Unraveling Gene–Gene Interactions Regulated by Ligands of the Aryl Hydrocarbon Receptor*

Charles D. Johnson,1,2 Yoganand Balagurunathan,3 Mahlet G. Tadesse,4 M. Hadi Falahatpisheh,1,2 Marcel Brun,1,3 Mary K. Walker,5 Edward R. Dougherty,2 and Kenneth S. Ramos1,2

1Department of Biochemistry and Molecular Biology and Center for Genetics and Molecular Medicine, University of Louisville, Louisville, Kentucky, USA; Departments of 2Veterinary Physiology and Pharmacology, 3Electrical Engineering, and 4Statistics, Texas A&M University, College Station, Texas, USA; 5Department of Pharmacology and Toxicology, University of New Mexico, Albuquerque, New Mexico, USA

The co-expression of genes coupled to additive probabilistic relationships was used to identify gene sets predictive of the complex biological interactions regulated by ligands of the aryl hydrocarbon receptor (Ahr). To maximize the number of possible gene–gene combinations, data sets from murine embryonic kidney, fetal heart, and vascular smooth muscle cells challenged in vitro with ligands of the Ahr were used to create predictor/training data sets. Biologically relevant gene predictor sets were calculated for Ahr, cytochrome P450 1B1, insulin-like growth factor–binding protein-5, lysyl oxidase, and osteopontin. Transcript levels were categorized into ternary expressions and target genes selected from the data set and tested for all possible combinations using three gene sets as predictors of transitional level. The goodness of prediction for each set was quantified using a multivariate nonlinear coefficient of determination. Evidence is presented that predictor gene combinations can be effectively used to resolve gene–gene interactions regulated by Ahr ligands. Key words: aryl hydrocarbon receptor, bioinformatics, gene networks, genomics. Environ Health Perspect 112:403–412 (2004). doi:10.1289/ehp.6758 available via http://ehp.niehs.nih.gov/ [Online 14 January 2004]

The assumption that deconstructive methodologies accurately describe the complexity of biological processes is inadequate at best. Until recently, functional genetic studies have been of limited scope and able to elucidate the role of only one or a few genes at a time. Several limitations render these techniques nonconducive to large-scale expression analysis; therefore, nucleotide hybridization technologies are now used to monitor the expression of thousands of genes at a given point in time. A fundamental goal of genomics research is to understand individual gene expression patterns within the symphonic context of the transcriptome and to unravel the genetic networks responsible for health and disease. However, the interactive gene networks responsible for expression of altered cellular phenotypes have not been fully defined.

To date, most microarray experiments have used correlation analysis to identify common genomic responses to a particular stimulus, or multivariate methodologies to examine more complex gene–gene interactions. Univariate methodologies can be used to identify common genomic responses to a particular stimulus but do not account for multiple influences on gene expression. Logistic regression and stepwise regression, on the other hand, are multivariate approaches successfully used to examine more complex genomic interactions but require prior knowledge of the system and assume linearity in assigning biological relatedness.

To understand the complex nature of cellular transformation in cancer, computational prediction methodology has been used to examine global patterns of gene expression (Kim et al. 2000a, 2000b). This method identifies associations between the expression patterns of individual genes by determining whether knowledge of the transcriptional levels of a small gene set predicts the associated transcriptional state of another gene. Although mRNA is not the final product of a gene, transcription is a critical component in the regulatory cascade and therefore provides an ideal point of investigation. A key goal in networks analysis is the development of analytical tools to delineate how individual gene actions are integrated into complex biological systems at the organelle, cell, organ, and organism levels.

The goal of this study was to unravel biological networks regulated by ligands of the aryl hydrocarbon receptor (Ahr). Ahr is a ligand-activated transcription factor involved in the regulation of cellular growth, differentiation, and metabolism in all species examined (Carlson and Perdew 2002). Ahr is a member of the large basic helix–loop–helix–PAS (bHLH-PAS) homology domain family of transcription factors that includes proteins involved in myoblast differentiation, such as myogenic differentiation antigen 1; the cellular response to hypoxia, such as Ahr nuclear translocator (Arnt) and hypoxia-inducible factor-1; the Drosophila neurogenic protein Sim (single-minded), and the Drosophila circadian rhythm protein Per (period). bHLH-PAS proteins generally form heterodimeric transcription factors, of which Ahr is the only member conditionally activated in response to ligand binding. Polycyclic aromatic hydrocarbon contaminants and by-products of aspartate aminotransferase metabolism are now recognized as ligands of the Ahr (Bittiger et al. 2003). After ligand binding within the PAS domain, the cytosolic Ahr undergoes a conformational change, dissociates from two 90-kDa heat shock proteins and the hepatitis B virus X-associated protein 2, and translocates to the nucleus where it dimerizes with Arnt (Carver and Bradfield 1997).

The Ahr Arnt heterodimer interacts with Ahr-responsive elements (5’TNGCGTG-3’) upstream of target genes to activate/ repress transcription of target genes. Several drug-metabolizing enzymes (Nebert 1994) are regulated by Ahr, but key molecular targets involved in regulation of cellular differentiation and growth have remained largely elusive. The complexity of Ahr signaling is emphasized by recent studies showing that Ahr also participates in post-transcriptional regulation of gene expression (Falahatpisheh and Ramos 2003).

The target genes chosen for study included Ahr, cytochrome P450 1B1 (Cyp1b1), insulin-like growth factor-binding protein-5 (Igfbp-5), lysyl oxidase (Lox), and osteopontin (Opn). Gene networks were defined on the basis of the co-determination

Address correspondence to K.S. Ramos, Department of Biochemistry and Molecular Biology, University of Louisville Health Sciences Center, Louisville, KY 40292 USA. Telephone: (502) 852-5217, Fax: (502) 852-6222, E-mail: kenneth.ramos@louisville.edu

*The online version of this article (available at http://www.ehponline.org) contains Supplemental Material.

This research was supported by National Institutes of Health grants ES04849, ES09106, ES07273 to K.S.R and ES09804 and ES12072 to M.K.W. C.D.J was supported by National Research Service Award (NSRA) ES012117, M.G.T. by National Cancer Institute CA90301, and M.H.F. by NRSA ES012542.

The authors declare they have no competing financial interests.

Received 23 September 2003; accepted 14 January 2004.
of transcriptional states resolved by statistical evaluation of data sets derived from large-scale simultaneous measurements made using Affymetrix microarray technology.

**Materials and Methods**

**Model Systems**

Gene transcription information from three independent Affymetrix microarray experiments was used for comprehensive computational analysis. The model systems used included mouse embryonic heart, kidney, and thoracic aorta challenging with hydrocarbon ligands to activate Ahr signaling. Vascular smooth muscle cells (vSMC) and fetal kidneys were exposed in vitro, whereas fetal hearts were exposed in vivo. This approach enabled us to identify common gene sets across tissue types without regard for contextual differences in transcriptional status under basal conditions. The objective was to identify highly conserved interactions regulated by Ahr regardless of genomic context.

**Data set 1.** Vascular smooth muscle cells were isolated from the thoracic aorta of adult C57BL/6J untreated mice (Jackson Laboratories, Bar Harbor, ME) (6 weeks of age) and maintained in serial culture as described (Ramos and Cox 1993). All studies were initiated using cells seeded in 150-mm plates at 75% confluence. To induce C0 synchronization, cultures were incubated for 72 hr in 0.1% fetal bovine serum in Medium 199 (InVitrogen Corp., Carsbad, CA). Cells were challenged for 8 or 24 hr with dimethyl sulfoxide (DMSO) or 3 μM benzo[a]pyrene (BP) and RNA was isolated. BP is a hydrocarbon ligand of the Ahr that modulates growth and differentiation of vascular cells (Kerzec and Rames 2000). Normalized data from 12 chips were used for the analysis.

**Data set 2.** Day 11.5 mouse embryos were surgically resected from C57BL/6J wild-type or Ahr knockout mice and placed in Hanks’ balanced salt solution. Embryonic kidneys (approximately 0.5 mm x 1 mm) were isolated by microsurgical dissection and deposited on culture inserts. BP was added to the medium at 3 μM daily for 4 days, and an equivalent volume of DMSO was added to controls. On day 4, kidneys were harvested for RNA isolation. Normalized data from a total of 8 chips were used for analysis.

**Data set 3.** After exposure to 1.5, 3, and 6 μg/kg 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in utero on gestation day 14.5 (n = 4 litters/treatment), gestation day 17.5 fetal mouse hearts were surgically resected and RNA was isolated. TCDD is a potent hydrocarbon ligand that binds and activates Ahr signaling. Normalized data from a total of 20 chips were used from this study.

**RNA Isolation**

Total RNA was extracted using TRI Reagent (Molecular Research Center, Inc., Cincinnati, OH) according to manufacturer’s specifications.

**Affymetrix GeneChip**

The Affymetrix Murine Genome U74A Array (Affymetrix, Inc., Santa Clara, CA) used in these studies represents all functionally characterized sequences (approximately 6,000) in the Mouse UniGene database (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene). In addition, approximately 6,000 expressed sequence tag clusters are included. Experimental procedures including double-stranded cDNA synthesis and biotinylated cRNA preparation were conducted as recommended in the Affymetrix GeneChip Expression Analysis Technical Manual (Affymetrix, Inc., 2003).

**Double-Stranded cDNA Synthesis**

Total RNA was processed using Qiagen RNeasy kit (Qiagen, Inc., Valencia, CA) according to manufacturer’s specifications. For double-stranded cDNA synthesis, 15–20 mg total RNA was first hybridized with 100 pmol T7- (dT)24 primer (5’-[biotin]-GTCGTCAAGATGCTACGGTTCAGCA-3’) and high performance liquid chromatography (GENSET Corp., La Jolla, CA) purified at 70°C for 10 min. Primer hybridization was completed in a 20-μL reaction containing a final concentration of 10 mM dithiothreitol (DTT), 500 mM each of deoxyribonucleoside triphosphate (dNTP) mix, and 1× first-strand cDNA buffer. The reaction was incubated at 42°C for 2 min followed by addition of 400–600 U SuperScript II Reverse Transcriptase (Gibco Life Technologies, Rockville, MD) to synthesize first-strand cDNA. After 1 hr, second-strand cDNA synthesis was carried out by adding 200 mM each dNTP, 10 U Escherichia coli DNA ligase, 40 U E. coli DNA polymerase I, 20 U E. coli RNase H, and 1× second-strand reaction buffer in a 150-μL volume and incubated at 16°C for 2 hr. T4 DNA polymerase was added at the end of the reaction for an additional 5 min and soaked in 10 mL 0.5 M EDTA. Phase Lock Gel (Eppendorf Scientific, Inc., Westbury, NY) extraction with phenol/chloroform followed by ethanol precipitation was subsequently performed to clean up the double-stranded cDNA. The cDNA pellet was resuspended in 12 mL RNase-free water (Ambion, Inc., Austin, TX).

**Biotin-Labeled cRNA Preparation**

Biotin-labeled cRNA target for hybridization to GeneChip Array (Affymetrix, Inc.) was prepared by in vitro transcription using BioArray High Yield RNA Transcript Labeling Kit (Affymetrix, Inc.). Briefly, 3.3–5 mL double-stranded cDNA was mixed gently with 4 mL each of 10× high-yield reaction buffer, 10× biotin-labeled ribonucleotides, 10× DTT, 10× RNase inhibitor mix, and 10× T7 RNA polymerase provided by the kit and incubated at 37°C for 4–5 hr, with gentle mixing every 30 min. Labeled cRNA was then cleaned with RNeasy Mini Kit (Qiagen, Inc.) to obtain an accurate quantification of the labeled cRNA. Twenty to 30 mg labeled cRNA was then fragmented to 35–200 bp with 8 mL 5× fragmentation buffer containing 200 mM Tris-acetate, pH 8.1; 500 mM potassium acetate; and 150 mM magnesium acetate in a total volume of 40 mL for 35 min at 94°C. Before hybridization onto GeneChip Array, the quality of labeling and fragmentation was verified on agarose gel, transferred onto nylon membrane, and detected with alkaline phosphatase streptavidin and DuolUX Chemiluminescent/Fluorescent Substrate using UltraSNAP Biotinylated Nucleic Acid Detection Kit (Vector Laboratories, Inc., Burlingame, CA).

**Hybridization to GeneChip Array**

After labeling and fragmentation, 15 μg fragmented biotinylated cRNA was hybridized to the Affymetrix GeneChip Array in a 300-μL cocktail containing 5 mL of 3 nM control oligonucleotide B2, 15 mL 20× eukaryotic hybridization controls, and 150 mL 2× hybridization buffer provided in the GeneChip Eukaryotic Hybridization Control Kit (Affymetrix, Inc.) together with 3 mL 10 mg/mL herring sperm DNA and 3 mL 50 mg/mL acetylated bovine serum albumin (BSA). The hybridization was carried out at 45°C and 60 rpm in GeneChip Hybridization Oven 640 (Affymetrix, Inc.) for 16 hr.

**Washing, Staining, and Scanning the Array**

After hybridization the washing and staining procedures were carried out on a GeneChip Fluidics Station 400 in conjunction with Affymetrix Microarray Suite 5.0 software (MAS 5.0; Affymetrix, Inc.). Briefly, the array was first washed with 10 cycles of 2 filling and draining cycles in a nonstringent buffer containing 6× SSPE (52.9 g sodium chloride; 8.28 g sodium phosphate; monobasic, 2.82 g EDTA, pH 7.9) and 0.01% Tween 20 at 25°C, followed by 4 cycles of 15 filling and draining cycles in a
All genes were sorted by logical relevance and included predictors. The genes to be predicted (i.e., with the greatest computational constraints, the 200 clones Agilent GeneArray Scanner (Affymetrix, nonstringent buffer and scanned by the 15 cycles of 4 filling and draining cycles in 25°C, followed by a second SAPE staining (Vector Laboratories, Inc.) for 10 min at 25°C. The stained ternary states that describe their behavior, x

\[ x_j = x - \bar{x}/s_d. \]  

\[ cv = sd/\bar{x}. \]  

were calculated for each gene across all samples, and the intensity level (\( x \)) for each gene was standardized across all arrays by

All genes were sorted by \( cv \), and, because of computational constraints, the 200 clones with the greatest \( cv \) were selected for use as predictors. The genes to be predicted (i.e., targets) were selected on the basis of biological relevance and included Ahr, Cyp1b1, Isgbp-5, Lox, and Otn. A heuristic method was used to discretize the data into ternary states that describe their behavior, with -1 used for downregulated genes, 0 for invariant genes, and +1 for upregulated genes. Each array was ordered from lowest to highest standardized values \( x \) without regard to gene identity and the means across all arrays at the 33rd and 66th percentile were used as the cutoffs for the different values \( x \) of the standardized data set. These percentiles were chosen because they divided the data evenly to provide sufficient variation throughout the computational analysis. Our goal was to yield good estimates of the transcriptional state of the target using ternary discrete functions. The computation of prediction strengths was completed using a stochastic model of gene intensities, defining \( X_i \) as the random variable that represents the ternary intensity (or transcription rate) of gene \( G_i \).

Realizations of \( X_i \) were denoted by lower case \( x_i \). A microarray experiment consisted of a realization \( x_1 \) to \( x_\nu \) of \( \nu \) random variables \( X_1 \) to \( X_\nu \), representing the intensities of the \( \nu \) genes \( G_1 \) to \( G_\nu \). Given \( N \) microarrays, the values \( x_\nu (j) \) to \( x_\nu (j) \) represented the values for the microarray \( j \), with \( j = 1 \) to \( n \). Without loss of generality, the predictive strength of genes \( G_1 \), \( G_2 \), and \( G_3 \) over gene \( G_4 \) was unknown and therefore needed to be computed. In this context, genes \( G_1 \), \( G_2 \), and \( G_3 \) were regarded as predictors, and \( G_4 \) as target. An example of ternary intensity values \( x_1 (j) \), \( x_2 (j) \), \( x_3 (j) \), and \( x_4 (j) \) from 1 to 9 (or realizations of \( X_1 \), \( X_2 \), \( X_3 \), and \( X_4 \)), for \( G_1 \), \( G_2 \), \( G_3 \), and \( G_4 \), across nine conditions (microarrays) is illustrated in Figure 1 (rows 1–4). The function with the smaller expected error was denoted by \( \psi_\text{opt} \). The optimal constant function was denoted by \( \psi_0 \), and was the one with minimal error over the three possible constant functions.

The optimal functions \( \psi_{\text{opt}} \) and \( \psi_0 \), and their errors \( \varepsilon[\psi_{\text{opt}}] \) and \( \varepsilon[\psi_0] \), respectively, were estimated from the microarray data by splitting them randomly in two sets of conditions, with \( n \) conditions for test (estimation of the error) and \( N-n \) conditions for train (estimation of the function), where \( N \) was the total number of conditions. The optimal function \( \psi_{\text{opt}} \) and the optimal constant function \( \psi_0 \) were estimated from the \( N-n \) examples using plug-in rule (Devroye et al. 1996). Let \( \psi_{\text{opt}} \) and \( \psi_0 \) be the estimates of \( \psi_{\text{opt}} \) and \( \psi_0 \), respectively, the errors \( \varepsilon[\psi_{\text{opt}}] \) and \( \varepsilon[\psi_0] \) were estimated using the following equations:

\[
\varepsilon[\psi_{\text{opt}}] = \frac{1}{n} \sum_{j=1}^{n} \left| \psi_{\text{opt}}(x_1(j), x_2(j), x_3(j), x_4(j)) - x_4(j) \right|.
\]

\[
\varepsilon[\psi_0] = \frac{1}{n} \sum_{j=1}^{n} |x_4(j)| - x_4(j)|.
\]

Figure 1. Schematic representation of gene ternary values and predictor functions \( \psi \). The intensity of genes \( G_1 \), \( G_2 \), and \( G_3 \) can be used to predict the values of the target \( G_4 \) via a ternary function \( \psi \). An arbitrary ternary function \( \psi \) defined by \( \psi(X_1, X_2, X_3) = X_1 + X_2 - X_3 \) (with the sum defined inside \([-1, 0, 1]\)) is shown on line 5. The quality of \( \psi \) to explain \( G_4 \) from \( G_1 \), \( G_2 \), and \( G_3 \) can be measured by its expected absolute error \( \varepsilon[\psi] = E[|\psi(X_1, X_2, X_3) - X_4|]\) where | | denotes absolute value, and the expectation is computed over the joint distribution of the random variables \( X_1 \), \( X_2 \), \( X_3 \), and \( X_4 \). In line 6, the difference \( |\psi_0| \) is shown to exemplify absolute differences. An estimation of the value of \( X_4 \) assuming a constant value, is shown on line 7.
**Table 1. Genes that predict selected target gene behavior.**

| Targets                  | Predictor                                      | Probe set ID<sup>2</sup> | Percentage<sup>3</sup> | Correlation<sup>3</sup> |
|--------------------------|------------------------------------------------|---------------------------|-------------------------|--------------------------|
| Ahr                      | Lymphocyte antigen 6 complex, locus e          | 101488_r_at               | 18                      | 0.77                     |
| 160496_at                | Insulin-like growth factor binding protein 3  | 95082_at                  | 16                      | 0.83                     |
|                          | Tumor necrosis factor receptor superfamily, member 1b | 94928_at                  | 14                      | 0.70                     |
|                          | Insulin-like growth factor binding protein 6  | 103904_at                 | 13                      | 0.72                     |
|                          | Brain derived neurotrophic factor             | 102727_at                 | 13                      | 0.80                     |
|                          | Secreted phosphoprotein 1                     | 97519_at                  | 9                       | 0.87                     |
|                          | Stratifin                                      | 96704_at                  | 7                       | 0.23                     |
|                          | Small proline-rich protein 2f                  | 94120_s_at                | 6                       | 0.18                     |
|                          | Actin, alpha, cardiac                          | 101028_i_at               | 6                       | –0.84                    |
|                          | NADH dehydrogenase (ubiquinone) 1             | 100628_at                 | 6                       | –0.84                    |
|                          | Fibroblast growth factor receptor 4           | 92937_at                  | 5                       | 0.31                     |
|                          | Ras-related protein                            | 97319_at                  | 5                       | –0.83                    |
|                          | Small proline-rich protein 2b                  | 99701_f_at                | 5                       | 0.20                     |
| Cyp1b1                   | Spleen tyrosine kinase                         | 162363_at                 | 31                      | –0.44                    |
| 99979_at                 | Squalene epoxide                              | 94322_at                  | 18                      | 0.37                     |
|                          | NADH dehydrogenase                             | 92330_r_at                | 14                      | 0.64                     |
|                          | Calbindin-28k                                  | 98133_at                  | 10                      | –0.20                    |
|                          | Renin 1 structural                             | 98480_s_at                | 10                      | –0.10                    |
|                          | Carboxylesterase 3                             | 101539_f_at               | 10                      | 0.86                     |
|                          | Cytokine receptor-like factor 1                | 161046_at                 | 9                       | 0.92                     |
|                          | Mitogen regulated protein, proliferin 3        | 93929_s_at                | 9                       | 0.92                     |
|                          | Tumor-associated calcium signal transducer 2   | 160651_at                 | 9                       | –0.12                    |
|                          | Fox-like antigen 1                             | 99903_at                  | 9                       | 0.89                     |
|                          | Forkhead box a2                                | 93950_at                  | 8                       | 0.86                     |
|                          | Follistatin                                    | 99817_at                  | 6                       | 0.73                     |
|                          | Igfbp-5                                        | 93078_at                  | 5                       | 0.37                     |
| Isg15                    | Secreted phosphoprotein 1                      | 97519_at                  | 14                      | –0.64                    |
| 100566_at                | Matrix metalloproteinase 3                     | 98833_at                  | 9                       | –0.61                    |
|                          | RNA binding motif, single stranded interacting protein 1 | 96207_at                  | 8                       | –0.57                    |
|                          | Alpha-2-hs-glycoprotein                        | 98882_at                  | 7                       | 0.27                     |
|                          | Angiotensinogen                                | 101987_at                 | 7                       | –0.86                    |
|                          | Actin, alpha, skeletal muscle                   | 100381_at                 | 7                       | 0.68                     |
|                          | Interferon 1 structural                        | 98480_s_at                | 7                       | 0.23                     |
|                          | Interferon activated gene 202a                 | 94774_at                  | 6                       | 0.04                     |
|                          | Retinol binding protein, plasma                | 96047_at                  | 6                       | 0.40                     |
|                          | Integrin beta 4                                | 103305_at                 | 6                       | 0.43                     |
|                          | Insulin-like growth factor binding protein 3   | 95082_at                  | 5                       | –0.40                    |
|                          | Hydroxysteroid 11-beta dehydrogenase 2         | 100493_at                 | 5                       | 0.24                     |
|                          | Small proline-rich protein 2c                  | 101761_f_at               | 5                       | 0.09                     |
|                          | Follistatin                                    | 99817_at                  | 5                       | –0.57                    |
|                          | Integrin 6                                     | 95511_at                  | 5                       | 0.85                     |
|                          | Carbonic anhydrase 3                           | 100375_at                 | 5                       | 0.26                     |
|                          | Lox                                            | 103487_at                 | 36                      | –0.37                    |
|                          | Lympocyte antigen 6 complex, locus H           | 96207_at                  | 36                      | 0.55                     |
|                          | Single stranded interacting protein 1          | 102373_at                 | 23                      | –0.55                    |
|                          | Peripherin                                     | 161462_f_at               | 21                      | –0.35                    |
|                          | Cadherin 16                                    | 93516_at                  | 16                      | –0.40                    |
|                          | Coagulation factor II                          | 101989_at                 | 11                      | –0.30                    |
|                          | Fibroblast growth factor 7                     | 99435_at                  | 11                      | 0.95                     |
|                          | Forkhead box a2                                | 93950_at                  | 11                      | 0.95                     |
|                          | Thrombomodulin                                 | 104601_at                 | 11                      | 0.91                     |
|                          | Alcohol dehydrogenase 1, complex               | 94906_at                  | 9                       | –0.35                    |
|                          | Fox-like antigen 1                             | 99903_at                  | 9                       | 0.97                     |
|                          | Follistatin                                    | 93983_at                  | 9                       | 0.93                     |
|                          | Forkhead box f2                                | 99846_at                  | 7                       | 0.87                     |
|                          | Actin, alpha, cardiac                          | 101028_i_at               | 5                       | –0.60                    |
|                          | Cyclin-dependent kinase inhibitor 1a (p21)     | 94881_at                  | 5                       | –0.53                    |
|                          | High mobility group a-hook 2                   | 99058_at                  | 5                       | 0.85                     |
|                          | Integrin alpha 3                               | 104211_at                 | 5                       | 0.32                     |
|                          | Interferon gamma                               | 99334_at                  | 5                       | 0.53                     |
|                          | Mitogen regulated protein, proliferin 3        | 93929_s_at                | 5                       | 0.98                     |
|                          | Phospholipase a2 group VII                     | 101923_at                 | 5                       | –0.48                    |
|                          | Ras-related                                    | 97319_at                  | 5                       | –0.56                    |

(Continued)
In our study with $n = N/3$, $n$ was defined as $n = N/3$, so $2n$ was used to train the functions and $n$ to estimate the error.

The performance of the predictor set was defined by the coefficient of determination (CoD), that is, the degree to which the transcriptional state of a given set of genes improves the prediction of the transcriptional state of a target gene relative to the best possible constant function (defined by a fixed value) (Devroye et al. 1996) and was defined by

$$\hat{\theta} = \frac{\epsilon(\psi_{\text{opt}}) - \epsilon(\psi)}{\epsilon(\psi)}.$$  

[5]

The CoD was estimated from the data using the following equation:

$$\theta = \frac{\epsilon(\psi_{\text{opt}}) - \epsilon(\psi)}{\epsilon(\psi)}.$$  

[6]

This process defines an estimate $\hat{\theta}$ of the CoD for the set of predictor genes $G_1, G_2, G_3$ and the target gene $G_t$. The process was repeated 1,000 times with different random splitting and the final estimates $\hat{\theta}$ as the average value of $\hat{\theta}$ over the 1,000 replications. The mean error $\epsilon(\psi_{\text{opt}})$ was also estimated as the average of the 1,000 estimates $\epsilon(\psi_{\text{opt}})$. The higher the CoD, the more accurate the prediction of the target gene transcriptional state derived from the transcriptional state of the three predictor genes. All possible combinations of three predictor genes were studied for each target, where the number of combinations was one from the order of 8 million sets of 3 selected from 200, and the estimated CoD $\hat{\theta}$ was computed for all combinations. The predictor sets (or combinations) were ordered from best to worst based on $\hat{\theta}$, and the analysis focused on the predictor combinations with $\hat{\theta} > 0.9$ and $\epsilon(\psi_{\text{opt}}) < 0.05$. The cutoff used was selected heuristically to restrict the number of sets to those with good prediction potential. Because of the nature of the analysis, there was no measure of false-positive rates other than future biological experimentation and scrutiny of the literature. Each gene $G_t$ was ranked by $f_j$, the percentage of sets in this family of good combinations that contain such a gene.

**Network Plot**

The possible relationships predicted by the predictor algorithm were illustrated using the program developed by Breitkreutz et al. (2003). All three gene combinations that met the cutoff criteria on the basis of test errors and CoD values ($\hat{\theta} > 0.9$ and $\epsilon(\psi_{\text{opt}}) < 0.05$) were linked to their respective targets. These schematics were overlaid to form a linkage diagram showing the relationship of target genes to predictors, as well as connections among target genes. The diagrams generated represent the total hypothetical interrelationships between all 200 predictor genes and the five target genes used in the analysis.

**Results**

The expression of 200 genes across 40 different Affymetrix arrays from three different tissues after activation of Ahr signaling was used to predict the behavior of selected targets. Ahr activation by a wide range of structurally divergent agents plays a major role in vascular, renal, cardiac, skin, bone marrow, liver, eye, ovary, and immune system function, as well as in carcinogenesis (Denison and Nagy 2003). Table 1 ranks predictor genes by the $f_j$ commonly identified for each target. The Ahr receptor itself was most commonly predicted by lymphocyte antigen 6 complex, locus e (selected 18% of the time for all good three-gene combinations), Igfbp-3 (16%), and tumor necrosis factor receptor superfamily-member 1b (14%). Cyp1b1 was best predicted by spleen tyrosine kinase (31%), squaene epoxidase (16%), and nicotinamide adenine dinucleotide (NADH) dehydrogenase (Ndufc1) (14%). Igfbp-5 was most frequently predicted by Opn (14%), matrix metalloproteinase 3 (9%), and RNA binding motif–single-stranded interacting protein 1 (8%). Lox was best predicted by lymphocyte antigen 6 complex, locus H (Ly6b) (36%), single-stranded interacting protein 1 (36%), and glutamyl aminopeptidase (32%). Finally, Opn was most often predicted by brain-derived neurotrophic factor (Bdnf) (15%), interleukin 6 (14%), and prolferin (14%). The frequency of predictor usage was influenced by the cutoffs used, thus emphasizing the importance of arbitrary cutoff selection. For comparison, each target was also correlated to the complete 12,000 Affymetrix gene data set using normalized intensity values. The results indicated that in many instances genes frequently identified as predictors in the CoD algorithm also exhibited strong correlation to their targets. This is best exemplified for Ahr and Opn, where Ndufc1 and α-actin (Actc1) exhibited strong correlations and

### Table 1. Continued.

| Target | Predictor                        | Probe set ID | Percentage | Correlation |
|--------|----------------------------------|--------------|------------|-------------|
| Ndufc1 | Brain derived neurotrophic factor | 102727_at    | 15         | 0.73        |
| 97519_at | Interleukin 6                     | 102128_at    | 14         | 0.70        |
|         | Proliferin                        | 94838B_r_at  | 14         | 0.52        |
|         |         | 94928_at         | 13         | 0.76        |
|         | Tumor necrosis factor receptor superfamily, member 1b | 95082_at    | 12         | 0.92        |
|         |         | 9534B_at         | 10         | 0.73        |
|         | Insulin-like growth factor binding protein 3 | 99802_at    | 9          | 0.71        |
|         | Epipepin                            | 97319_at     | 8          | –0.91       |
|         | Ras-related protein                | 100622_at    | 7          | –0.91       |
|         | NADH dehydrogenase                 | 100622_at    | 7          | –0.91       |
|         | Actin, alpha, cardiac              | 101082_s_at  | 7          | –0.82       |
|         | Actin, alpha 1, skeletal muscle    | 100081_at    | 7          | –0.85       |
|         | Cyclin-dependent kinase inhibitor 1a (p21) | 94881_at    | 6          | 0.04        |
|         | Interferon gamma                   | 99334_at     | 6          | 0.38        |
|         | Angiopoietin-like 4                | 96119_s_at   | 6          | 0.69        |
|         | Cd44 antigen                       | 103005_s_at  | 5          | 0.71        |
|         | Cytokine receptor-like factor 1    | 101046_at    | 5          | 0.72        |
|         | High mobility group at-hom 2       | 94935_at     | 5          | 0.67        |
|         | Tumor-associated calcium signal transducer 2 | 100586_at   | 5          | 0.60        |
|         | Forkhead box a2                    | 100586_at    | 5          | 0.80        |
|         | Fos-like antigen 1                 | 99835_at     | 5          | 0.72        |
|         | Matrix metalloproteinase 3         | 98833_at     | 5          | 0.55        |

*Murine genome U74v2 (for complete details: http://www.affymetrix.com). *Percentage representing the number of times a given predictor was used for given target. *Linear correlation of the predictor genes to the given target.
were good predictors. As demonstrated by CoD, however, genes exhibiting low correlation to specific targets may play an important role in predicting target behavior. This is exemplified by the best predictor for *Lox*, with a frequency of 36% but a correlation coefficient of only −0.37. Thus, the CoD methodology identified interactions between targets and predictors that could potentially be missed when examined on a gene-by-gene basis.

The complex nature of biological interactions between targets and predictors is illustrated in Figure 2. The linking diagram was built using the $f_i$ ranks for each predictor gene and each target gene. The inclusion of any singular gene does not represent a causal relationship between predictor and target, but rather that gene $G_i$ with two other genes was also a good predictor of a specific target. To some degree, all targets were predictive of each other, a relationship not surprising given that targets were selected *a priori* on the basis of their responsiveness to Ahr ligands. Several genes predominated as predictors of each target. These relationships were denoted by the thickness of the connective lines (Figure 2) and the percentage of time used in three gene combinations (Table 1). The number of unique predictors varied with target and cutoff range, with 85% of all possible predictors used to estimate Ahr, 63% Cyp1b1, 98% Igfbp-5, 22% Lox, and 53% Opn.

Given our cutoff range, a total of 34 genes were resolved to provide a comprehensive association of all targets and predictors (Figure 3). The predominant linkages were between all five targets, particularly Opn as a predictor of Ahr and Cyp1b1. The best nontarget predictors included Bdnf, Actc1, c-myc single-strand-binding protein, Igfbp-6, and integrin beta 4. Figure 4A–E shows the use of individual gene predictor values for construction of three gene predictor set models. It should be noted that predictor genes may lie upstream or downstream from the target or be part of a chain of interaction among various intermediates within the biological network. A large number of three-clone combinations met

---

**Figure 2.** Gene–gene interaction networks activated by ligands of the Ahr. All three gene combinations for each target that met the cutoff of COD > 0.9 and error < 0.5 were individually plotted using a program developed by Breitkreutz et al. (2003). The thickness of the line denotes the selection frequency for individual gene for each target.
the selection criteria \((\theta > 0.9\) and \(\varepsilon(\psi_{\text{opt}}) < 0.05)\), with one or two clones identified as predominant predictors within the sample pool. A complete listing of target–predictor clone combinations is presented in Table 1 of the Supplemental Material (http://www.epsonline.org).

When examined as three-gene predictor sets, the best combinations for \textit{Abr} were \textit{Igfbp-3}, lymphocyte antigen 6, and \textit{Actc1} with CoD of 0.963; \textit{Ras}-related, proiferin, and \textit{Igfbp-6} with CoD of 0.96; and thrombomodulin, lymphocyte antigen 6, locus, and \textit{Igfbp-6} with CoD of 0.953 (Figure 4A).

The best combinations for \textit{Cyp1b1} were \textit{Abr}, melanoma growth-simulating activity-alpha (\textit{Gro1}) oncogene, and \textit{Cd44} antigen with CoD of 0.944; \textit{Actc1}, \textit{Bdnf}, and epiregulin with CoD of 0.939; and \textit{Abr}, \textit{Cd44} antigen, and proiferin 3 with CoD of 0.944 (Figure 4B). The best combinations for \textit{Igfbp-5} were signal transducer and activator of transcription 1 (\textit{Stat1}), \textit{Opn}, and procollagen-lysine, 2-oxoglutarate-5-dioxygenase 3 (\textit{Pld3}) with CoD of 0.98; \textit{Opn}, \textit{Igfbp-1}, and interleukin 1 alpha with CoD of 0.97; and insulin II, \textit{Opn}, and \textit{Igfbp-1} with CoD of 0.97. These combinations performed best when used collectively (Figure 4C). When examined as three gene predictor sets, the best predictor combinations for \textit{Lox} were lymphocyte antigen 6, locus H, \textit{Nd4h1}, and peripherin with CoD of 0.902; peripherin, lymphocyte antigen 6, locus H, and proiferin 2 with CoD of 0.923; and caderhin 16, single-stranded interacting protein 1, and \textit{Cd44} antigen with CoD of 0.924 (Figure 4D). For \textit{Opn} the best three-gene predictor combinations were \textit{Abr}, \textit{Gro1} oncogene, and \textit{Cd44} antigen with CoD of 0.944; \textit{Actc1}, \textit{Bdnf}, and epiregulin with CoD of 0.939; and \textit{Abr}, \textit{Cd44} antigen, and proiferin 3 with CoD of 0.944 (Figure 4E).

**Discussion**

The identification of relevant components of the biological response to \textit{Abr} ligands was modeled using the transcriptional profiles of cells from murine embryonic heart and kidney or vasculature. Genes whose transcription levels exhibited coupling by CoD methodology were hypothesized to be predictive of one another, whether lying upstream or downstream within the biological network or distributed about the network in such a way that their relation to the target gene was only based on chains of interactions among various intermediate genes. Linear correlation was found to identify predominant genomic responses to a particular stimulus but was unable to identify sets of genes whose actions and interactions drive the transcriptional level of a target. This is problematic, as a gene in and of itself may not be highly correlated to a target but in combination with other genes may be predictive of the behavior of the target. For example, \textit{Ly6h} was identified as the most frequently used \textit{Lox} predictor but showed only modest correlation with \textit{Lox} as a target. Similarly, the small proline-rich proteins 2b and 2f were only poorly correlated to \textit{Abr} but rank among the top predictors of \textit{Abr} behavior.

When used individually, each predictor gene exhibited a CoD measure comparable to the value obtained using a linear correlation coefficient model. In contrast the CoD method detected multivariate nonlinear influences on gene expression in complex genetic networks and enabled the calculation of a CoD value that mathematically reflected interactions among all predictors. Use of different treatment modalities was advantageous and allowed identification of genes that move across the transcriptome in unison after activation of \textit{Abr} signaling regardless of contextual differences in gene expression or tissue-specific patterns of differentiation. For example, the constitutive and inducible expression of \textit{Cyp1b1}, a gene involved in steroidogenesis, is highly tissue specific. \textit{Cyp1b1} is expressed constitutively in the adrenal gland, ovary, and testes and is highly inducible by adrenocorticotropin, cAMP, peptide hormones, and \textit{Abr} ligands (Buters et al. 2002). \textit{Cyp1b1} is a complex gene, as \textit{Abr}-independent induction is observed in certain tissues (Kerzee and Ramos 2001), and the gene is overexpressed as a function of transformation (Vondracek et al. 2002). Similarly, \textit{Opn} is differentially expressed under constitutive and inducible conditions and interacts with surface receptors in a context-specific manner. In the normal human kidney, \textit{Cd44} is the primary receptor partner of \textit{Opn}, whereas integrin binding and activation are predominant under stressful conditions (Xie et al. 2001). This dichotomy is also observed within the context of injury, as \textit{Opn} promotes accumulation of macrophages and participates in macrophage-mediated renal injury but exerts renoprotective actions during acute ischemia. Such differences give rise to unique fingerprints that influence gene–gene interactions. Thus, combined data sets within different contextual networks increased the power of the prediction and allowed better integration of biological complexity.

As a biological target, \textit{Abr} was best predicted by lymphocyte antigen 6c (\textit{Ly6c}), a retinoic acid–responsive gene involved in T-cell activation (Mao et al. 1996; Schedlich...
and Graham 2002). Ahr has been implicated in the immune suppressant activity of aromatic hydrocarbons (Kerkvliet et al. 2002) and in the regulation of retinoic acid metabolism. Two other predictor genes for Ahr, Igfbp-3, and Igfbp-6 have also been identified as retinoic acid–regulated genes (Dailly et al. 2001; Schedlich and Graham 2002), suggesting that Ahr may couple to predictor genes through retinoic acid linkage (Andreola et al. 1997). Recent studies demonstrate that retinoids activate Ahr signaling (Soprano and Soprano 2003), whereas Ahr controls the expression of genes involved in retinoic acid metabolism (Gonzalez and Fernandez-Salguero 1998). Thus, computational strategies delineated connections between Ahr and retinoic acid that otherwise could not have been predicted in the absence of biological information.

Cyp1b1 was predicted by syk tyrosine kinase (Syk) in all cases, a relationship consistent with work by Mouh and Burchiel (1998) showing that oxidative metabolites of B[a]P increase tyrosine phosphorylation of Syk. Syk participates in integrin signaling, leading to nuclear factor-kappa B activation and increased levels of cytokine mRNAs (Lin et al. 1995) and regulation of cancer cell growth and metastasis (Coopman et al. 2000). Similarly, Ndufc1 appeared in each predictor set, with recent microarray experiments showing common responses in cancer cells for Cyp1b1 and Ndufc1 (Huang et al. 2003). Cyp1b1 catalyzes activation of molecular oxygen in an NADH-dependent electron transport pathway, so a connection with Ndufc1 and the mitochondrial electron transport chain is consistent with the two enzyme systems (Johnston and Bronsky 1995).

In several instances, the algorithm selected a predominant gene, such as Opn, when Igfbp-5 was chosen as a target. Opn is a secreted acidic phosphoprotein containing a conserved GRGDS sequence that regulates a variety of cellular processes, primarily through the αvβ3 integrin. Opn is a component of the human atherosclerotic plaque that promotes adhesion and spreading of vascular cells and is a potent chemotactic factor for vSMC (Wilson et al. 2002). The strong relationship between Igfbp-5 and Opn is in keeping with studies showing that Opn binds to Igfbp-5 with high affinity (Nam et al. 2000) and that this interaction concentrates Igfbp-5 on the matrix and modulates cooperative interactions between Igf-1 receptor and integrin αvβ3 in atherosclerotic lesions (Nichols et al. 1999). Igfbp-5 is the most conserved member of the IGFBP family and a regulator of bone, kidney, and mammary gland function. Igfbp-5 plays a decisive role in tumor cell proliferation (Schedlich and Graham 2002) and together with Opn promotes IGF-1–inducible biological effects. Other genes selected as predictors have been linked to Igfbp-5, such as signal transducer and activator of transcription (Chapman et al. 1999) and insulin (Duan et al. 1996).

For other targets, such as Lox, the connections between predictor and target gene were less clear. Lox is an extracellular and intracellular copper-containing enzyme that initiates cross-linking of collagen and elastin by catalyzing oxidative deamination of the epsilon-amino group in certain

![Diagram](https://via.placeholder.com/150)

**Figure 4.** Predictor/target relationships identified using modeling methodology. Three gene combinations were used to predict the behavior of selected target genes: (A) Ahr, (B) Cyp1b1, (C) Igfbp-5, (D) Lox, and (E) Opn. The values shown are CoD values for each individual predictor and the amplified effect of combining multiple predictors for one target.
lysine and hydroxylysine residues of colla-
gens and lysine residues of elastin. Lox is
found in smooth muscle cell nuclei and is
speculated to be involved in oxidative
deamination of peptidyl lysine (Kuviainen
et al. 1984; Li et al. 1997). As such, the
finding that matrix-related genes such as
periphrin and proliferin2 were selected as
predictors is of interest. Integrins may link
Lox and periphrin, with both collagen and
periphrin modulated by the α2β1 integrin
(Khalsa et al. 2000). The connection with
immune-related genes such as Cad44 and
Ly6b is intriguing and may involve inter-
eron. Interferon-gamma, a proinflamma-
tory cytokine, downregulates Lox gene
expression in rat aortic smooth muscle cells
(Song et al. 2000), and Ly6b has also been
identified as a target of interferon (Horie
et al. 1998). Lox and Nfatc1 are coregu-
lated under shear stress (Ando et al. 1996),
so a link between these two genes may
also exist.

The relationship between Opn and Ahr
is intriguing. Recent studies have shown
that resveratrol (3,5,4’-trihydroxystilbene),
an Ahr antagonist, modifies the inhibitory
effect of TCDD on Opn expression (Singh
et al. 2000). Opn is a ligand for the Cad44
receptor and stimulates Cad44 expression on
the osteoclast surface (Chellaiah et al.
2003; Sodek et al. 2002). The relationship
between Cad44 and Opn is well established
(Denhardt et al. 2001). The analysis also
predicted a relationship between Opn and
Acte1, a connection consistent with reports
by Shu and co-workers (2002) showing that
Opn-expressing tubuli in kidney were
tightly associated with a peritubular influx
of alpha-smooth muscle actin. It is also not
surprising that Opn expression was pre-
dicted by two growth factors such as
epiregulin and Bdnf (Lee et al. 2001; 
Toyota et al. 1995). The observed associa-
tions indicate that measurement of the
transcriptional state of multiple gene sets
can be combined to better predict the
expression level of a target gene. The
strongest interaction between targets was
the interplay between Ahr, Cyp1b1, and
Opn (Figures 2 and 3). Although these bio-
ological connections are novel, they are in
keeping with studies showing that Ahr sig-
alling is critical to Opn regulation (Singh
et al. 2000) and that ligands of the Ahr
modulate Opn expression. For some of the
target/predictor relationships, no direct
linkages could be established. An example is
Bdnf, a protein that allows the survival of
certain neuronal populations and found to
be highly predictive of Opn. As both Bdnf
and Opn are involved in matrix regulation,
the relationship could be an interesting
area for future investigation.

Conclusions

Novel putative interactions were defined to
investigate gene–gene interactions regu-
lated by the Ahr. In many instances, the
relevance of the biological relationships
uncovered using CoD methodology was
ratified by published reports, whereas in
others, novel interactions were identified.
The computational approach used afforded
us the ability to begin constructing gene
networks that define broad-ranging interac-
tive biological relationships. Although the
biological bases for these theoretical rela-
tionships must be investigated further, the
number of possible combinations is now
reduced to a manageable size that can be
systematically scrutinized using established
molecular methodologies.

References

Affymetrix, Inc. 2003. GeneChip Expression Analysis
Technical Manual, Rev. 4. Santa Clara, CA:Affymetrix, Inc.
Ando J, Tsbohi K, Korenaga R, Takahashi K, Kosaki K,
Ishikhi M, et al. 1996. Differential display and
cloning of shear stress-responsive messenger RNAs in human endothelial cells.
Biochim Biophys Res Commun 225:347–351.
Andreola F, Fernandez-Salguario PM, Chaintre MV,
Pertovich MP, Gonzalez-FJ, De Luca LM. 1997.
Aryl hydrocarbon receptor knockout mice (AHR–/–) exhibit liver retinoid accumulation and
reduced retinoid acid metabolism. Cancer Res
57:2385–2388.
Bittinger MA, Nguyen LP, Bradford CA. 2003.
Aspartate aminotransferase generates pro-
gonists of the aryl hydrocarbon receptor. Mol
Pharmacol 64:550–556.
Breitkreutz BJ, Stoy J, Tyers M. 2003. Osprey: a net-
work system that integrates published protein
interactions and systematic gene perturbations.
Bioinformatics 19:1837–1838.
Buters JT, Mahadevan B, Quintanilla-Martinez L,
Gonzalez FJ, Geim H, Baird WM, et al. 2002.
Cytocrome P450 1B1 determines susceptibility to
dibenzo(a,j)pyrene-induced tumor formation.
Chem Res Toxicol 15:1127–1135.
Carlson DB, Perdew GH. 2002. A dynamic role for the
Ah receptor in cell signaling? Insights from a
diverse group of Ah receptor interacting pro-
teins. J Biochem Mol Toxicol 16:317–325.
Carver LA, Bradford CA. 1997. Ligand-dependent
interaction of the aryl hydrocarbon receptor with
a novel immunophilin homolog in vivo. J Biol
Chem 272:11452–11456.
Chapman BS, Lauroncic PC, Tenner E, Flint DJ,
Selbert S, Takeda K, et al. 1999. Suppression of
epithelial apoptosis and delayed mammary gland
involution in mice with a conditional knockout of
Stat3. Genes Dev 13:2604–2616.
Chellaiah MA, Kizer N, Biswas R, Alvarez U, Strauss-
Schoenberger J, Rivas L, et al. 2003. Osteopontin
deficiency produces osteoclast dysfunction due to
reduced Cad44 surface expression. Mol Biol
Cell 14:173–180.
Cooper PJ, Do MT, Barth M, Bowden ET, 
Hayes AJ, Bassey E, et al. 2000. The Syk tyrosine
kinase suppresses malignant growth of human breast
cancer cells. Nature 406:742–747.
Dailly YP, Zhou Y, Linkhart TA, Baylink DJ, Strong DD.
2001. Structure and characterization of the
human insulin-like growth factor binding protein
(Igfbp)-6 promoter: identification of a functional
retinoid response element. Biochim Biophys Acta
1518:145–151.
Denhardt DT, Giachetti CM, Rittling SR. 2001. Role of
osteopontin in cellular signaling and toxicant
injury. Annu Rev Pharmacol Toxicol 41:723–749.
Denison MS, Nagy SR. 2003. Activation of the aryl
hydrocarbon receptor by structurally diverse
exogenous and endogenous chemicals. Annu
Rev Pharmacol Toxicol 43:309–334.
Devroye L, Gyorffy L, Lugosi G. 1996. A Probabilistic
Theory of Pattern Recognition. New York: 
Springer-Verlag.
Duan C, Hawes SB, Pravette T, Clemmons DR. 1996.
Insulin-like growth factor-I (IGF-I) regulates IGF-
binding protein-5 synthesis through transcrip-
tional activation of the gene in aortic smooth
muscle cells. J Biol Chem 271:4280–4288.
Falafatpisheh MH, Ramos KS. 2003. Ligand-activated
Ahr signaling leads to disruption of nephrogene-
sis and altered Wilms’ tumor suppressor mRNA
splicing. Oncogene 22:2160–2171.
Gonzalez FJ, Fernandez-Salguario P. 1998. The aryl
hydrocarbon receptor: studies using the AHR-
null mice. Drug Metab Dispos 26:1194–1198.
Horie M, Okutomi K, Taniguchi Y, Ohbuchi Y,
Suzuki M, Takahashi E. 1998. Isolation and char-
acterization of a new member of the human Ly6
gene family (Ly6h). Genomics 35:365–368.
Huang E, Ishaia S, Pittman J, Dressman H, Bild A,
Kloos M, et al. 2003. Gene expression phenotypic
models that predict the activity of oncogenic
pathways. Nat Genet 34:226–230.
Johnston MC, Bronsky PT. 1995. Prenatal craniofa-
cial development: new insights on normal and
abnormal mechanisms. Crit Rev Oral Biol Med
6:368–422.
Kerkvliet NI, Shepherd DM, Baecher-Steppan L 2002.
T lymphocytes are direct, aryl hydrocarbon
receptor (AhR)-dependent targets of 2,3,7,8-
tetrachlorodibenzo-p-dioxin (TCDD): AhR expres-
sion in both CD4+ and CD8+ T cells is necessary
for full suppression of a cytotoxic T lymphocyte
response by TCDD. Toxicol Appl Pharmaco-
185:146–152.
Kerze JK, Ramos KS. 2000. Activation of c-Ha-ras by
benz[a]pyrene in vascular smooth muscle cells
involves redox stress and aryl hydrocarbon
receptor. Mol Pharmacol 58:152–158.
2001. Constitutive and inducible expression of
Cyp1a1 and Cyp1b1 in vascular smooth muscle
cells: role of the Ahr BH1/BH3/5AS transcription
factor. Circ Res 89:573–582.
Khalsa PS, Zhang C, Sommerfeldt D, Hadijarjegyou M.
2000. Expression of integrin α2β1 in axons and
receptive endings of neurons in rat, hairy skin.
Neurosci Lett 293:13–26.
Kim S, Dougherty ER, Bittner ML, Chen Y, Sivakumar K,
Meltzer PS, et al. 2000a. Multivariate measure-
ment of gene expression relationships. Genomics
67:201–209.
2000b. General nonlinear framework for the
analysis of gene interaction via multivariate
expression arrays. J Biomed Optics 5:411–424.
Kuviainen H, Savolainen ER, Kivirikko KI. 1984.
Human placental lysyl oxidase: purification, par-
tial characterization, and preparation of two spe-
cific antisera to the enzyme. J Biol Chem
259:6996–7002.
Lee R, Kermani P, Teng KK, Hampstead BL. 2001.
Regulation of cell survival by secreted prone-
urotrophins. Science 294:1945–1948.
Li W, Nellaiappan K, Strasser T, Graham L,
Thomas KM, Kagan HM. 1997. Localization and activity of lysyl oxidase within nuclei of fibrogenic cells. Proc Nat Acad Sci USA 94:12817–12822.

Lin TH, Rosales C, Mondal K, Bolen JB, Haskill S, Juliano RL. 1995. Integrin-mediated tyrosine phosphorylation and cytokine message induction in monocyctic cells. A possible signaling role for the Syk tyrosine kinase. J Biol Chem 270:16189–16197.

Mao M, Yu M, Tong J-H, Ye J, Zhu J, Huang Q-H, et al. 1996. RIG-E, a human homolog of the murine Ly-6 family, is induced by retinoic acid during the differentiation of acute promyelocytic leukemia cell. Proc Nat Acad Sci USA 93:5910–5914.

Mounho BJ, Burchiel SW. 1998. Alterations in human B cell calcium homeostasis by polycyclic aromatic hydrocarbons: possible associations with cytochrome P450 metabolism and increased protein tyrosine phosphorylation. Toxicol Appl Pharmacol 149:80–89.

Nam TJ, Busby WH Jr, Rees C, Clemmons DR. 2000. Thrombospondin and osteopontin bind to insulin-like growth factor (IGF)-binding protein-5 leading to an alteration in IGF-I-stimulated cell growth. Endocrinology 141:1100–1106.

Nebert DW. 1994. Drug metabolism and signal transduction: possible role of Ah receptor and arachidonic acid cascade in protection from ethanol toxicity. EXS 71:231–240.

Nichols TC, du Laney T, Zheng B, Bellinger DA, Nickols GA, Engleman W, et al. 1999. Reduction in atherosclerotic lesion size in pigs by alphaVbeta3 inhibitors is associated with inhibition of insulin-like growth factor-1-mediated signaling. Circ Res 85:1040–1045.

Ramos KS, Cox LR. 1993. Aortic endothelial and smooth muscle cell cultures. Vol 1 (Tyson CA, Frazier JM, eds). In: Methods in Toxicology: In Vitro Biological Systems. St. Louis, MO:Academic Press, 159–168.

Schedlich LJ, Graham LD. 2002. Role of insulin-like growth factor binding protein-3 in breast cancer cell growth. Microsc Res Tech 59:12–22.

Shu Y, Hoshi S, Tomari S, Watanabe T, Nagata M. 2002. Phenotypic changes and cell cycle activation in early tubulointerstitial injury of rat adriamycin nephrosis. Pathol Int 52:214–223.

Singh SU, Casper RF, Fritz PC, Sukhu B, Ganss B, Girard B Jr, et al. 2000. Inhibition of dioxin effects on bone formation in vitro by a newly described aryl hydrocarbon receptor antagonist, resveratrol. J Endocrinol 167:183–195.

Sodek J, Zhu B, Huynh MH, Brown TJ, Ringuette M. 2002. Novel functions of the matricellular proteins osteopontin and osteonectin/SPARC. Connect Tissue Res 43:308–319.

Song YL, Ford JW, Gordon D, Shanley CJ. 2000. Regulation of lysyl oxidase by interferon-gamma in rat aortic smooth muscle cells. Arterioscler Thromb Vasc Biol 20:962–968.

Soprano DR, Soprano KJ. 2003. Pharmacological doses of some synthetic retinoids can modulate both the aryl hydrocarbon receptor and retinoid receptor pathways. J Nutr 133:277S–281S.

Toyoda H, Komurasaki T, Uchida D, Takayama Y, Isobe T, Okuyama T, et al. 1995. Epiregulin: a novel epidermal growth factor with mitogenic activity for rat primary hepatocytes. J Biol Chem 270:7495–7500.

Vondracek M, Weaver DA, Sarang Z, Hedberg JJ, Willey JC, Warngard L, et al. 2002. Transcript profiling of enzymes involved in detoxification of xenobiotics and reactive oxygen in human normal and simian virus 40 T antigen-immortalized oral keratinocytes. Int J Cancer 99:776–782.

Wilson E, Parrish AR, Bral CM, Williams ES, Ramos KS. 2002. Collagen suppresses the proliferative phenotype of allylamine-injured vascular smooth muscle cells. Atherosclerosis 162:289–297.

Xie Y, Sakatsume M, Nishi S, Narita I, Arakawa M, Gejyo F. 2001. Expression, roles, receptors, and regulation of osteopontin in the kidney. Kidney Int 60:1645–1657.