysis. *Science* **285**, 901-906 (1999).

Wodicka, L. *et al*. A Genome-Wide Expression Monitoring in Saccharomyces Cerevisiae. *Nature Biotechnology* **15**, 1359-1367 (1997).

Appendix A: Glossary

NOTE:

▲ = MAS 4.0-Specific Terms (Empirical Algorithms)

◆ = MAS 5.0-Specific Terms (Statistical Algorithms)

Absolute Analysis: The qualitative analysis of a single array to determine if a transcript is Present, Absent or Marginal.

Array: A collection of probes on glass encased in a plastic cartridge.

▲ **Average Difference:** A quantitative relative indicator of the level of expression of a transcript (Σ(PM-MM)/pairs in the average).

Background: A measurement of signal intensity caused by autofluoresence of array surface and non-specific binding of target/stain molecules (SAPE).

Baseline Array: An array designated as the baseline when being analyzed in comparison analysis with which the experimental array is compared to detect changes in expression. For example, if the baseline file is derived from a treated sample and the experiment from an untreated sample, all genes activated by the treatment will have decrease calls.

◆ **Biweight Estimate:** An estimate of the central value of a sample used by the Affymetrix® Statistical Algorithms.

◆ **Change:** A qualitative call indicating an Increase (I), Marginal Increase (MI), No Change (NC), Marginal Decrease (MD) or Decrease (D) in transcript level between a baseline array and an experiment array.

◆ **Change *p*-value:** A *p*-value indicating the significance of the Change call. The change *p*-value measures the probability that the expression levels of a probe set in two different arrays are the same or not. When the *p*-value is close to 0.5, they are likely to be the same. When the *p*-value is close to 0, the expression level in the experiment array is higher than that of the baseline array. When the *p*-value is close to 1, the expression level in the experiment arrays is lower than that of the baseline.

Chip: See Array.

Comparative Analysis: The analysis of an experimental array compared to a baseline array.

▲ **Decision Matrix:** An algorithm that examines a collection of metrics used to determine the status of a hybridized transcript.

◆ **Detection:** A qualitative measurement indicating if the transcript is detected (Present), not detected (Absent), or marginally detected (Marginal).

◆ **Detection *p*-value:** A *p*-value indicating the significance of the Detection call. A Detection *p*-value measures the probability that the discrimination scores of all probe pairs in the probe set are above a certain level (Tau), and that the target is likely to be Present.

◆ **Discrimination Score [R]:** The relative difference between a Perfect Match and its Mismatch (R=(PM-MM)/(PM+MM)).

▲ **Empirical Algorithms:** The algorithms contained in GeneChip® Analysis Suite and Microarray Suite 4.0 based on empirical data generated by Affymetrix.

Experimental Array: An array that is used in comparison analysis to be compared to the baseline array to detect changes in expression. For example, if the baseline file is derived from an untreated sample and the experiment from a treated sample, all genes activated by the treatment will have increase calls.

Feature: A single square-shaped probe cell on an array (another term for probe cell). A feature ranges in size from 18 to 50 microns depending on the array type.

Hybridization Controls: Controls added to the sample before hybridization to the array (refer to Chapter 1 for more information).

◆ **Idealized Mismatch:** A value used in place of the Mismatch intensity when Rules 2 and 3 are used in the Signal Algorithm (refer to Chapter 2 for more information on Rules in the Statistical Algorithms).

◆ **Latin Square:** An experimental design used to monitor the ability to detect a transcript accurately over a range of concentrations. It also allows the statistical analysis of patterns and variability in repeated measurements in a systematic fashion.

Mask: Filter used during synthesis of a GeneChip® array that exposes discreet areas of a wafer to ultraviolet light.

Metric: The calculated answer of mathematical equations used by the GeneChip® algorithms.

Mismatch Probe (MM): A 25-mer oligonucleotide designed to be complementary to a reference sequence except for a single, homomeric (nucleotide mismatch that contains the complementary base to the original) base change at the 13th position. Mismatch probes serve as specificity controls when compared to their corresponding Perfect Match probes.

Noise: The result of small variations in digitized signals in the scanner as it samples the probe array surface and is measured by examining the pixel-to-pixel variations in signal intensities.

Non-parametric Test: A statistical test without the assumption of a particular distribution of the data, also known as a distribution-free test.

Normalization: Adjusting an average value of an experimental array equal to that of the baseline array so that the arrays can be compared (refer to Algorithms description for more information).

◆ ***p*-value:** The probability that a certain statistic is equal or more extreme to the observed value when the null hypothesis is true. The null hypothesis is that the two samples are the same.

Parametric Test: A statistical test that assumes that the data sampled is from a population that follows a Gaussian or normal distribution.

Perfect Match Probe (PM): A 25-mer oligonucleotide designed to be complementary to a reference sequence. The probe sequence that is complementary to the sequence to be hybridized.

◆ **Perturbation:** The range by which the normalization factor is adjusted up or down by the user.

Photolithography: The process used to manufacture probe arrays in conjunction with combinatorial chemistry through a series of cycles. Using light, photolabile protecting groups are removed from linkers bound to the glass substrate (wafer) to enable nucleoside phosphoramidite addition in specific deprotected locations. Each light exposure and subsequent phosphoramidite addition is equal to one cycle. Typically, probe arrays are synthesized in about 80 cycles.

Probe: A 25-mer oligonucleotide designed to be complementary to a reference sequence. It is synthesized on the surface of the array using photolithography and combinatorial chemistry. Hybridization to probes provides intensity data used in both Empirical and Statistical algorithms.

Probe Array Tiling: The spatial organization of probe array features into probe pairs and sets.

Probe Cell: A single square-shaped feature on an array containing probes with a unique sequence. A probe cell ranges in size from 18 to 50 microns per side depending on the array type (refer to Figure 1).

Probe Pair: Two features within a probe set (refer to Figure 1). Each probe of a probe pair is designed to differ only at the nucleotide base interrogation position. The probe pair is designed to detect a Perfect Match (PM) and a Mismatch (MM).

Probe Set: A collection of probe pairs which interrogates the same sequence, or set of sequences. A probe set typically contains between 11 to 20 probe pairs (refer to Figure 1).

SAPE: Streptavidin-phycoerythin dye used to bind the biotin. In the GeneChip® Expression Assay, the biotinylated nucleotides are incorporated into the cRNA during the *in vitro* transcription (IVT) reaction.

Scaling: Adjusting the average intensity or signal value of every array to a common value (target intensity) in order to make the arrays comparable.

◆ **Signal:** A quantitative measure of the relative abundance of a transcript.

◆ **Signal Log Ratio:** The change in expression level for a transcript between a baseline and an experiment array. This change is expressed as the $\log_2$ ratio. A signal log ratio of 1 is the same as a Fold Change of 2.

◆ **Signal Log Ratio High:** The upper limit of the Signal Log Ratio within a 95% confidence interval.

◆ **Signal Log Ratio Low:** The lower limit of the Signal Log Ratio within a 95% confidence interval.

Single Array Analysis: See Absolute Analysis.

Spike Controls: Controls that are added to the sample before cDNA synthesis (refer to Chapter 1 for more information).

◆ **Stat Pairs:** The number of probe pairs in the probe set.

◆ **Stat Common Pairs:** The number of common probe pairs on two arrays (experiment versus baseline) after saturation across the probe set is determined.

◆ **Stat Pairs Used:** The number of probe pairs in the probe set used in the Detection call.

◆ **Statistical Algorithms:** The algorithms contained in Microarray Suite Version 5.0. This algorithm was developed using standard statistical methods.

◆ **Tau:** A user-definable threshold used to determine the detection call.

Target: The sample applied as labeled (biotinylated), fragmented cRNA to a GeneChip® probe array for hybridization.

Wafer: The glass substrate onto which probes are synthesized during the manufacturing of probe arrays.

◆ **Wilcoxon's Signed Rank Test:** A non-parametric pair-wise comparison test. This test is used to determine the Detection and Change calls for analysis.

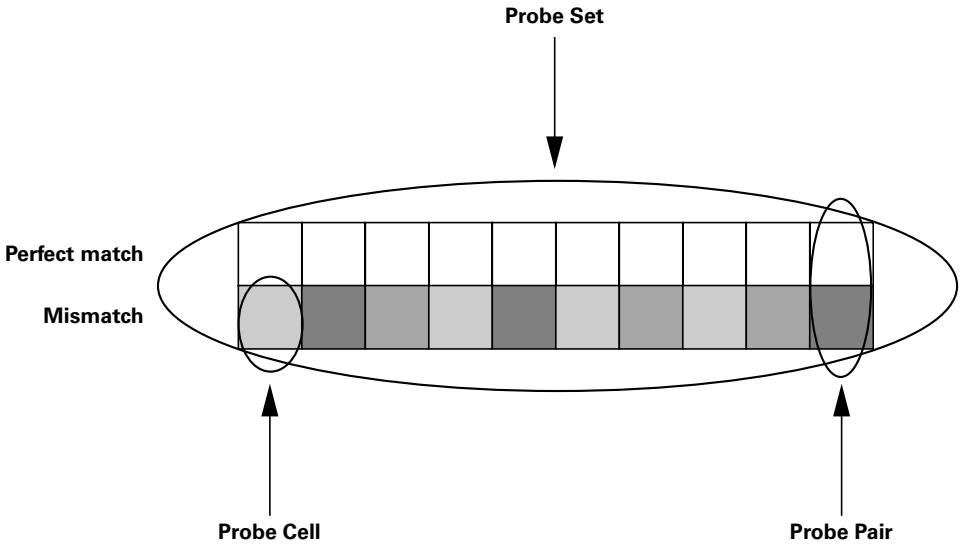


Figure 1

Appendix B: GeneChip Probe Array Probe Set Name Designations

In addition to the _at ("antisense target") and _st ("sense target") probe set name designations, there are other designations that reflect special characteristics of a particular probe set based on probe design and selection criteria. These designations are listed below.

Probe Set Name Designations Prior to HG-U133 Set:

_f_at (sequence family):

Probe set that corresponds to sequences for which it was not possible to pick a full set of 16-20 unique and/or shared similarity-constrained probes. Some probes in this set are similar (e.g., polymorphic) but not necessarily identical to other gene sequences. Some family members overlap a portion of the probe set. Family members can be singleton or an Affymetrix designated group of sequences.

--- --- --- --- ---	12345_f_at probes
----------	transcript #1
----------	transcript #2
----------	transcript #3
-X------X------	transcript #4 (w/polymorphisms)

_s_at (similarity constraint):

Probe set that corresponds to a small number of unique genes (<5%) that share identical sequence. Probes were chosen from the region that is common to these genes. Group members can be singleton or a group of sequences. For _s probe sets, there is not enough unique sequence to design a separate _at probe set.

--- --- --- ---	23456_s_at probes
----------	transcript #5
----------	transcript #6
----------	transcript #7

_g_at (common groups):

Probes chosen in region of overlap. To differentiate from an _s group, the sequences are represented as singletons (_at probe sets either have the same probe set ID number or the preceding probe set ID number) on the same probe array as well. In other words, for _g probe sets, there is enough unique sequence to design a separate _at probe set.

--- --- ---	34567_at probes
----------	transcript #8
--- --- --- ---	34568_g_at probes
----------	transcript #9

_r_at (rules dropped):

Designates sequences for which it was not possible to pick a full set of unique probes using Affymetrix' probe selection rules. Probes were picked after dropping some of the selection rules.

_i_at (incomplete):

Designates sequences for which there are fewer than the required numbers of unique probes specified in the design.

_b_at (ambiguous probe set):

All probe selection rules were ignored. Withdrawn from GenBank.

_l_at (long probe set):

Sequence represented by more than 20 probe pairs.

Probe Set Name Designations for HG-U133 Set (These are the only probe set extensions used in the HG-U133 Set)

_s_at:

Designates probe sets that share all probes identically with two or more sequences. The _s probe sets can represent shorter forms of alternatively polyadenylated transcripts, common regions in the 3' ends of multiple alternative splice forms, or highly similar transcripts. Approximately 90% of the _s probe sets represent splice variants. Some transcripts will also be represented by unique _at probe sets.

_x_at:

Designates probe sets that share some probes identically with two or more sequences. Rules for cross-hybridization were dropped in order to design the _x probe sets.

Appendix C: Microarray Suite Expression Defaults

MAS 5.0 Expression Analysis Default Settings

Parameter	# Probe Pairs/Probe Set	
	16-20	11
Alpha1	0.04	0.05
Alpha2	0.06	0.065
Tau	0.015	0.015
Gamma1L	0.0025	0.0045
Gamma1H	0.0025	0.0045
Gamma2L	0.003	0.006
Gamma2H	0.003	0.006
Perturbation	1.1	1.1

MAS 4.0 Expression Analysis Default Settings

Parameter	Value
SDT Multiplier	4.0*
Ratio Threshold	1.50
Ratio Limit	10.00
Pos/Neg Min	3.0
Pos/Neg Max	4.0
Pos Ratio Min	0.33
Pos Ratio Max	0.43
Avg Log Ratio Min	0.90
Avg Log Ratio Max	1.30
STP	3.0
CT Multiplier	<compute>
% Change Threshold	80
Inc/Dec Min	3.0
Inc/Dec Max	4.0
Inc Ratio Min	0.33
Inc Ratio Max	0.43
Dpos-Dneg Ratio Min	0.20
Dpos-Dneg Ratio Max	0.30
Avg Log Ratio Change Min	0.90
Avg Log Ratio Change Max	1.30

* The default SDT Multiplier value is 4.0 for antibody-stained arrays. The default SDT Multiplier value for non-antibody-stained arrays is 2.0.

Appendix D: File Types

File Types in Microarray Suite

Experiment Data File Name	File Extension	Description
Experiment Information File	*.exp	Contains information about the experiment name, sample, and probe array type. The experiment name also provides the name for subsequent test data files generated during the analysis of the experiment.
Data File	*.dat	The image of the scanned probe array.
Cell Intensity File	*.cel	The software derives the *.cel file from a *.dat file and automatically creates it upon opening a *.dat file. It contains a single intensity value for each probe cell delineated by the grid (calculated by the Cell Analysis algorithm).
Chip File	*.chp	The output file generated from the analysis of a probe array.
Report File	*.rpt	The report generated from the analysis output file (*.chp).
Experiment Information File	*.tif	A standard file format for graphic images. The Microarray Suite software exports graphic images in this file format.
Data File	*.txt, *.xls	A standard format for text files. The Microarray Suite software exports text in this file format. A standard format for Excel files.
Library Files	*.cif, *.cdf, *.psi	The probe information or library files contain information about the probe array design characteristics, probe utilization and content, and scanning and analysis parameters. These files are unique for each probe array type.
Fluidics Files	*.bin, *.mac	The fluidics files contain information about the washing, staining, and/or hybridization steps for a particular array format.

AFFYMETRIX, Inc.

3380 Central Expressway
Santa Clara, CA 95051 U.S.A.
Tel: 1-888-362-2447 (1-888-DNA-CHIP)
Fax: 1-408-731-5441
sales@affymetrix.com
support@affymetrix.com

AFFYMETRIX UK Ltd

Voyager, Mercury Park,
Wycombe Lane, Wooburn Green,
High Wycombe HP10 0HH
United Kingdom
Tel: +44 (0)1628 552550
Fax: +44 (0)1628 552585
saleseurope@affymetrix.com
supporteurope@affymetrix.com

www.affymetrix.com

For research use only.
Not for use in diagnostic procedures.

Part No. 701190 Rev. 2

AFFYMETRIX®