Metadata of the chapter that will be visualized online

Chapter Title	Genomic Microarray Analysis	
Copyright Year	2014	
Copyright Holder	Springer Science+Business Media Dordrecht	
Corresponding Author	Prefix	Dr.
Family Name	Hamlet	
	Particle	
	Given Name	Stephen
	Suffix	
	Division	Griffith Health Institute, Gold Coast Campus
	Organization/University	Griffith University
	City	Nathan
	State	QLD
	Postcode	4222
	Country	Australia
Author	Prefix	Dr.
Family Name	Petcu	
	Particle	
	Given Name	Eugen
	Suffix	
	Division	Griffith Health Institute and School of Medicine, Gold Coast Campus
	Organization/University	Griffith University
	City	Nathan
	State	QLD
	Postcode	4222
	Country	Australia
Author	Prefix	Prof.
Family Name	Ivanovski	
	Particle	
	Given Name	Saso

Suffix	
Division	Griffith Health Institute and School of Dentistry and Oral Health, Gold Coast Campus
Organization/University	Griffith University
City	Nathan
State	QLD
Postcode	4222
Country	Australia

Abstract	<p>Microarrays utilize classical hybridization techniques to measure the levels of expression of genomic sequences. Advances in microarray fabrication using robotic technology and bioinformatics, now means entire genomes can be interrogated with a high degree of sensitivity, specificity and throughput. To achieve this, RNA collected from samples is first transcribed into complimentary RNA (cRNA) that is subsequently fragmented and labeled with biotin. These fragments are allowed to hybridize with their complementary oligonucleotide sequence located in microscopic features on the array support media. Each of these features consists of hundreds of short oligonucleotide probes that match the target mRNA sequence. This results in an increase in abundance of biotinylated cRNA molecules on the feature to which a fluorescent dye is subsequently bound. The fluorescence levels can then be used to infer the relative abundance of specific mRNA sequences in the original sample. The microarray analysis of gene expression is thus a key weapon in the molecular biology arsenal that can open windows to the molecular processes manifested by the phenotypic differences observed between normal and diseased states.</p>
----------	--

Keywords (separated by “ - ”)	Microarray - RNA - Affymetrix
-------------------------------	-------------------------------

AUTHOR QUERIES

Q1 Please provide E-mail address for “Stephen Hamlet”.

Chapter 30

Genomic Microarray Analysis

[AU1] Stephen Hamlet, Eugen Petcu, and Saso Ivanovski

30.1 Introduction

New blood vessel formation i.e. angiogenesis, is a fundamental biological process involved in tissue development, maintenance, repair and disease. The use of high throughput approaches such as DNA microarray technology that enables the parallel analysis of multiple genes has identified an increasing number of genes that have been linked to angiogenesis in recent years [1–3]. New bioinformatic approaches, such as the meta-analysis of this publicly available microarray data, are now being investigated to potentially identify ‘common higher-level’ transcriptional profiles in different disease states. The inclusion of such large sample numbers promises to detect the differential expression of genes that may not be otherwise found in individual studies and to minimize the rate of false positive findings [4].

Gene expression profiling therefore provides a means to make valuable insights into the effect(s) of differential gene expression that are ultimately manifested by the phenotypic differences observed between normal and diseased states. Simply put, the ‘expression’ of the inherited set of instructions that an organism carries within its genetic code in response to stimuli or instruction, is facilitated by the transcription of messenger RNA (mRNA), which is subsequently translated into the proteins that ultimately determine the phenotype or observable characteristics and traits of an organism (Fig. 30.1).

S. Hamlet (✉)
Griffith Health Institute, Gold Coast Campus, Griffith University,
Nathan, QLD 4222, Australia

E. Petcu
Griffith Health Institute and School of Medicine, Gold Coast Campus, Griffith University,
Nathan, QLD 4222, Australia

S. Ivanovski
Griffith Health Institute and School of Dentistry and Oral Health, Gold Coast Campus,
Griffith University, Nathan, QLD 4222, Australia

Fig. 30.1 The central dogma of genetics

23 Molecular methods of gene expression analysis based on the DNA sequence
24 have evolved from the seminal work by Southern [5] and typically include hybrid-
25 ization techniques using DNA probes to identify and quantitate complementary
26 DNA sequences. Microarray technology however is able to facilitate high through-
27 put whole genome analysis of gene expression at both low cost and most impor-
28 tantly, increased sensitivity and specificity compared with these traditional
29 methods.

30 In general terms, a microarray is a collection of microscopic features immobi-
31 lized on a solid support surface. These features contain millions of immobilized
32 identical probes that can hybridize with fluorescently labeled target molecules to
33 produce an increase in fluorescence intensity over a background level that can be
34 measured using a fluorescent scanner to produce both quantitative (gene expression)
35 and qualitative (diagnostic) data. Microarrays can thus be categorized based upon
36 the solid-surface support used, the nature of the probe and the methods used for
37 probe addressing and/or target detection as illustrated in Table 30.1.

t1.1 **Table 30.1** Comparison of microarray platforms

t1.2	Array type	Array support	Probe type	Probe density
t1.3	Printed	Glass slides	cDNA, PCR amplicons 200–	Moderate
t1.4			800 bp Oligonucleotides	
t1.5	In situ synthesized	Quartz wafer	25–80 bp	High
t1.6			Affymetrix – GeneChips®	
t1.7			20–25 bp	
t1.8	High-density bead arrays	Fiber optic bundles or silicon slides	Roche – NimbleGen® 60–100 bp	15,000 – >10 ⁶
t1.9			Agilent – SurePrint® – 60 bp	
t1.10			50-mer gene-specific probes are attached to beads, which are then assembled into the arrays	
t1.11	Electronic	Hydrogel permeation layer loaded with streptavidin over an electrode array	Nanogen – NanoChip® uses electronic addressing of charged biomolecules on the electrode array to separate and concentrate analyte targets	High
t1.12				50,000 – >10 ⁶
t1.13	Liquid-bead suspension	Microspheres	Non-specific capture sequences	Low
t1.14				400
t1.15				Low
t1.16				100
t1.17				
t1.18				
t1.19				
t1.20				

GeneChip® (Affymetrix, Santa Clara, CA) microarrays, arguably the most widely utilized for research purposes and therefore the focus of the methodology described in this chapter, use oligonucleotide probes which are synthesized using a semiconductor-based photochemical process directly onto the surface of the microarray, which is typically a 1.2 cm² quartz wafer (Fig. 30.2). These probes are typically short (20–25 base pairs) so multiple probes per target are included to improve sensitivity, specificity, and statistical accuracy. Classically, 11 probes are used per 600 bases being examined [6]. The gene sequence data for these probes is typically obtained from repositories of collected experimental data such as GenBank and the European Molecular Biology Laboratory (EMBL).

The use of ‘probe sets’ further increases the assay specificity. A probe set includes one perfect-match probe and one mismatch probe that contains a 1-bp difference in the middle position of the probe (i.e., position 13 of a 25-bp probe). Each member of the probe set is located in a separate feature, which allows the mismatch probe to act as a negative control to identify possible nonspecific cross-hybridization events. Affymetrix GeneChips typically have >10⁶ features per microarray depending on the inter-feature distance [7, 6].

Technical issues (discussed in the methodology) aside, a poorly designed experiment is probably the next most common reason that microarray experiments fail to yield results that are interpretable. Success with microarray approaches therefore requires a sound experimental design and a coordinated and appropriate use of statistical tools (Fig. 30.3). Most microarray experiments for example can be classified into one of three types. The first is the comparison of the same sample type before or after some defined treatment. Issues such as the experimental time course or duration of the treatment and the dose, etc., that can have dramatic effects all need

Fig. 30.2 Affymetrix GeneChip® preparation

Fig. 30.3 Gene profiling workflow using microarray analysis

to be considered in these experiments. Second is a comparison of samples that are isogenic except for a limited number of genetic changes, such as a single over-expressed or mutated gene. Again, care in interpretation is needed as the loss of one gene may be compensated by the up-regulation of other genes. Finally, the third experiment type is the comparison of normal or diseased samples from multiple individuals. This type of study is particularly prone to problems with unbalanced cohorts.

30.2 Methodology 70

30.2.1 Double Stranded cDNA Synthesis 71

Notes 72

- SuperScript Double-Stranded cDNA Synthesis Kit (Life Technologies). 73
- The starting material should be 5–20 µg of *total RNA*. 74
- The RNA quality should be checked using a spectrophotometer, a purity value of 1.9–2.1 is acceptable. It can also be checked on a 1–3 % agarose gel where there should be no smearing of the ribosomal bands or analyzed on the ‘Lab Chip’ from Agilent Technologies. 75 76 77 78

1st strand cDNA synthesis: 79

- (a) In an RNase-free 1.5-mL microcentrifuge tube, add oligo-dT according to the table below. Incubate at 70 °C for 10 min then quick-chill on ice. 80 81

	µg of total RNA		
t2.1		5–8	8.1–16
t2.2	T7-oligo dT	2 µL	2 µL
t2.3	RNA in DEPC water	9 µL	8 µL
t2.4	Total volume	11 µL	10 µL

- (b) Briefly centrifuge then add the following to the tube, mix well and incubate at 42 °C for 2 min to equilibrate sample. 82 83 84

	Reagents	Volume
t3.1	5× first strand buffer	4 µL
t3.2		
t3.3	0.1 M DTT	2 µL
t3.4	dNTP mix (10 mM)	1 µL
t3.5		

- (c) Add SuperScript II RT according to the following table. Mix gently, and incubate at 45 °C for 1 h. The total volume should now be 20 µL. Place the tube on ice to terminate the reaction.

		µg of total RNA		
		5–8	8.1–16	16.1–20
SuperScript II RT	2 µL	3 µL	4 µL	
Total volume	20 µL	20 µL	20 µL	20 µL

2nd strand cDNA synthesis:

Notes

- Place the first strand reaction on ice; centrifuge briefly to bring down the condensation on the sides of the tubes.
- Transfer the first stand mixture to a larger tube and add:

Reagent	Volume (µL)
First strand cDNA synthesis	20
DEPC water	91
5× second strand buffer	30
10 mM dNTP mix	3
10U/µL <i>E. coli</i> DNA ligase	1
10U/µL <i>E. coli</i> DNA polymerase I	4
2U/µL <i>E. coli</i> RNase H	1
Total volume	150 µL

- (a) Gently mix the tube (don't vortex) then centrifuge to bring down any residue on the sides.
- (b) Incubate at 16 °C for 2 h
- (c) Add 2 µL of T4 DNA polymerase (5U/µL) and incubate at 16 °C for 5 min
- (d) Add 10 µL of 0.5 M EDTA
- (e) Then clean up the sample or freeze at –20 °C.

30.2.2 Clean Up of Double Stranded cDNA

Notes

- MinElute Reaction Cleanup Kit (Qiagen)
 - All steps in this procedure are to be done at room temperature and the sample should be in a 1.5 or 2 mL eppendorf tube.
- (a) Add 600 µL of cDNA binding buffer to the 162 µL final double stranded cDNA synthesis product. Mix by vortexing for 3 s. This should result in a solution that is slightly yellow, about the same colour as the cDNA binding buffer before the solution was added. If this is not the case and the solution is purple, then add 10 µL of 3 M sodium acetate to correct the pH.

- (b) Apply 500 µL of the sample to the cDNA cleanup spin column. Have the column in a 2 mL collection tube, and centrifuge at 10,000 rpm for 1 min. Discard flow through.
- (c) Reload the column with the remaining 262 µL and centrifuge as above and discard flow through.
- (d) Transfer the column to a new 2 mL tube and wash with 750 µL of cDNA wash buffer then spin as above. Discard flow through.
- (e) Open the cap of the spin column and centrifuge at 25,000 g and discard flow through and the tube. This is to dry the membrane.
- (f) Transfer the spin column into a 1.5 mL collection tube and add 14 µL of cDNA elution buffer onto the membrane. Centrifuge at 10,000 rpm for 1 min. The elution volume is about 12 µL.
- (g) It is not recommended to RNase treat the sample before cDNA labeling.
- (h) Analyse the sample on a 1 % agarose gel (about 1 µL of sample) then stain with EtBr; this isn't to be done if the RNA is from total RNA.
- (i) There is no need to try and quantify the amount of cDNA because the primers will also absorb at 260 nm.

30.2.3 Synthesis of Biotin Labeled cRNA

Notes

- BioArray HighYield RNA Transcript Labeling Kit (T7) (Enzo Life Sciences)
- All reagents should be briefly centrifuged to bring down the entire sample which may have collected on the sides of the tubes.
- If the sample is prepared correctly there will be enough RNA for at least three chips.

t6.1	Total RNA (µg)	Volume of cDNA to use
t6.2	5–8	10 µL
t6.3	8.1–16	5 µL
t6.4	16.1–20	3.3 µL

Add template cDNA and other reagents in order:

t7.1	Reagent	Volume
t7.2	Template cDNA	3.3–10 µL (see table above)
t7.3	Distilled/deionised water	18.7–12 µL (see table above)
t7.4	10× Hy reaction buffer	4 µL
t7.5	10× biotin-labelled ribonucleotides	4 µL
t7.6	10× DTT	4 µL
t7.7	10× RNase inhibitor mix	4 µL

(continued)

Reagent	Volume
20× T7 RNA polymerase	2 μL
Total volume	40 μL

- (a) Mix reagents then centrifuge to ensure everything is at the bottom of tube.
- (b) Place in 37 °C water bath and incubate for 4–5 h gently mixing the tube every 30–45 min.
- (c) Overnight incubation may produce shorter products which is less desirable.
- (d) Store product at –70 or –20 °C.

30.2.4 Cleanup and Quantification of Biotin-labelled cRNA

Notes

- GeneChip® IVT cRNA Cleanup Kit (Affymetrix)
 - All the nucleotides must be removed so the reading at 260 nm is accurate
 - Don't use phenol/chloroform extraction as this lowers the yield (biotin is organic)
 - Save aliquot to do gel electrophoresis analysis
 - Work at room temperature
- (a) Add 60 μL RNase free water to the in vitro transcription (IVT-step 2) reaction and mix well (vortex 3 s).
 - (b) Add 350 μL IVT cRNA binding buffer to sample and vortex.
 - (c) Add 250 μL ethanol (96–100 %) to the lysate and mix with a pipette.
 - (d) Apply sample (700 μL) to the IVT cRNA cleanup spin column sitting in a 2 mL tube. Spin for 15 s at 10,000 rpm. Discard the flow through.
 - (e) Transfer the spin column into a new 2 mL tube and wash with 500 μL wash buffer then spin for 15 s at 10,000 rpm. Discard the flow through.
 - (f) Put 500μl of 80 % ethanol onto the column then spin for 15 s at 10,000 rpm. Discard the flow through.
 - (g) Open the cap and spin for 5 min at maximum speed. Discard the flow through and the collection tube.
 - (h) Transfer the spin column into a new 1.5 mL tube and pipette 11 μL of RNase free water onto the middle of the membrane then centrifuge for I minute at maximum speed.

Notes

- When analysing use 1/100 or 1:200 dilutions
- To check the yield; Absorbance at 260 and 280 nm
- Adjusted cRNA yield=RNA_m - (total RNA_i)(y)
- RNA_m=amount measured (μg)
- Total RNA_i=starting amount of RNA
- Y=fraction of cDNA in IVT

30.2.5 Fragmentation of cRNA 174

Notes 175

- GeneChip® 3' IVT Express Kit Assay (Affymetrix) 176
- The minimum concentration of cRNA is 0.6 µg/mL 177
- (a) Mix 2 µL of 5× fragmentation buffer per 8 µL of RNA and water. The final concentration can be between 0.5 and 2 µg/mL. It is important to use the adjusted cRNA concentrations. 178 179 180
- (b) Incubate at 94 °C for 5 min then sit on ice. 181
- (c) Save some of the sample to run on a gel (1 µg if using Ethidium bromide). 182
- (d) Store at –20 °C until ready to hybridize. 183

Notes 184

- To check fragmentation, run 1 % of each sample on a 1 % agarose gel. 185
- (a) Mix RNA with loading dye and heat to 65 °C for 5 min 186
- (b) Stain after running 187
- (c) Use a denaturing gel. 188

30.2.6 Hybridisation 189

Notes 190

- GeneChip® Hybridization, Wash, and Stain Kit (Affymetrix) 191
- The stock of 20× Genechip eukaryotic hybridisation control must be heated to 65 °C for 5 min to ensure cRNA is resuspended before aliquoting. 192 193
- The array must be allowed to equilibrate to room temperature before use. 194

t8.1				
t8.2	Component	Test array	Standard array	Final concentration
t8.3	Fragmented cRNA	5 µg	15 µg	0.5 µg/mL
t8.4	Control oligonucleotide B2 (3nM)	1.7 µL	5 µL	50 pM
t8.5	20× Eukaryotic hybridization controls (bio B, bio C, bio D, Cre)	5 µL	15 µL	1.5, 5, 25, 100 pM
t8.6				
t8.7	Herring sperm DNA (10 mg/mL)	1 µL	3 µL	0.5 mg/mL
t8.8	2× hybridization buffer	50 µL	150 µL	1×
t8.9	Water	To 100 µL	To 300 µL	
t8.10	Total volume	100 µL	300 µL	

- (a) Heat the hybridization cocktail to 99 °C for 5 min. 195 196
- (b) Fill the array with the maximal volume (100 or 300 µL depending upon the chip) of 1× hybridization buffer to wet the array. 197 198

- (c) Incubate the array with buffer for 10 min at 45 °C rotating it in the oven.
- (d) Transfer the hybridization cocktail from (b) to 45 °C and leave for 5 min.
- (e) Spin cocktails (still in tube) at maximum speed for 5 min.
- (f) Remove the buffer solution from the probe array cartridge and fill with appropriate volume of the clarified hybridization cocktail avoiding anything which has been deposited on the bottom.
- (g) Place the probe array in the rotisserie box (make sure it is balanced) in the oven at 45 °C rotate at 60 rpm for 16 h.

30.2.7 Washing and Staining the Array

Notes

- An automated GeneChip® Fluidics Station (Fig. 30.4) is used for the wash and stain operation of Affymetrix GeneChip® arrays.
- After hybridization, remove the hybridization cocktail from the probe array and set the solution aside in a micro-centrifuge tube. Store the solution on ice at the moment, but for long term storage transfer to -80 °C.
- Fill the probe array completely with the appropriate volume of non-stringent wash buffer. Both high and low stringency buffers are used in the wash process.
- The probe array can be stored at 4 °C for a maximum of 3 h if essential, but it will need to be equilibrated to room temperature before processing on the fluidics station.

Fig. 30.4 GeneChip® Fluidics Station

- Don't freeze Streptavidin-phycoerythrin (SAPE); store in the dark wrapped in 219
alfoil; prepare just before use. 220
- Follow prompts on fluidic station for wash/stain protocol 221

Staining Procedure 1 Eukaryotic targets (~75 min) 222

t9.1	Components	Volume	Final concentration
t9.2	2× MES stain buffer	300 µL	1×
t9.3	50 mg/mL acetylated	24 µL	2 mg/mL
t9.4	BSA		
t9.5	1 mg/mL SAPE	6 µL	10 µg/mL
t9.6	Water	270 µL	
t9.7	Total volume	600 µL	

Staining Procedure 2 This protocol is recommended for use with probe arrays 224
with probe cells of 24 µm or smaller (~90 min). 225

t10.1	Components	Volume	Final concentrations
t10.2	2× MES stain buffer	600 µL	1×
t10.3	50 mg/mL acetylated	48 µL	2 mg/mL
t10.4	BSA		
t10.5	1mg/mL SAPE	12 µL	10 µg/mL
t10.6	Water	540 µL	
t10.7	Total volume	1,200 µL	

Antibody solution 227

t11.1	Components	Volume	Final concentration
t11.2	2× MES stain buffer	600 µL	1×
t11.3	50 mg/mL acetylated BSA	24 µL	2 mg/mL
t11.4	10 mg/mL normal goat IgG	6 µL	0.1 mg/mL
t11.5	0.5 mg/mL biotinylated antibody	3.6 µL	3 µL/mL
t11.6	Water	266.4 µL	
t11.7	Total volume	600 µL	

30.2.8 Scanning the Array 229

Notes 230

- Affymetrix GeneChip® Scanner (Fig. 30.5) 231
- Warm up the laser up for at least 15 min 232
- For probe arrays with probe cells 24 µm or less – pixel value 3 µm, wavelength 233
570 nm 234
- For a 50 µm probe array – pixel value 6 µm, wavelength 570 nm 235

Fig. 30.5 GeneChip® scanner

30.3 Data Analysis

30.3.1 Data Extraction

Data is provided as .DAT files. These files are the output of the scanner and contain information about each pixel in the raw images. They are imported into the Affymetrix GCOS software to calculate the intensities. The libraries required to process the specific chip sets can be obtained from Netafix (Affymetrix). These intensity (.CEL) files may subsequently be used as input into image analysis software such as GeneSpring.

30.3.2 Importing Data Into GeneSpring

The intensity (.CEL) files are imported into GeneSpring and preprocessed with a normalisation algorithm. Robust Multiarray Averaging (RMA) adjusts the signal intensity for systematic errors introduced by differences in procedures and dye intensity effects. For example, the signal strength for the mismatched probes can often be larger than that of the perfect match probes, implying that the mismatched probe is detecting true signal as well as background signal resulting in nonsensical negative expression values.

RMA is a normalisation procedure for microarrays that background corrects, normalises and summarises the probe level information without the use of the information obtained in the mismatched probes. We routinely use the GeneChip RMA (GC-RMA) algorithm, which uses sequence-specific probe affinities of the

GeneChip probes to attain more accurate gene expression values as the preferred method for analyzing Affymetrix data. This method has been described in the literature to show the greatest consistency.

30.3.3 Normalisation and Filtering 259

The preprocessed data is often further normalised per gene as the absolute level of expression among different genes varies dramatically. For example a small change in fluorescence of a high abundance transcript may represent a significant change in expression of a gene that is expressed at low levels. By default, most microarray data analysis programs calculate the mean expression level for each gene, then normalize each sample against that mean, or control value. Following this, the noise in the array can be filtered out by discarding any genes labeled as absent in any of the samples of the experiment, as well as discarding any gene which had a raw signal value under 100. Only genes that pass these criteria should be used onwards.

30.3.4 Finding Significant Genes 269

To determine genes that were significantly differentially regulated, i.e different between time points or between health and disease (Fig. 30.6), one way ANOVA is routinely performed on the filtered and normalized genes, followed by multiple testing correction to reduce the number of false positives. Post hoc analysis is then performed on the ANOVA results.

30.3.5 Functional Characterization 275

Significant genes as determined from the statistical analysis may then be further characterized functionally, by examining their over-representation in pathways and in gene ontology categories (Fig. 30.7).

1. Overrepresentation of genes in KEGG pathways:
We use a script from the R integration package in GeneSpring (Bioscript library 2.2) to map to KEGG pathways. This script is designed to compare a user-defined gene list to known biological pathways, identifying genes in common and producing a measurement of confidence on the over-representation.
2. Overrepresentation of genes in Gene Ontology categories:
Online tools such as The Database for Annotation, Visualization and Integrated Discovery (DAVID, <http://david.abcc.ncifcrf.gov>) inputs lists of significantly expressed genes to determine enriched functional-related gene groups based on Gene Ontology (GO) classifications.

Fig. 30.6 Clustering analysis of significant genes

Fig. 30.7 'DAVID' and 'KEGG' functional analysis tools

References

289

1. Costouros NG, Libutti SK (2002) Microarray technology and gene expression analysis for the study of angiogenesis. *Expert Opin Biol Ther* 2(5):545–556 290
2. Toft JH, Lian IA, Tarca AL et al (2008) Whole-genome microarray and targeted analysis of angiogenesis-regulating gene expression (ENG, FLT1, VEGF, PlGF) in placentas from pre-eclamptic and small-for-gestational-age pregnancies. *J Matern Fetal Neonatal Med* 21(4):267–273 291
3. Anders M, Fehlker M, Wang Q et al (2013) Microarray meta-analysis defines global angiogenesis-related gene expression signatures in human carcinomas. *Mol Carcinog* 52:29–38 292
4. Choi JK, Choi JY, Kim DG et al (2004) Integrative analysis of multiple gene expression profiles applied to liver cancer study. *FEBS Lett* 565:93–100 293
5. Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 98:503–517 294
6. Dalma-Weiszhausz DD, Warrington J, Tanimoto EY et al (2006) The affymetrix GeneChip platform: an overview. *Methods Enzymol* 410:3–28 295
7. Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D (1991) Light-directed, spatially addressable parallel chemical synthesis. *Science* 251:767–773 296

Uncorrected Proof