

CA). Touchdown PCR amplifications were performed as recommended¹⁸. Cycle sequencing protocols were used with ABI sequencers at the Hutchinson Center Biotechnology Facility.

DHPLC. Mutation detection was performed using the Transgenomic WAVE system. Following PCR amplification, the Pfu polymerase was inactivated, and the DNA samples were heated and cooled to form heteroduplexes¹⁸. For most fragments, the predicted WAVE (v.3.5) melting temperatures and separation gradients were used¹⁹.

Acknowledgments

We thank Bruce Draper for helpful discussions. This work was supported by grant RO1 GM29009 (to S.H.) from the National Institutes of Health. S.H. is an investigator of the Howard Hughes Medical Foundation, which also provided support for Karen Wolfe of the James Roberts lab, whom we thank for helping us with the screen.

- Koornneef, M., Dellaert, L.W. & van der Veen, J.H. EMS- and radiation-induced mutation frequencies at individual loci in *Arabidopsis thaliana* (L.) Heynh. *Mutat. Res.* **93**, 109–123 (1982).
- Underhill, P.A. et al. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res.* **7**, 996–1005 (1997).
- Henikoff, S. & Comai, L.A. DNA methyltransferase homolog with a chromodomain exists in multiple forms in *Arabidopsis*. *Genetics* **149**, 307–318 (1998).
- Platero, J.S., Hartnett, T. & Eissenberg, J.C. Functional analysis of the chromo domain of HP1. *EMBO J.* **14**, 3977–3986 (1995).
- Rose, T.M. et al. Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Res.* **26**, 1628–1635 (1998).
- <http://aims.cps.msu.edu/aims/> AIMS: *Arabidopsis* Information Management System.
- Gross, E. et al. A comparison of BRCA1 mutation analysis by direct sequencing, SSCP and DHPLC. *Hum. Genet.* **105**, 72–78 (1999).
- Li-Sucholeiki, X.C. et al. Applications of constant denaturant capillary electrophoresis/high-fidelity polymerase chain reaction to human genetic analysis. *Electrophoresis* **20**, 1224–1232 (1999).
- Postai, J. et al. Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Res.* **17**, 2421–2435 (1989).
- Liu, L.X. et al. High-throughput isolation of *Caenorhabditis elegans* deletion mutants. *Genome Res.* **9**, 859–867 (1999).
- Bowman, J.L., Smyth, D.R. & Meyerowitz, E.M. Genes directing flower development in *Arabidopsis*. *Plant Cell* **1**, 37–52 (1989).
- Schimmenti, J. & Bucan, M. Functional genomics in the mouse: phenotype-based mutagenesis screens. *Genome Res.* **8**, 698–710 (1998).
- Kheterpal, I. & Mathies, R.A. Capillary array electrophoresis DNA sequencing. *Anal. Chem.* **71**, 31A–37A (1999).
- Larsen, L.A. et al. High-throughput single-strand conformation polymorphism analysis by automated capillary electrophoresis: robust multiplex analysis and pattern-based identification of allelic variants. *Hum. Mutat.* **13**, 318–327 (1999).
- Oleykowski, C.A. et al. Mutation detection using a novel plant endonuclease. *Nucleic Acids Res.* **26**, 4597–4602 (1998).
- Aljanabi, S.M. & Martinez, I. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Res.* **25**, 4692 (1997).
- http://faculty.washington.edu/comai/lpg_miniprep.htm
- Kuklin, A. et al. Detection of single-nucleotide polymorphisms with the WAVE DNA fragment analysis system. *Genetic Testing* **1**, 201–206 (1997).
- Jones, A.C. et al. Optimal temperature selection for mutation detection by denaturing HPLC and comparison to single-stranded conformation polymorphism and heteroduplex analysis. *Clin. Chem.* **45**, 1133–1140 (1999).

High-fidelity mRNA amplification for gene profiling

Ena Wang^{1,3}, Lance D. Miller^{2,3}, Galen A. Ohnmacht¹, Edison T. Liu², and Francesco M. Marincola^{1*}

¹Surgery Branch, Division of Clinical Sciences, National Cancer Institute and the Department of Transfusion Medicine, Clinical Center, National Institutes of Health, Bethesda, MD. ²Medicine Branch, Division of Clinical Sciences, National Cancer Institute, National Institutes of Health, Gaithersburg, MD.

³These two authors contributed equally to this project. *Corresponding author (marincola@nih.gov)

Received 10 January 2000; accepted 18 February 2000

The completion of the Human Genome Project¹ has made possible the comprehensive analysis of gene expression^{2,3}, and cDNA microarrays are now being employed for expression analysis in cancer cell lines⁴ or excised surgical specimens⁵. However, broader application of cDNA microarrays is limited by the amount of RNA required: 50–200 µg of total RNA (T-RNA) and 2–5 µg poly(A) RNA⁶. To broaden the use of cDNA microarrays, some methods aiming at intensifying fluorescence signal^{7–9} have resulted in modest improvement. Methods devoted to amplifying starting poly(A) RNA^{10,11} or cDNA¹² show promise, in that detection can be increased by orders of magnitude. However, despite the common use of these amplification procedures^{11,13–16}, no systematic assessment of their limits and biases has been documented. We devised a procedure that optimizes amplification of low-abundance RNA samples by combining antisense RNA (aRNA) amplification¹⁰ with a template-switching effect (Clonetech, Palo Alto, CA). The fidelity of aRNA amplified from 1:10,000 to 1:100,000 of commonly used input RNA was comparable to expression profiles observed with conventional poly(A) RNA- or T-RNA-based arrays.

One round of amplification yielded ~10³-fold of the estimated amount of starting mRNA, and two rounds yielded ~10⁵-fold. Random bias resulting from RNA amplification or nonspecific hybridization was assessed by hybridizing differentially labeled aRNA-based targets from the same melanoma line to 2,008 human gene microarrays (NCI-OncoChip, NCI, Bethesda, MD). Scatter plots of Cy3 (green) versus Cy5 (red) signal reproducibly revealed a strong linear relationship ($R^2 = 0.99$) (Fig. 1A). Similar linearity was observed with aRNA from a renal cancer line ($R^2 = 0.96$). To assess systematic bias introduced by aRNA amplification, the expression profile of labeled aRNA-based targets was compared to that of conventional T-RNA and poly(A) RNA-based targets by identifying differentially expressed genes from two different sources (A375 and ML-1) (Fig. 1B). Highly reproducible “outliers” in four consecutive T-RNA-based arrays at optimized target concentration (100 µg for Cy3 and 50 µg for Cy5 targets) were considered to be differentially expressed. Outliers were defined as genes producing array spots that exhibit Cy3: Cy5 ratios significantly different from 1.0 at a 99.0% confidence level (cutoff ratio ranged from 1.7 to 2.1). To exclude labeling biases, T-RNA-based targets from either cell line were labeled with the reciprocal fluorochrome in every other duplicate experiment. Therefore, a green spot on one array would be red in the reciprocal. True (concordant) outliers were those that were positive using reciprocal fluorochrome and reproducible using the same fluorochrome. Results were analyzed using the hierarchical clustering technique of Eisen and coworkers¹⁷. Outliers were ranked into mutually exclusive confidence groups (Fig. 2A). The “4 Match” group consisted of concordant spots ($n = 267$) in all four hybridizations. The “3 Match” group represented concordance in three hybridizations ($n = 69$). The “2 Match rec.” group contained only reciprocal outliers ($n = 12$), and the “2 Match rep.” group reproducible outliers ($n = 311$) appearing twice in the four consecutive arrays but not in the reciprocal fluorochrome experiments. The fourth group (2 Match rep.) was believed to represent genes whose measurement of expression was confounded by labeling bias affecting low transcript levels in which background fluorescence intensity was higher with one but not the other dye.

Of the outliers identified by the aRNA amplified from 0.25 to 3.0 µg source RNA, 85–92% reproducibly matched “true outliers” identified by T-RNA. The level of concordance was identical comparing an additional hybridization using T-RNA or poly(A) RNA. Detection of true outliers degenerated in aRNA amplified from 0.125 to 0.031 µg total RNA (30–70%). However, a second amplification restored concordance in aRNA from 0.031 to 0.010 µg T-RNA (80–85%). Visual demonstration of the level of outlier concordance was accomplished by applying a high-stringency filter (Cy3: Cy5 or Cy5: Cy3 ratios above

TECHNICAL REPORTS

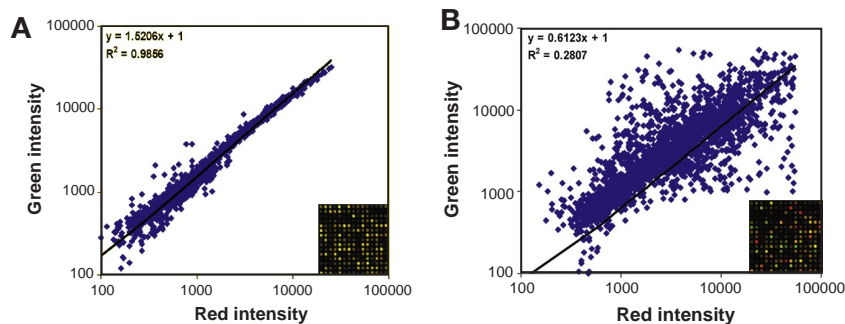


Figure 1. (A) Assessment of random or labeling bias by hybridization of differentially labeled aRNA target (3 µg) amplified from the same melanoma cell line to 2,008-gene OncoChip. (B) Comparison of aRNA targets from a melanoma cell line versus human myeloid leukemia cell line labeled with Cy3 and Cy5 respectively ($R^2 = 0.28$).

3, fluorescence intensity >300 in one channel unless the other channel was >1,000, and a spot size of ≤50 pixels). Genes satisfying these requirements in at least five experiments were clustered¹⁷ (Fig. 2B). Clustering revealed 251 outliers with strong concordance that decreased with reducing amounts of source RNA and could be reestablished by a second amplification of low source material.

To identify false positives, a more tolerant filter (Cy5: Cy3 or Cy3: Cy5 above 3, fluorescence intensity >150 in one channel in any of the experiments) was applied, allowing visualization of less reproducible outliers. Approximately 250 false positives biasing the Cy5 channel (blue bar in Fig. 3A) emerged with aRNA from <0.125 µg of source T-RNA. These false outliers were not detected with T-RNA or aRNA from 3.0 µg to 0.25 µg. We postulated that this Cy5 bias was related to differential optical detection of the red and green fluorochrome at low target concentration. One round of amplification from 62 to 31 ng T-RNA yielded quantities of aRNA (1.3–2.0 µg) below the standard target concentration (3 µg) used for aRNA- or poly(A) RNA-based arrays. Consequently, lower amounts of target were used in these arrays, decreasing fluorescence particularly in low-abundance transcripts. Reamplification of aRNA from 0.031 to 0.010 µg T-RNA permitted aRNA-based hybridization with optimal target concentration (3 µg) and restored the ability to detect outliers in each confidence group with percentages comparable to the 0.25–3.0 µg aRNA set. Furthermore, false positive signals in the Cy5 channel were suppressed (Fig. 3A), suggesting that the Cy5 bias was not due to molecular anomalies from RNA amplification but to a remediable postamplification artifact.

The number of experimental outliers discordant from the four confidence groups was summarized as percentage of the total number of genes on the array (Fig. 3B). This parameter, a reliable measure of nonreproducibility, was 4.5% when using labeled T-RNA-based targets. The percentage of nonreproducible outliers noted with aRNA-based hybridizations from 0.25 µg to 3.0 µg source RNA ranged from 3 to 6% similar to T-RNA-based arrays. This measure of nonreproducibility

increased in arrays using aRNA from 0.031 to 0.125 µg source T-RNA but was reduced to baseline levels by a second round of aRNA amplification. In vitro transcription has been utilized for differential gene expression studies^{11,13–16}. However, these studies have estimated the linearity and reproducibility of poly(A) RNA amplification in a limited number of genes by northern blot or in situ hybridization^{11,13–16}. Conventional antisense mRNA amplification can introduce biases in the amplified product because of a possible 5' underrepresentation and because of the low stringency temperature applied during double-stranded cDNA (ds-cDNA) synthesis. In this study, a modification of conventional antisense mRNA amplification¹⁶ exploiting a template-switching effect at the 5' end¹⁸ ensured the generation of full-length ds-cDNA. Furthermore, the template-switching primer-dependent second-strand cDNA synthesis occurs at 75°C. Thus, this modification overcomes potential 3' bias (useful when unmapped sequences are used for array spotting) and enhances sequence specificity by high-temperature cDNA synthesis. This technique yields up to 10⁵-fold linear amplification of high-fidelity aRNA from nanograms of T-RNA and is applicable whether T-RNA or poly(A) RNA is used. This has been confirmed by Mahadevappa and colleagues¹⁹, who showed that T-RNA-based oligonucleotide arrays yield results comparable to those obtainable using poly(A) RNA^{11,20}. Our combined results define the operational parameters of RNA amplification approaches and expand the utilization of cDNA microarrays to experimental conditions in which starting material is the limiting factor. Such conditions include clinical specimens from fine-needle aspirates or microdissection or experimental models studying embryonic tissue or small organisms.

Experimental protocol

A375 melanoma and ML-1 human myeloid leukemia cell lines were obtained

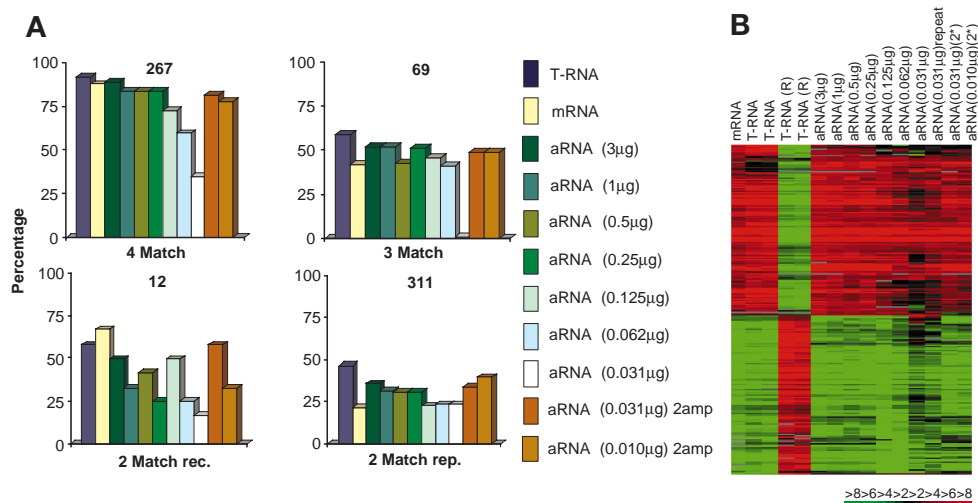


Figure 2. (A) Grading of outlier reproducibility in mRNA, T-RNA, and aRNA hybridizations. Percentages of the genes belonging to each confidence group identified as outliers in experimental conditions are shown as bars and the total number of outliers for each group is indicated. RNA concentrations in the labels refer to starting amount of source T-RNA. (B) High-stringency hierarchical cluster diagram of differentially expressed genes (outliers) in mRNA, T-RNA, and aRNA array hybridizations that encompasses all four confidence groups. Columns designate single-array hybridizations: targets from melanoma cell lines are Cy3 (green) biased except for T-RNA in which targets were reciprocally labeled: T-RNA (R). Numbers in parentheses (A and B) refer to amount of source T-RNA from which aRNA was amplified. 2* designates aRNA obtained after two rounds of amplification (2amp in A). Rows designate single genes (array probes). Green and red cells reflect genes expressed at higher levels in A375 and ML-1 cells, respectively. Black cells indicate genes with approximately equivalent expression levels, and gray cells indicate missing or filter-excluded data. The magnitude of the log-transformed ratio is reflected by the degree of color saturation (see color scale). Shown are 251 genes with expression ratios of threefold or greater in at least five hybridizations.

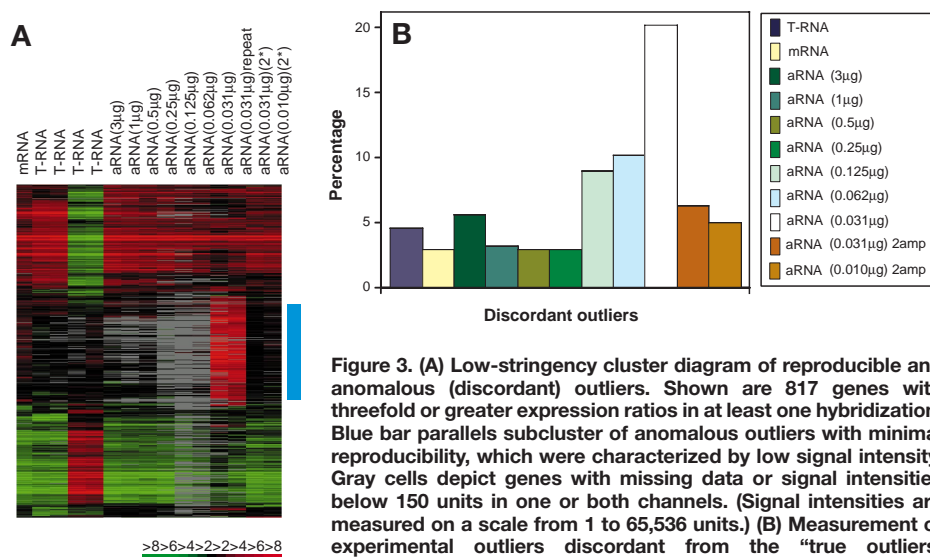


Figure 3. (A) Low-stringency cluster diagram of reproducible and anomalous (discordant) outliers. Shown are 817 genes with threefold or greater expression ratios in at least one hybridization. Blue bar parallels subcluster of anomalous outliers with minimal reproducibility, which were characterized by low signal intensity. Gray cells depict genes with missing data or signal intensities below 150 units in one or both channels. (Signal intensities are measured on a scale from 1 to 65,536 units.) (B) Measurement of experimental outliers discordant from the "true outliers" determined by the control T-RNA hybridizations presented as percentage of the total number of genes in the array.

from the American Type Culture Collection (ATCC; Rockville, MD) and the National Human Genome Research Institute, respectively, and maintained in RPMI supplemented with 10% fetal calf serum (Biofluids, Rockville, MD). Total RNA was isolated using RNeasy midi kits (Qiagen, Valencia, CA) and refined using TRIZOL reagent (Gibco-BRL, Gaithersburg, MD). mRNA was purified from T-RNA using Oligotex mRNA isolation kit (Qiagen). RNA concentrations were determined by OD₂₆₀ reading in 50 mM sodium hydroxide (GeneQuant, Clamart Cedex, France).

aRNA was prepared from T-RNA in 9 µl diethyl H₂O containing 1 µl (1 µg µl⁻¹) oligo-dT (15)-T7 (5'-AAA CGA CGG CCA GTG AAT TGT AAT ACG ACT CAC TAT AGG CGC-3') primer. Total RNA was denatured at 70°C for 3 min and primed while cooling to room temperature. T7 bacteria phage promoter was incorporated into cDNA synthesis in a reverse transcription (RT) reaction by adding 4 µl of first-strand reaction buffer, 2 µl 0.1 M dithiothreitol (DTT; Gibco-BRL), 2 µl 10 mM dNTP, 1 µl RNasin (Promega, Madison, WI), 1 µl (1 µg µl⁻¹) template switch primer (5'-AAG CAG TGG TAT CAA CGC AGA GTA CGC GGG-3') (Clontech, Palo Alto, CA), and 2 µl Superscript II (SSII) reverse transcriptase (Gibco-BRL). cDNA synthesis was completed at 42°C for 1 h. Full-length ds-cDNA was synthesized by adding 106 µl of DNase-free water, 15 µl Advantage PCR buffer (Clontech), 3 µl 10 mM dNTP, 1 µl RNase-H (Promega), 3 µl Advantage cDNA Polymerase (Clontech). The following temperature cycle was used: 2 min at 37°C for RNA digestion, 3 min at 94°C for denaturation, 3 min at 65°C for priming, and 30 min at 75°C for extension. Reactions were terminated by incubation in 7.5 µl 1 M NaOH with 2 mM EDTA at 65°C for 10 min. cDNA was extracted with phenol-chloroform-isoamyl and precipitated with ethanol in the presence of 0.1 µg linear acrylamide (0.1 µg µl⁻¹, Ambion, Austin, TX). cDNA, resuspended in 16 µl DEPC H₂O was passed through a Bio-6 chromatography column (Bio-Rad, Cambridge, MA) and washed three times with 700 µl DEPC-treated H₂O. Samples were lyophilized to 16 µl. For the second round of amplification, 16 µl of purified full-length ds-cDNA were incubated with 4 µl of each 75 mM NTP (ATP, GTP, CTP, and UTP), 4 µl of 10× reaction buffer, and 4 µl of transcription enzyme mixture (T7 Megascript Kit 1334, Ambion) in 40 µl volume at 37°C for 5 h. RNA recovery and removal of template DNA was achieved by TRIZOL purification. Aliquots of aRNA (1.3 µg prepared from 31 ng of source T-RNA, and 0.65 µg prepared from 10 ng source T-RNA) were reverse transcribed into cDNA using 2 µg of random hexamer with 5 µl first-strand buffer, 2 µl 0.1 M DTT, 1 µl RNasin, 2 µl of 10 mM dNTP, and 2 µl of SSII. The reaction mixture was heated to 65°C for 10 min before adding SSII; then synthesis was continued at 42°C for 1 h. Second-strand cDNA synthesis was initiated by 1 µg oligo-dT-T7 primer in the conditions used in the first round. In vitro transcription of aRNA was carried out as for the first round.

We labeled 50 µg (for Cy3 labeling) or 100 µg (for Cy5 labeling) T-RNA and 3 µg aRNA or nonamplified mRNA in a RT reaction employing 8 µg of random hexamer primer in the presence of Cy3- or Cy5-labeled dUTP (Amersham, Piscataway, NJ) and SSII (Gibco-BRL). Reaction products were purified in a Bio-6 chromatography column followed by Microcon concentra-

tion (purified and labeled cDNA target in 20 µl containing 2.6 µl 20× SSC (sodium chloride, sodium citrate buffer), 8 µg of poly(dA), 4 µg yeast Trna, and 10 µg of human Cot 1 DNA (Gibco-BRL). Before hybridization, the mixture was heated to 99°C for 2 min and then cooled to room temperature. At that point, 0.46 µl of 10% sodium dodecyl sulfate (SDS) were added. Hybridization was carried out at 65°C for 12–18 h in a waterbath. Before scanning, slides were washed in 2× SSC with 0.1% SDS for 2 min, then 1× SSC, 0.2× SSC, and 0.05× SSC, sequentially for 1 min each.

A total of 2,008 named cDNAs were spotted onto poly-L-lysine-coated slides using an OmniGrid arrayer (GeneMachines, San Carlos, CA). Hybridized arrays were scanned at 10 µm resolution on a GenePix 4000 scanner (Axon Instruments, Inc., Foster City, CA) at variable PMT voltage to obtain maximal signal intensities with <1% probe saturation. Resulting tiff images were analyzed via ArraySuite software (National Human Genome Research Institute, Bethesda, MD)²¹.

Acknowledgments

The authors wish to acknowledge John Powell (CIT) for establishing the NCI microarray database (mAdb) and accessibility of analytical tools, and Louis Staudt (NCI) for array clones.

- Collins, F.S. et al. New goals for the U.S. Human Genome Project. *Science* **282**, 682–629 (1998).
- Lander, E.S. The new genomics: global view of biology. *Science* **274**, 536–539 (1996).
- Brown, P.O. & Botstein, D. Exploring the new world of the genome with DNA microarrays. *Nat. Genet.* **21**, 33–37 (1999).
- DeRisi, J. et al. Use of cDNA microarray to analyze gene expression patterns in human cancer. *Nat. Genet.* **14**, 457–460 (1996).
- Perou, C.M. et al. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc. Natl. Acad. Sci. USA* **96**, 9212–9217 (1999).
- Duggan, D.J., Bittner, M., Chen, Y., Meltzer, P. & Trent, J.M. Expression profiling using cDNA microarrays. *Nat. Genet.* **21**, 10–14 (1999).
- Chen, J.J. et al. Profiling expression patterns and isolating differentially expressed genes by cDNA microarray system with colorimetry detection. *Genomics* **51**, 313–324 (1998).
- Rajeevan, M.S., Dimulescu, I.M., Unger, E.R. & Vernon, S.D. Chemiluminescent analysis of gene expression on high-density filter arrays. *J. Histochem. Cytochem.* **47**, 337–342 (1999).
- Zeje, Y., Xiaoyi, W., Yu, T. & Huan, H. The method of micro-displacement measurement to improve the space resolution of array detector. *Med. Engineer. Phys.* **20**, 149–151 (1998).
- Phillips, J. & Eberwine, J.H. Antisense RNA amplification: a linear amplification method for analyzing the mRNA population from single living cells. *Methods* **10**, 283–288 (1996).
- Lockhart, D.J. et al. Expression monitoring of hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* **14**, 1675–1680 (1996).
- Trenkle, T., Welsh, J., Jung, B., Mathieu-Daude, F. & McClelland, M. Non-stoichiometric reduced complexity probes for cDNA arrays. *Nucleic Acids Res.* **26**, 3883–3891 (1998).
- Luo, L. et al. Gene expression profiles of laser-captured neuronal subtypes. *Nat. Med.* **5**, 117–122 (1999).
- Van Gelder, R.N. et al. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc. Natl. Acad. Sci. USA* **87**, 1663–1667 (1990).
- Eberwine, J.H. et al. Analysis of gene expression in single live neurons. *Proc. Natl. Acad. Sci. USA* **89**, 3010–3014 (1992).
- Kacharina, J.E., Crino, P.B. & Eberwine, J.H. Preparation of cDNA from single cells and subcellular regions. *Methods Enzymol.* **303**, 3–19 (1999).
- Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**, 14863–14868 (1998).
- Matz, M. et al. Amplification of cDNA ends based on template-switching effect and step-out PCR. *Nucleic Acids Res.* **27**, 1558–1560 (1999).
- Mahadevappa, M. & Warrington, J.A. A high-density probe array sample preparation method using 10- to 100-fold fewer cells. *Nat. Biotechnol.* **17**, 1134–1136 (1999).
- Schena, M. et al. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc. Natl. Acad. Sci. USA* **93**, 10614–10619 (1996).
- National Human Genome Research Institute. Division of Intramural Research. *Microarray Project (µAP)*. http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/img_analys.html.