Expression Profiling Using Affymetrix GeneChip Microarrays

Herbert Auer, David L. Newsom and Karl Kornacker

Columbus Children's Research Institute, 700 Children's Drive, Columbus, Ohio 43205, USA

Phone: ++ 614 355 3516

Fax: ++ 614 722 5895

Email: auerh@ccri.net

Abstract

The approximately 25,000 genes in mammalian genomes can be transcribed at different levels. Measurements of gene expression for ten thousands of genes in parallel gives the most comprehensive picture of steady-state levels of transcripts and is used in basic and applied research. Microarrays are the most frequently used technology for genome-wide expression profiling; from the various available microarray platforms, Affymetrix GeneChips are most frequently used for expression profiling and over 3,000 scientific publications describe results of this technology. In medical research, expression profiling by microarrays holds great promises for better understanding of diseases, identification of new therapeutic targets and subclassification of diseases to identify individualized treatment strategies.

Key Words: Expression profiling; microarrays; RNA; differential gene expression; genechips

Running title: GeneChip expression profiling

1. Introduction

High-density DNA microarrays provide an important tool to study the global patterns of gene expression. One of the earliest descriptions of today's most frequently used platform for expression profiling, Affymetrix GeneChips, is from 1996 [1] and many advances have been made over the last decade. The latest generations of GeneChips interrogate over 40,000 transcripts (Genome arrays), analyze individual exons of genes separately (Exon arrays), or even more advanced, the entire genome for transcriptional activity without focusing on predicted genes (Tiling arrays). Genome arrays providing one measurement of transcriptional activity per gene became tools to generate data reproducible across laboratories [2] and results from genome arrays can be reproduced by other measurements of gene expression like quantitative PCR [3]. Exon arrays and tiling arrays hold great promises, but reproducibility has not been shown yet, sample processing recommendations get modified every few months and an agreement on data analysis methods does not even appear on the horizon.

One of the best promises hold by expression profiling is in the field of medicine. Physicians hope for a detailed understanding of the molecular characteristics of the specific disease of a certain patient. This knowledge would allow them to provide an individualized treatment based on the molecular profiles of the affected tissue. As an example, expression profiles could provide information on the success of leukemia treatment in children [4].

2. Materials

2.1. RNA isolation

1. Tissue grinder "Pellet pestle" for microtubes (Kimble/Kontes, Vineland, NJ)

2. Absolutely RNA Nanoprep Kit (Stratagene, La Jolla, CA)

3. Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, Wilmington, DE)

4. 96% Ethanol (Fisher Scientific, Chicago, IL), 70% Ethanol (see **Note 1**)

5. Agilent 2100 Bioanalyzer (Agilent, Waldbronn, Germany)

6. Agilent RNA 6000 Nano Kit (Agilent, Waldbronn, Germany)

2.2. RNA amplification and labeling

1. Ovation Biotin RNA Amplification System and Labeling System 2300 (NuGEN Technologies, San Carlos, CA)

2. 80% Ethanol

3. DNA Clean & Concentrator - 25 (Zymo Research, Orange, CA)

4. Dye Ex 2.0 Spin Kit (Qiagen, Valencia, CA)

2.3. Microarray hybridization, washing, staining, scanning

1. Prepare 12X MES stock buffer (1.22 M MES, 0.89 M Na^+ from 61.61 g MES hydrate (4-Morpholineethanesulfonic acid monohydrate) for molecular biology and 193.3 g MES sodium salt (4-Morpholineethanesulfonic acid sodium salt), biotechnology performance certified (both Sigma-Aldrich) per liter. The pH should be between 6.5 and 6.7. Store at 4°C, shield from light and discard if solution turns yellow.
2. Prepare 50 ml 2X hybridization buffer from 8.3 ml 12X MES stock buffer, 17.7 ml 5M NaCl (Ambion, Austin, TX), 4.0 ml 0.5M EDTA Disodium Salt, (Sigma-Aldrich, St. Louis, MO), 0.1 ml 10% Tween-20, (Surfact-Amps 20, Pierce Chemicals, Rockford, IL) and 19.9 ml water. Store at 4°C and shield from light.

3. BSA, 50 mg/ml (Invitrogen Life Technologies, Carlsbad, CA)

4. Herring sperm DNA, 10 mg/ml (Promega, San Luis Obispo, CA)

5. GeneChip Eukaryotic Hybridization Control Kit containing 20X eukaryotic hybridization controls and control oligo B2 (Affymetrix, Santa Clara, CA)

6. Dimethyl sulfoxide (DMSO), (Sigma-Aldrich, St. Louis, MO)

7. R-Phycoerythrin Streptavidin (Molecular Probes, Carlsbad, CA)

8. Prepare wash buffer A (6X SSPE, 0.01% Tween-20) from 300 ml 20X SSPE (3M NaCl, 0.2M NaH_2PO_4, 0.02M EDTA, BioWhittaker Molecular Applications/Cambrex, Rockland, MA), 1 ml 10% Tween-20 and 699 ml water per liter. Filter through a 0.2 μm filter and store at room temperature.

9. Prepare wash buffer B (100 mM MES, 0.1M Na^+, 0.01% Tween-20) from 83.3 ml 12X MES stock buffer, 5.2 ml 5M NaCl, 1.0 ml 10% Tween-20 and 910.5 ml water per liter. Filter through a 0.2 μm filter, store at 4°C and shield from light.

10. Prepare 250 ml 2X stain buffer (200 mM MES, 2 M Na+, 0.1 % Tween-20) from 41.7 ml 12X MES stock buffer, 92.5 ml 5 M NaCl, 2.5 ml 10% Tween-20 and 113.3 ml water. Filter through a 0.2 μm filter, store at 4°C and shield from light.

11. Resuspend 50 mg Goat IgG, Reagent Grade (Sigma-Aldrich, St. Louis, MO) in 5 ml 150 mM NaCl.

12. Anti-streptavidin antibody (goat), biotinylated (Vector Laboratories, Burlingame, CA)

2.4. Affymetrix Equipment:

1. GeneChip 3000 7G (Affymetrix, Santa Clara, CA)

2. Fluidics Station 450 (Affymetrix, Santa Clara, CA)

3. Hybridization Oven 640 (Affymetrix, Santa Clara, CA)

2.5. Microarray quality control, data analysis

1. GCQC software (http://www.dnaarrays.org/downloads.php)

2. ArrayAssist software (http://www.stratagene.com/tradeshows/feature.aspx?fpId=91)

3. Methods

Microarray expression profiling works most effectively when samples are compared to each other instead of attempting to measure gene expression in an absolute manner [5]. These measurements of differential expression should represent biological differences between samples and the influence of technical variation on results should be kept as small as possible. A plethora of factors can contribute to measurements of differential expression caused by technical variation. Some examples are different RNA isolation protocols, variable amounts of starting material, variability RNA integrity between samples, different labeling chemistries used within a group of samples which should be compared afterwards, and different technicians processing samples. As an example, up to 75% of genes measured as differentially expressed can be false positive measurements (not representing biological differences between samples), if two samples with pronounced differences in RNA integrity are compared [6]. Many factors contributing to technical variation in expression profiling are most likely unknown; therefore, within a project (i.e. a group of samples where differential expression between samples should be calculated afterwards), technical variation should be kept as small as possible. This is

best achieved by parallel processing of samples by the same laboratory personnel within the shortest possible time; all chemistry and microarrays should have the same lot number and the identical equipment (like hybridization oven, washing station and scanner) should be used for all samples within a project.

In many cases only limited cell numbers are available for expression profiling. Isothermal amplification [7], the herein described method utilizes chemistry applicable to very small numbers of cells. It can start from as low as 5 ng total RNA and therefore can be used for a broad range of projects. In many cell types, this amount of RNA can be isolated from several thousand cells. This chemistry generates cDNA as an amplification product instead of cRNA (as generated by many other chemistries used for RNA labeling). cDNA hybridization to microarrays has been shown to generate more information on differential expression than cRNA [8] and less technical variability [9].

3.1. RNA isolation

1. Adherent cells are lysed in 100 µl lysis buffer containing β-mercaptoethanol without prior washing with PBS. Cells grown in suspension are spun down before lysis. Tissue should be snap-frozen in liquid nitrogen immediately after biopsy and stored at -80ºC until lysis. For homogenization, tissue is placed in a 1.5 ml tube containing 100 µl lysis buffer and immediately homogenized using the tissue grinder until no pieces of tissue are visible anymore. For further isolation steps see technical manual. The therein described optional step of DNAse treatment is performed.

2. Use 1.25 µl of RNA for quantification by the ND-1000 spectrophotometer (see **Note 2**). The blank measurement is performed using the elution buffer. The ratio of absorption at 260/280 nm should be above 1.8.
3. Perform RNA integrity control using RNA Nanochips 6000 according to the technical manual. This step consumes 1 µl RNA (see **Note 3**).

3.2. Amplification and labeling

3.2.1. Thermal cycler programming

Program 1 Primer Annealing: 65°C 5 minutes, 4°C forever

Program 2 First Strand Synthesis: 48°C 60 minutes, 70°C 15 minutes, 4°C forever

Program 3 Second Strand Synthesis: 37 °C 30 minutes,75 °C 15 minutes, 4°C forever

Program 4 Amplification: 48°C 60 minutes, 95°C 5 minutes, 4°C forever

Program 5 Fragmentation and Labeling: 50 °C 30 minutes, 4°C forever

3.2.2. cDNA synthesis

1. Place A3 (First Strand Enzyme Mix) on ice and thaw A1 (First strand Primer) and A2 (First Strand Buffer) at room temperature. Once thawed, mix A1 and A2 by vortexing, A3 by flicking the tube several times; place all reagents on ice.
2. Add 30 ng of total RNA sample to a 0.2 ml tube and add water to 5 µl. Add 2 µl of A1 to the RNA. Vortex briefly, spin down shortly.
3. Incubate 5 minutes at 65°C (program 1); as soon as completed place samples immediately on ice (see **Note 4**).

4. Prepare a master-mix of 12 µl A2 plus 1 µl A3 per sample on ice, add 13 µl of this per sample, mix by pipetting and spin down briefly.
5. Incubate in thermal cycler using program 2 (first strand synthesis).
6. Place B2 (Second Strand Enzyme Mix) on ice and thaw B1 (Second Strand Buffer) at room temperature. Once thawed, mix B1 by vortexing, B2 by flicking the tube several times; place all reagents on ice.
7. Prepare a master-mix of 2 µl B2 plus 18 µl B1 per sample on ice, add 20 µl of this per sample, mix by pipetting and spin down briefly.
8. Incubate in thermal cycler using program 3 (second strand synthesis); after completion, proceed immediately with amplification.

3.2.3. Amplification

1. Place C3-C (SPIA Enzyme Mix) on ice and thaw C2-C (SPIA Buffer) and C1 (SPIA Primer) at room temperature. Once thawed, mix C2-C and C1 by vortexing, C3-C by flicking the tube several times; place all reagents on ice. Place one additional 0.2 ml tube per sample on ice.
2. Prepare a master-mix of 4 µl water, 4 µl C1, 40 µl C3-C and 72 µl C2-C per sample, add 120 µl of this per sample (see **Note 5**). Mix by pipetting on ice.
3. Split reaction by removing 80 µl in pre-chilled 0.2ml tube.
4. Start program 4 (amplification) at thermal cycler; once cycler reaches 48°C, place tubes in cycler. After amplification, samples can be stored at -20°C.

3.2.4. Purification of amplified cDNA

1. Per sample, add 320 µl DNA binding buffer into a 1.5 ml tube.
2. Combine both 80 µl aliquots of each sample into one of these tubes, mix by vortexing and spin down briefly.
3. Place one Zymo 25 column per sample into a collection tube.
4. Load the entire volume of the sample (480 µl) onto column.
5. Centrifuge column in the collection tube for 10 seconds at > 10,000 g.
6. Discard flow-through. Place the column back in the same collection tube.
7. Wash sample by adding 200 µl of room temperature 80% ethanol. Do not use the wash buffer provided with the columns.
8. Centrifuge column in the collection tube for 10 seconds at > 10,000 g.
9. Discard flow-through. Place the column back in the same collection tube.
10. Add 200 µl of room temperature 80% ethanol.
11. Centrifuge column in the collection tube for 30 seconds at > 10,000 g.
12. Place the column in a clean 1.5 ml tube.
13. Add 30 µl of room temperature water to the center of each column. Let columns stand for 1 minute at room temperature.
16. Centrifuge column in 1.5 ml tube for 30 seconds at > 10,000 g.
17. Discard column; there should be approximately 30 µl of purified cDNA.
18. Mix sample by vortexing, then spin briefly.
19. Use 1.25 µl of cDNA for quantification by the ND-1000 spectrophotometer. The conversion factor is 33 for single stranded DNA. The blank measurement is performed using water. The ratio of absorption at 260/280 nm should be above 1.8.

20. Analyze cDNA molecular weight distribution using RNA Nanochips 6000 according to the technical manual. An example of molecular weight distribution is shown in **Fig. 1.**

3.2.5. Fragmentation and labeling

1. Place F2 (Fragmentation Enzyme) on ice and thaw F1 (Fragmentation Buffer Mix) at room temperature. Once thawed, mix F1 by vortexing, F2 by flicking the tube several times; place all reagents on ice.
2. Pipette 4 µg of purified cDNA into 0.2 ml tube and fill up to 25 µl with water.
3. Add 5 µl F1 per sample, mix by pipetting.
4. Add 5 µl F2 per sample, mix by pipetting and spin down briefly.
5. Incubate in thermal cycler using program 5 (fragmentation and labeling).
6. Thaw F3 (Labeling Buffer Mix), F4 (Biotin Reagent) and F5 (Stop Buffer) at room temperature. Once thawed, mix by vortexing, place F4 on ice and keep F3 and F5 at room temperature.
7. Add 5 µl F3 per sample, mix by pipetting.
8. Add 2.5 µl F4 per sample, mix by vortexing and spin down briefly.
9. Incubate in thermal cycler using program 5 (fragmentation and labeling).
10. Add 7.5 µl F5, mix by vortexing and spin down briefly.
11. After labeling, samples can be stored at -20°C.

3.2.6. Purification of Biotin labeled cDNA

1. Vortex DyeEx columns for 3 seconds.

2. Loosen column cap, snap off bottom closure, and place into a 2ml collection tube.

3. Spin at 750 g (2,800 rpm in a standard benchtop microcentrifuge) for 3 min, discard flow-through, transfer column to a clean collection tube.

4. Apply labeled cDNA to the center of the resin surface.

5. Spin at 750 g for 3 min. Discard column.

6. Purified labeled cDNA can be stored at -20°C.

3.3. Microarray hybridization, washing, staining, scanning

See **Note 7**

1. Heat frozen 20X Eukaryotic hybridization controls 5 minutes at 65°C.

2. Prepare hybridization cocktail using the entire volume of labeled and purified cDNA in 1.5 ml tubes following **Table 1**.

3. Equilibrate arrays in the packing at the bottom of the hybridization oven at 45°C for at least 10 minutes and set two heat blocks to 99°C and 45°C.

4. Heat denature hybridization cocktail for 2 minutes at 99°C, equilibrate for 5 minutes at 45°C.

5. Centrifuge tubes for 10 seconds at >10,000 g.

6. Place one pipette tip into one of the two septa of each array.

7. Load hybridization cocktail through second septum into array.

8. Hybridize arrays for 16 to 24 hours in hybridization oven at 45°C, rotating at 60 rpm.

9. Remove hybridization cocktail from arrays, fill arrays with wash buffer A and store hybridization cocktails at -20°C. For washing and staining, use EukGE-WS2v4_450 for standard arrays and Midi_Euk-2v3 for midi arrays (see **Note 8**).

10. Before scanning, control arrays for presence of air bubbles. If bubbles are present, remove them by adding wash buffer A through septum.
11. Scan arrays; CEL files are generated automatically after the scan is finished (see **Note 9**).

3.4. Microarray quality control and data analysis

See **Note 10**

1. Copy CEL files into the folder containing the GCQC.exe file and press enter. Two files are generated by GCQC, GCQCL and Raw_AFFX_PM_Intensities (see **Note 11**).
2. In the GCQCL file, PM95 should be > 1000, PM95/MM95 should be > 4, PM95/PM50 should be > 10 and MM05 should be < 100 (see **Note 12**). Some examples of GCQC results are shown in **Table 2**.
3. Load CEL files in ArrayAssist software. Use RMA from the options for data analysis algorithms (see **Note 13**).
4. Perform log transformation of expression estimates and calculate differential gene expression (see **Note 14**).

Notes

1. All solutions used for RNA have to be RNAse free. The frequently used practice of treating water with DEPC can inhibit enzymatic reactions. Therefore, nuclease free water provided with the labeling chemistries should be used or nuclease-free water (Non-DEPC treated, Ambion, Austin, TX). Pipette tips and tubes must be RNAse-free too. Do not perform DEPC treatment of tips and tubes but order instead specified as "RNAse free". All buffers for washing of arrays should be prepared in water that has a resistance of 20 $M\Omega cm^{-1}$ or more; buffers should be filtered through 0.2 µm filter. Water for buffers does not have to be nuclease free. Solutions as indicated to be purchased as solutions should not be prepared from powder. The chance of making errors is too big in relation to the potential financial loss from non-usable array data.
2. Many other photometers consume big volumes of RNA for quantification. Therefore, a large portion of RNA would have to be used for quantification or the sample would be that diluted that the absorption reading is below accuracy of the photometer. For the ND-1000 spectrophotometer, 1.25 µl is sufficient; the 1 µl mentioned in Nanodrop's manual is sometimes not sufficient for accurate readings.
3. Agilent Expert software version 1.2.3.4 or higher provide quantification of RNA integrity by the "RNA Integrity Number" (RIN). For cell cultures, RIN should be above 9 and for primary cells RIN should be above 7.
4. Leaving the samples in the thermal cycler until it reaches 4°C, generates subsequently shorter cDNAs products.

5. The amplification enzyme should not reach temperature above 4°C before the reaction is placed in the thermal cycler. All pipetting must be performed on ice and as quickly as possible.
6. Standard arrays are for example 133 Plus 2.0 or 430 2.0 arrays, midi arrays are for example 133A 2.0 or 430A 2.0 arrays.
7. This protocol contains deviations from the protocol provided by Affymetrix. Follow the instructions as given here.
8. For detailed instructions on array washing, and staining refer to Affymetrix GeneChip Expression Analysis Technical Manual (revision 4). Streptavidin phycoerythrin is unstable and should be protected from light; a once opened tube should not be used longer than 2 months.
9. CEL files contain information on the 75^{th} percentile of intensities of each probe. Almost all downstream data analysis is performed using CEL files.
10. Report Files generated by GCOS and ArrayAssist software provide information about data quality derived from scaled values and therefore are not very informative.
11. The file Raw_AFFX_PM_Intensities contains raw (non-scaled) probe intensities of probe sets whose identifiers start with AFFX. Using Excel, graphs of intensities for these probe sets can be drawn. These graphs are helpful in troubleshooting.
12. MM05 reports the 5^{th} percentile of intensity of Mismatch probes; this parameter is used to assess background signals. PM95 reports the 95^{th} percentile of Perfect Match probes; this parameter reports the upper end of signal intensity. The ratio PM95/MM95 reports the signal over background ratio. The PM95/PM50 reports the slope of PM intensities at the upper end; high background compresses this ratio.

13. ArrayAssist can be downloaded from Stratagene's website for free. After a test period, the software downgrades to ArrayAssist Light, still capable of calculating RMA expression estimates. Other data analysis algorithms (GCOS, PLIER) can generate high percentages of false positive reports of differential gene expression. Any kind of software for generating expression estimates summarizes the eleven or more probes within one probe set to provide one estimate for quantification of the measured transcript.
14. When replicates of experiments have been performed, a T-test for statistical significance of differential expression should be applied. $P < 0.05$ is considered as significant and corrections for multiple testing should not be applied. When only one sample has been processed per biological group, differential expression measured above two-fold usually can be considered as reliable since the majority of these measurements can be confirmed by Real-Time PCR. If more than one probe set on the array measures a certain transcript, the probe set with the highest absolute value of differential expression should be used for further analysis.

References

1. Lockhart DJ, Dong H, Byrne MC, et al. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotechnol 1996;14(13):1675-80.

2. Shi L, Reid LH, Jones WD, et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nat Biotechnol 2006;24(9):1151-61.

3. Canales RD, Luo Y, Willey JC, et al. Evaluation of DNA microarray results with quantitative gene expression platforms. Nat Biotechnol 2006;24(9):1115-22.

4. Yeoh EJ, Ross ME, Shurtleff SA, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. Cancer Cell 2002;1(2):133-43.

5. Bammler T, Beyer RP, Bhattacharya S, et al. Standardizing global gene expression analysis between laboratories and across platforms. Nat Methods 2005;2(5):351-6.

6. Auer H, Lyianarachchi S, Newsom D, Klisovic MI, Marcucci G, Kornacker K. Chipping away at the chip bias: RNA degradation in microarray analysis. Nat Genet 2003;35(4):292-3.

7. Dafforn A, Chen P, Deng G, et al. Linear mRNA amplification from as little as 5 ng total RNA for global gene expression analysis. Biotechniques 2004;37(5):854-7.

8. Barker CS, Griffin C, Dolganov GM, Hanspers K, Yang JY, Erle DJ. Increased DNA microarray hybridization specificity using sscDNA targets. BMC Genomics 2005;6(1):57.

9. Eklund AC, Turner LR, Chen P, et al. Replacing cRNA targets with cDNA reduces microarray cross-hybridization. Nat Biotechnol 2006;24(9):1071-3.

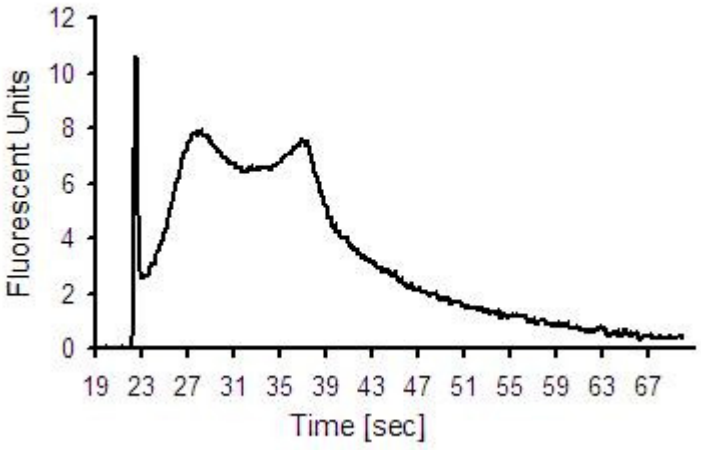

Fig. 1. Size distribution of amplified cDNA measured using the 2100 Bioanalyzer.

Table 1. Preparation of the hybridization cocktail

Component	Standard array	Midi array (see **Note 6**)
Labeled cDNA	50 µl	34 µl
Control oligonucleotide B2	3.7 µl	2.5 µl
20X Eukaryotic hybridization controls	11 µl	7.5 µl
Herring sperm DNA	2.2 µl	1.5 µl
Acetylated BSA	2.2 µl	1.5 µl
2x Hybridization buffer	110 µl	75 µl
100% DMSO	22 µl	15 µl
Water	19 µl	13 µl
Final Volume	220 µl	150 µl

Table 2. Quality control parameters provided by GCQC software

SAMPLE	MM05	MM25	MM50	MM75	MM95	PM05	PM25	PM50	PM75	PM95	PM95/PM50	PM95/MM95
A	29	31	33	39	75	29	32	35	49	**176**	5.029	2.347
B	**279**	380	486	629	1004	301	427	572	830	2920	5.105	**2.908**
C	57	72	97	154	468	60	82	124	256	1489	12.008	**3.182**
D	97	133	173	238	600	108	156	221	418	2661	12.041	4.435

For each sample, 5^{th}, 25^{th} 50^{th}, 75^{th} and 95^{th} percentile of intensities for Mismatch (MM) Perfect Match (PM) probes are provided respectively. Sample A shows results of a microarray where phycoerythrin-staining worked poorly, sample B shows results of a microarray with high background, sample C shows results of cRNA hybridization (instead of cDNA), resulting in lower PM95/MM95 ratios, sample D shows results of a successful microarray experiment. Problematic quality control parameters are highlighted in bold.