Quantitative PCR Deconstruction of Discrepancies between Results Reported by Different Hybridization Platforms

Federico M. Goodsaid, Roger J. Smith, and I.Y. Rosenblum

Division of Drug Safety and Metabolism, Schering-Plough Research Institute, Lafayette, New Jersey, USA

Differences in hybridization platforms used in gene array analysis experiments can lead to significant differences in hybridization results. In this study we used quantitative reverse transcription–polymerase chain reaction (qRT–PCR) to investigate discrepant results between the National Institute of Environmental Health Sciences cDNA and Affymetrix oligo platforms used to evaluate hepatic gene expression changes in rats exposed to methapyrilene. Caldesmon cDNA platform hybridization results showed decreases in gene expression levels for the high-dose methapyrilene 7-day pooled samples compared with their controls. By contrast, the Affymetrix oligonucleotide platform showed increases in expression levels for these samples. Quantitative gene expression measurements provide an explanation for the discrepancies observed for these samples. In the case of caldesmon, there is a 74-base sequence in the cDNA clone that is absent in the Affymetrix sequence. The amplicon based on the cDNA clone shows > 100-fold suppression relative to the day 7 high-dose methapyrilene-pooled control. These data demonstrate the importance of using a "gold standard," such as qRT-PCR to confirm key hybridization results as well as to understand the sources of discrepancies resulting from different hybridization platforms.
Key words: **cDNA arrays, gene expression, oligo arrays, quantitative real-time PCR, toxicogenomics.** ***Environ Health Perspect*** **112:456–459 (2004). doi:10.1289/txg.6695 available via *http://dx.doi.org/* [Online 15 January 2004]**

Several possible sources of discrepancies between hybridization results can be identified using different gene array platforms. Among these are differences in low-end sensitivity, dynamic range, and linearity (Aach et al. 2000). Such differences are likely to influence the magnitude of gene expression but not account for qualitative shifts from induction to repression reported by two different hybridization platforms. Discrepancies in the absolute values reported for induction or downregulation may simply result in differences in the statistical significance of a set of hybridization results, but a qualitative shift from induction to repression in the results reported from two different hybridization platforms can lead to an incorrect biological interpretation (Kuo et al. 2002).

A cDNA hybridization platform (NIEHS; National Institute of Environmental Health Sciences, Research Triangle Park, NC) and the Affymetrix (Santa Clara, CA; http://www.affymetrix.com/) synthetic oligo hybridization platforms were used in a series of rat toxicogenomic studies coordinated by the International Life Sciences Institute (ILSI) Health and Environmental Sciences Institute (HESI) (Ulrich et al. 2004). One of these studies evaluated the effects of methapyrilene on rat hepatic gene expression as measured by both cDNA and synthetic oligo hybridization platforms. The results of the NIEHS cDNA platform for caldesmon in these studies showed a decrease (Hamadeh et al. 2002) in gene expression levels for the high-dose methapyrilene 7-day pooled samples compared with their controls. However, the Affymetrix platform showed increases (Waring et al. 2004) in expression levels for caldesmon in these samples.

Quantitative reverse transcription–polymerase chain reaction (qRT–PCR) is a gene expression measurement platform that can be used to independently assess discrepancies between different hybridization platforms reported for specific genomic sequences. This platform is considered a "gold standard" for its analytical sensitivity, dynamic range, and linearity of gene expression measurements (Bustin 2000; Dooley et al. 2003). Multiple primer pairs were designed to probe specific sections of genomic sequences for caldesmon to help identify the reason for the discrepant hybridization results. We show here the results of our analysis with qRT–PCR.

Materials and Methods

RNA samples and conversion to cDNA. Pooled RNA samples were obtained from the methapyrilene study as described in Waring et al. (2004). Total RNA was converted to cDNA with the RT–PCR Miniprep Kit (Applied Biosystems, Palo Alto, CA).

Real-time PCR analysis of gene expression. Primers (Table 1) for the high efficiency detection of rat caldesmon mRNA and 18S ribosomal RNA were designed using Primer Express (Applied Biosystems, Foster City, CA). Primer pairs were selected with a single peak in the melting curve of the resulting PCR products. Real-time PCR data and analyses were collected on an Applied Biosystems 7700 Sequence Detection System instrument. These data were analyzed using the comparative C_T method as described in the instructions of the ABI PRISM 7700 User Bulletin #2 (P/N 4303859) from Applied Biosystems. The amount of each amplicon was normalized to an endogenous control (18S RNA) whose expression was proportional to the total amount RNA in each sample.

Results

The UniGene (http://www.ncbi.nih.gov/UniGene/) sequence for caldesmon used in the assembly of the NIEHS cDNA chip is shown in Figure 1. The NIEHS chip used in the studies reported by Ulrich et al. (2004) included other sequences identified by the same UniGene number Rn.33965 and NIEHS accession number AI044091. Sequences with this UniGene accession number are shown in Table 2 with their corresponding unique Affymetrix ID (http://www.affymetrix.com), reference sequences and descriptions, and their GenBank accession numbers (http://www.ncbi.nih.gov/GenBank). The GenBank accession numbers in this table corresponding to caldesmon are AL180288 and U18419. The other GenBank accession numbers listed in this table correspond to rat alpha-2 microglobulin (X14552, J00738). Each Affymetrix reference sequence aligns with the sequence corresponding to its GenBank accession number, and no homology was found between sequences corresponding to these accession numbers (data not shown).

The UniGene accession number sequence AI044091 differs from U18419 in the addition of 74 bp to the original U18419 sequence between positions 276 and 351 (Figure 1). This 74-bp sequence

This article is part of the mini-monograph "Application of Genomics to Mechanism-Based Risk Assessment."

Address correspondence to F.M. Goodsaid, Schering-Plough Research Institute, 144 Rt. 94, PO Box 32, Lafayette, NJ 07848 USA. Telephone: (973) 940-4612. Fax: (973) 940-4183. E-mail: federico.goodsaid@spcorp.com

The authors declare they have no competing financial interests.

Received 22 August 2003; accepted 15 December 2003.

is completely homologous with the alpha-2 microglobulin accession number X14552 and J00738 sequences. qRT-PCR of an amplicon overlapping this 74-bp sequence, as well as for the alpha-2 microglobulin sequences X14552 and J00738, showed a substantial reduction (> 100-fold) in message levels for the day 7 high-dose methapyrilene sample, whereas a similar analysis of amplicons designed for the amplification of U18419 and AL180288 showed induction levels above 5-fold (Figure 2). Data for hybridization (Waring et al. 2004) and for qRT-PCR were collected from RNA samples from pooled replicate rats. An arbitrary threshold value of 1.5 was used for the analysis of the qRT-PCR data.

Figure 3 shows the baseline expression for these amplicons relative to U18149. A comparison of this type is possible for amplicons such as these designed for similar PCR efficiencies and amplified in similar chemical matrices. These data show differences of 10- to 100-fold for the alpha-2 microglobulin amplicons compared with the caldesmon amplicons.

Discussion

The analysis of discrepancies in hybridization results between platforms requires knowledge of the original sequences used on each chip platform as well as independent confirmation using a gold standard such as quantitative PCR. The results of this analysis for the caldesmon discrepancy associated with the studies reported by Ulrich et al. (2004) show how these differences may be resolved. In this case the inclusion of a 74-bp sequence in the UniGene caldesmon sequence (Figure 1) accounted for the measurement of an incorrect caldesmon message. Although this sequence analysis showed how the sequences differed from each other, further analysis with qRT-PCR demonstrated how a discrepant result between expression changes measured by the cDNA and Affymetrix platforms was actually obtained.

qRT-PCR data in Figures 2 and 3 suggest a possible explanation for the discrepancy between the NIEHS and Affymetrix platform results for the day 7 high-dose methapyrilene sample. Results in Figure 2 indicate that the message level for the amplicon corresponding to the 74-bp sequence was reduced by > 100-fold compared with its baseline expression level. This reduction in message level was also observed for the amplicons in the alpha-2 microglobulin accession number X14552 and J00738 sequences. However, amplicons in the caldesmon sequences with accession numbers U18419 and AL180288 both show detectable levels (> 1.5-fold) of induction as a result of methapyrilene treatment.

Results in Figure 3 show that baseline expression levels for the 74-bp and alpha-2 microglobulin amplicons were 10- to 100-fold higher than the corresponding levels for the caldesmon amplicons. These data predict that the NIEHS cDNA chip would show a significant reduction in expression level for the hybridization of caldesmon with the 74-bp additional sequence as well as for the alpha-2 microglobulin sequences that were incorrectly labeled in this platform as caldesmon (Figure 3). Not only would we expect a significant reduction in expression levels, but this reduction would also occur from baseline expression levels at least an order of magnitude higher than those for the true caldesmon sequence.

Affymetrix caldesmon probe sequences were homologous with sequences for accession numbers U18419 and AL180288. These sequences correctly probe expression levels for the caldesmon gene. Affymetrix probes for the alpha-2 microglobulin sequences with accession numbers X14552 and J00738 were reported (Waring et al. 2004) to have reduced expression levels (5- and 7-fold, respectively) for the day 7 high-dose methapyrilene sample compared with its baseline expression levels.

Other artifacts can lead to discrepancies between the results reported by different hybridization platforms. The studies reported by Ulrich et al. (2004) had a similar discrepant result for PCTAIRE expression. This particular result was traced to a UniGene sequence for this gene that was the reverse complement of the original sequence (data not shown). In this example where the sequence is completely reversed, accurate hybridization results would not be possible.

The results presented by Ulrich et al. (2004) reported a discrepant call level of approximately 5% between the Affymetrix and cDNA platforms. Analysis of all of these discrepancies is beyond the scope of the present report. However, it is clear that such an analysis is essential at least for individual gene expression results that are used, for example, for the assembly of mechanistic models of toxicity or for the identification of genomic biomarkers of toxicity.

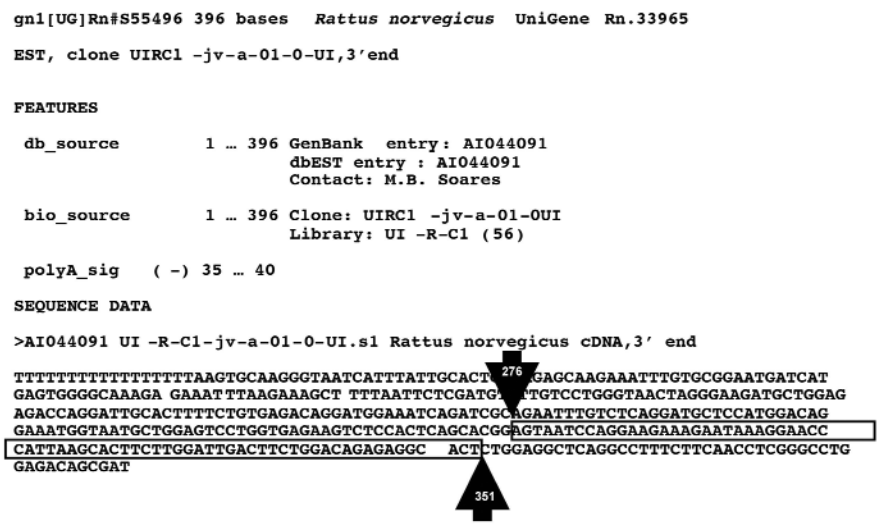

Figure 1. Sequence for UniGene accession number AI044091 (UniGene Rn.33965). The 74-bp sequence insert highlighted in this sequence runs from 276 to 351 bp. From UniGene (http://www.ncbi.nih.gov/UniGene/).

Table 1. Sequences for primers of amplicons tested.

GenBank accession no.[a]	Sequence	Forward primer	Reverse primer
AI044091	UniGene caldesmon	CCATTAAGCACTTCTTGGATTGACT	GCCTCCAGAGTGCCTCTCTGT
AI180288	Caldesmon	TCATGTTTCTGCACTTTAGACTAAAGC	TGCCTACTTAGATAATTTCTTCCATGCT
U18419	Caldesmon	CACTGACCACGTCCATTCCTT	CTGGTCAGTGCATGCGTTTATAA
X14552	Alpha-2-microglobulin	GACCCCTCTTTCCCATTTCC	GACTTTCATACATTGCCTGAGTGAAG
J00738	Alpha-2-microglobulin	AGTGCCTCTCTGTCCAGAAGTCA	GAACCCATTAAGCACTTCTTGGAT

[a]From GenBank (http://www.ncbi.nih.gov/GenBank/).

Table 2. Affymetrix sequences with caldesmon UniGene accession number.

UniGene ID[a]	NIEHS accession no.[b]	Affymetrix ID[c]	GenBank accession no.[d]	Affymetrix sequences	Affymetrix description
Rn.33965	AI044091	rc_AI180288_s_at	AI180288	CAGACATCATGTTTCTGCACTTTAGACTAA AGCATGGAAGAAATTATCTAAGTAGGCAA TCAAAATTCTCTGAAAGTGATCCACTTCAG ATCTGATATAGGGCAGTGATGATTGTCTTT TTTTTAAAAAAGAAGATGTACTGTTGACAT ATTGCTTTTCTTCTATGCTGATTCATACCTA GATTGGGTGATTATTTTAGCTGACAGTGGT ACTGATTTTTTTCTTCAGGTTAGTTGCTTTG TGGATTTCTCTGGT	EST224031 *Rattus norvegicus* cDNA, 3′ end /clone=RSPCS84 /clone_end=3′ /gb=AI180288 /µg=Rn.10621 /len=417
Rn.33965	AI044091	U18419_at	U18419	GGTTATAGCGCATGCACTGACCACGTCCAT TCCTTAACGCTGAGGTTATAAACGCATGCA CTGACCAGCTCCATTCCTTAACCCTGAGGT TATAGCGAATGCACTGACCAGCTCCATTCC TTAACGCTGAGGTTATAGCGCATGCACTGA CCAGCTCCATTCCTTAACCCTGAGGTTATA GCGCATGCACTGACCAGCTCCATTCCTTAA CCCTGAGGTTATAGCGCATGCACTGACCAG CTCCATTCCTTAACG	*Rattus norvegicus* nonmuscle caldesmon mRNA, complete cds /cds=(723,2318) /gb=U18419 /gi=622966 /µg=Rn.10621 /len=5541
Rn.33965	AI044091	X14552_at	X14552	GTCTGGAGAGCACACTCCTCTGACCCCTCT TTCCCATTTCCCAAGACTTCACTCAGGCAA TGTATGAAAGTCTTTAAAAGTGGCAAGGTT TTCACCATTATTCCTCAAGGCAATGACCAT TCTTCAGAGCTTCTTATGCCGAAGGTTGTG GAAACAAGCCTCACCTTTCGTTACTTCATT TTTCATAGGCCTTCCATAAGGAAAGAGTCA TTTATTCTATGCCTTTCCTCCCTGTTTTCTG ACAAATAATGTT	Rat salivary gland mRNA for (alpha)2(mu) globulin, type 1 /cds=(54,593) /gb=X14552 /gi=55569 /µg=Rn.10203 /len=1710
Rn.33965	AI044091	J00738_s_at	J00738	AGAGTGCCTCTCTGTCCAGAAGTCAATCCA AGAAGTGCTTAATGGGTTCCT	*Rattus norvegicus* submaxillary gland alpha-2µ globulin mRNA, complete cds /cds=(58,603) /gb=J00738 /gi=204262 /µg=Rn.10203 /len=1003

[a]From UniGene (http://www.ncbi.nih.gov/UniGene/). [b]From NIEHS (http://dir.niehs.nih.gov/microarray/). [c]From Affymetrix (http://www.affymetrix.com). [d]From GenBank (http://www.ncbi.nih.gov/GenBank/).

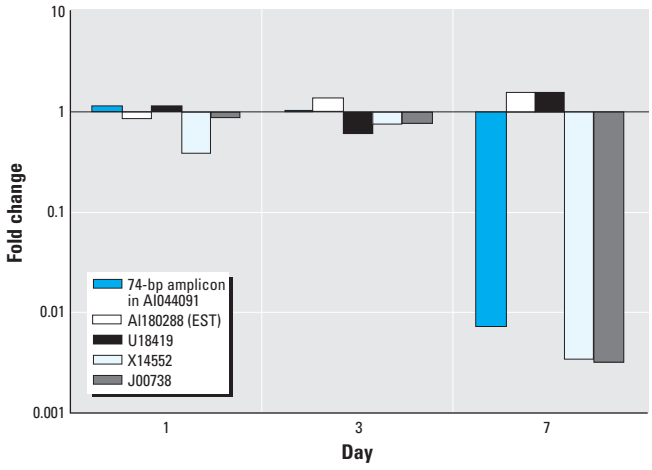

Figure 2. Relative quantification of amplicons tested. The graph shows the fold change relative to basal expression levels in pooled rat liver RNA samples for the high-dose methapyrilene-pooled rat liver RNA samples from Waring et al. (2004) for quantitative gene expression of the amplicons defined by the primers in Table 2. Accession numbers are from GenBank (http://www.ncbi.nih.gov/GenBank/).

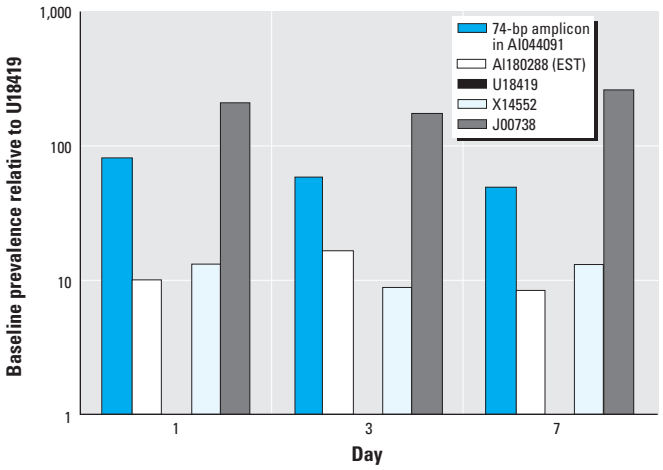

Figure 3. Normalized baseline expression of amplicons tested. The graph shows the baseline gene expression levels sampled for amplicons defined by the primers in Table 2. These data have been normalized to 18S total RNA and plotted relative to the baseline expression levels for U18419. Accession numbers are from GenBank (http://www.ncbi.nih.gov/GenBank/).

References

Aach J, Rindone W, Church GM. 2000. Systematic management and analysis of gene expression data. Genome Res 10:431–445.

Bustin SA. 2000. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. J Mol Endocrinol 25:169–193.

Dooley TP, Reddy SP, Wilborn TW, Davis RL. 2003. Biomarkers of human cutaneous squamous cell carcinoma from tissues and cell lines identified by DNA microarrays and qRT-PCR. Biochem Biophys Res Comm 306:1026–1036.

Hamadeh HK, Knight BL, Haugen AC, Sieber S, Amin RP, Bushel P, et al. 2002. Methapyrilene toxicity: anchorage of pathologic observations to gene expression alterations. Toxicol Pathol 30:470–482.

Kuo WP, Jenssen T-K, Butte AJ, Ohno-Machado L, Kohane IS. 2002. Analysis of matched mRNA measurements from two different microarray technologies. Bioinformatics 18:405–412.

Ulrich RG, Rockett JC, Gibson GG, Pettit SD. 2004. Overview of an interlaboratory collaboration on evaluating the effects of model hepatotoxicants on hepatic gene expression. Environ Health Perspect 112:423–427.

Waring JF, Ulrich RG, Flint N, Morfitt D, Kalkul A, Staedtler F, et al. 2004. Interlaboratory evaluation of rat hepatic gene expression changes induced by methapyrilene. Environ Health Perspect 112:439–448.