Are Gene Expression Microarray Analyses Reliable? A Review of Studies of Retinoic Acid Responsive Genes

Peter J. van der Spek[1], Andreas Kremer[1], Lynn Murry[2], and Michael G. Walker[2]*
[1] *Janssen Pharmaceutica, Beerse, Belgium;* [2] *Incyte Genomics, Inc., CA 94087, USA*

Microarray analyses of gene expression are widely used, but reports of the same analyses by different groups give widely divergent results, and raise questions regarding reproducibility and reliability. We take as an example recent published reports on microarray experiments that were designed to identify retinoic acid responsive genes. These reports show substantial differences in their results. In this article, we review the methodology, results, and potential causes of differences in these applications of microarrays. Finally, we suggest practices to improve the reliability and reproducibility of microarray experiments.

Key words: microarrays, retinoic acid response gene, reliability case studies

Introduction

Microarrays are widely used to identify genes of putative medical or industrial value, to characterize the possible mode of action of drugs, and in many other applications. However, disinterested researchers have rarely evaluated the reproducibility and reliability of microarray results. Published reports of the same analyses by different researchers often give widely divergent results. We wish to understand the possible sources of these differences, and to identify practices that may improve the quality of microarray studies. We take as an example recent published reports on microarray experiments that were designed to identify retinoic acid responsive genes. Retinoic acid (RA) is an anti-tumor agent that induces differentiation or apoptosis (*1 – 7*). Several retinoic acid related compounds have demonstrated anti-tumor effects, and some are in clinical use or in trials for human therapy (*7 – 14*). Thus, there is good reason for wishing for reliable identification of the genes involved.

In the next section, we give a chronological review of microarray expression experiments to identify retinoic acid responsive genes. In the discussion section, we examine the methodology of microarray experiments exemplified by these studies.

Expression Analyses of RA-responsive Genes

Tamayo and colleagues (*15*) studied acute promyelocytic leukemia NB4 cells stimulated with 1 μM RA for 0, 6, 24, 48, or 72 h. Their study also included comparative analysis of three other cell types (HL-60, U937, and Jurkat) treated with PMA at 0, 0.5, 6, or 24 h. They used the Affymetrix HU6000 array, which contains probes for 6,416 genes, and identified genes that showed a three-fold relative change in expression and a 100-unit absolute change in expression. They confirmed results with Northern blots. In a cluster analysis, they identified a group of 154 genes induced by RA in NB4 cells that were not regulated in the other three cell lines. The differentially expressed genes included neutrophil differentiation markers (granulocyte colony stimulating factor receptor, CD59, and defensin alpha 4), some proteins known to be RA-induced (RIG-E and the interferon-inducible genes IFI56, INP10, and IRF1), and genes not previously linked to RA, including G0S2, and two proteolysis-related genes, LMP7 and UBE1L.

Liu and colleagues (*16*) examined RA-induced genes in the same NB4 cell line as did Tamayo and colleagues. They identified differentially expressed genes using microarrays, subtractive hybridization, and differential-display-polymerase chain reaction. For their analysis, they used Clontech membranes containing 588 genes in triplicate experiments. They found 99 of the 588 genes had greater than two-fold changes in expression; 37 were up-regulated, and 62 were downregulated. They confirmed expression of 8 up-regulated and 2 down-regulated genes by RT-PCR. Subtractive hybridization yielded 703 clones, which assembled into 316 contigs. 159 were known genes, 25 were mitochondrial, ribosomal or ALU repeats, and

* **Corresponding author.**
E-mail: mwalker@stanfordalumni.org

132 were novel fragments. Of these 159, RT-PCR confirmed 63 (55 induced and 8 repressed), and the remainder showed no change in expression or no signal. Differential display yielded 30 bands, of which 13 were subsequently confirmed by Northern blots to be induced genes. Merging the 99 genes from the microarray analysis with the genes identified by subtraction or differential display yielded 169 genes modulated by RA. 8 of these were previously uncharacterized genes. The 100 up-regulated genes included transcription factors (c-JUN, ETR, ID-2, HOX-A1, ACTR), signal transduction genes (JAKs/STAT, cAMP/PKA, PKC, calmodulin) protein modulation (UAE1, SUMO1), apoptosis resistance (DAD-1, Bfl-1, GADD153), cell cycle exit (p19INK, p21WAF), proliferation suppression (BTG1, src-like adapter protein, FGR, LIMK) and neutrophil-related genes (MCP-1, defensin, X-CGD). The 69 down-regulated genes included transcription factors involved in cell proliferation (c-MYC, NF kappa B, GATA2), DNA synthesis proteins, cell cycle checkpoint genes (cyclin A, B1), map kinases (p38, JNk2, ERK3), apoptosis agonists (ICH-1L, FAST kinases) and others. Their study included time-course analysis of gene induction; of 100 up-regulated genes, expression induction was most frequent in the period from 12 to 48 h after treatment, and of the 69 down-regulated genes 59 were suppressed within 8 hours. They further characterized the role of these genes in RA signaling by treatment with cyclohexamide and found that expression of 8 induced and 24 repressed genes was not blocked. Liu compared their results to those of Tamayo, who had used the same treatment of the same cell line, but had used a different array. Liu reports that, of their differentially expressed genes, 48 up-regulated genes and 39 down-regulated genes were present on the Tamayo array. There were similar changes of expression for 42 of 48 up-regulated genes and 20 of 39 down-regulated genes. There were 6 genes with opposite changes between the two studies; for the two discrepant up-regulated genes (CD11A and LIMK1), RT-PCR confirmed Liu's results. All the genes with discrepant results were in the 2-fold to 5-fold range.

Rusiniak and colleagues (*17*) studied the fusion oncoprotein promyelocytic leukemia-retinoic acid receptor alpha (PML-RARalpha) which blocks myeloid differentiation in acute promyelocytic leukemia (APL). They identified genes induced by RA in TF1 myeloid leukemia cells expressing PML-RARalpha that were not induced in TF1-neo control cells. Among the genes they found was B94 (TNFAIP2). They further characterized B94 induction by RA in a series of experiments with other cell lines and examined its expression in APL patients. These studies indicated that B94 plays a role in myeloid development and is a target gene of PML-RARalpha. An article is in preparation with full details of the differentially expressed genes.

Soref and colleagues (*18*) studied RA-induced genes in primary human tracheobronchial epithelial (TBE) cells. They prepared a microarray membrane comprising 30,000 cDNA clones derived from primary human TBE cells that had been cultured for more than 30 d under an air-liquid culture containing all-trans-RA and hormonal supplements. The 30,000-clone array was hybridized with cDNA from RA-treated and untreated cultures of primary human TBE cells. Of 79 clones with 5-fold or greater differential expression, 76 were confirmed by Northern blot. 14 of the 76 clones encoded novel cDNA sequences; 5 of the 14 had identical 3'-ends and were determined to be fragments of a single novel gene, which they named hRDH-TBE. A time course study showed RA-induced expression of hRDH-TBE two hours after treatment. An article is in preparation with details of the differentially expressed clones.

Gupta and colleagues (*19*) identified genes regulated by the nuclear hormone peroxisome proliferator-activator receptor gamma (PPARgamma) agonist in the Moser S (M-S) colon carcinoma cell line that could be regulated independently by a retinoid X receptor specific agonist. The treatments in their study were synthetic PPAR and RXR ligands from GlaxoSmithKline and included rosiglitazone, PPAR-subtype selective agonists and antagonists, and the RXR agonist LG100268. These authors used two microarray platforms. One experiment used Research Genetics GeneFilters microarrays with 5,184 cDNA clones to compare expression in M-S cells treated for 24 h or 6 d to either rosiglitazone or a control. The second experiment used Affymetrix oligonucleotide Human GeneFL GeneChip arrays with 5,600 unique genes to compare expression in M-S cells exposed for 24 h to either the PPARgamma ligand GW7845 or a control. They reported that the cDNA filters gave a large number of false positives and were not used further other than to aid interpretation of the oligonucleotide array data. The Affymetrix software determined that 89 of 5,600 genes (1.6%) were induced or repressed greater than 2.5 fold by the PPARgamma ligand. Of these 89, 14 were selected for further confirmation; only 9 of the 14 were con-

firmed by Northern blot hybridization. Of these 9, the up-regulated genes were adipophilin, liver fatty acid binding protein (L-FABP), neutrophil gelatinase-associated lipocalcin (NGAL), keratin 20, carcinoembryonic antigen (CEA), non-specific cross-reacting antigen (NCA) and biliary glycoprotein (BGP), the latter three being members of the CEA family. The down-regulated genes were regenerating protein 1A (Reg1A) and gob-4. Further experiments showed that, of the nine genes, only the three CEA family members were not regulated by the RXR-specific agonist LG100268.

Kerley and colleagues (*20*) studied gene expression during RA-induced differentiation of human embryonal carcinoma cells. Among the regulated genes they found induction of the nuclear receptor coregulator receptor interacting protein 140 (RIP140) within three hours of treatment. They further characterized the gene by determining that a construct comprising a luciferase gene reporter with RIP140 upstream regulatory elements supported RA-induced transcription and that overexpression of RIP140 repressed ligand-activated retinoid receptors. An article is in preparation with full details of the differentially expressed genes.

Udea (*21*) constructed a microarray comprising 2007 cDNAs from a primary neuroblastoma library and identified RA-regulated genes in a RTBM1 neuroblastoma cell line. In two replicates, 7 genes were upregulated and 5 genes were down regulated. The results for 6 genes were confirmed by semi-quantitative reverse transcriptase PCR analysis. Of these 6 genes, amyloid precursor-like protein 2 (APLP2), P311, dihydropyrimidinase related protein 3 (DRP3) and RGP4 were upregulated, and Id-2 and vimentin were downregulated.

In a collaboration among the present authors, we examined genes responding to two concentrations of RA in two cell types at two time points. We compared the expression levels of genes in a bone-marrow-derived cell line (from a metastatic bone marrow neuroblastoma) and in a fibroblast-derived cell line cells (from breast fibroblast tissue) before and after treatment with RA. The neuroblastoma samples were either untreated or treated with 10 μM RA and cultured for 4 d. The fibroblast samples were either untreated or treated with 1 μM RA for 20 h. Approximately 5000 cDNA's from each library were sequenced by gel electrophoresis. We found 13 genes expressed at high levels in the RA-treated libraries but not detected in the untreated libraries. Of these 13 genes, eight are known to be associated with RA and differentiation: P450 Retinoic Acid Induced 1, retinoic acid hydroxylase, laminin, retinol-binding-protein receptor p63, lamin A, beta-2-microglobulin, and APLP2. Five differentially expressed genes were not previously linked to RA. Two are related to previously uncharacterized cDNA clones (Genbank accession numbers NM_022749 and AB033050). The three known genes not previously linked to RA are integrase interactor 1a (INI1A) protein, a tumor suppressor for rhabdoid tumors (*22 – 27*); endosialin/tumor endothelial marker 1 (TEM1), a possible receptor for unknown ligands that is expressed in tumor blood vessel endothelium (*28 – 30*); and Mlx, a transcription factor associated with cell cycle progression (*31*). Retinoic-acid-binding protein was detected at low levels in the untreated libraries and significantly upregulated in the RA-treated libraries. Vimentin and fibronectin, both known to be associated with RA, were greatly downregulated in the RA-treated libraries in comparison with the untreated libraries.

Discussion

The genes identified as responsive to retinoic acid are, to a large degree, different in each experiment. Many genes identified as differentially expressed by microarray analysis are not confirmed by RT-PCR, Taqman or Northern blot analyses. The experiments of Gupta, Soref, Chaib (*18, 19, 32*) and other researchers show that microarray results, and in particular quantitative estimates of changes in expression, are only partially consistent with RT-PCR, Taqman or Northern blot analyses. In Chaib's study of differences of expression of 588 genes in normal versus malignant prostate, microarray analysis identified 15 genes that changed expression greater than two-fold, but RT-PCR verified the fold difference for only 6 of the 15 and Northern blot for 7 of 8 genes examined. Such results support a view that microarray experiments are not reproducible and are not reliable. Two questions arise: 1) Can we account for these differences in results among these studies, and 2) are there practices that can make microarray experiments reproducible and reliable?

Genes that change expression in response to a stimulus in one cell line or tissue type may show no change in an experiment with another, closely related type, because different receptors, transcription factors, and other factors are present. The experiments reviewed here did not always use the same tissue type

or cell line, and those that used the same cells and the same treatment still produced significant differences in results. These differences may be ascribed to culture conditions, diurnal influences such as feeding time, or other factors. Differences in mRNA stability may cause differences in mRNA detection if the time between sample collection and sample assay differs between experiments (*33*, *34*).

Differences in the threshold for declaring a change in expression (such as 2-fold, 2.5-fold, 3-fold, or 5-fold) may contribute substantially to differences in results. The widespread practice of using an arbitrary threshold of fold change in expression (2-fold, 3-fold or higher) is probably inappropriate. Consider a gene whose measured expression varies greatly in replicates (for example, from no change to 3-fold change) versus a gene whose measured expression is consistent across replicates (for example, consistently in the range 1.7-fold to 1.9-fold change). The former gene may pass the 2-fold threshold in some replicates, but not be a reliable indicator of the differences resulting from the treatment, while the latter fails the 2-fold threshold, but is a more reliable indicator of the differences resulting from the treatment. An arbitrary threshold of 2-fold change is likely to yield abundant false positive and false negative results. Arbitrary fold differences should be replaced by measures of statistical significance using replicates. Unfortunately, microarray studies often do not include replicates. Few studies compare the variation between replicates to the variation across treatments to determine if the supposed differences in expression are statistically significant (for example with a t-test or analysis of variance). In part, the lack of replicates may be attributable to the expense of microarrays, but the expense of replicates is small compared to the expense of months of research effort wasted because of spurious microarray results.

Some differences in results are due to differences in the genes on the arrays, particularly if the genes were selected for likely response to the stimulus, as in the study by Soref (*18*). Some differences are likely due to differences in the array technologies (cDNA versus synthesized oligomers or glass versus membrane). There is substantial variability in the quality control of microarray manufacturing and measurement. Academic and pharmaceutical researchers who have manufactured their own microarrays give anecdotal reports of serious difficulties in ensuring quality and reproducibility. For example, many brands of microscope slides commonly used to prepare arrays to fluoresce in the absence of any DNA at the wavelengths used to detect DNA hybridisation. Similar anecdotal reports from commercial manufacturers indicate that substantial resources are necessary to ensure quality. Sequence verification of cDNA arrays has shown that up to 30% of the supposed sequences on an array may not, in fact, be the sequence of the desired gene, but are instead from some other gene, or possibly even from some other species. Errors in clone labeling and retrieval from cold storage, errors in presenting 96-well plates to a robot, contamination, and other sources contribute to this problem.

Differences may be introduced by the use of different reference standards in two-color arrays. The current standard practice for two-color arrays is to compare mRNA levels in the experimental samples in one channel of the array (cy3) to a reference standard, often made pooling several samples (for example, from cell lines), in the second channel (cy5). If particular genes are expressed at low levels, or not at all, in the reference standard, then the apparent relative expression of those genes can be quite variable, or in some cases not measurable. A universal reference standard, made by pooling mRNA from diverse tissues, is sometimes proposed. This proposal is not necessarily a good solution, because detection of subtle changes in expression is more reliable if gene expression in the reference standard is very similar to that in the experimental sample. For example, subtle differences between normal and neoplastic tissues are more reliably detected by direct comparison of the two samples rather than by comparison of each against a dissimilar expression pool.

Cross-hybridization occurs commonly between genes with greater than 80 percent identity to probe cDNA, so a positive result may be due to a homologous gene. Splice variants can also yield false negative results if the splice variant does not share a substantial region with the DNA on the array.

The experiments by Liu, by Soref (*16*, *18*) and by other researchers show that expression varies substantially with time following a treatment, but microarray studies often do not examine expression changes at multiple time points. Without temporal data, it is difficult to have confidence that a microarray analysis correctly identifies more than a small portion of a regulatory pathway or to estimate with any confidence the quantitative connections among the genes.

A further difficulty arises during the data analysis. In an experiment involving tens of thousands of genes measured in only a handful of samples, we would ex-

pect many genes to show apparent differences just by chance. Adjustment for multiple comparisons, such as a Bonferroni correction, is desirable, to reduce the large number of false-positive results.

The substantial differences in results lead to legitimate concerns about the validity of the microarray approach. The current best practices would appear to be to verify the sequence of clones on the array, select sequences to recognize splice variants and homologs, and ensure quality control in manufacturing of the array. Samples should be collected and analysed at multiple time-points after treatment, with replicates that allow determination of the statistically significant differences in expression rather than an arbitrary fold difference, corrected for multiple comparisons, and confirmed with other expression measurement methods. With appropriate good practices, we can use microarrays to help determine the true functions of genes.

Acknowledgements

We thank our colleagues at Janssen Phamaceutica and Incyte Genomics for their contributions and discussions of this work.

References

1. Akie, K., *et al.* 2000. A combination treatment of c-myc antisense DNA with all-trans-retinoic acid inhibits cell proliferation by downregulating c-myc expression in small cell lung cancer. *Antisense Nucleic Acid Drug Dev.* 10: 243-249.
2. Eck-Enriquez, K., *et al.* 2000. Pathways through which a regimen of melatonin and retinoic acid induces apoptosis in MCF-7 human breast cancer cells. *Breast Cancer Res. Treat.* 61: 229-239.
3. Hansen, L.A., *et al.* 2000. Retinoids in chemoprevention and differentiation therapy. *Carcinogenesis* 21: 1271-1279.
4. Manna, S.K. and Aggarwal, B.B. 2000. All-trans-retinoic acid upregulates TNF receptors and potentiates TNF-induced activation of nuclear factors-kappaB, activated protein-1 and apoptosis in human lung cancer cells. *Oncogene* 19: 2110-2119.
5. Martinet, N., *et al.* 2000. Retinoic acid receptor and retinoid X receptor alterations in lung cancer precursor lesions. *Cancer Res.* 60: 2869-2875.
6. Okamoto, K., *et al.* 2000. Differences in uptake and metabolism of retinoic acid between estrogen receptor-positive and -negative human breast cancer cells. *Cancer Chemother. Pharmacol.* 46: 128-134.
7. Zhang, D., *et al.* 2000. Retinoids and ovarian cancer. *J. Cell Physiol.* 185: 1-20.
8. De Coster, R., *et al.* 1996. P450-dependent enzymes as targets for prostate cancer therapy. *J. Steroid Biochem. Mol. Biol.* 56: 133-143.
9. Douer, D., *et al.* 2001. Treatment of newly diagnosed and relapsed acute promyelocytic leukemia with intravenous liposomal all-trans retinoic acid. *Blood* 97: 73-80.
10. Goncalves, A., *et al.* 2001. Phase II study of a combination of cisplatin, all-trans-retinoic acid and interferon-alpha in squamous cell carcinoma: clinical results and pharmacokinetics. *Anticancer Res.* 21: 1431-1437.
11. Krekels, M.D., *et al.* 1997. Induction of the oxidative catabolism of retinoid acid in MCF-7 cells. *Br. J. Cancer* 75: 1098-1104.
12. Krekels, M.D., *et al.* 1996. Analysis of the oxidative catabolism of retinoic acid in rat Dunning R3327G prostate tumors. *Prostate.* 29: 36-41.
13. Lawrence, J.A., *et al.* 2001. Phase I clinical trial of alitretinoin and tamoxifen in breast cancer patients: toxicity, pharmacokinetic, and biomarker evaluations. *J. Clin. Oncol.* 19: 2754-2763.
14. Miller, W.H., Jr., *et al.* 2000. A phase I-II study of 9-cis retinoic acid and interferon-alpha2b in patients with advanced renal-cell carcinoma: an NCIC Clinical Trials Group study. *Ann. Oncol.* 11: 1387-1389.
15. Tamayo, P., *et al.* 1999. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci. USA* 96: 2907-2912.
16. Liu, T.X., *et al.* 2000. Gene expression networks underlying retinoic acid-induced differentiation of acute promyelocytic leukemia cells. *Blood* 96: 1496-1504.
17. Rusiniak, M.E., *et al.* 2000. Identification of B94 (TNFAIP2) as a potential retinoic acid target gene in acute promyelocytic leukemia. *Cancer Res.* 60: 1824-1829.
18. Soref, C.M., *et al.* 2001. Characterization of a novel airway epithelial cell-specific short chain alcohol dehydrogenase/reductase gene whose expression is up-regulated by retinoids and is involved in the metabolism of retinol. *J. Biol. Chem.* 276: 24194-24202.
19. Gupta, R.A., *et al.* 2001. Target genes of peroxisome proliferator-activated receptor gamma in colorectal cancer cells. *J. Biol. Chem.* 276: 29681-29687.
20. Kerley, J.S., *et al.* 2001. Transcriptional activation of the nuclear receptor corepressor rip140 by retinoic acid: a potential negative-feedback regulatory mechanism. *Biochem. Biophys. Res. Commun.* 285: 969-975.
21. Ueda, K. 2001. Detection of the retinoic acid-regulated genes in a RTBM1 neuroblastoma cell line using cDNA microarray. *Kurume Med. J.* 48: 159-

164.

22. Biegel, J.A., *et al.* 1999. Germ-line and acquired mutations of INI1 in atypical teratoid and rhabdoid tumors. *Cancer Res.* 59: 74-79.
23. DeCristofaro, M.F., *et al.* 1999. Alteration of hSNF5/INI1/BAF47 detected in rhabdoid cell lines and primary rhabdomyosarcomas but not Wilms' tumors. *Oncogene* 18: 7559-7565.
24. Manda, R., *et al.* 2000. Absence of hSNF5/INI1 mutation in human lung cancer. *Cancer Lett.* 153: 57-61.
25. Sevenet, N., *et al.* 1999. Spectrum of hSNF5/INI1 somatic mutations in human cancer and genotype-phenotype correlations. *Hum. Mol. Genet.* 8: 2359-2368.
26. Sevenet, N., *et al.* 1999. Constitutional mutations of the hSNF5/INI1 gene predispose to a variety of cancers. *Am. J. Hum. Genet.* 65: 1342-1348.
27. Suzuki, A., *et al.* 1997. Gene expression of malignant rhabdoid tumor cell lines by reverse transcriptase-polymerase chain reaction. *Diagn. Mol. Pathol.* 6: 326-332.
28. Christian, S., *et al.* 2001. Molecular cloning and characterization of endosialin, a C-type lectin-like cell surface receptor of tumor endothelium. *J. Biol. Chem.* 276: 7408-7414.
29. Opavsky, R., *et al.* 2001. Molecular characterization of the mouse Tem1/endosialin gene regulated by cell density in vitro and expressed in normal tissues in vivo. *J. Biol. Chem.* 276: 38795-38807.
30. St Croix, B., *et al.* 2000. Genes expressed in human tumor endothelium. *Science* 289: 1197-1202.
31. Meroni, G., *et al.* 2000. Mlx, a new Max-like bHLHZip family member: the center stage of a novel transcription factors regulatory pathway? *Oncogene* 19: 3266-3277.
32. Chaib, H., *et al.* 2001. Profiling and verification of gene expression patterns in normal and malignant human prostate tissues by cDNA microarray analysis. *Neoplasia* 3: 43-52.
33. Johnson, M.A., *et al.* 2000. Mutants of Arabidopsis defective in a sequence-specific mRNA degradation pathway. *Proc. Natl. Acad. Sci. USA* 97: 13991-13996.
34. Ohme-Takagi, M., *et al.* 1993. The effect of sequences with high AU content on mRNA stability in tobacco. *Proc. Natl. Acad. Sci. USA* 90: 11811-11815.

Received: 28 December, 2002
Accepted: 13 January, 2003