user manual

Lucidea Microarray ScoreCard

User's Guide v1.1

um 301-416

molecular dynamics
part of Amersham Pharmacia Biotech

The Microarray System is for research purposes only. It is not intended or approved for diagnosis of disease in humans or animals.

All goods and services are sold subject to the terms and conditions of sale of the company within the Amersham Pharmacia Biotech group that supplies them. A copy of these terms and conditions is available on request.

April 2001

Limited license notice

The Microarray ScoreCard (and/or its methods of use) is covered by one or more pending US and foreign patent applications.

Nucleic acid arrays, array makers, systems (instruments, software, and reagents) are licensed for research use only. No implied right to make, use, have made, offer to sell, lease, distribute, sell, or import nucleic acid probe arrays or any other product in which Affymetrix or Molecular Dynamics 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 or licensed and sold by Molecular Dynamics, or permitted licensees, and may not be used in database generation for external license or sale, or for service-based pharmaceutical research.

Amersham Pharmacia Biotech UK Limited Amersham Place Little Chalfont Buckinghamshire England HP7 9NA

Amersham Pharmacia Biotech AB SE-751 84 Uppsala Sweden

Amersham Pharmacia Biotech Inc 800 Centennial Avenue PO Box 1327 Piscataway NJ 08855 USA

Amersham Pharmacia Biotech Europe GmbH Munzinger Strasse 9 D-79111 Freiburg Germany

Molecular Dynamics Inc 928 East Arques Avenue Sunnyvale CA 94085-4520 USA

Table of contents

Preface

About this guide ix
Related publications ix
Special notices x
Assumptions x
Assistance xi

Chapter 1 Introduction

1.1 About the Lucidea Microarray ScoreCard control plate 1-1
1.2 About the Lucidea Microarray ScoreCard software 1-1
1.3 Workflow overview 1-3
1.4 Example files on the Lucidea Microarray ScoreCard CD 1-4

Chapter 2 The Lucidea Microarray ScoreCard control reagents

2.1 Control samples and functions 2-1
2.2 Spike mix composition and use 2-4
2.3 Storing the control plate 2-5
2.4 Preparing the control plate for spotting 2-5
2.5 Setting up the Generation III Array Spotter 2-6
2.6 Setting up the microscanner 2-7
2.7 Control element positions on a slide 2-7

Chapter 3 Preparing Lucidea Automated Spotfinder analysis files for import

3.1 Image analysis requirements 3-1
3.2 Analyzing image files from the Array Scanner 3-2
3.3 Analyzing image files from the GenePix 4000B Microscanner 3-2
3.4 Arranging the image analysis files 3-2

Chapter 4 Preparing ArrayVision 4.x analysis files for import

4.1 The image and analysis data files 4-1
4.2 Setting the display format 4-2
4.3 Retrieving the image files 4-2

4.4 Selecting the spot-labeling method 4-3
4.5 Selecting the background removal method 4-3
4.6 Defining the references 4-3
4.7 Defining the analysis settings 4-3
4.8 Analyzing the image files 4-3
4.9 Exporting the analysis data 4-4

Chapter 5 Preparing ArrayVision 5.1 analysis files for import

5.1 The image and analysis data files 5-1
5.2 The Protocol Editor window 5-2
5.3 Selecting the analysis parameters 5-3
5.4 Loading the images 5-3
5.5 Setting up the layout of the array 5-4
5.6 Setting the spot-labeling method 5-5
5.7 Selecting the analysis measures 5-7
5.8 Selecting background subtraction methods 5-7
5.9 Defining the references (optional) 5-8
5.10 Analyzing the image files 5-8
5.11 Exporting the analysis data 5-8

Chapter 6 Preparing GenePix Pro analysis files for import

6.1 The image and analysis data files 6-1
6.2 Opening the image files in GenePix Pro 6-1
6.3 Selecting the feature layout 6-2
6.4 Saving the results file 6-3

Chapter 7 Using the Lucidea Microarray ScoreCard software

7.1 Starting the Lucidea Microarray ScoreCard software 7-1
7.2 Getting software help 7-2

Chapter 8 Viewing the analysis results

8.1 Control data scatter plot 8-1
8.2 Ratio Analysis table 8-3
8.3 Ratio Precision table 8-4
8.4 Detection Limits table 8-5
8.5 Actin Gene Performance table 8-6
8.6 Normalized control plate data 8-6

8.7 System validation information 8-9
8.7.1 Pen-to-pen variability 8-9
8.7.2 Spot-set to spot-set variability 8-10
8.8 Quality report file content 8-11
8.9 Normalized data file content 8-13

Chapter 9 Comparing data from multiple slides

9.1 Assessing data quality and statistical significance 9-1
9.1.1 Signal level and precision 9-1
9.1.2 Sensitivity and Specificity 9-2
9.1.3 Gene Expression data Accuracy and Precision 9-2
9.2 Interpreting the Gene Expression data from single and multiple hybridizations 9-2
9.2.1 Calculating gene expression ratios 9-3
9.2.2 Combining gene expression ratios from multiple spots within a slide 9-3
9.2.3 Combining data from replicate slides 9-3
9.2.4 Comparing data using the common reference approach 9-4

Appendix A Data normalization and statistical significance

A.1 About the normalization method A-1
A.2 Normalizing the data A-2
A.2.1 Working with the data in logarithmic scale A-2
A.2.2 Calculating the uncorrected log ratio A-3
A.2.3 Calculating the normalized log ratios A-3
A.2.4 Adjusting the normalized log ratios A-3
A.2.5 Verifying the normalized results A-4
A.3 Estimating the confidence intervals for the gene expression data A-4
A.4 Calculating the default threshold for the mean NLR SD A-5

Appendix B Using the Amersham Pharmacia Biotech reflective slides with Lucidea Microarray ScoreCard

B.1 Effects of the reflective slide on data normalization B-1
B.2 Detecting the saturated pixels B-1
B.3 Eliminating the saturated pixels B-2

Glossary

Preface

About this guide

The *Lucidea Microarray ScoreCard User's Guide* explains how to use Lucidea™ Microarray ScoreCard™ v1.1 so that accurate comparisons can be made within an experiment and across multiple experiments.

A PDF version of this guide is provided. The PDF file resides in the …\Molecular Dynamics\Microarray\Microarray ScoreCard folder.

The software is provided with the control plate for processing the control plate data. A Help file for the software is available. To access the Help file, in the software, choose **Help Topics** from the Help menu.

Related publications

In addition to the *Lucidea Microarray ScoreCard User's Guide*, the following publications are provided with the Microarray System:

- *Microarray System Applications Guide: Gene Expression*—Provides the latest information on protocols. Contact Amersham Pharmacia Biotech for more information. See Assistance for the contact information.
- *Microarray System Training Manual*—Provides training material to be used during the training course. Contact Amersham Pharmacia Biotech for more information. See Assistance for the contact information.
- *Microarray System Setup Instructions*—Provides instructions on how to set up the Microarray System database and how to set up the computer monitor for the analysis workstation.
- *Microarray System Generation III User's Guide*—Explains how to use Gen3DB, the Array Spotter, and the Array Scanner for microarray experiments.
- *Microarray System ASC-XT User's Guide for the Generation III Array Spotter*—Explains how to use the ASC-XT software to produce higher density arrays with the Generation III Array Spotter instrument.
- Lucidea Automated Spotfinder Software Help—Explains how to use the Lucidea Automated Spotfinder to quantitate microarray images and produce analysis reports.

- *ArrayVision Operations Manual*—Explains how to use the ArrayVision™ software to quantitate microarray images and produce analysis reports.
- *GenePix 4000B Microarray Scanner User's Guide*—Explains how to use the GenePix™ 4000B microarray scanner to produce images of microarray slides.
- *GenePix Pro 3.0 Microarray Acquisition and Analysis Software User's Guide*—Explains how to use the GenePix Pro 3.0 software to quantitate images of slides that are scanned on the GenePix 4000 microscanner.
- *IQ Solutions Documentation*—Explains how to use the ImageQuant™ software to view and prepare the images for publication.

Special notices

Make sure you follow the precautionary statements presented in this guide.

Caution **Indicates that loss of data or invalid data could occur if the user fails to comply with the advice given.**

Important **Highlights information that is critical for optimal performance of the system.**

Note: Identifies items of general interest.

Assumptions

This guide assumes that you are familiar with the Molecular Dynamics Generation III Microarray System, the GenePix microscanner, and one of the following image analysis software: Lucidea Automated Spotfinder, ArrayVision, or GenePix Pro. For information about the Microarray System, Lucidea Automated Spotfinder, ArrayVision, or the GenePix microscanner, see the documents listed in the Related publications section.

This guide assumes that you have a basic understanding of the methods used in gene expression studies. This guide also assumes that you have a basic understanding of mathematical terms and calculations commonly used in scientific experiments.

The software-related instructions in this guide assume that you have basic computer skills. You should be familiar with the Microsoft™ Windows NT™ or Windows™ 2000 graphical user interface. If you do not have these skills, consult the appropriate documentation or refer to the Help file for the operating system.

Assistance

If you have questions about or require assistance with the Lucidea Microarray ScoreCard control plate or the Lucidea Microarray ScoreCard software, contact your local Amersham Pharmacia Biotech service office. For the local office nearest you, see the Amersham Pharmacia Web site. You can also contact one of the following offices.

US and Canada
Molecular Dynamics Inc.
928 East Arques Avenue
Sunnyvale, CA 94085-4520
USA
Tel: (1) (800) 743-7782 or (1) (408) 773-1222
Fax: (1) (408) 773-1493
Email: masupport@am.apbiotech.com

International
Amersham Pharmacia Biotech UK Limited
Amersham Place, Little Chalfont, Buckinghamshire
England HP7 9NA
Tel: (44) (1494) 544000
Fax: (44) (1494) 542266

Europe
Amersham Pharmacia Biotech Europe GmbH
Munzinger Strasse 9, D-79111 Freiburg
Germany
Tel: (49) (76) 1490-30
Fax: (49) (76) 1490-3159

Japan
Amersham Pharmacia Biotech KK
Otsuka Daiichi Seimei Bldg.
32-22, Higashi Ikebukuro 2-chome
Toshima-ku
Tokyo 170, Japan
Tel: (81) (3) 5331-9317
Fax: (81) (3) 5331-9372

Asia Pacific

Amersham Pharmacia Biotech Asia Pacific Ltd
15/F, North, Cornwall House, 979 King's Road
Taikoo Place, Quarry Bay
Hong Kong
Tel: (852) 2811-8693
Fax: (852) 2811-5251

Web site

www.apbiotech.com/applications/microarray

Chapter 1 Introduction

Lucidea Microarray ScoreCard v1.1 consists of the control plate and the software. This chapter provides an overview of the control plate and the software. The topics are—

- About the Lucidea Microarray ScoreCard control plate (section 1.1)
- About the Lucidea Microarray ScoreCard software (section 1.2)
- Workflow overview (section 1.3)
- Example files on the Lucidea Microarray ScoreCard CD (section 1.4)

1.1 About the Lucidea Microarray ScoreCard control plate

In evaluating microarray experiment results, you must first minimize variations in the measurements so that accurate comparisons can be made within an experiment and across multiple experiments. One method of minimizing the variations is the use of control samples so that you can relate the experiment data to the control data in a quantitative manner.

You use the control plate in your experiments for this purpose. The control plate is a 384-well microplate that contains 12 replicates of 32 control samples. For more information about the control samples, see chapter 2.

When using the Microarray System, you can set up your experiment to include the control plate. The control samples are placed on each slide to allow spot-to-spot comparisons within and across multiple slides.

1.2 About the Lucidea Microarray ScoreCard software

Caution **The software must be used only with the control plate. Using the software with other control samples will not provide accurate results.**

The software processes microarray experiment data from one slide. Using the results from the software, you can compare the data from one slide with the results from other slides within and across multiple experiments.

The software—

- Allows you to import two-color image analysis data from Lucidea Automated Spotfinder, ArrayVision, or GenePix Pro.
- Processes the experiment data from one slide by calculating the quality measures.
- Normalizes the data on the slide.
- Correlates the sample information to the experiment values.
- Performs additional calculations so that you can assess the system performance.
- Prints and exports all the processed results to tab-delimited text files.

Important **The software does not provide an absolute scale for the assessment of the experiment data. You must accept or reject the data based on your interpretation of the calculated results.**

The software requires—

- 233 MHz or faster Pentium™ or equivalent computer, 64 MB RAM or more, Windows NT 4.0 (Service Pack 5 or better) or Windows 2000, and a minimum of 50 MB of free hard disk space
- Internet Explorer 5.5 or higher
- Two-column arrays produced with the Molecular Dynamics Microarray System Generation III Array Spotter, running either the Generation III Array Spotter Control software or the ASC-XT software
- Lucidea Automated Spotfinder 1.0 or later analysis data in tab-delimited text format
- ArrayVision 4.0 or later analysis data in tab-delimited text format
- GenePix Pro 3.0 or later analysis data in the GenePix Pro results (.gpr) format

1.3 Workflow overview

Figure 1-1 shows the overall workflow for the Lucidea Microarray ScoreCard.

Caution **The image analysis files must be properly set up for input into the Lucidea Microarray ScoreCard software. Be sure to follow the instructions in chapters 3 through 6 to correctly generate the image analysis data for the software.**

For details on how to set up the analysis in Lucidea Automated Spotfinder, ArrayVision, or GenePix, see the user documentation for the image analysis software.

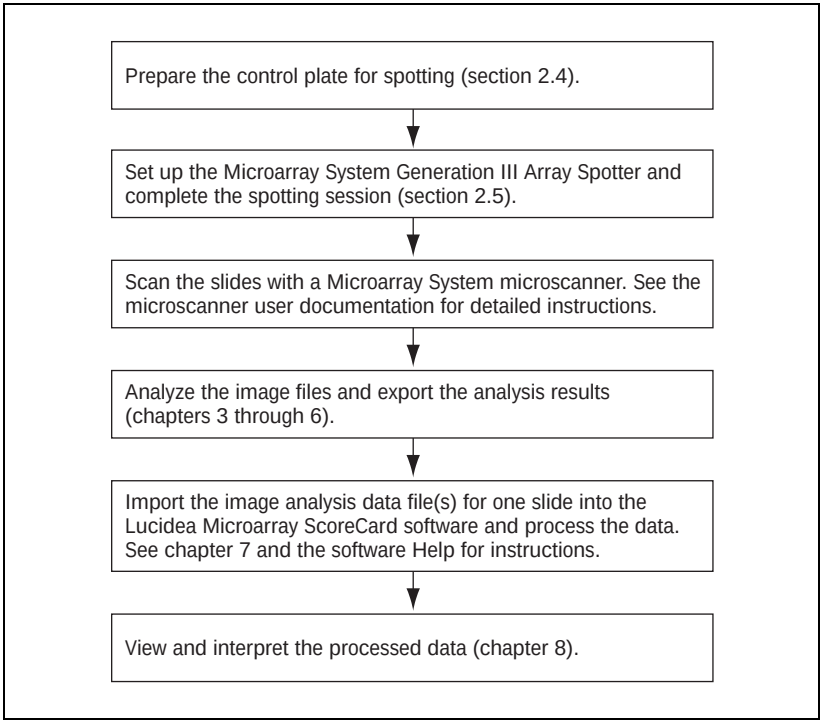


Figure 1-1. Lucidea Microarray ScoreCard workflow.

1.4 Example files on the Lucidea Microarray ScoreCard CD

In addition to the software and documentation, the Lucidea Microarray ScoreCard CD contains the following example files in the …/Microarray folder:

- **Control_elements_image.tif—**An image file that shows three pen areas. Each pen area consists of three rows (or three microplates) of control elements.
- **Lucidea Automated Spotfinder 1.0_lr_1/2.txt—**A set of four text files containing the output from Lucidea Automated Spotfinder 1.0.
- **415 ArrayVision 4.0.txt—**A text file containing the output from ArrayVision 4.0, using the required setup as described in chapter 3.
- **415 ArrayVision 5.1.txt—**A text file containing the output from ArrayVision 5.1, using the required setup as described in chapter 4.
- **GenePix Pro 3.0.gpr—**A GenePix Pro results file containing the output from GenePix Pro 3.0.
- **Control_Plate_content.txt—**A text file containing clone information.

Note: The same image file was loaded into ArrayVision 4.0 and 5.1 so the analysis results from both versions are almost identical.

Chapter 2 The Lucidea Microarray ScoreCard control reagents

This chapter describes the control samples in the Lucidea Microarray ScoreCard control plate, how to use the control samples, and the layout of the control samples on a slide. The topics in this chapter are—

- Control samples and functions (section 2.1)
- Spike mix composition and use (section 2.2)
- Storing the control plate (section 2.3)
- Preparing the control plate for spotting (section 2.4)
- Setting up the Generation III Array Spotter (section 2.5)
- Setting up the microscanner (section 2.6)
- Control element positions on a slide (section 2.7)

Note: Currently, the control plate requires and is limited to two-color analysis.

2.1 Control samples and functions

Table 2-1 lists the control samples, their positions in the control plate, and the corresponding identification in the software. For detailed instructions on how to prepare the control samples for spotting, see the instructions shipped with the control reagents.

Table 2-1. Control samples and positions in a microplate

Well set	(well positions)	Control sample type	Description	ID in ScoreCard
1	(A1 through A12)	Positive control	Total human DNA/complex target	1PC
2	(B1 through B12)	Negative control	*Arabidopsis thaliana* protein G1p	1NC
3	(C1 through C12)	Dynamic range control	Yeast Intergenic Region from Chrom XI	1DR
4	(D1 through D12)	Dynamic range control	Yeast Intergenic Region from Chrom XIV	2DR
5	(E1 through E12)	Dynamic range control	Yeast Intergenic Region from Chrom VII	3DR
6	(F1 through F12)	Dynamic range control	Yeast Intergenic Region from Chrom VII	4DR
7	(G1 through G12)	Dynamic range control	Yeast Intergenic Region from Chrom XII	5DR
8	(H1 through H12)	Dynamic range control	Yeast Intergenic Region from Chrom XII	6DR
9	(I1 through I12)	Ratio control	Yeast Intergenic Region from Chrom XII	1RC
10	(J1 through J12)	Ratio control	Yeast Intergenic Region from Chrom XII	2RC
11	(K1 through K12)	Ratio control	Yeast Intergenic Region from Chrom XII	3RC
12	(L1 through L12)	Ratio control	Yeast Intergenic Region from Chrom XIII	4RC
13	(M1 through M12)	Negative control	PolydA oligonucleotide	2NC
14	(N1 through N12)	Housekeeping gene	Actin gamma-1 Hs.14376	1HG
15	(O1 through O12)	Housekeeping gene	Glyceraldehyde-3-phosphate dehydrogenase Hs.169476	2HG
16	(P1 through P12)	Housekeeping gene	Ubiquinol-cytochrome c reductase core protein II Hs.173554	3HG
17	(A13 through A24)	Housekeeping gene	Casein kinase II beta polypeptide Hs.165843	4HG
18	(B13 through B24)	Reserved for future use	Empty	1Reserved
19	(C13 through C24)	Reserved for future use	Empty	2Reserved
20	(D13 through D24)	Reserved for future use	Empty	3Reserved
21	(E13 through E24)	Reserved for future use	Empty	4Reserved
22	(F13 through F24)	Housekeeping gene	Est highly similar to NY-REN-37 antigen Hs.173684	5HG
23	(G13 through G24)	Negative control	Empty	3NC
24	(H13 through H24)	Housekeeping gene	Human hydroxymethyl glutaryl-CoA lyase Hs.831	6HG

Table 2-1. Control samples and positions in a microplate (continued)

Well set	(well positions)	Control sample type	Description	ID in ScoreCard
25	(I13 through I24)	Housekeeping gene	Neuroblastoma RAS viral oncogene homolog Hs.69855	7HG
26	(J13 through J24)	Housekeeping gene	Eukaryotic translation initiation factor 4A, isoform 1 Hs.129673	8HG
27	(K13 through K24)	Housekeeping gene	Ubiquinol-cytochrome c reductase core protein II Hs.173554	9HG
28	(L13 through L24)	Housekeeping gene	Enoyl Coenzyme A hydrolase, short chain 1 Hs.76394	10HG
29	(M13 through M24)	Housekeeping gene	Actin gamma-1 Hs.14376	11HG
30	(N13 through N24)	Negative control	*Bacillus subtilis* diaminopimelate decarboxylase lysA	4NC
31	(O13 through O24)	Negative control	*Bacillus subtilis* phenylalanine biosynthesis associated protein pheB	5NC
32	(P13 through P24)	Positive control	Total human DNA/complex target	2PC

The total human DNA/complex targets, present in well sets 1 and 32, are good positive control (PC) elements because they always produce measurable signals regardless of the tissue-specific mRNA used for hybridizing sample synthesis. These positive controls are useful in ArrayVision analysis because their positive signals serve as grid anchors for array template creation. They are also useful indicators of the quality of the hybridizing mRNA samples.

The various housekeeping genes (HG) present in well sets 14 through 17 and well sets 24 through 29 serve as additional positive control elements. The expression of housekeeping genes is unlikely to fluctuate significantly within a given tissue. Therefore, they are used for the comparative analysis of the slides within a single experiment or across multiple experiments.

Note: The housekeeping gene, Actin gamma-1, is used twice in the control plate, in well set 14 (1HG) and well set 29 (11HG). It is used to calculate the detection limit to actin ratio and the percent coefficient of variation (%CV) for each dye. For details about these calculations, see chapter 6.

The *A. thaliana* protein G1p in well set 2 and the bacterial genes from *B. subtilis* in well sets 30 and 31 are not present in human-derived cDNAs.

Therefore, they are useful as negative control (NC) elements for evaluating the degree of the background signal levels of nonspecific hybridization. The plant gene is also used to determine the detection limit.

The polydA oligonucleotide in well set 13 does not lead to high levels of hybridization. Therefore, it is also a good negative control (NC) element for verifying the degree of nonspecific hybridization in an experiment.

In well sets 3 through 8, six sequences derived from intergenic regions in yeast have been shown not to hybridize with mammalian cDNAs. Using a spike mix of the corresponding mRNA in the labeling reactions and hybridizing the mix on yeast targets produces signals that serve as dynamic range (DR) controls in the experiment (section 2.2). Dynamic range controls allow you to evaluate the dynamic range and sensitivity of the system.

In well sets 9 through 12, four other sequences derived from intergenic regions in yeast have also been shown not to hybridize with mammalian cDNAs. Using a spike mix of the corresponding mRNA in the labeling reactions and hybridizing the mix on yeast targets produces signals that serve as ratio controls (RC) in the experiment (section 2.2). Ratio controls are references for measuring differential expression.

2.2 Spike mix composition and use

Important **Avoid repeated freezing and thawing of the spike mixes.**

Provided with the control plate are 20 tubes of spike mixes:

- 10 tubes, each containing 500 µl of the Cy™3 spike mix
- 10 tubes, each containing 500 µl of the Cy5 spike mix

For the dynamic range control and ratio control samples, mRNA spikes are added in the Cy3 or Cy5 labeling reactions. Table 2-2 shows the concentrations and the relative abundance (compared to mRNA in the labeling reactions) of the spike mixes that correspond to the dynamic range control elements and the ratio control elements in the control plate. These values can be used as quality measures that help you assess the output data in the software.

Table 2-2. Spike mix composition

Control sample	Cy3:Cy5 Ratio	Conc. in mix (pg/5µl mix) Cy3	Cy5	Relative abundance
1DR	1:1	33 000	33 000	3.3%
2DR	1:1	10 000	10 000	1%
3DR	1:1	1 000	1 000	0.1%
4DR	1:1	330	330	0.033%
5DR	1:1	100	100	0.01%
6DR	1:1	33	33	0.0033%
1RC	1:3	1 000	3 000	NA
2RC	3:1	3 000	1 000	NA
3RC	1:10	1 000	10 000	NA
4RC	10:1	10 000	1 000	NA

For the labeling reactions, add 5 µl of the appropriate spike mix per microgram of mRNA. Use the spiked mRNA in the first-strand cDNA synthesis reaction. The spiked mRNA can be labeled using oligo dT and/or random primers.

Important **If you plan to use the reflective slides from Amersham Pharmacia Biotech, see appendix B for instructions on how to adjust the spike mix concentration to reduce pixel saturation and eliminate inaccurate data.**

2.3 Storing the control plate

You can store the control plate for future experiments. Before you store a used control plate, cover the plate with a plastic or foil adhesive film. Store the control plate at -20 °C.

2.4 Preparing the control plate for spotting

Caution **Make sure you follow the instructions shipped with the Microarray ScoreCard control reagents for diluting and preparing the control samples for spotting.**

To use the control plate—

1. Remove the control plate from storage. Leave the adhesive cover on the plate.
2. Warm the plate to room temperature.
3. Spin the plate at 3000 x g for 3–5 minutes.

Caution **Amersham Pharmacia Biotech recommends a long spin to collect condensate from the adhesive film and to prevent cross-sample contamination.**

4. Peel off the adhesive cover.
5. Prepare the control plate exactly as described in the instructions shipped with the Microarray ScoreCard control reagents.
6. Load the control plate into the Generation III Array Spotter instrument. For instructions on how to properly load the microplate into the Array Spotter, see the *Microarray System Generation III User's Guide.*

Important **To facilitate image analysis, Amersham Pharmacia Biotech recommends you place the control plate in slot 1 in the microplate elevator so that the control samples are placed in the first row of each pen area on the slides.**

7. Close the Array Spotter door(s).

2.5 Setting up the Generation III Array Spotter

When you set up the spotting session, make sure you—

- Choose the Normal spotting mode in the Generation III Array Spotter Control software. **Note:** For system verification only, you can use the Single Plate spotting mode in Array Spotter Control to spot one control plate repeatedly.
- Choose the Normal spotting mode or the High Density spotting mode, Duplicate spotting pattern in the ASC-XT software. **Note:** Make sure you set the spotting parameters so that all the control elements from each pen are spotted in sequence.
- Select the Complete Last Row option in the ASC-XT software so that each row contains an equal number of spots.
- Use only full microplates (32 well sets).

Important **Using partial microplates causes calculation errors in the software.**

For detailed instructions on how to set the spotting parameters, see the *Microarray System Generation III User's Guide.* If you are using the ASC-XT software, see the *Microarray System ASC-XT User's Guide for the Generation III Array Spotter.*

After you set up the Array Spotter, you can begin the spotting session.

2.6 Setting up the microscanner

To scan the slides containing the spotted control samples, you use the typical procedure for your microscanner.

Important **If you plan to use reflective slides from Amersham Pharmacia Biotech, see appendix B for instructions on how to adjust the PMT setting to reduce pixel saturation and eliminate inaccurate data.**

2.7 Control element positions on a slide

When you specify the normal spotting mode or the duplicate spotting pattern in the high-density spotting mode, each spot set includes the positive control elements, negative control elements, dynamic range control elements, and ratio control elements.

Figure 2-1 shows an example of the placement of the control samples on a slide. In this example—

- The normal spotting mode is specified in the Generation III Array Spotter Control software.
- Four plates are spotted (four rows).
- The control samples are always in the first row of every pen area.
- The first and last control elements in the first row have the strong signals.

For detailed information about the Normal spotting pattern, see the *Microarray System Generation III User's Guide.*

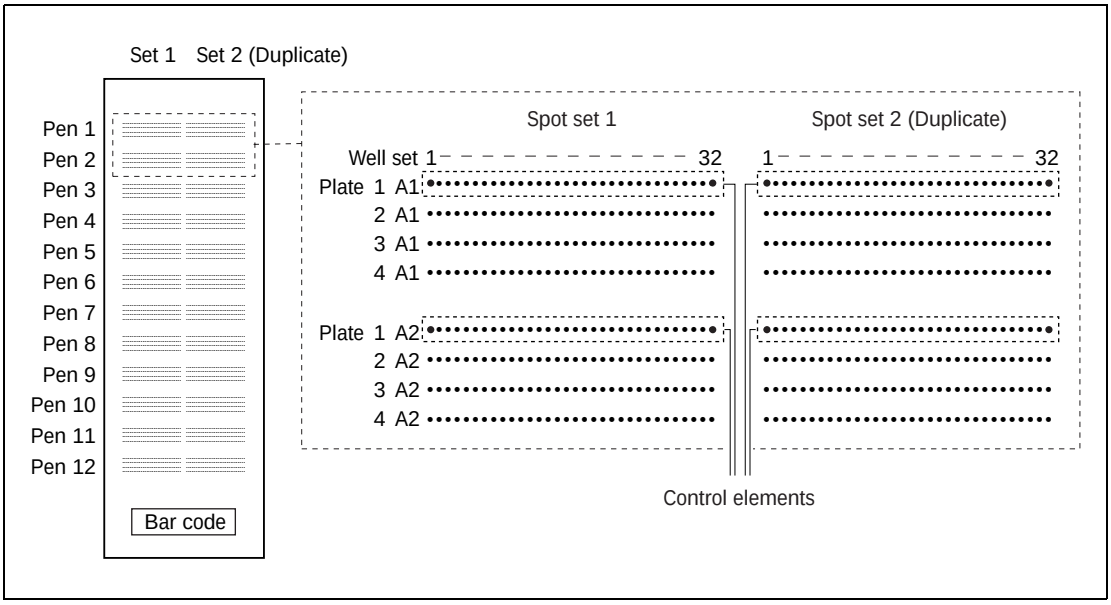


Figure 2-1. The control samples on a slide.

The Control_elements_image.tif file on the Lucidea Microarray ScoreCard CD shows an array of control samples that are spotted repeatedly on the slide. You can use this image to estimate the appearance (signal strength, color, and sequence) of the control elements on your slides.

Chapter 3 Preparing Lucidea Automated Spotfinder analysis files for import

This chapter describes the image analysis requirements for the Lucidea Automated Spotfinder to generate output that can be imported into the Lucidea Microarray ScoreCard software. The topics in this chapter are—

- Image analysis requirements (section 3.1)
- Analyzing image files from the Array Scanner (section 3.2)
- Analyzing image files from the GenePix 4000B Microscanner (section 3.3)
- Arranging the image analysis files (section 3.4)

This chapter provides general guidelines required for the correct analysis output. For detailed instructions on how to set up the analysis, refer to the Lucidea Automated Spotfinder Help.

3.1 Image analysis requirements

For two-color microarray experiments, Lucidea Automated Spotfinder requires four individual image files:

- Spot set 1 (left), dye 1
- Spot set 1 (left), dye 2
- Spot set 2 (right), dye 1
- Spot set 2 (right), dye 2

Lucidea Automated Spotfinder produces a set of four individual analysis files, each containing the data from one of the four image files. The Lucidea Microarray ScoreCard software can open these files directly and perform the calculations.

3.2 Analyzing image files from the Array Scanner

In two-color microarray experiments, the Generation III Array Scanner produces four image files in the following order:

- Spot set 1 (left), dye 1
- Spot set 1 (left), dye 2
- Spot set 2 (right), dye 1
- Spot set 2 (right), dye 2

These files match the requirements for Lucidea Automated Spotfinder. Therefore, you can scan the slides without using any special settings.

3.3 Analyzing image files from the GenePix 4000B Microscanner

If you want to use Lucidea Automated Spotfinder to analyze the image files from the GenePix 4000B Microscanner, you need to set up the scan as follows to produce four separate image files:

- Scan spot set 1 and spot set 2 independently.
- Scan each of the two dyes independently.

For instructions on how to scan a slide using the GenePix 4000B Microscanner, see the *GenePix 4000B Microarray Scanner User's Guide.*

3.4 Arranging the image analysis files

Lucidea Automated Spotfinder creates four analysis files that are in the tab-delimited text format. You need to keep these files in the same folder to import them into the Lucidea Microarray ScoreCard software.

To automatically store the analysis files in the same folder after analysis, in the Preferences window (figure 3-1), select **Results stored in default output directly** in the Results Location area. Alternatively, you can move the four files into a new folder after you finish the analysis.

For information on how to use Lucidea Automated Spotfinder, see the Lucidea Automated Spotfinder Help.

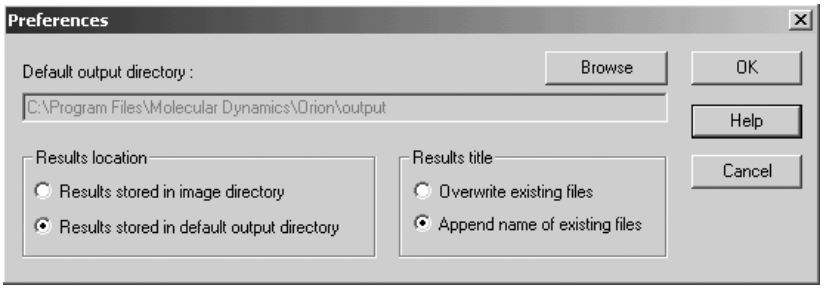


Figure 3-1. The Preferences window.

Chapter 4 Preparing ArrayVision 4.x analysis files for import

This chapter describes the analysis settings that must be used in ArrayVision 4.x to generate output that can be imported into the Lucidea Microarray ScoreCard software. The topics in this chapter are—

- The image and analysis data files (section 4.1)
- Setting the display format (section 4.2)
- Retrieving the image files (section 4.3)
- Selecting the spot-labeling method (section 4.4)
- Selecting the background removal method (section 4.5)
- Defining the references (section 4.6)
- Defining the analysis settings (section 4.7)
- Analyzing the image files (section 4.8)
- Exporting the analysis data (section 4.9)

Depending on which 4.x version of ArrayVision you are using, the exact method for setting up the analysis can vary. This chapter provides general guidelines required for the correct analysis output. For detailed instructions on how to set up the analysis, refer to the *ArrayVision Operations Manual.*

This chapter also provides analysis settings that are compatible with Gen3DB. For more information about Gen3DB, see the *Microarray System Generation III User's Guide.*

Note: An example analysis file generated by ArrayVision 4.0 is in the …\Molecular Dynamics\Microarray\Microarray ScoreCard folder.

4.1 The image and analysis data files

In two-color microarray experiments, the Generation III Array Scanner produces four image files in the following order:

- Spot set 1 (left), dye 1
- Spot set 1 (left), dye 2
- Spot set 2 (right), dye 1
- Spot set 2 (right), dye 2

When you perform image analysis in ArrayVision, you can produce one of the following:

- A single analysis file containing data from all four image files
- A set of four individual analysis files, each containing the data from one of the four image files (compatible with Gen3DB)

The Lucidea Microarray ScoreCard software can analyze the data in either file type. However, the software processes only files with specific field names, and the fields must be in a specific order. Therefore, you must make sure that the analysis itself is set up properly for producing the correct output file(s).

4.2 Setting the display format

In ArrayVision, how you display the images in the window affects how the image analysis data is produced: a single file that contains data from all four images or four individual files each containing data from a single image.

When you set up the display format, be sure to use one of the following settings:

- For the single-file output setting, choose—
  - Total # of channels—4
  - Total # of views—1 or 4
- For the four-file output setting (Gen3DB-compatible)—
  - Total # of channels—1
  - Total # of views—1

4.3 Retrieving the image files

You should retrieve all four image files associated with the slide. When you retrieve the image files, you must retrieve them in the order they were generated from the Generation III Array Scanner (section 4.1):

- L1—Spot set 1 (left), channel 1
- L2—Spot set 1 (left), channel 2
- R1—Spot set 2 (right), channel 1
- R2—Spot set 2 (right), channel 2

4.4 Selecting the spot-labeling method

The Lucidea Microarray ScoreCard software requires that the array elements be ordered in the horizontal sequence in the image analysis file (figure 4-1). In ArrayVision, when you create an array template in the Template Definition window, be sure to choose the Sequential Horizontal labeling method.

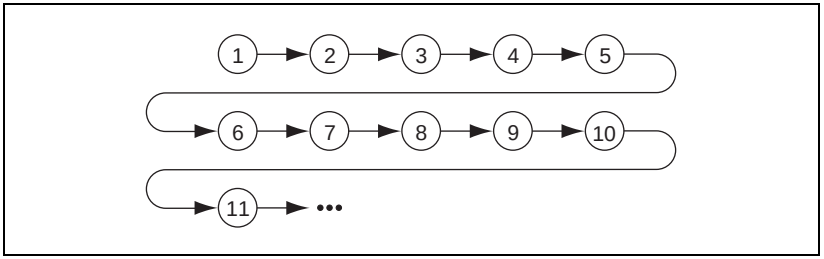


Figure 4-1. The sequential horizontal order.

4.5 Selecting the background removal method

When you set up the analysis for background removal, you can select any location for background removal methods except the Selected Array Elements option. You select the background removal method when you define the array template.

4.6 Defining the references

A reference is the calculated value to which all array elements are normalized. However, do not define a reference using an array element because the software will exclude the array element from the exported data and misalign the remaining data.

4.7 Defining the analysis settings

When you define the analysis settings, choose one of the following array types:

- For the single-file output setting, select **Multiple arrays** and set the Array channels range from 1 to 4.
- For the four-file output setting (Gen3DB-compatible), select **Single array.**

4.8 Analyzing the image files

After you finish setting up the parameters as described in this chapter, analyze the image files using your usual process.

4.9 Exporting the analysis data

After ArrayVision finishes the analysis, export the analysis data to tab-delimited text file(s). These files can be imported into the Lucidea Microarray ScoreCard software.

To make sure you have set up the analysis correctly, compare your exported file with the example input file in the …\Microarray folder. Note that the example file was generated by ArrayVision 4.0.

Chapter 5 Preparing ArrayVision 5.1 analysis files for import

This chapter describes the required analysis settings for ArrayVision 5.1 to generate output that can be imported into the Lucidea Microarray ScoreCard software. The topics in this chapter are—

- The image and analysis data files (section 5.1)
- The Protocol Editor window (section 5.2)
- Selecting the analysis parameters (section 5.3)
- Loading the images (section 5.4)
- Setting up the layout of the array (section 5.5)
- Setting the spot-labeling method (section 5.6)
- Selecting the analysis measures (section 5.7)
- Selecting background subtraction methods (section 5.8)
- Defining the references (optional) (section 5.9)
- Analyzing the image files (section 5.10)
- Exporting the analysis data (section 5.11)

This chapter provides only the instructions required for the correct analysis output. For detailed instructions on how to set up the analysis, refer to the *ArrayVision Operations Manual.*

This chapter also provides analysis settings that are compatible with Gen3DB. For more information about Gen3DB, see the *Microarray System Generation III User's Guide.*

Note: An example analysis file generated by ArrayVision 5.1 is in the …\Microarray folder.

5.1 The image and analysis data files

In two-color microarray experiments, the Generation III Array Scanner produces four image files in the following order:

- Spot set 1 (left), dye 1
- Spot set 1 (left), dye 2
- Spot set 2 (right), dye 1
- Spot set 2 (right), dye 2

When you perform image analysis in ArrayVision, you can produce one of the following:

- A single analysis file containing data from all four image files
- A set of four individual analysis files, each containing the data from one of the four image files (compatible with Gen3DB)

The Lucidea Microarray ScoreCard software can analyze the data in either file type. However, because the software can process only files with specific field names and the fields must be in a particular order, you must make sure that the analysis itself is set up properly for producing the correct output file.

5.2 The Protocol Editor window

In ArrayVision 5.1, you must set up the analysis protocol (equivalent to the analysis template in version 4.x) before you can begin the image analysis. To set up an analysis protocol, you must be in the Protocol Editor window (figure 5-1).

Note that the Protocol Editor window is split into two panes:

- **Navigation pane—**Displays an explorer tree.
- **Display pane—**Displays the content of the item you choose from the explorer tree.

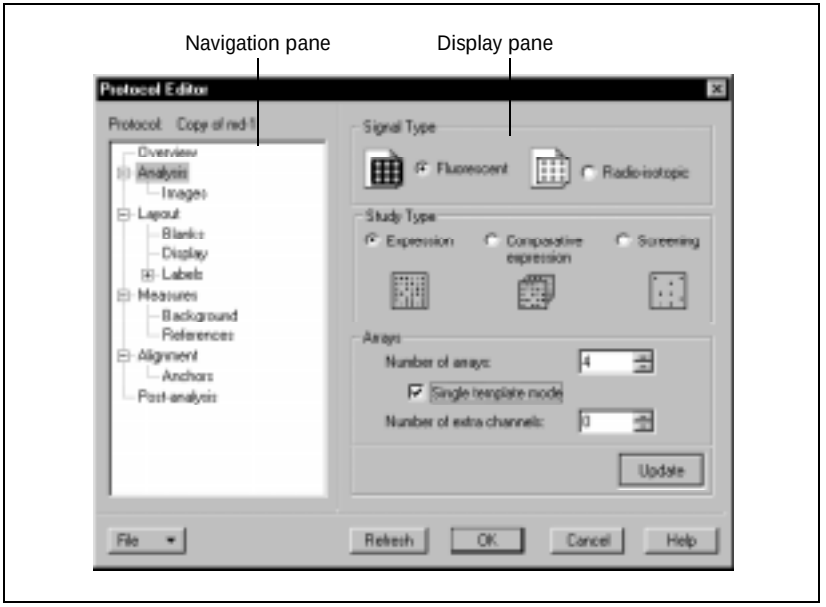


Figure 5-1. The Protocol Editor window.

5.3 Selecting the analysis parameters

From the navigation pane (figure 5-1), choose **Analysis.** The Analysis pane appears. In the Analysis pane—

- In the Signal Type area, select **Fluorescent.**
- In the Study Type area, select **Expression.**
- In the Arrays area, select **4** from the Number of arrays list.

5.4 Loading the images

From the navigation pane, choose **Images** (under Analysis). The Images pane appears (figure 5-2). In the Images pane, retrieve the four image files in the following order:

- L1—Spot set 1 (left), channel 1
- L2—Spot set 1 (left), channel 2
- R1—Spot set 2 (right), channel 1
- R2—Spot set 2 (right), channel 2

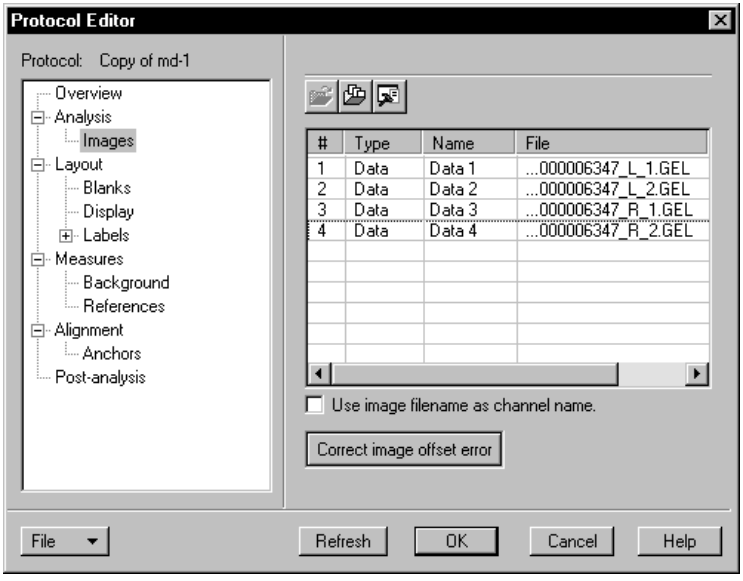


Figure 5-2. Loading images in ArrayVision 5.1.

5.5 Setting up the layout of the array

Use the guidelines in this section when you set up the array layout for analysis. From the navigation pane, choose **Layout**. The Layout pane appears (figure 5-3).

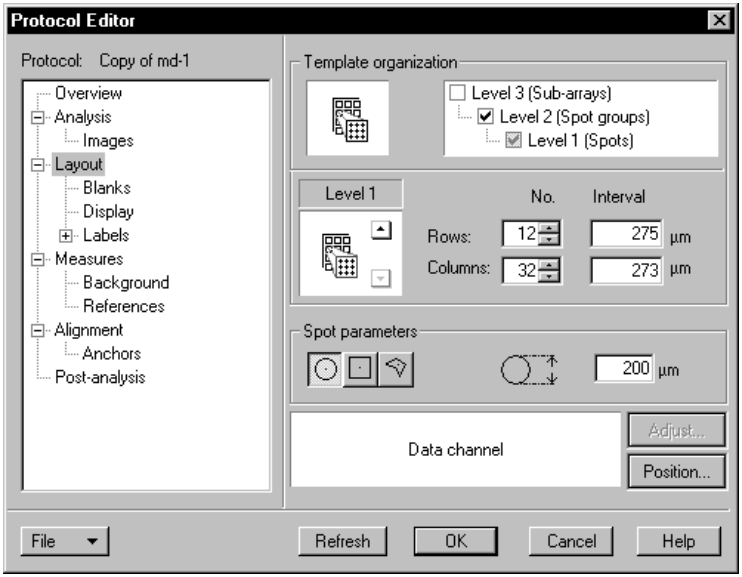


Figure 5-3. The Layout pane in the Protocol Editor window, displaying the Level 1 array layout parameters.

In the Layout pane—

1. In the Template organization area, select Level 2 (Spot groups). Note that Level 1 (Spots) should already be selected.
2. In the Level 1 area, make sure the number of rows is 12 or fewer and the number of columns is 32.
3. Click the up arrow in the Level 1 area icon. The Level 2 layout information appears (figure 5-4).
4. In the Level 2 area, make sure the number of rows is 12 and the number of columns is 1.
5. Select the remaining parameters that are best suited for your experiment.

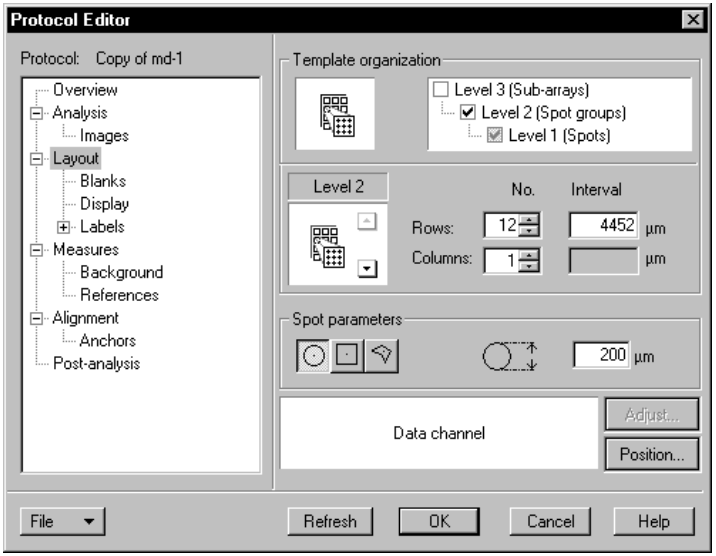


Figure 5-4. Setting Level 2 array layout parameters in the Layout pane.

5.6 Setting the spot-labeling method

From the navigation pane, choose **Labels** (under Layout). The Labels pane appears (figure 5-5).

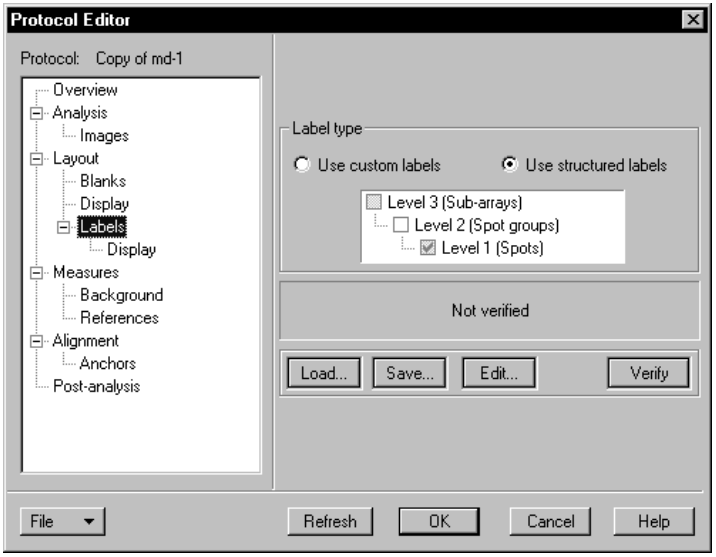


Figure 5-5. Setting the spot-labeling method in the Labels pane.

In the Labels pane—

1. In the Label type area, select **Use structured labels.**
2. Deselect Level 2 (Spot groups). Note that Level 1 is preselected.
3. Click **Edit** to define the labels. The Label Editor window appears (figure 5-6).
4. In the Label Editor window, click the **Auto-number spots** button, point to **Autosequence Row-wise**, and choose **Numeric labels (1, 2, …).** The Label Editor window displays the correct spot sequence for Lucidea Microarray ScoreCard (figure 5-6).

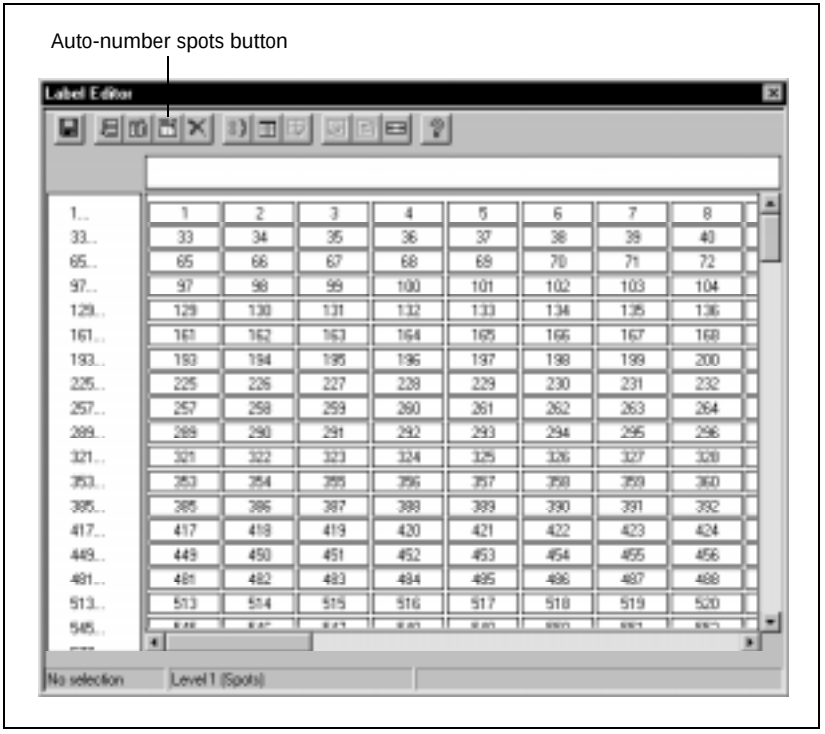


Figure 5-6. The Label Editor window, displaying the correct spot sequence for import into the Lucidea Microarray ScoreCard software.

5.7 Selecting the analysis measures

From the navigation pane, choose **Measures.** The Measures pane appears (figure 5-7). In the Measures pane—

1. Choose **Volume** or **Density** from the Principle Measure list.
2. Choose **Mean** or **Median** from the Background list.
3. Select any measure from the **Name** list.

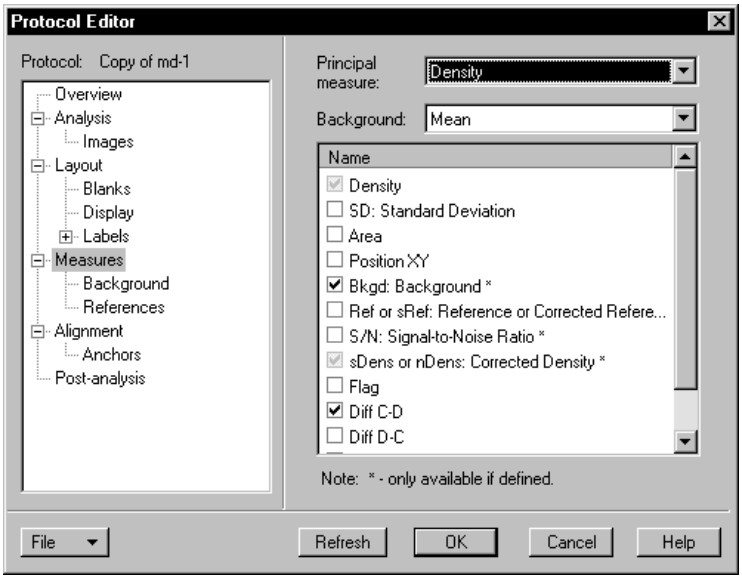


Figure 5-7. The Measures pane.

5.8 Selecting background subtraction methods

From the navigation pane, choose **Background** (under Measures). The Background pane appears (figure 5-8). In the Background pane, you can select any of the available background subtraction methods except the following: Selected spots and Selected spot groups.

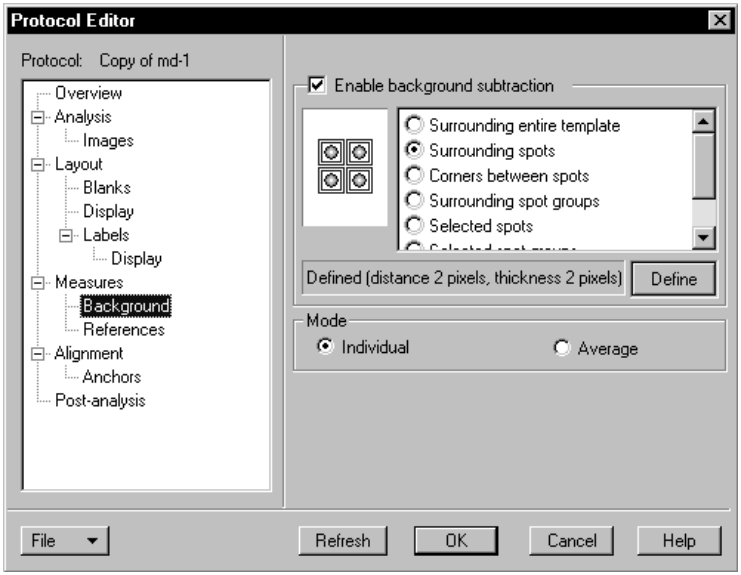


Figure 5-8. The Background pane.

5.9 Defining the references (optional)

A reference is the calculated value to which all array elements are normalized. You do not need to define a reference because Lucidea Microarray ScoreCard will normalize the data for you. However, if you choose to enable reference use, select only one of the following:

- All spots
- User defined areas

5.10 Analyzing the image files

After you finish setting up the parameters as described in this chapter, analyze the image files using your usual process.

5.11 Exporting the analysis data

After ArrayVision finishes the analysis, export the analysis data to tab-delimited text file(s). These files can be imported into the Lucidea Microarray ScoreCard software.

To make sure you have set up the analysis correctly, compare your exported file with the example input file in the …\Microarray folder.

Chapter 6 Preparing GenePix Pro analysis files for import

This chapter describes the analysis settings that must be used in GenePix Pro 3.x to generate output that can be imported into the Lucidea Microarray ScoreCard software. The topics in this chapter are—

- The image and analysis data files (section 6.1)
- Opening the image files in GenePix Pro (section 6.2)
- Selecting the feature layout (section 6.3)
- Saving the results file (section 6.4)

Depending on which 3.x version of GenePix Pro you are using, the exact method for setting up the analysis can vary. This chapter provides general guidelines required for the correct analysis output. For detailed instructions on how to set up the analysis, refer to the *GenePix Pro 3.x Microarray Acquisition and Analysis Software User's Guide.*

Note: An example analysis file generated by GenePix Pro 3.x is in the …\Microarray folder.

6.1 The image and analysis data files

In two-color microarray experiments, you can use the GenePix 4000B Microscanner to scan a slide and produce image files in a number of different ways. As long as you scan the 12 pen areas and both spot sets on the slide, you can produce the image files and the analysis results that are required by the Lucidea Microarray ScoreCard software.

6.2 Opening the image files in GenePix Pro

To perform image analysis in GenePix Pro, you must open all the image files associated with the slide. The number of image files you open depends on how you scanned the slide and how many image files were produced. As long as all of the images are opened when you perform analysis, GenePix Pro will produce the required data for the Lucidea Microarray ScoreCard software.

6.3 Selecting the feature layout

In GenePix Pro, you define the array geometry before you begin the image analysis. To define the array geometry, open the New Blocks window (figure 6-1). When you define the array geometry, be sure to select **Rectangular** in the Feature Layout area.

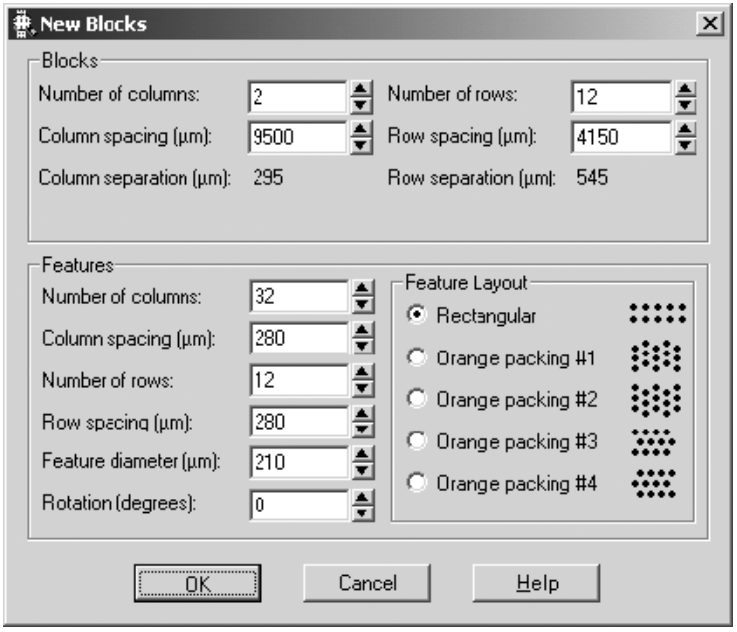


Figure 6-1. The New Blocks window.

6.4 Saving the results file

After you finish the image analysis, be sure to save the results in a single .gpr format file (figure 6-2). The Lucidea Microarray ScoreCard software cannot open other formats from GenePix Pro.

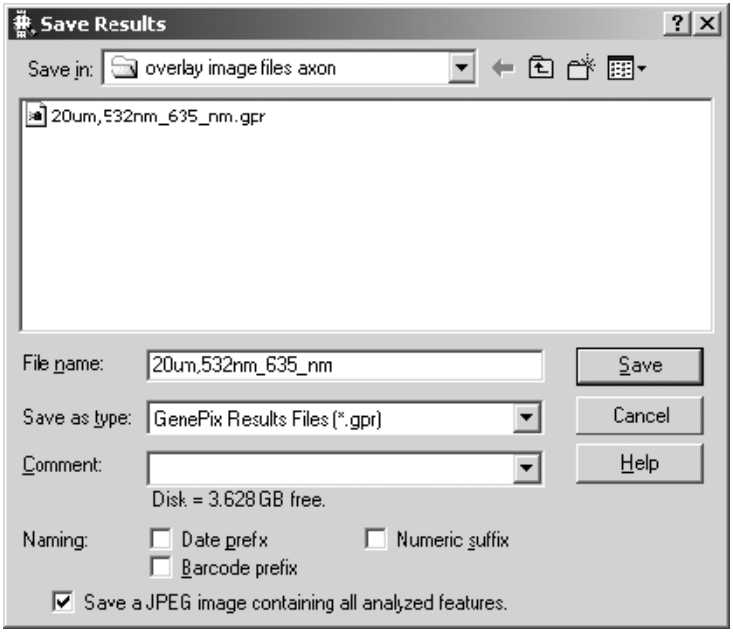


Figure 6-2. The Save Results window.

Chapter 7 Using the Lucidea Microarray ScoreCard software

This chapter describes how to start the software and how to get instructions for the software. The topics in this chapter are—

- Starting the Lucidea Microarray ScoreCard software (section 7.1)
- Getting software help (section 7.2)

7.1 Starting the Lucidea Microarray ScoreCard software

To start the software, from the Windows desktop, click **Start,** point to **Programs,** point to **MD APPS**, point to **Microarray**, and choose **ScoreCard.** The Lucidea Microarray ScoreCard window appears (figure 7-1).

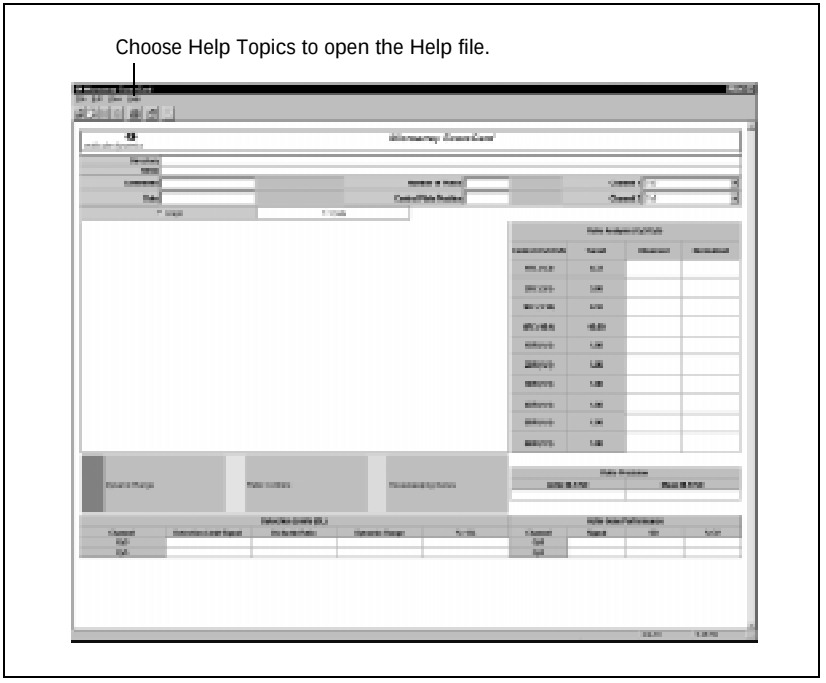


Figure 7-1. The Lucidea Microarray ScoreCard window.

7.2 Getting software help

For instructions on how to use the Lucidea Microarray ScoreCard software, see the provided Help. To access the Help, choose **Help Topics** from the Help menu.

If you have questions about or require assistance with the software, contact Microarray System Technical Support. See Assistance in the Preface for contact information.

Chapter 8 Viewing the analysis results

This chapter explains how to view the analyzed data so that you can interpret the experiment results. The topics in this chapter are—

- Control data scatter plot (section 8.1)
- Ratio Analysis table (section 8.2)
- Ratio Precision table (section 8.3)
- Detection Limits table (section 8.4)
- Actin Gene Performance table (section 8.5)
- Normalized control plate data (section 8.6)
- System validation information (section 8.7)
- Quality report file content (section 8.8)
- Normalized data file content (section 8.9)

For a detailed description of the calculations, see appendix A.

Caution **Make sure you are using the correct settings in the Lucidea Microarray ScoreCard software before you evaluate the results.**

Note: This chapter assumes that you are familiar with statistical and mathematical terms and concepts commonly used in scientific experiments.

8.1 Control data scatter plot

The control data scatter plot (figure 8-1) shows the relationship between the averaged Cy3 signal values (x-axis) and the averaged Cy5 signal values (y-axis) for the following controls: dynamic range, ratio, and housekeeping genes. To see the signal values for a data point, click the data point in the graph. The signal values for that data point appear below the graph.

The graph provides an overall view of the dynamic range of the experiment and the performance of the ratio controls. The plot of the dynamic range controls should appear approximately linear over three logs. Under perfect experimental conditions, the slope would be approximately 1.

The example in figure 8-1 shows the relationship between the averaged Cy3 sDxA values (x-axis) and the averaged Cy5 sDxA values (y-axis) for the dynamic range, ratio, and housekeeping gene controls. Notice that the dynamic range controls appear linear.

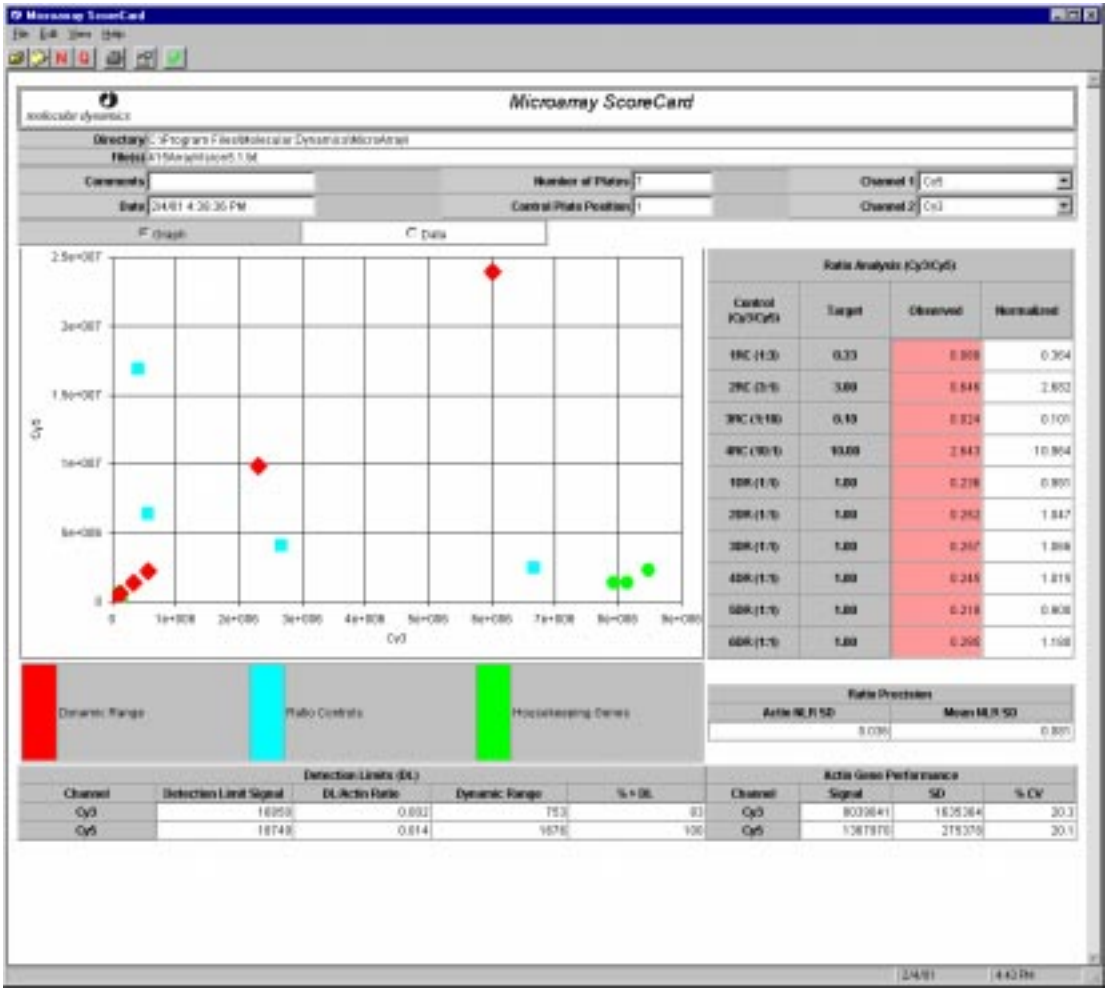


Figure 8-1. The scatter plot of the control plate data in the Lucidea Microarray ScoreCard window.

8.2 Ratio Analysis table

The Ratio Analysis table (figure 8-2) allows you to compare the target ratios against the observed ratios for the control elements. The table shows the following output data:

- **Observed—**The observed average ratio of the Cy3 signal to the Cy5 signal of the particular control element. The value is obtained by taking the antilog of the average uncorrected log ratio (section 8.6).
- **Normalized—**The observed average normalized ratio of the Cy3 signal to the Cy5 signal of the particular control element. The value is obtained by taking the antilog of the average normalized log ratio (section 8.6).

Ratio Analysis (Cy3/Cy5)			
Control (Cy3/Cy5)	Target	Observed	Normalized
1RC (1:3)	0.33	0.088	0.364
2RC (3:1)	3.00	0.646	2.682
3RC (1:10)	0.10	0.024	0.101
4RC (10:1)	10.00	2.643	10.964
1DR (1:1)	1.00	0.236	0.981
2DR (1:1)	1.00	0.252	1.047
3DR (1:1)	1.00	0.257	1.066
4DR (1:1)	1.00	0.245	1.015
5DR (1:1)	1.00	0.218	0.900
6DR (1:1)	1.00	0.295	1.180

Figure 8-2. The Ratio Analysis table.

The software highlights in red the observed or normalized values if the fold difference between the observed or normalized value and the target value is greater than the threshold you set. The default threshold setting is 1.5 fold difference.

The example in figure 8-2 uses the default threshold and shows that the fold difference between every observed value and its corresponding target value is greater than 1.5.

Because of the differences in Cy3 and Cy5 (the rate of incorporation, stability, and fluorescence), the type of slide used, and experimental errors, the observed values do not necessarily match the corresponding target values. On the other hand, the normalization calculations should correct for these differences in the observed values. Therefore, the normalized values should closely match the corresponding target values. The example in figure 8-2 demonstrates this point: the target 1RC value is 0.33 and the normalized value is 0.364.

8.3 Ratio Precision table

The Ratio Precision table (figure 8-3) allows you to estimate the precision of the gene expression ratios in your experiment. The table shows the following output data:

- **Actin NLR SD—**The standard deviation of the normalized log ratio for the 48 replicate actin spots.
- **Mean NLR SD—**The average standard deviation of the normalized log ratio for the following control types: dynamic range (DR), ratio control (RC), housekeeping genes (HG), and positive control (PC). The standard deviation for each control is calculated from 24 replicates.

Ratio Precision	
Actin NLR SD	Mean NLR SD
0.036	0.081

Figure 8-3. The Ratio Precision table.

The two NLR SD values can indicate the reproducibility of the gene expression ratios for the hybridization experiment. Smaller values for the NLR SD represent more precise gene expression ratios.

The NLR SD values can also be used to estimate confidence intervals for the gene expression data. For information about the calculation, see appendix A.

The software highlights in red the Mean NLR SD value if it is greater than the threshold you set. The default threshold value is 0.150. The example in figure 8-3 uses the default threshold and shows that the Mean NLR SD is less than the default threshold.

To determine the threshold value to use and to understand how this value relates to the precision of the differentially expressed ratios, see appendix A.

8.4 Detection Limits table

The Detection Limits table (figure 8-4) shows the following output data:

- **Detection Limit Signal—**The average signal value of the plant-derived negative control for Cy3 and for Cy5 plus 3 times standard deviation (mean + 3SD). The detection limit data allows you to compare the detection limits across multiple slides within an experiment or across multiple experiments.
- **DL/Actin Ratio—**The ratio of the detection limit to the average actin gene signal (well sets 14 and 29) for the particular dye. The detection limit to actin gene ratio measures sensitivity relative to the actin gene. It can also be used to check detection limits from one or more experiments. Note that the smaller the ratio, the greater the sensitivity.

  The software highlights in red the values that are greater than the threshold you set. The default threshold value is 0.025. The example in figure 8-4 uses the default threshold and shows that all the ratios are less than the default threshold.
- **Dynamic Range—**The ratio of the maximum sDxA value to the detection limit signal for the particular dye. The dynamic range data is an indicator of the quality of the data because higher dynamic range values mean lower background noise in the data. However, do not use the dynamic range data as an absolute measure of quality.

  Note that the Lucidea Microarray ScoreCard software does not evaluate the upper end of the dynamic range for deviations from linearity or the effects of saturated pixels.

  The software highlights in red the dynamic range values that are less than the threshold you set. The default threshold is 50. The example in figure 8-4 uses the default threshold and shows that all the dynamic range values are greater than the default threshold.

  Note: If the dynamic range calculation results in the division by a 0 detection limit signal, the software will display a value of 5 000.
- **%>DL—**The percent of all signal values (all samples, including the control elements) that are greater than the detection limits of the particular dye. Low %>DL can indicate high background noise in the data or poor hybridizing sample quality. A low %>DL can also indicate that you incorrectly specified the control plate position in the software.

Caution **Be sure to check the control plate position specification in the software before you evaluate the detection limit data.**

Detection Limits (DL)				
Channel	Detection Limit Signal	DL/Actin Ratio	Dynamic Range	% > DL
Cy3	16850	0.002	753	83
Cy5	18749	0.014	1676	100

Figure 8-4. The Detection Limits table.

8.5 Actin Gene Performance table

Data in the Actin Gene Performance table (figure 8-5) are calculated from the actin gene control element in well sets 14 and 29. This gene has the most replicates on the slide (a total of 48). Therefore, it can be a good indicator of system performance. The Actin Gene Performance table shows the following output data:

- **Signal—**The average actin gene signal for the particular dye.
- **SD—**The corresponding standard deviation value.
- **%CV—**The corresponding percent coefficient of variation, which can indicate spot-to-spot variation and data reproducibility. For example, a high %CV can indicate a damaged pen or high local background.

  The software highlights in red the %CV values that are greater than the threshold you set. The default threshold value is 50. The example in figure 8-5 uses the default threshold and shows that all the %CV values are less than the default threshold.

Actin Gene Performance			
Channel	Signal	SD	% CV
Cy3	8039041	1635364	20.3
Cy5	1367970	275378	20.1

Figure 8-5. The Actin Gene Performance table.

8.6 Normalized control plate data

The normalized control plate data grid (figure 8-6) shows the following columns:

- **Control—**The abbreviated name for the type of control used (table 2-1). To see the associated clone used as the control, rest the pointer on the control abbreviation and the clone information appears below the pointer.
- **Cy3 signal—**The average Cy3 signal values for replicates of the particular control element.

- **Cy3 %CV—**The percent coefficient of variation for the Cy3 signal values.
- **Cy5 signal—**The average Cy5 signal values for replicates of the particular control element.
- **Cy5 %CV—**The percent coefficient of variation for the Cy5 signal values.
- **Cy3/Cy5 Uncorrected Log Ratio—**The average uncorrected log ratio of the Cy3 signal to the Cy5 signal of the particular control element.

  The log ratios for the dynamic range control elements and most housekeeping gene control elements should be close to 0 because no differential expression is expected for these controls. However, because of the differences in the Cy3 and Cy5 labeling reactions, the uncorrected log ratios for these controls are frequently different than 0. In addition, there is frequently a deviation from the constant log ratio at low signal values.
- **Cy3/Cy5 Normalized Log Ratio—**The average normalized log ratio of the Cy3 signal to the Cy5 signal of the particular control element.

  The normalized log ratios are calculated using a new proprietary method that considers signals from every array element (not including the controls) and compensates for differences in dye performance. The normalized log ratio for the dynamic range control elements should be close to 0. You can use these values as references to compare data with high background signals, artifacts, systematic errors with probe labeling, and other experiment errors.
- **Cy3/Cy5 SD for the Uncorrected Log Ratio and Normalized Log Ratio—**The standard deviations of the uncorrected log ratio and normalized log ratio values.
- **Mean—**The overall mean and the mean standard deviations of the uncorrected and normalized log ratios for the dynamic range, ratio, and housekeeping gene control elements. These values appear at the bottom of the grid.

  Comparing the mean standard deviations of the uncorrected and normalized signal ratios allows you to evaluate the normalization process. The mean standard deviations provide an estimate of the precision of the log ratios and can be used to estimate confidence intervals for the gene expression ratios. The mean standard deviations of the normalized log ratio should be less than that of the uncorrected log ratio. If this is not the case, check and make sure that you have the correct dye-channel assignment in the software.

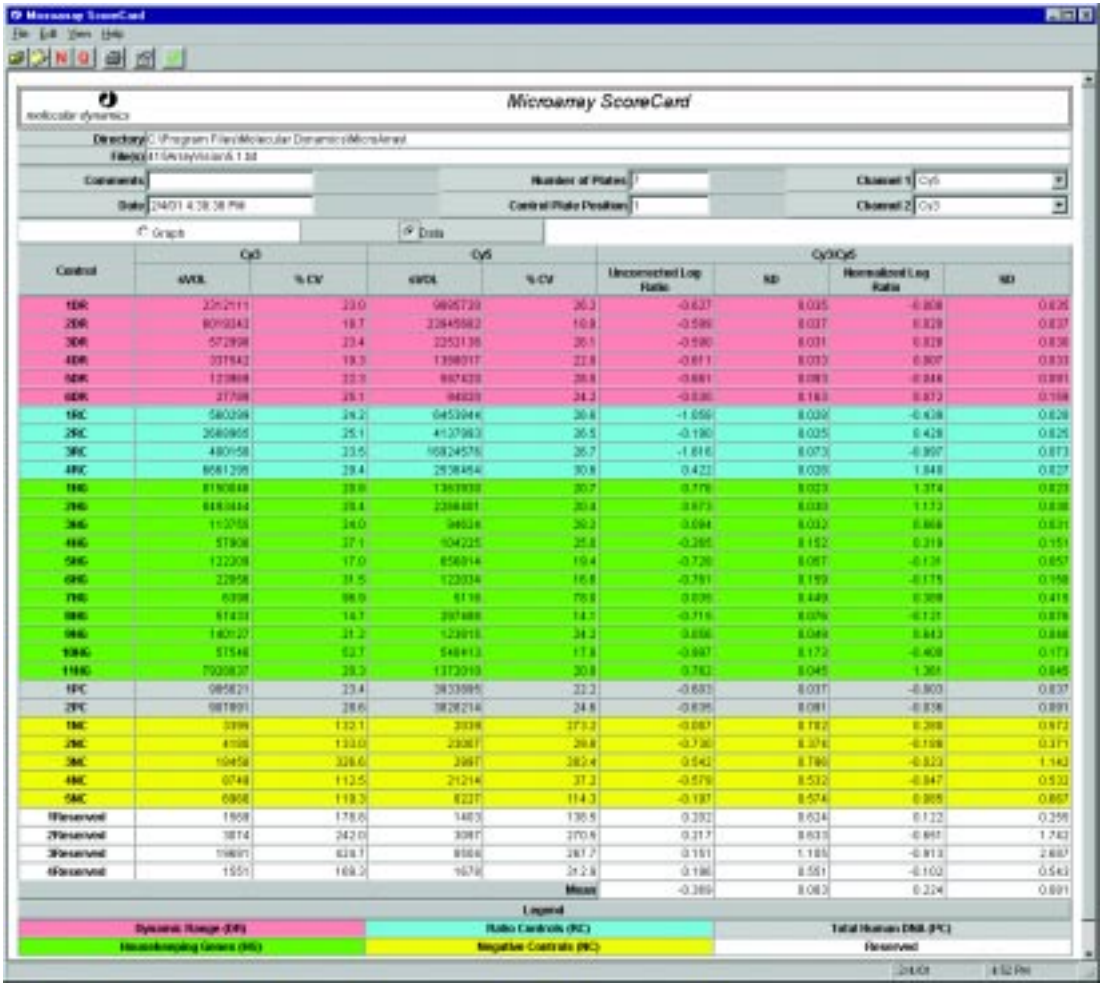
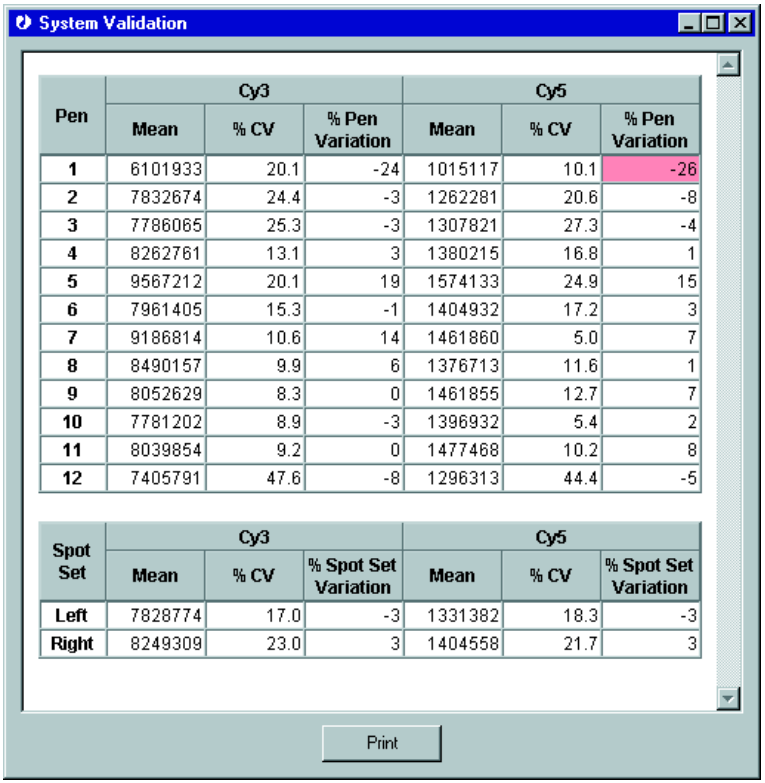


Figure 8-6. The data grid in the Lucidea Microarray ScoreCard window. Note that the grid in this example is color-coded by the type of control element.

8.7 System validation information

You can use the information in the System Validation window (figure 8-7) to check the following:

- Pen-to-pen variability (section 8.7.1)
- Spot-set to spot-set variability (section 8.7.2)



Pen	Cy3			Cy5		
	Mean	% CV	% Pen Variation	Mean	% CV	% Pen Variation
1	6101933	20.1	-24	1015117	10.1	-26
2	7832674	24.4	-3	1262281	20.6	-8
3	7786065	25.3	-3	1307821	27.3	-4
4	8262761	13.1	3	1380215	16.8	1
5	9567212	20.1	19	1574133	24.9	15
6	7961405	15.3	-1	1404932	17.2	3
7	9186814	10.6	14	1461860	5.0	7
8	8490157	9.9	6	1376713	11.6	1
9	8052629	8.3	0	1461855	12.7	7
10	7781202	8.9	-3	1396932	5.4	2
11	8039854	9.2	0	1477468	10.2	8
12	7405791	47.6	-8	1296313	44.4	-5

Spot Set	Cy3			Cy5		
	Mean	% CV	% Spot Set Variation	Mean	% CV	% Spot Set Variation
Left	7828774	17.0	-3	1331382	18.3	-3
Right	8249309	23.0	3	1404558	21.7	3

Figure 8-7. The System Validation window.

8.7.1 Pen-to-pen variability

The first table in the System Validation window (figure 8-7) shows pen-to-pen variability using the following data:

- **Pen—**Pen position in the pen assembly. The pen assembly consists of 12 pens, where pen 1 spots from well A1 through P1 and A13 through P13.
- **Mean—**For each pen, the average of all Cy3 or Cy5 signal values from the actin gene in well sets 14 and 29 (1HG and 11HG).

- **%CV—**Percent coefficient of variation (standard deviation/mean) of the average of all Cy3 or Cy5 signal values for the particular pen. Note that there are four replicates per pen per slide, or a total of 48 elements per slide.
- **% Pen Variation—**Percent variation of a particular pen:

  $$((\text{Mean Pen}_i - \text{Mean}_{\text{Actin}}) / \text{Mean}_{\text{Actin}}) * 100$$

  $\text{Mean}_{\text{Actin}}$ is the average actin gene signal as calculated in the Actin Gene Performance table (section 8.5).

The software highlights in red all % Pen Variation results with absolute values exceeding the threshold you set. The default threshold setting is 25%.

For example, if you use the default threshold for pen variability, all the results with absolute values exceeding the 25% threshold are highlighted in red (figure 8-7). The highlights indicate possible problems and should be considered when you evaluate the data.

Negative % Pen Variation values indicate pen areas with mean signal below the overall mean housekeeping gene values for the entire slide.

A large percent variation that occurs consistently in the same pen across multiple slides from the same spotting session can indicate a problem with that pen.

Important **Relatively large percent pen variation values may indicate problems with pen performance. However, process steps after spotting can also contribute to the variation. Therefore, you should take other factors into consideration when reviewing the large percent variation.**

8.7.2 Spot-set to spot-set variability

The second table in the System Validation window (figure 8-7) shows spot-set to spot-set variability using the following data:

- **Mean—**For each spot set, the average of all Cy3 or Cy5 signal values from the actin gene in well sets 14 and 29.
- **%CV—**The corresponding percent coefficient of variation for each average value.
- **% Spot Set Variation—**Percent variation of the signal from the spot set:

  $$((\text{Mean Spot Set}_i - \text{Mean}_{\text{Actin}}) / \text{Mean}_{\text{Actin}}) * 100$$

  $\text{Mean}_{\text{Actin}}$ is the average actin gene signal as calculated in the Actin Gene Performance table (section 8.5).

The software highlights in red all % Spot Set Variability results with absolute values exceeding the threshold you set. The default threshold setting is 10%.

For example, if you use the default threshold for spot set variability, all the results with absolute values exceeding the 10% threshold are highlighted in red. Figure 8-7 does not show any value exceeding the default threshold for spot set variability.

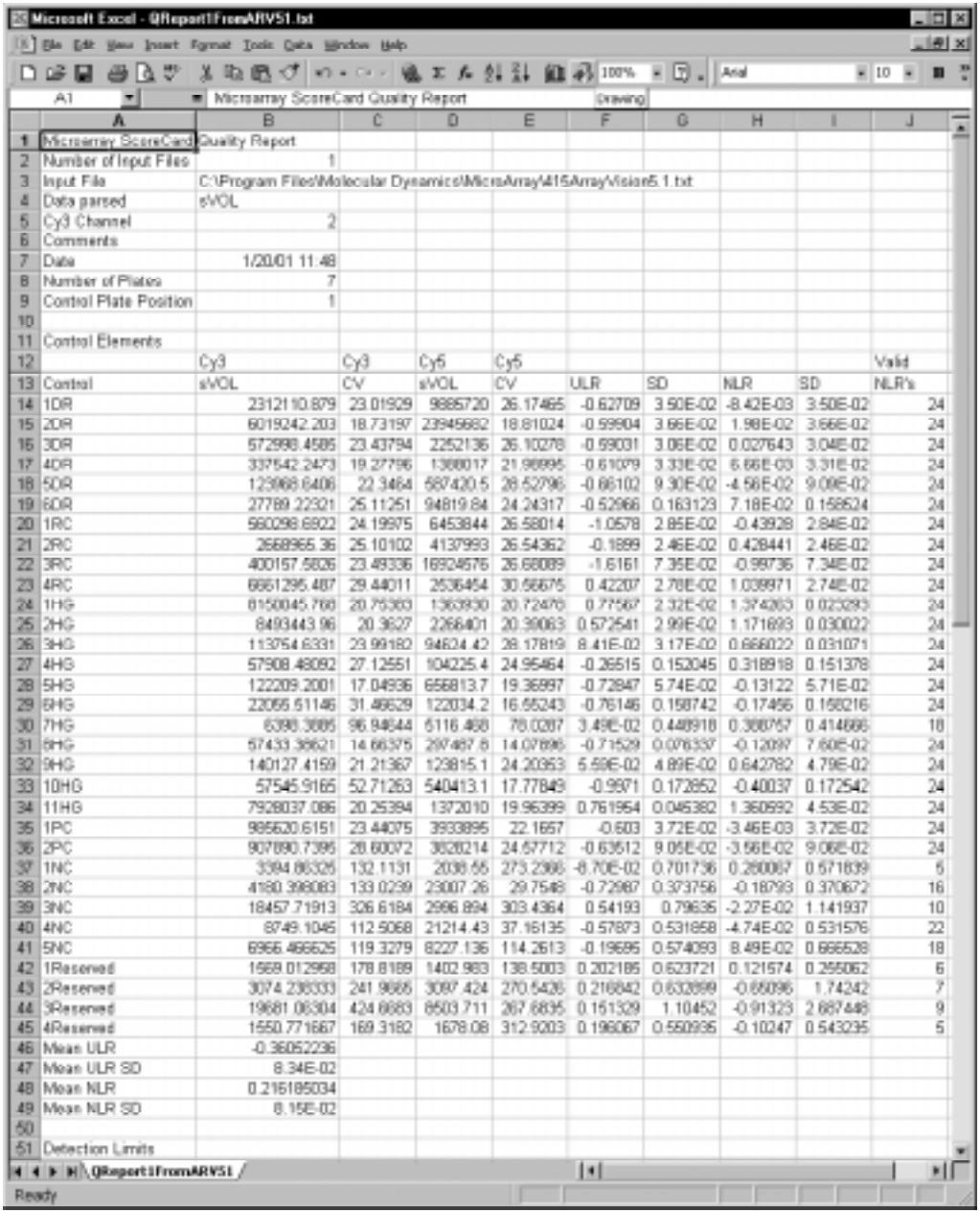
If you see highlighted values in the table, the highlights can indicate possible problems with slide surface chemistry and should be considered when you evaluate the data.

8.8 Quality report file content

The quality report that you export contains all the data you see in the Lucidea Microarray ScoreCard window, including the system validation data. Figure 8-8 shows an example of a quality report file. In this example, the tab-delimited text file is viewed in Excel for clarity.

The quality report shows the following information:

- **Header information—**The type of report file, the number of input files, the input file name(s) and location(s), the signal data processed, the channel assignment of the Cy3 data, comments about the analysis, the date of the analysis, the total number of microplates in the experiment, and the position of the control plate in the Array Spotter.
- **Normalized control plate data—**All the information from the data grid. In addition, for each control element, the report shows the number of valid normalized log ratios used in the calculations.
- **Detection limits—**The data from the Detection Limits table.
- **Housekeeping gene performance—**The data from the Actin Gene Performance table.
- **Ratio analysis—**The data from the Ratio Analysis table.
- **System validation—**The data from the System Validation window.
- **Normalization Constants—**The values used in the normalization calculations. For information about the normalization calculation, see appendix A.
- **Mean 1 through 5DR NLS—**The mean normalized log ratios for the dynamic range control elements 1 through 5. This value is used to adjust the normalized log ratios for the dynamic range (DR) and ratio control (RC) elements.



Microsoft Excel - QReport1FromARV51.txt

File Edit View Insert Format Tools Data Window Help

A1 — Microarray ScoreCard Quality Report

	A	B	C	D	E	F	G	H	I	J
1	Microarray ScoreCard	Quality Report								
2	Number of Input Files	1								
3	Input File	C:\Program Files\Molecular Dynamics\MicroArray\415ArrayVision5.1.txt								
4	Data parsed	sVOL								
5	Cy3 Channel	2								
6	Comments									
7	Date	1/20/01 11:48								
8	Number of Plates	7								
9	Control Plate Position	1								
10										
11	Control Elements									
12		Cy3	Cy3	Cy5	Cy5					Valid
13	Control	sVOL	CV	sVOL	CV	ULR	SD	NLR	SD	NLR's
14	1DR	23121.10.879	23.01929	9885720	26.17465	-0.62709	3.50E-02	-8.42E-03	3.50E-02	24
15	2DR	6019242.203	18.73197	23945682	18.81024	-0.99904	3.66E-02	1.98E-02	3.66E-02	24
16	3DR	572998.4585	23.43794	2252136	26.10278	-0.69031	3.06E-02	0.027643	3.04E-02	24
17	4DR	337542.2473	19.27796	1388017	21.98995	-0.61079	3.33E-02	6.66E-03	3.31E-02	24
18	5DR	123968.6406	22.3464	567420.5	28.52796	-0.66102	9.30E-02	-4.56E-02	9.09E-02	24
19	6DR	27789.22321	25.11251	94819.84	24.24317	-0.52966	0.163123	7.18E-02	0.158524	24
20	1RC	980298.6822	24.19975	6453844	26.58014	-1.0578	2.85E-02	-0.43928	2.84E-02	24
21	2RC	2668965.36	25.10102	4137993	26.54362	-0.1899	2.46E-02	0.428441	2.46E-02	24
22	3RC	400157.5826	23.49336	16924576	26.68089	-1.6161	7.35E-02	-0.99736	7.34E-02	24
23	4RC	6661295.487	29.44011	2536454	30.56675	0.42207	2.78E-02	1.039971	2.74E-02	24
24	1HG	8150045.768	20.75393	1363930	20.72476	0.77567	2.32E-02	1.374263	0.023293	24
25	2HG	8493443.96	20.3627	2266401	20.39063	0.572541	2.99E-02	1.171693	0.030022	24
26	3HG	113754.6331	29.99182	94624.42	28.17819	8.41E-02	3.17E-02	0.666022	0.031071	24
27	4HG	57908.48092	27.12551	104225.4	24.95464	-0.26515	0.152045	0.318918	0.151378	24
28	5HG	122209.2001	17.04936	656813.7	19.36997	-0.72847	5.74E-02	-0.13122	5.71E-02	24
29	6HG	22055.51146	31.46629	122034.2	16.55243	-0.76146	0.158742	-0.17456	0.158216	24
30	7HG	6398.3885	96.94644	5116.468	78.0287	3.49E-02	0.448918	0.388757	0.414666	18
31	8HG	57433.39621	14.66375	297497.6	14.07896	-0.71529	0.076337	-0.12097	7.60E-02	24
32	9HG	140127.4159	21.21367	123815.1	24.20353	5.59E-02	4.89E-02	0.642782	4.79E-02	24
33	10HG	57545.9165	52.71263	540413.1	17.77849	-0.9971	0.172852	-0.40037	0.172542	24
34	11HG	7928037.086	20.25394	1372010	19.96399	0.761954	0.046382	1.360992	4.53E-02	24
35	1PC	985620.6151	23.44075	3933895	22.1657	-0.603	3.72E-02	-3.46E-03	3.72E-02	24
36	2PC	907890.7395	28.60072	3628214	24.57712	-0.63512	9.05E-02	-3.56E-02	9.06E-02	24
37	1NC	3394.86325	132.1131	2038.55	273.2366	-8.70E-02	0.701736	0.280067	0.571839	5
38	2NC	4180.398083	133.0239	23007.26	29.7548	-0.72987	0.373756	-0.18793	0.370672	16
39	3NC	18457.71913	326.6184	2996.894	303.4364	0.54193	0.79635	-2.27E-02	1.141937	10
40	4NC	8749.1045	112.5068	21214.43	37.16135	-0.57873	0.531858	-4.74E-02	0.531576	22
41	5NC	6966.466625	119.3279	8227.136	114.2613	-0.19695	0.574093	8.49E-02	0.666528	18
42	1Reserved	1568.012958	178.8188	1402.983	138.5003	0.202185	0.623721	0.121574	0.255062	6
43	2Reserved	3074.230333	241.9666	3097.424	270.5426	0.216842	0.632899	-0.66096	1.74242	7
44	3Reserved	19681.06304	424.6683	8503.711	267.6835	0.151329	1.10452	-0.91323	2.687446	9
45	4Reserved	1550.771667	169.3182	1678.08	312.9203	0.196067	0.560936	-0.10247	0.543235	5
46	Mean ULR	-0.36052236								
47	Mean ULR SD	8.34E-02								
48	Mean NLR	0.216185034								
49	Mean NLR SD	8.15E-02								
50										
51	Detection Limits									

QReport1FromARV51

Ready

Figure 8-8. The quality report file in Excel. Because of its size, only the top half of the report is shown.

8.9 Normalized data file content

The normalized data file that you export contains the original image analysis data, the uncorrected and normalized log ratio for every sample, the normalized and mean values that were calculated for every sample, and the clone information. Figure 8-9 shows an example of a normalized file. In this example, the tab-delimited text file is viewed in Excel for clarity.

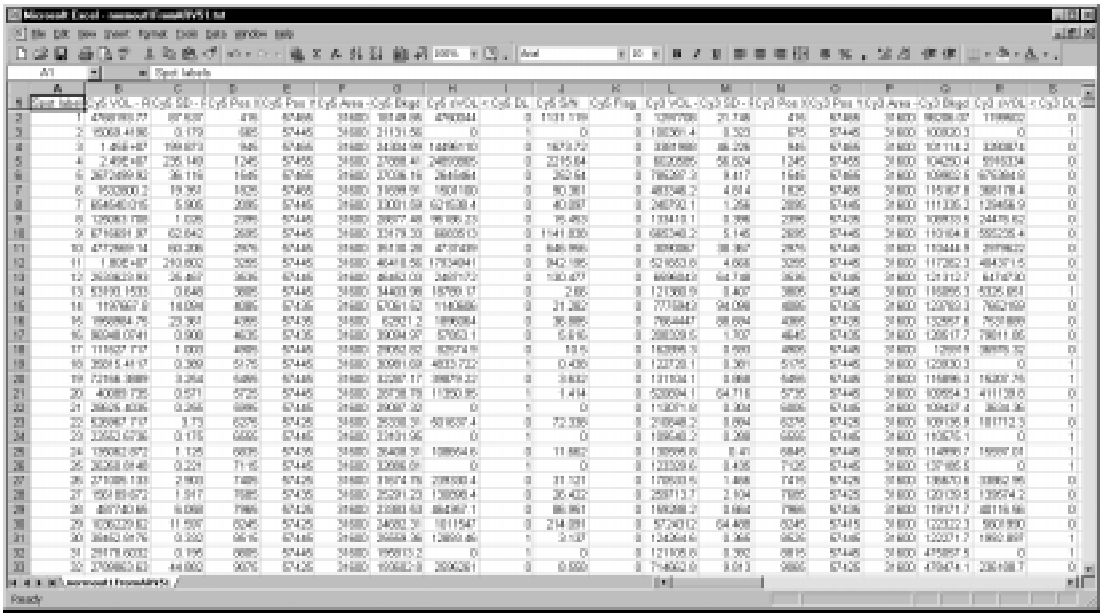


Figure 8-9. The normalized data file in Excel. Because of its size, only a small portion of the report is shown.

Included with the image analysis output measures are—

- **Detection limit flags (<Cy3 DL, <Cy5 DL)**—The flags indicate whether the signals fall below the detection limit. A 1 indicates the signal is below the detection limit. A 0 indicates the signal is above the detection limit.
- **Spot ULR**—The uncorrected log ratio for the particular spot.
- **Spot NLR**—The normalized log ratio for the particular spot.
- **Norm Cy3**—The normalized Cy3 signal for the particular spot, which is Cy5 signal * $10^{(\text{Spot NLR})}$.
- **ULR 1**—The uncorrected log ratio for the particular spot in spot set 1.
- **ULR 2**—The uncorrected log ratio for the particular spot replicate in spot set 2.

- **NLR 1**—The normalized log ratio for the particular spot in spot set 1.
- **NLR 2**—The normalized log ratio for the particular spot replicate in spot set 2.

For an explanation of how these values are calculated, see appendix A.

The normalized data file structure matches the requirements for Spotfire Pro™ and ArrayStat™ Statistical Informatics™ software. You can open this file directly in Spotfire Pro to display the Cy5 versus normalized Cy3 scatter plots and visualize the spot set 1 and spot set 2 comparisons. You can open this file directly in ArrayStat Statistical Informatics to identify outliers and determine which genes are significantly differentially expressed.

Chapter 9 Comparing data from multiple slides

This chapter explains how to assess the Microarray ScoreCard results and compare the data from multiple slides. The topics in this chapter are—

- Assessing data quality and statistical significance (section 9.1)
- Interpreting the Gene Expression data from single and multiple hybridizations (section 9.2)

9.1 Assessing data quality and statistical significance

Before you evaluate Microarray experimental results, you should use a common set of criteria across a batch of slides to accept or reject a slide in the experiment. The criteria are based on the quality measures that the Microarray ScoreCard software calculates. The software uses the thresholds you set to determine whether or not the quality measurements are within the acceptable range.

This section provides some guidelines you can use to accept or reject the slides when you evaluate the Microarray ScoreCard results. The quality attributes that should be taken into consideration and the corresponding quality measures calculated by the software are discussed. For information about how to use the software, see the Help available in the software. For a description of the data that the software generates, see chapter 8.

9.1.1 Signal level and precision

To determine the reproducibility of signals within one channel, you can use the signal level and precision information in the Actin Gene Performance table (section 8.5). Smaller %CV values for actin indicate more reproducible actin signals across the slide and more uniform background and hybridization. If actin appears to be differentially expressed, use another housekeeping gene, such as enoyl coenzyme A hydrolase (10HG), to determine the reproducibility of the signal levels across slides.

9.1.2 Sensitivity and Specificity

To assess the system sensitivity and to check for non-specific binding, use the information in the Detection Limits table (section 8.4). The detection limit signal information can indicate the sensitivity of the Microarray system and the overall quality of each hybridization:

- The lower the signal detected, the greater the sensitivity.
- The higher the signal detected, the lower the specificity of hybridization.

The dynamic range controls (DRs) correlate with the relative abundance of the mRNA spike and can indicate the expression level that can be detected. For example, if dynamic range 6 (6DR) is not above the threshold signal level for the detection limit, you can state with confidence that genes with expression levels below the 6DR relative abundance are not detectable.

9.1.3 Gene Expression data Accuracy and Precision

To determine the accuracy and precision of the ratios, examine the Ratio Analysis table (section 8.2) and the Ratio Precision table (section 8.3). Consider the following when you inspect the information in these tables:

- In the Ratio Precision table, the average standard deviation of the normalized log ratio (Mean NLR SD) should be acceptable. Typically, the values should be less than the default threshold value of 0.150.
- In the Ratio Analysis table, the values of the normalized ratio and dynamic range controls should be acceptable based on your threshold difference from the target values.
- The normalized log ratios (NLRs) may be less accurate when one or more signals fall below the detection limit.

To assess the performance of the normalization process, see appendix A.

9.2 Interpreting the Gene Expression data from single and multiple hybridizations

This section discusses how to use the output from Lucidea Microarray ScoreCard to obtain gene expression ratios (n-fold up-regulation or down-regulation) and how to combine data across multiple hybridizations.

9.2.1 Calculating gene expression ratios

For each spot, the Spot NLR value in the normalized data file (section 8.9) represents the log (base 10) of the Normalized Cy3 signal to Cy5 signal ratio. To calculate a gene expression ratio from a single spot, take the antilog of Spot NLR by raising 10 to the power of (Spot NLR).

$$\text{Gene expression ratio from a single spot} = 10^{\text{Spot NLR}}$$

This represents the ratio of the amount of the message in the Cy3-labeled mRNA to the amount of the message in the Cy5-labeled mRNA complementary to the individual spot. For example, if the Spot NLR value is 0.50, the gene expression ratio is 10 raised to the power 0.50, which is 3.16. Thus, there is approximately a 3.16-fold more complementary message in the Cy3-labeled mRNA than in the Cy5-labeled mRNA.

If you want to report the ratio relative to the Cy3-labeled mRNA because the Cy3-labeled mRNA is your reference tissue, reverse the sign of the NLR value. The ratio becomes 10 to the power of negative Spot NLR:

$$\text{Reverse ratio} = 10^{-\text{Spot NLR}}$$

9.2.2 Combining gene expression ratios from multiple spots within a slide

To combine data from multiple spots, calculate the average normalized log ratio (NLR) values for each spot from a particular DNA sequence and raise 10 to the power of the average NLR:

$$\text{Combined gene expression ratio} = 10^{\text{Average NLR for each spot}}$$

Lucidea Microarray ScoreCard calculates and reports the average NLR for the duplicate spots in the normalized data file (section 8.9).

Important **You should calculate the average value from the NLR before taking the antilog to transform to the combined gene expression ratio. This also applies to calculating confidence intervals (see section A.3).**

9.2.3 Combining data from replicate slides

To combine data from multiple replicate slides, average the NLR values for all the replicate spots of each DNA sequence before transforming the values from NLRs to gene expression ratios. If you performed replicate hybridizations using the dye-reversal method, you have to correct for the dye reversal in your calculations. In dye-reversal experiments, on the first slide, the experimental RNA is labeled with Cy3 and the reference RNA is labeled with Cy5, and on the second slide the experimental RNA is labeled with Cy5 and the reference

RNA is labeled with Cy3. When you perform the calculations, you have to correct for the dye reversal by reversing the sign of the NLR values for the second slide before calculating the average with the NLR values from the first slide. Then you can transform the NLR values to the gene expression ratios as described in section 9.2.2.

9.2.4 Comparing data using the common reference approach

You can compare data from the common reference RNA preparation for several hybridizations. For example, you can perform one experiment using Cy3-labeled brain mRNA and Cy5-labeled skeletal muscle mRNA. In a second experiment, the hybridization involves Cy3-labeled kidney mRNA and Cy5-labeled skeletal muscle mRNA. You can estimate the relative gene expression between brain and kidney from these data as follows:

- Suppose the ratio of the average NLR for a given DNA sequence from the brain to the skeletal muscle hybridization is 1.23.
- Suppose the ratio of the average NLR for a given DNA sequence from the kidney to the skeletal muscle hybridization is -0.27.
- The NLR for brain over kidney can be calculated as the NLR for brain minus the NLR for kidney, or 1.23-(-0.27)= 1.50. Raising 10 to the power 1.50 gives a brain to kidney gene expression ratio of 31.6.
- In general, the gene expression ratio of any condition to another when a common reference is used can be calculated from the following equation:

The normalized log ratio between two conditions = $NLR_{condition1}$-$NLR_{condition2}$

The gene expression ratio = $10^{((NLR_{condition1}) - (NLR_{condition2}))}$

Important **You should always interpret the gene expression ratios with caution if one or more of the signals used to calculate them fall below the detection limit.**

Appendix A Data normalization and statistical significance

This appendix describes the Lucidea Microarray ScoreCard normalization method. The topics in this appendix are—

- About the normalization method (section A.1)
- Normalizing the data (section A.2)
- Estimating the confidence intervals for the gene expression data (section A.3)
- Calculating the default threshold for the mean NLR SD (section A.4)

Note: This appendix assumes that you are familiar with statistical and mathematical terms and concepts commonly used in scientific experiments.

A.1 About the normalization method

Normalizing microarray data is an essential first step in the correct interpretation of the gene expression experiment results. For two-color experiments, the Lucidea Microarray ScoreCard uses the exponential normalization method to correct for artifacts in studies using Cy3 and Cy5 dyes. Experiments show that—

- The required normalization factor is not constant, but is a function of the Cy5 signal.
- The magnitude of the artifact varies among all the hybridized slides. Therefore, the normalization equation parameters must be determined empirically for each slide.

The exponential normalization method is advantageous because it is based on all the data, not just a single control gene or a series of external spikes. The method improves normalization, particularly for weakly expressed genes, and results in better centering and tighter distribution for the non-differentially expressed genes.

A.2 Normalizing the data

This section describes the following:

- Working with the data in logarithmic scale (section A.2.1)
- Calculating the uncorrected log ratio (section A.2.2)
- Calculating the normalized log ratios (section A.2.3)
- Adjusting the normalized log ratios (section A.2.4)
- Verifying the normalized results (section A.2.5)

A.2.1 Working with the data in logarithmic scale

The software works with the data in logarithmic scale (log base 10) because the distribution of the data becomes normal (symmetrically centered about x=0, when the genes are not differentially expressed) in logarithmic space. This transformation facilitates the statistical analysis of the data.

Table A-1 demonstrates the differences between the untransformed and log-transformed ratios of the Cy3 to Cy5 signals. The untransformed ratios show that the corresponding pairs of up- and down-regulated genes ratios do not center about 1, the ratio of the genes when they are not differentially expressed. After the log transformation, all the corresponding pairs of up- and down-regulated ratios are distributed symmetrically about 0, the log of the ratio of the non-differentially expressed genes.

Table A-1. Transforming ratios into logarithmic scale

Gene	Ratio	Log_{10} (Ratio)
A: not differentially expressed	1	0.00
B: 2-fold up-regulated	2	0.30
C: 2-fold down-regulated	0.50	-0.30
D: 4-fold up-regulated	4	0.60
E: 4-fold down-regulated	0.25	-0.60
F: 10-fold up-regulated	10	1.00
G: 10-fold down-regulated	0.10	-1.00

A.2.2 Calculating the uncorrected log ratio

For each spot on the slide, the software calculates the uncorrected log ratio of Cy3 signal to Cy5 signal as follows:

1. For each spot, the software calculates the log of the ratio of the Cy3 signal to the Cy5 signal. The software does not calculate the log ratio if the Cy3 or the Cy5 signal is less than or equal to zero.

   Log ((Cy3 signal) / (Cy5 signal))

2. The software calculates the log of the Cy5 signal.

   Log (Cy5 signal)

3. After excluding all the spots from control plate elements, the software finds the best fit constants (a, b, and c) for the following exponential normalization equation, which is defined as the uncorrected log ratio (ULR):

   ULR = a + b*exp(-(log(Cy5 signal)/c))

The software reports the best-fit constants (a, b, and c) and the correlation coefficient (R-squared) in the quality report. For more information about the quality report, see section 8.8.

A.2.3 Calculating the normalized log ratios

For each spot, the software calculates the normalized log ratio (NLR) as follows:

NLR = log((Cy3 signal)/(Cy5 signal)) - (a + b*exp(-log(Cy5 signal)/c))

A.2.4 Adjusting the normalized log ratios

Experiments have shown that when the ratio of the mRNA concentrations at the start of the labeling reaction is not 1:1, a constant offset results in the normalized log ratios for the dynamic range controls (DR1 through DR6) and the ratio controls (RC1 through RC4). To adjust for this offset, the mean normalized log ratio for the dynamic range controls 1 through 5 (DR1 through DR5) is subtracted from each dynamic range and ratio control normalized log ratio:

Adjusted NLR for a DR or RC = NLR - (Mean NLR for DR1 to DR5)

Note: The normalized log ratio of dynamic range control 6 (DR6) is not used in this calculation because it is frequently below the detection limit.

The software reports the average unadjusted NLR for DR1–5 in the quality report. For more information about the quality report, see section 8.8.

A.2.5 Verifying the normalized results

You can verify the normalized results as follows:

- After you specify the dye-channel assignment in the software, display the normalized control plate data in the data grid (section 8.6). Compare the standard deviations of the average normalized log ratio (NLR SD) to the standard deviations of the average uncorrected log ratio (ULR SD). The exponential normalization method improves the precision performance of the replicate groups if the average NLR SD is less than the average ULR SD.
- After you specify the dye-channel assignment in the software, verify that in the Ratio Analysis table, the normalized values closely match the corresponding target values.
- After you specify the dye-channel assignment in the software, examine the standard deviations of the average normalized log ratio (NLR SD). Reverse the dye-channel assignment and observe that the average NLR SD will typically increase, indicating a better correlation between the normalization equation with the Cy5 signals than with the Cy3 signals.
- You can use the information in the normalized data file (section 8.9) to plot and view the relationship of the following:
  - The normalized log ratio for each spot (Spot NLR) and the log of the Cy5 signals (log(Cy5))
  - The normalized log ratio of each spot (Spot NLR) and the log of the Cy3 signals (log(Cy3)).

  The plots should show that the Spot NLR values center around x=0, with no significant relationship between the signal level and NLR.

  Alternatively, you can plot the normalized Cy3 signal values (Norm Cy3) against the normalized Cy5 signal values (Norm Cy5). The plot should form a line with a slope of 1.0.

A.3 Estimating the confidence intervals for the gene expression data

The software calculates the average standard deviation of the normalized log ratio for each control element, from 24 replicates. Experiments have shown that for most hybridizations, the standard deviation for each grouping is approximately the same for those groups whose signal is above the detection limit. Therefore, the software calculates and reports the Mean NLR SD as an approximation of the variation in gene expression ratios across the hybridizations.

In statistical analysis, you can improve the confidence in the estimated standard deviation by using many replicate samples. In addition, you can pool variance estimates from many replicate groups, assuming that the variance is the same among all the replicate groups. With these principles, you can calculate the confidence intervals based on the Z or normal distribution:

$$CI = \text{Mean NLR} \pm \frac{Z(\alpha) \times \text{Mean NLR SD}}{\sqrt{n}}$$

Where—

CI is the confidence interval for the mean NLR for a given gene, based on n replicates. In the Generation III Microarray System, most genes are spotted in duplicate (n=2). CI specifies a range of values within which the mean NLR SD may lie. You can calculate the intervals for different confidence levels. For example, a CI calculated at 99.5% is the range of values that will contain the true mean 99.5% of the time.

α is the probability of values falling outside the confidence interval because of random variation. α complements the confidence level for the mean NLR SD. For example, a 99.5% confidence interval has an α value of 0.005.

Z is a value that is a function of α. Based on a given α value, you can look up the corresponding Z value in a Z-test table. The Z value for an α of 0.005 is 2.807.

Note: The confidence interval equation assumes that the control samples are spotted in duplicate. If one of the duplicates is not available, drop the $\sqrt{2}$ from the equation.

A.4 Calculating the default threshold for the mean NLR SD

A gene is significantly differentially expressed if the confidence interval for the Mean NLR does not include 0. Using the α value of 0.005 (section A.3), you can approximate the significance in the difference in NLR by calculating the half-width for the confidence interval:

$$CI_{Half} = Z * (\text{Average NLR SD})/\sqrt{2}$$

Raise the CI_{Half} result to the power of 10 to convert the result from logarithmic space to n-fold space. Table A-2 lists the results of these calculations for an α value of 0.005. The calculations assume the controls are spotted in duplicate on the slide.

Table A-2. Mean NLR SD values and the corresponding detectable changes

Mean NLR SD	Detectable Change (n-fold)
0.050	1.26
0.100	1.60
0.150	1.98
0.200	2.49
0.250	3.13

Table A-2 shows that a two-fold or greater change is considered significant when α is 0.005 and if the mean NLR SD is less than or equal to 0.150. The software uses this value as the default mean NLR SD threshold.

By selecting a different α value, a different level of gene expression difference to detect, and a different number of replicates, you can calculate a different threshold to set in the software.

Note: The confidence intervals calculated in this section are based on the Z distribution and are good approximations. You can use the ArrayStat Statistical Informatics software for the exact statistical analyses, involving the normality test of residuals and the identification and removal of outliers.

Appendix BUsing the Amersham Pharmacia Biotech reflective slides with Lucidea Microarray ScoreCard

This appendix explains the effects of reflective slides on data normalization and how to adjust your slide scanning techniques to produce usable data for analysis. The topics in this appendix are—

- Effects of the reflective slide on data normalization (section B.1)
- Detecting the saturated pixels (section B.2)
- Eliminating the saturated pixels (section B.3)

B.1 Effects of the reflective slide on data normalization

The reflective slide technology from Amersham Pharmacia Biotech increases the signal strength and sensitivity in microarray experiments. Although these slides allow you to detect very weak signals, spots with extremely high signals can become saturated in the image.

For the Generation III Array Scanner, saturated pixels are pixels that reach 100 000 rfu, the maximum signal measurement. Because the true signal levels are underestimated, saturated pixels result in inaccurate spot quantitation. This phenomenon impacts the validity of the data normalization results regardless of the method used. The greater the number of saturated spots, the larger the impact on data normalization.

Caution **Quantitation results will not be accurate for any spot that contains saturated pixels, even if the saturated pixels occur in one channel only.**

B.2 Detecting the saturated pixels

You can use ImageQuant to detect saturated pixels. When you view the image, use the Grey/Color Adjust feature to change the low-end limit of the gray scale to 99998, and click OK. The spots containing saturated pixels will appear black in the image.

B.3 Eliminating the saturated pixels

Amersham Pharmacia Biotech recommends that you choose one of the following methods to eliminate saturated pixels in your images:

- **Reduce the PMT setting**—Increased signals can saturate the scanner's PMT and produce saturated pixels. Therefore, reducing the PMT voltage setting can reduce or eliminate the pixel-saturation problem.

  Decrease the PMT voltage setting from the default value of 700 V to approximately 500 V. Decreasing the PMT setting will not alter the detection limits (the detection of low expressors) because the background signals will also be reduced.

- **Reduce the amount of dye used per slide**—If pixel saturation occurs for both Microarray ScoreCard housekeeping genes and non-ScoreCard targets, reduce the amount of the dye (labeled probe) used per slide. Amersham Pharmacia Biotech recommends using 10–15 pmol of each dye per slide.

- **Reduce the amount of spike mix**—If pixel saturation occurs only with the Microarray ScoreCard dynamic range or ratio control targets, use 60% of the recommended amount of spike mix to add to the labeling reaction. Reducing the spike mix will not alter the detection limit or perturb the ratios.

When performing microarray experiments, you should always review the results of each experiment to eliminate inaccurate data. When you use Microarray ScoreCard, make sure you understand the outcome of the ScoreCard analysis.

Glossary

array elements—the cloned DNA fragments or PCR products that are spotted on the specially coated microscope slides.

DxA—density value of an array element multiplied by its area. Also referred to as volume, this measurement is the total amount of signal in an array element. For more information about DxA, see the *ArrayVision Operations Manual* or Help file.

Dens—density value for each spot. The value represents the average of all the pixels contained in the spot. For more information about Dens, see the *ArrayVision Operations Manual* or Help file.

dynamic range control—control elements used to evaluate the dynamic range and sensitivity of the system.

channel—in ArrayVision, a storage area for a single image. Each image is associated with a dye used in the microarray experiment.

clone library—a set of cloned DNA fragments.

coefficient of variation—standard deviation divided by the mean, often expressed in terms of percent:

$$\%CV = (\text{Standard Deviation}/\text{Mean}) * 100$$

control elements—the control samples that are spotted on the specially coated microscope slides.

fluorescence—the emission of light by a dye molecule as a result of the absorption of laser light.

gene expression—the manifestation of a characteristic that is specified by a gene, such as the production of a protein by a gene.

housekeeping genes—ubiquitously expressed genes found in all cell types. In a microarray experiment, the housekeeping genes are those with minimal expression variation and are used for normalization of the samples.

hybridizing samples—the single-stranded DNA or RNA fragments that bind to the array elements in the hybridization experiments.

microarray—an orderly arrangement of DNA or RNA fragments with a spot size approximately 100–400 µm in diameter. Microarrays, usually containing thousands of spots, are used in high-throughput gene mapping, mutation detection studies, and gene expression analysis.

negative control—the array element used for evaluating the degree of background signal levels of nonspecific hybridization.

normalization—the process of calculating the ratio or the difference between a set of standard array element signal values and the sample signal values to bring the data to the same scale for comparative expression studies.

positive control—the array element used for the normalization of the samples.

quantitation—the process in which signal intensities of the array spots and other values are calculated.

ratio control—control elements that are used as reference for measuring differential expressions.

relative fluorescence units (rfu)—the arbitrary units in which fluorescence intensity is reported by the Array Scanner.

SD—standard deviation of the pixel density values. For more information about SD, see the *ArrayVision Operations Manual* or Help file.

sDens—subtracted density value. Density value of the spot minus the background density value. For more information about sDens, see the *ArrayVision Operations Manual* or Help file.

sDxA—background subtracted DxA, or an array element's total signal (volume) minus the background volume:

sDxA = (Densitydata - Densitybkgd) * (Areadata)

For more information about sDxA, see the *ArrayVision Operations Manual* or Help file.

spot buffer—the distance between the edges of consecutive spots on the slide.

spot diameter—the length of the chord that passes through the center of a circular region within which a spot will be placed. The spot diameter added to the spot buffer produces the center-to-center distance between consecutive spots.

spot set—a group of spots on a slide occupying an area covered by a single scan pass. Each slide can have up to two spot sets: spot set 1 is on the left side, spot set 2 is on the right side. In a normal mode spotting pattern, the array elements in spot set 2 are the duplicates of spot set 1.

spotting area—the area of the slide within which the Array Spotter spots the DNA or RNA material. The spotting area on the slide is approximately 18.36 mm by 54.25 mm.

spotting mode—the logical flow in which the Array Spotter spots the DNA or RNA material on the slides.

spotting pattern—the pattern on the slide that results from a particular spotting mode.

sVOL—subtracted volume value. Volume value of the spot minus the background volume value. For more information about sVOL, see the *ArrayVision Operations Manual* or Help file.

VOL—density value of each spot multiplied by its area. For more information about VOL, see the *ArrayVision Operations Manual* or Help file.

well set—a group of 12 consecutive wells in a microplate. Each well set corresponds to a set of wells spotted together by 12 spotting pens. Each 384-well microplate contains 32 well sets.

Index

A

actin NLR SD 8-4
analysis measures 5-7
analyzing image files 4-3, 5-8
array element sequence for analysis 4-3
Array Scanner image files 3-2, 4-1, 5-1, 6-1
ArrayStat Statistical Informatics software 8-14
ArrayVision
    analysis 2-3, 4-1, 5-1, 6-1
    output 1-2
Automated Spotfinding Application
    analysis 3-1
average signals 8-6, 8-7
A. thaliana 2-3

B

background removal method 4-3, 5-7
bacterial genes 2-3
B. subtilis 2-3

C

comparative analysis 2-3
confidence intervals A-4
control plate 1-1, 2-5
control samples 1-1, 2-1, 2-3
Cy3/Cy5 normalized log ratio 8-7
Cy3/Cy5 uncorrected log ratio 8-7

D

defining references 4-3, 5-8
detection limits 8-5, 8-13
Detection Limits table 8-5
duplicate spotting pattern 2-6
dynamic range
    controls 2-4
    evaluation 2-4
    of experiment 8-1
    values 8-5

E

excluding references 4-3
exponential normalization A-1
exporting image analysis data 4-4, 5-8

F

feature layout 6-2

G

GenePix 4000B image files 3-2
graph, interpreting 8-1

H

Help file 7-2
high-density spotting mode 2-6
housekeeping gene
  function as control 2-3
  performance 8-6
  ratio 8-5
Housekeeping Gene Performance table 8-6

I

image analysis
  exporting 4-4, 5-8
  output files 3-1, 3-2, 4-2, 5-2, 6-1
  settings 4-3, 5-3
image display format 4-2
image files
  Array Scanner 3-2, 4-1, 5-1, 6-1
  GenePix 4000B 3-2
  retrieving 4-2, 5-3
  saving 6-3
input files 3-1, 3-2, 4-2, 5-2, 6-1

L

loading image files in ArrayVision 5-3
logarithmic scale A-2

M

mean NLR SD 8-4, A-5
mean standard deviations, comparing 8-7
Microarray ScoreCard
  describing 1-1
  requirements 1-2
  starting 7-1
  workflow 1-3
microscanner settings 2-7

N

negative controls 2-3, 2-4
nonspecific hybridization 2-4
normal distribution A-5
normal spotting mode 2-6
normalization calculation A-1
normalized data grid 8-6
normalized log ratio 8-13, A-3
normalized signal ratio 8-3, 8-13

O

observed signal ratio 8-3

P

pen variability 8-9
pixels, saturated B-1
plant genes 2-3
poly A sequences 2-4
positive controls 2-3
Protocol Editor window 5-2

Q

quality report 8-11

R

Ratio Analysis table 8-3
ratio controls, description 2-4
ratio controls, performance 8-1
Ratio Precision table 8-4
reflective slides 2-5, 2-7, B-1
retrieving image files in ArrayVision 4-2, 5-3

S

saturated pixels B-1
saving image files in GenePix Pro 6-3
scanner settings 2-7
scatter plot, interpreting 8-1
sensitivity evaluation 2-4
sequence of array elements 4-3, 5-4, 5-5
signal data 8-6, 8-7
single-plate spotting mode 2-6
slides, reflective 2-5, 2-7, B-1
spike mix 2-4
Spotfire Pro software 8-14
spot-set variability 8-10
spotting control samples 2-7
spotting mode
    high-density 2-6
    normal 2-6
    single-plate 2-6
spotting pattern, duplicate 2-6
Statistical Informatics software 8-14
system validation 8-9
system verification 2-6

U

uncorrected log ratio 8-3, 8-13, A-3

Y

yeast sequences 2-4