DRUG DISCOVERY
AND GENOMIC TECHNOLOGIES

Research Report

Hybridization Cross-Reactivity within Homologous Gene Families on Glass cDNA Microarrays

BioTechniques 31:1182-1192 (November 2001)

E.M. Evertsz, J. Au-Young, M.V. Ruvolo, A.C. Lim, and M.A. Reynolds
Incyte Genomics, Palo Alto, CA, USA

ABSTRACT

Glass cDNA microarrays can be used to profile the expression of thousands of gene targets in a single experiment. However, the potential for hybridization cross-reactivity needs to be considered when interpreting the results. Here, we describe hybridization experiments with a model array representing four distinct functional classes (families): chemokines, cytochrome P-450 isozymes, G proteins, and proteases. The cDNA clones selected for this array exhibited pairwise sequence identities ranging from 55% to 100%, as determined by a homology scoring algorithm (LALIGN). Targets for microarraying were amplified by PCR and spotted in 4-fold replication for signal averaging. One designated target from each family was further amplified by PCR to incorporate a T7 promoter sequence for the production of synthetic RNA transcripts. These transcripts were used to generate fluorescent hybridization probes by reverse transcription at varying input concentrations. As expected, hybridization signals were highest at the matching target elements. Targets containing less than 80% sequence identity relative to the hybridization probe sequences showed cross-reactivities ranging from 0.6% to 12%. Targets containing greater than 80% identity showed higher cross-reactivities (26%–57%). These cross-reactive signals were analyzed for statistical correlation with the length of sequence overlap, percent sequence identity, and homology score determined by LALIGN. Overall, percent sequence identity was the best predictor of hybridization cross-reactivity. These results provide useful guidelines for interpreting glass cDNA microarray data.

INTRODUCTION

The natural progression of sequencing entire genomes has been to develop parallel hybridization technologies to profile gene expression in biological tissues. These include glass cDNA microarrays that are capable of analyzing the expression of thousands of genes in a single hybridization experiment (3,14, 15,17). In this approach, cDNA clones selected from mRNA expression libraries are amplified by PCR and spotted robotically onto a chemically modified glass surface. Experimental RNA samples are labeled by reverse transcription with the incorporation of fluorescent dyes for hybridization onto the array. Hybridization experiments typically employ a two-color format for the analysis of differential gene expression, for example, between diseased versus normal tissues (5). This technique has been used successfully in human, plant, and microbial gene expression studies (10,11,13). Studies have been reported using large compendiums of microarray expression data for gene target validation and the elucidation of disease-relevant gene pathways (2,9,16).

The purpose of the present study was to establish guidelines for interpreting microarray results where spotted cDNA targets are known or suspected to share moderate to high sequence homology. This problem becomes increasingly important as gene content is expanded toward genome-wide expression analysis. There are numerous examples of functionally related genes sharing high sequence homology within and between genomes (4,6). In addition, an increasing number of alternatively spliced gene transcripts have been identified in the course of sequencing and aligning expressed sequence tags (12), demonstrating that cDNAs derived from mRNA transcripts can also share exact identity over a portion of their sequence.

To our knowledge, no specific cross-hybridization data on microarrays has been published to date. However, hybridization experiments on nylon membranes have shown that cross-reactivity is possible when two different gene targets share 77%–100% sequence identity (19). It has also been stated that cross-reactivity can become significant on glass cDNA microarrays when gene targets share greater than 75% sequence identity (18). Here, we describe a series of cross-hybridization experiments with cDNA sequences selected from four different functional classes and sharing 55%–100% sequence identities.

MATERIALS AND METHODS

Clone Nomination and Homology Scoring

Six gene families were selected for cross-hybridization experiments: chemokines, cytochrome P450 (CYP), G protein gamma subunits, and serine, threonine, and metalloproteases. These genes function primarily in the areas of signal transduction and drug metabolism and have been widely studied as potential drug targets and/or toxicity markers. Representative sequence-verified cDNA clones (Incyte Genomics,

Palo Alto, CA, USA) were queried by BlastN sequence comparison against other cDNA sequences sharing the same class of protein functional annotation. Clones sharing greater than 50% sequence identity respectively by BlastN were then selected for hybridization experiments. These clones were further scored for sequence homology using a pairwise sequence alignment algorithm, LALIGN (8). Corresponding GenBank® accession nos. were determined by matching the 5′ sequence read of each clone against the top hit in GenBank 117 (BlastN). Each clone contained the full-length coding insert, with the exception of the CYP family. The CYP family included examples of alternatively spliced genes. Twelve partial insert clones represented eight CYP homologs, three of which comprised splice variants.

To investigate hybridization cross-reactivity, an individual cDNA clone was selected from each gene family and designated as parent. This clone was used to generate a probe sequence for comparison against each cDNA clone sequence in the family. Pairwise homology scores were then calculated using a Smith-Waterman alignment tool, LALIGN (8). We decided to use this algorithm because it provides scores that represent both percent sequence identity and length of sequence alignment, and we hypothesized that such scores might provide a better prediction of cross-hybridization than sequence identity scoring alone. The reported alignment scores all exhibited greater than 50% sequence identity.

Preparation of PCR Targets, RNA Transcripts, and Fluorescent cDNA Hybridization Probes

The procedure used to generate PCR targets, parent RNA transcripts, and fluorescent hybridization probes is summarized in Figure 1. Each cDNA clone was amplified by PCR using vector-specific primers to generate a dsDNA target to be arrayed. To obtain RNA transcripts for the preparation of hybridization probes, designated targets were further amplified by PCR using gene-specific primers containing a T7-promoter sequence (5′-NNNNNNTAATACGACTCACTATAGGGAG-3′) attached to the forward primer and a poly-dT_{30} sequence attached to the reverse primer, respectively. The T7 PCR templates were used to generate cRNA transcripts by incubation with T7 RNA polymerase (MEGAscribe™ kit; Ambion, Austin, TX, USA). Each RNA transcript was quantitated by fluorescence spectroscopy using an RNA-specific dye (RiboGreen®; Molecular Probes, Eugene, OR, USA). The fluorescence measurements were run in triplicate at two different dilutions, and the values were applied to a seven-step standard curve for quantitation. The integrity of each transcript was confirmed by agarose gel electrophoresis.

For hybridization experiments, pooled cRNA transcripts (typically 100 pg each unless indicated otherwise) were fluorescently labeled by reverse transcription with cyanine-3 or cyanine-5 dyes (GEMbright™ microarray labeling kits; Incyte Genomics). The resulting fluorescently labeled cDNA probes were purified by size exclusion chromatography (TE-30 column; BD Biosciences Clontech, Palo Alto, CA, USA), ethanol precipitated, and then resuspended in 24 µL hybridization buffer (5× SSC, 0.2% SDS, 1 mM DTT).

Microarray Fabrication

The PCR targets were purified using gel filtration over Sephacryl-400™ (Amersham Pharmacia Biotech, Piscataway, NJ, USA). The targets were then concentrated and resuspended in 2× SSC for arraying. To minimize spot-to-spot variations, all PCR targets were normalized to a concentration of 100 ng/µL. In addition, the target elements were printed in quadruplicate onto each array for signal averaging. Microarraying was performed in the Incyte Genomics microarray facility (80–100 µm spots at 180-µm spacing, center to center). After arraying, the slides were irradiated at 120 mJ in a UV Stratalinker® model 2200 instrument (Stratagene, La Jolla, CA, USA) and then rinsed in 0.2% SDS for 2 min, followed by three 1-min rinses in deionized water. Slides were then treated with 0.2% I-Block® reagent (Tropix, Bedford, MA, USA)

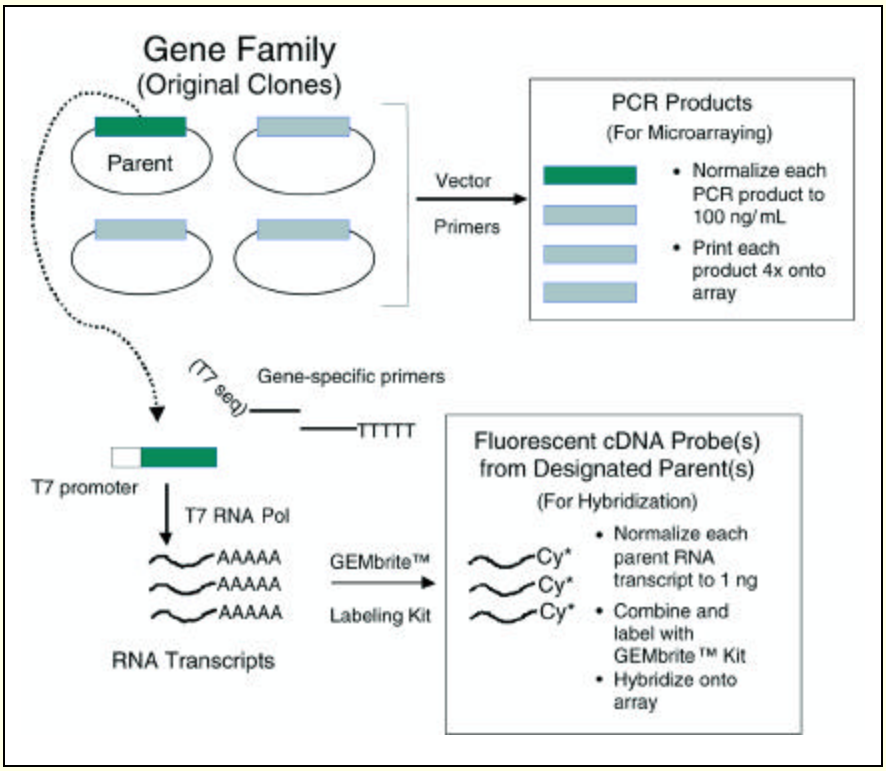

Figure 1. Scheme for the preparation PCR targets and fluorescent hybridization probes. Clones selected to represent each functional family were amplified by PCR using vector primers. One designated parent clone from each family was further amplified by PCR using chimeric gene-specific primers to produce RNA transcripts for use in the generation of fluorescent hybridization probes (see Materials and Methods).

in 1× Dulbecco's PBS (Invitrogen, Carlsbad, CA, USA) at 60°C for 30 min. The arrays were then rinsed again in 0.2% SDS and water as described above.

Hybridization and Data Analysis

Fluorescently labeled hybridization probes were applied to the microarrays under a 22 × 22 mm glass cover slip, placed in a sealed chamber to prevent evaporation, and then incubated at 60°C for 6 h. After hybridization, the microarrays were washed in 1× SSC/0.1% SDS/1 mM DTT at 45°C for 10 min and then in 0.1× SSC/0.2% SDS/1 mM DTT at 25°C for 3 min. The microarrays were then imaged using a GenePix™ dual-laser confocal scanner (Axon Instruments, Foster City, CA, USA) at 10 µm resolution. The scanned images were converted into 16-bits-per-pixel resolution, yielding a 65 536 count dynamic range. GEMtools™ software (Incyte Genomics) was used for image analysis. Four independent microarray hybridizations were conducted for each experiment. Target signals were corrected for local background and then averaged (four spots per array × number of arrays per experiment). Unless stated otherwise, four hybridization reactions were run per condition. Error bars shown in the figures are one standard deviation from the averaged signal values. Correlation coefficients between averaged hybridization signals and alignment parameters (clone length, overlap between the probe sequence and cDNA clone sequence, percent sequence identity, LALIGN score, etc.) were determined using Statistica® software (StatSoft, Tulsa, OK, USA).

RESULTS AND DISCUSSION

As stated in the Introduction, the purpose of this study was to measure hybridization cross-reactivity between cDNA targets sharing moderate to high sequence homology and to determine whether these cross-reactive signals could be correlated with their primary sequences (i.e., percent sequence identity, physical length, percent GC content, etc.). In particular, we considered that percent sequence identity might be a good predictor of cross-reactivity, based on reported cross-hybridization results on nylon membranes (15). Therefore, cDNA clones were selected from different functional gene families based on sequence identity scoring (i.e., the highest scoring clones available). In addition, one clone from each family was designated as a "parent" for hybridization comparisons based on the highest combined identity score.

We next considered whether the homology-scoring algorithm LALIGN could be used as a predictor of cross-reactivity between known sequences. In this case, we took into account the actual sequences present in the hybridization reaction. As described in the Materials and Methods section, a synthetic RNA transcript was generated from each designated parent using chimeric gene-specific primers. Whenever possible, these primers were designed near the ends of the first-round PCR products. However, this was not always practicable, so a number of the resulting transcripts span only a portion of the designated cDNA clone. These synthetic RNA transcript sequences were used for homology scoring by LALIGN against each cDNA clone of the family because they represent the actual templates used in the preparation of fluorescent hybridization probes. The respective homology scores and other sequence parameters are summarized in Table 1.

Hybridization Cross-Reactivity in the Serine Protease, Metalloprotease, Chemokine, and G-Protein Families

Hybridization probes generated from selected parent cDNA sequences were hybridized to a spotted cDNA microarray comprising all of the cDNA clones chosen for this study (see Materials and Methods). In the first hybridization experiment, replicate hybridizations were conducted using probe sequences from the serine protease, metalloprotease, chemokine, and G-protein families. The results from this experiment are summarized in Figure 2. Within each family, the targets corresponding to the designated parents were well discriminated from the other targets (cross-reactivities between 0.6% and 6%). As shown in Table 1, the highest cross-reactivities corresponded generally to the clones exhibiting the highest homology scores. However, despite similar homologies, the magni-

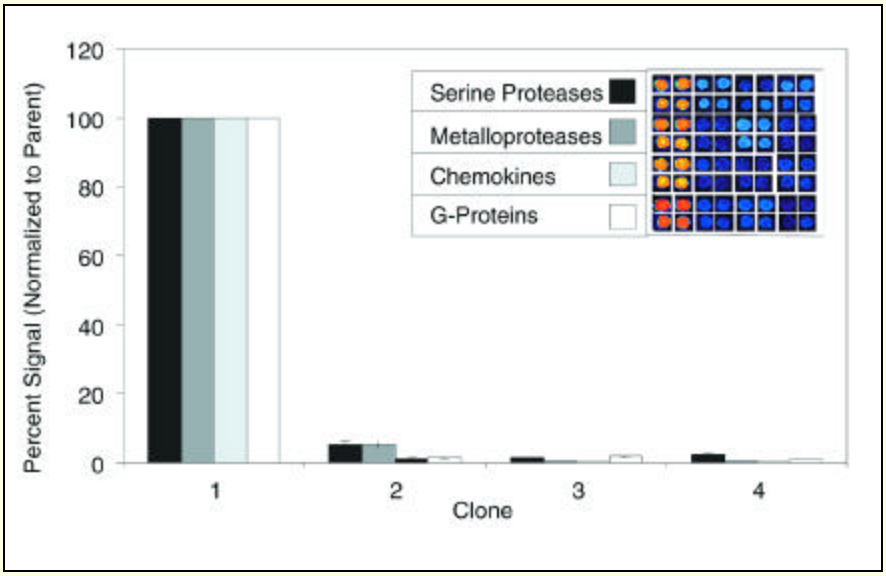

Figure 2. Hybridization cross-reactivity for the serine protease, metalloprotease, chemokine, and G-protein families. For these hybridization experiments, the probe reactions contained a pool of parent RNA transcripts (1 ng). This RNA concentration was selected to provide hybridization signals within the top third of the linear response curve for the scanner. The clones used to generate microarray targets for this experiment are summarized in Table 1. Inset: Hybridization signals shown as false color images (blue = bottom of the response curve < green < yellow < red < white = top of the response curve; note that not all colors may be represented in the figure insert).

Table 1. GenBank Accession Nos. and Alignment Scores for Spotted cDNA Clones Representing Homologous Gene Family Members[a]

Family	Clone	GenBank AccessionNo.	Transcript Overlap (bp)	Percent Identity	LALIGN Score	Percent Signal[b]
Serine Proteases	**1**	**g181189**	**816**	**100**	**3867**	**100**
	2	g4503136	464	65	721	5.5
	3	g35282	748	54	296	1.5
	4	g4504578	140	60	112	2.6
Metallo-proteases	**1**	**g4505204**	**1424**	**100**	**7063**	**100**
	2	g36632	1419	85	5079	5.2
	3	g903981	50	68	82	1.0
	4	g348020	139	55	77	0.9
Chemo-kines	**1**	**g1924937**	**497**	**100**	**2485**	**100**
	2	g4506840	373	68	713	2.9
	3	g4885586	405	64	603	0.9
	4	g3928270	181	80	572	0.6
G-Proteins	**1**	**g616866**	**164**	**100**	**820**	**100**
	2	g6912393	164	68	352	1.5
	3	g7022042	163	67	329	1.9
	4	g3329379	151	59	197	1.3
Threonine Proteases	**1**	**g4506204**	**243**	**100**	**1194**	**100**
	2	g565646	112	56	83	0.7
	3	g4506196	51	71	96	0.7
	3	**g4506196**	**533**	**100**	**2665**	**100**
	2	g565646	53	62	73	6.3
	1	g4506204	131	54	91	12.3
p450 (CYP)	**1**	**g181299**	**484**	**100**	**2420**	**100**
	2	g2618613	489	85	1638	56.9
	3	g2618613	500	82	1492	28.1
	4	g181343	477	86	1779	37.1
	5	g181327	476	80	1552	26.2
	6	g6470140	433	71	1019	8.1
	7	g181357	416	60	595	5.7
	8	g181299	106	100	530	33.6
	9	g6470140	110	76	316	4.4
	10	g181293	243	58	285	10.6
	11	g181269	55	60	80	6.2
	12	g6470140	39	67	66	6.0

[a]Bold, selected parent clone.

[b]Signal intensities as a percentage of the selected parent clone. For the CYP family, values are shown for the experiment summarized in Figure 3d that was hybridized at 60°C.

tudes of cross-hybridization differed between gene families. For example, in the chemokine family, an averaged cross-hybridization signal of 2.9% was observed between the parent (monocyte chemotactic protein-2, g1924937) and a second family member (monocyte chemoattractant protein-1, g4506840). The latter target exhibited 68% sequence identity with an LALIGN score of 713. In the serine protease family, clone g4503136 scored similarly (65% sequence identity with an LALIGN score of 721), yet gave a slightly higher cross-reactivity (5.5%). In the case of the metalloprotease family, comparably low cross-reactivity (5.2%) was observed for a clone exhibiting considerably higher homology (85% sequence identity over 99% of the designated parent probe sequence, with an LALIGN score of 5079).

Crossover Hybridization Experiment in the Threonine Protease Family

Hybridization probes from two clones of the threonine protease family (g4506204 and g4506196) were hybridized separately to the same cDNA microarray described above to determine whether alternative parent clone nominations could yield similar results.

In both assays, the non-parent targets were well discriminated from the selected parents (Table 1). However, the magnitudes of cross-hybridization were different in each case. Cross-reactivities measured for clone g4506204 were low (about 0.7%), whereas those measured for clone g4506196 were slightly higher (6.3% and 12.3%). This difference cannot be explained by the homology scores, which were similar, although the RNA transcripts generated from these two clones had different lengths (243 and 533 bp, respectively). Therefore, according to these results, homology scoring does not appear to be an exact predictor of cross-hybridization.

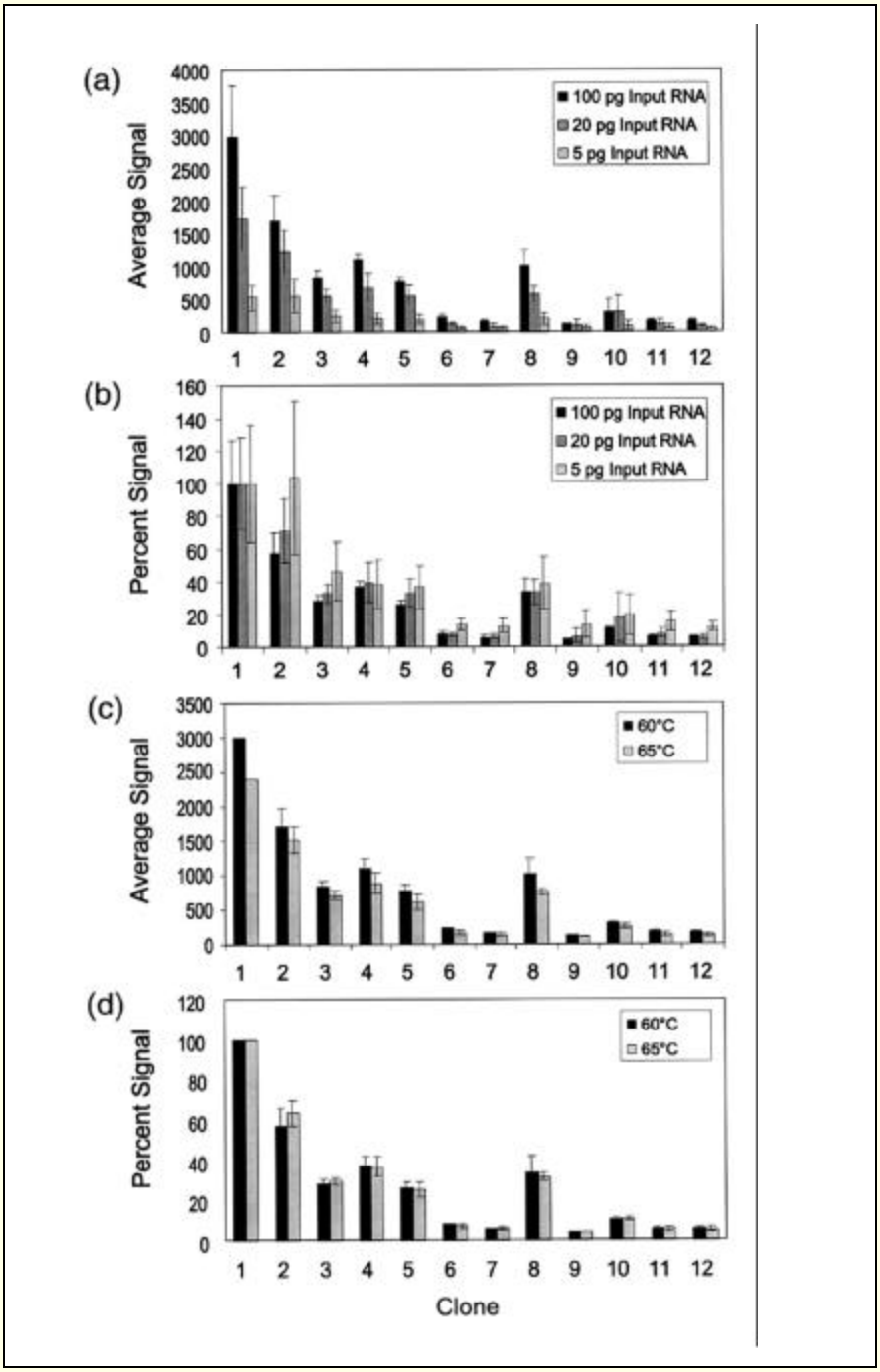

Figure 3. Hybridization cross-reactivity for the cytochrome P450 family. (a and b) Three separate sets of hybridization experiments are summarized using low (5 pg), medium (20 pg), and high (100 pg) input RNA transcript for probe generation (n = 2 per condition). (c and d) Hybridization results from 100 pg input RNA transcripts following hybridizations at standard (60°C) and elevated (65°C) temperatures (n = 3 per condition).

Cross-Hybridization in the CYP Family

Additional experiments were conducted with the CYP family to examine hybridization cross-reactivity over a range of input RNA concentrations, as would be expected from a complex biological sample, and also at different hybridization temperatures. Hybridization probes generated from the CYP parent clone (g181299) were hybridized to the same cDNA microarray described in the previous experiments. This family comprised a wider range of sequence homologies, including multiple examples of sequence identities greater than 80%. In several cases, the clones represented splice variants or different subclones of the same gene. One of these clones (clone 8) comprised a splice variant of the designated parent clone. The results from these experiments are summarized in Figure 3. The magnitudes of cross-reactive signals increased with input RNA concentrations (Figure 3a); however, the percent signal was relatively constant between these inputs (Figure 3b).

Raising the hybridization temperature from 60°C to 65°C resulted in slightly lower averaged signals (Figure 3c), but again the percent signals relative to the parent were similar (Figure 3d). These data suggest that cross-hybridization on microarrays is relatively constant and reproducible over the range of RNA inputs and temperatures investigated.

These cross-reactive signals showed a reasonable correspondence to the homology scores. Clones with less than 80% sequence identity (clones 6, 7, and 9–12) were well discriminated from the parent clone (cross-reactivity ranged between 4.4% and 10.6%). On the other hand, clones with 80%–86% sequence identity (clones 2–5) showed less discrimination (cross-reactivity > 25%). Additionally, one clone (clone 8) scored 100% sequence identity over about one-quarter of the parent transcript and showed about 33.6% cross-reactivity. This clone represents a splice

Table 2. Correlation of Cross-Hybridization Signals with Percent Sequence Identity and LALIGN Homology Scores Correlations of Cross-Reactivity Across families (All Data, I = 27)

Correlations of Cross-Reactivity across Families (*n* = 11)			
	Overlap	Identity	LALIGN
Percent Signal	0.19	**0.65**	0.34
Correlations of Cross-Reactivity for CYP Family Only *n* = 11)			
	Overlap	Identity	LALIGN
Percent Signal	0.52	**0.75**	**0.76**

Bolded values are significant at $P < 0.05$.

variant of the gene corresponding to the parent clone.

Correlation of Hybridization Cross-Reactivities with Sequence Overlaps and Homology Scores

Statistical analysis of cross-reactive hybridization signals is summarized in Table 2. Across all families, only sequence identity showed significant correlation with signal intensity. Assuming that LALIGN might show a stronger correlation within a family (since sequence overlaps would be expected to vary widely between families), the analysis was repeated including only data for the p450 (CYP) family. Again, sequence identity showed the strongest correlation, although LALIGN also showed a significant correlation in this case. Taken together, these results indicate that hybridization cross-reactivity can best be predicted by sequence identity and that it is relatively independent of sequence overlap (at least above 100 bp, based on our data). It is interesting to note that the correlation score for sequence overlap increased in significance when the data for the CYP family was treated separately (Table 2). Five of the 11 clones in this family share greater than 80% sequence identity relative to the designated parent, including one splice variant (clone 8) that is 100% identical over a region about one-fifth of the parent transcript sequence. Therefore, the length of sequence overlap can have a greater effect on cross-hybridization in cases of high sequence homology, particularly in cases of sequences that represent alternatively spliced genes.

The results obtained from this study demonstrate that glass cDNA microarrays can generally distinguish hybridization between cDNA targets sharing less than 80% sequence identity. These results are in good agreement with previous data on nylon membranes where hybridization cross-reactivity was not observed until the alignments exceeded 77% sequence identity (15).

In the case of the CYP family data, substantial cross-reactivity was observed in two instances: the first case resulted from sequences exhibiting greater than 80% identity over of the majority of the probe sequence; the second case was for one sequence having 100% identity over about one-fifth of the probe sequence. These data indicate that higher levels of hybridization cross-reactivity can be expected in cases of high sequence identity over a majority of the sequence or exact identity over a portion of the sequence. Thus, splice variants are also predicted to show hybridization cross-reactivity, depending on the extent of sequence overlap with the clone represented on the array.

Application of Cross-Hybridization Results to Microarray Design and Gene Expression Analysis

Pairwise sequence alignment algorithms are widely used to elucidate conserved structural, functional, and evolutionary relationships among gene families (6,7). The data presented here demonstrate that sequence alignments and identity scoring can also be used to estimate hybridization cross-reactivity within gene families. When performing these calculations, it is important to consider the sequences that are actually arrayed. For example, clones representing a particular target gene may only contain a portion of the consensus gene

sequence that is available. Regions of the gene that are outside of the clone sequence should be excluded from these calculations. In the present study, the sequences of the cRNA transcripts generated from the selected parent clones were used for homology scoring by LALIGN because these were the templates that were used for probe generation. In the case of a biological sample, the full-length target sequence (if available) would be preferred.

With one exception (i.e., the 65°C hybridization temperature investigated with the CYP family), the conditions of the microarray experiments reported here were conducted according to the standard conditions of probe labeling, hybridization, and data analysis that are employed in our production process. In practice, hybridization experiments under different conditions (buffer, temperature, denaturants, etc.) have failed to yield improved discrimination (data not shown). However, it is possible that other stringency conditions not yet attempted could yield reduced cross-reactive signals. These studies are ongoing.

While the results presented here suggest that homologous regions of genes should be avoided when selecting clones for cDNA microarrays, they also indicate that cross-species hybridizations could be conducted on microarrays (i.e., where expressed gene sequences between the two species share 80% sequence identity or higher). This suggests that human cDNA microarrays could be used to analyze tissues from animal models of human disease. For example, a recent comparative analysis of orthologous genomic loci from human and mouse showed that genes could be identified in the latter based on conserved exon features (1). This predicts that expressed sequences in the human and the mouse will have sufficient homology for pathway analyses using cross-species microarray hybridizations. Indeed, preliminary investigations in our laboratory suggest that this is the case (P. Scott Eastman, unpublished results).

It is interesting to note that of the two gene families displaying greater than 80% sequence identity (metalloproteases and CYP) only the CYP family showed significant cross-hybridization from clones mapping to different genes. Therefore, as discussed earlier, pairwise sequence identity scoring is not an exact predictor of cross-hybridization but certainly needs to be considered when designing microarrays and interpreting the results.

ACKNOWLEDGMENTS

We thank Steven Daniel for technical contributions. We also thank Lyle Arnold, Rick Johnston, and P. Scott Eastman for many useful technical discussions and support, and Zora Modrusan and Mike Walker for critically reviewing this manuscript.

REFERENCES

1. **Batzoglou, S., L. Pachter, J.P. Mesirov, B. Berger, and E.S. Lander.** 2000. Human and mouse gene structure: comparative analysis and application to exon prediction. Genome Res. *10*:950-958
2. **Brown, P.O. and D. Botstein.** 1999. Exploring the new world of the genome with DNA microarrays. Nat. Genet. *21*:33-37.
3. **DeRisi, J., L. Penland, P.O. Brown, M.L. Bittner, P.S. Meltzer, M. Ray, Y. Chen, Y.A. Su, and J.M. Trent.** 1996. Use of a cDNA microarray to analyze gene expression patterns in human cancer. Nat. Genet. *14*:457-460.
4. **Doyle, J.J. and B.S. Gaut.** 2000. Evolution of genes and taxa: a primer. Plant Mol. Biol. *42*:1-23.
5. **Duggan, D.J., M. Bittner, Y. Chen, P. Meltzer, and J.M. Trent.** 1999. Expression profiling using cDNA microarrays. Nat. Genet. *21*:10-14.
6. **Gogarten, J.P. and L. Olendzenski.** 1999. Orthologs, paralogs and genome comparisons. Curr. Opin. Genet. Dev. *9*:630-636.
7. **Gotoh, O.** 1991. Multiple sequence alignment: algorithms and applications. Adv. Biophys. *36*:159-206.
8. **Huang, X. and W. Miller.** 1991. A time-efficient, linear-space local similarity algorithm. Adv. Appl. Math. *12*:337-357.
9. **Hughes, T.R., M.J. Marton, A.R. Jones, C.J. Roberts, R. Stoughton, C.D. Armour, H.A. Bennett, E. Coffey et al.** 2000. Functional discovery via a compendium of expression profiles. Cell *102*:109-126.
10. **Iyer, V.R., M.B. Eisen, D.T. Ross, G. Shuler, T. Moore, J.C.F. Lee, J.M. Trent, L.M. Staudt et al.** 1999. The transcriptional program in the response of human fibrolasts to serum. Science *283*:3-87.
11. **Lashkari, D., J. DeRisi, J. McCusker, A. Namath, C. Gentile, S. Hwang, P.O. Brown, and R. W. Davis.** 1997. Yeast microarrays for genome wide parallel genetic and gene expression analysis. Proc. Natl. Acad. Sci. USA *94*:13057-13062.
12. **Liang, F., I. Holt, G. Pertea, S. Karamycheva, S.L. Salzberg, and J. Quackenbush.** 2000. Gene index analysis of the human genome estimates approximately 120,000 genes. Nat. Genet. *25*:239-240.
13. **Ruan, Y., J. Gilmore, and T. Conner.** 1998. Towards Arabidopsis genome analysis: monitoring expression profiles of 1400 genes using cDNA microarrays. Plant J. *15*:821-833.
14. **Schena, M., D. Shalon, R.W. Davis, and P.O. Brown.** 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science *270*:467-470.
15. **Schena, M., D. Shalon, R. Heller, A. Chai, P.O. Brown, and R.W. Davis.** 1996. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. Proc. Natl. Acad. Sci. USA *93*:10614-10619.
16. **Scherf, U., D.T. Ross, M. Waltham, L.H. Smith, J.K. Lee, L. Tanabe, K.W. Kohn, W.C. Reinhold et al.** 2000. A gene expression database for the molecular pharmacology of cancer. Nat. Genet. *24*:236-244.
17. **Shalon, D., S.J. Smith, and P.O. Brown.** 1996. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res. *6*:639-645.
18. **Spellman, P.T., G. Sherlock, M.Q. Zhang, V.R. Iyer, K. Anders, M.B. Eisen, P.O. Brown, D. Botstein, and B. Futcher.** 1998. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. Mol. Biol. Cell *9*:3273-3297.
19. **Vernier, P., R. Mastrippolito, C. Helin, M. Bendali, J. Mallet, and H. Tricoire.** 1996. Radioimager quantification of oligonucleotide hybridization with DNA immobilized on transfer membrane: application to the identification of related sequences. Anal. Biochem. *235*:11-19.

Received 7 March 2001; accepted 1 August 2001.

Address correspondence to:

Dr. Mark A. Reynolds
Incyte Genomics
3160 Porter Avenue
Palo Alto, CA 94304, USA
e-mail:markr@incyte.com