Exploiting Genome Data to Understand the Function, Regulation, and Evolutionary Origins of Toxicologically Relevant Genes

Nazzareno Ballatori,[1,2] James L. Boyer,[2,3] and John C. Rockett[4]

[1]Department of Environmental Medicine, University of Rochester School of Medicine, Rochester, New York, USA; [2]Mount Desert Island Biological Laboratory, Salisbury Cove, Maine, USA; [3]Liver Center, Department of Medicine, Yale University School of Medicine, New Haven, Connecticut, USA; [4]Reproductive Toxicology Division, National Health and Environmental Effects Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, USA

The wealth of new information coming from the many genome sequencing projects is providing unprecedented opportunities for major advances in all areas of biology, including the environmental health sciences. To facilitate this discovery process, experts in the fields of functional genomics and informatics and the emerging field of toxicogenomics recently gathered at the Mount Desert Island Biological Laboratory in Salisbury Cove, Maine, site of a National Institute of Environmental Health Sciences Marine and Freshwater Biomedical Science Center, to share their ideas and latest research findings. The goal of the symposium was to highlight approaches that may be used to identify and characterize toxicologically relevant genes being discovered in the genome sequencing projects. Many of the approaches rely heavily on comparative models as a way of identifying gene homology, ontology, and physiologic function, and on the availability of databases that facilitate storage, analysis, interpretation, and widespread dissemination of relevant data. *Key words:* comparative genomics, databases, mechanisms of toxicity, nontraditional animal models, toxicogenomics. *Environ Health Perspect* 111:871–875 (2003). doi:10.1289/txg.5961 available via *http://dx.doi.org/* [Online 5 November 2002]

This is a report on the Ninth Mount Desert Island Biological Laboratory (MDIBL) Environmental Health Sciences Symposium "Exploiting Genome Data to Understand the Function, Regulation and Evolutionary Origins of Toxicologically Relevant Genes" held 10–11 July 2002 at the MDIBL in Salisbury Cove, Maine, USA. The Symposium was sponsored by the National Institute of Environmental Health Sciences (NIEHS) Center at the MDIBL, the National Center for Research Resources, the Yale University Liver Center, the Kinter Memorial Lectureship Fund, and the MDIBL.

Toxicogenomics is a relatively new and developing field of study that combines clinical, genomic, and proteomic data into a unified framework for understanding the biochemical and genetic basis for various diseases (Hamadeh et al. 2002a; Olden et al. 2001; Simmons and Portier 2002; Tennant 2002). Although toxicogenomics has sometimes been defined rather blandly as a combination of toxicology and genomics, most investigators realize it involves much more. A more apt description is provided by Selkirk and Tennant (2002):

Toxicogenomics is a new scientific field that elucidates how the entire genome is involved in biological responses of organisms exposed to environmental toxicants/stressors. It combines information from studies of genomic-scale mRNA profiling, cell-wide or tissue-wide protein profiling (proteomics), genetic susceptibility, and computational models to understand the roles of gene-environment interactions in disease.

J.K. Selkirk (NIEHS, Research Triangle Park, North Carolina, USA) reflected the view of many when he suggested the scientific community is on the threshold of extremely important discoveries in understanding mechanisms of disease development and progression. He gave much of the credit to the various genome programs ongoing over the past few years. Selkirk described two papers published recently by members of the National Center for Toxicogenomics (NCT) of NIEHS (Table 1). In one article, Hamadeh et al. (2002b) compared gene expression profiles in the livers of rats fed with several compounds (Clofibrate, Wyeth 14,643, gemfibrozil) from a common class of toxicants, the peroxisome proliferators, and one (phenobarbital) from a different class, the enzyme inducers. Using clustering, simple correlation, or principal component analyses, this study demonstrates cDNA microarrays can be used to generate chemical-specific gene expression profiles that can be distinguished across and within these classes. The expression profiles were used as a training set to identify highly discriminant genes using linear discriminant analysis and genetic algorithm/K-nearest neighbors. Using these genes in the analysis of blinded liver RNA samples exposed to phenytoin, diethylhexylphthalate, or hexobarbital allowed the

authors to successfully predict whether these samples were derived from livers of rats exposed to enzyme inducers or peroxisome proliferators (Hamadeh et al. 2002c). Selkirk went on to discuss the strategy of the NCT and how short-term objectives such as the aforementioned prediction assays (signature patterns of exposure and adverse effects) would eventually lead to gene expression databases with advanced query tools, relational interfaces, and comprehensive annotation. The efforts of the NCT will be assisted by the formation of the NIEHS-funded Toxicogenomics Research Consortium (TRC; *http://ehpnet1.niehs.nih.gov/docs/ 2002/110-2/extram-speaking.html*), a program designed for advancing environmental health sciences research into the frontier of toxicogenomics research. The overall goal of the NCT and TRC is to conduct a coordinated, multidisciplinary, multi-institutional effort to define how the entire genetic complement of relevant organisms responds to environmental agents, including chemicals, physical agents, and physiologic stresses.

The power and limitations of gene expression profiling were also described in presentations by J.C. Rockett (U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, USA) and C.A. Bradfield (McArdle Laboratory, University of Wisconsin, Madison, Wisconsin, USA). Gene expression profiling has been used increasingly as a tool for monitoring the impact of toxicants on human

and wildlife populations. The profiling ("fingerprinting") of genes expressed in a given cell, population of cells, tissue, or organ promises a number of uses including *a*) providing a tool for discriminating between classes of toxicants; *b*) assisting in the early detection of toxicant exposure; *c*) helping to elucidate mechanisms or modes of action of environmental toxicants in individual species; and *d*) identifying common modes of action across species.

Bradfield's laboratory has developed an approach to classify toxicants on the basis of their influence on profiles of mRNA transcripts (Thomas et al. 2001). Changes in liver gene expression were examined after exposure of mice to 24 model treatments that fall into five well-studied toxicologic categories: peroxisome proliferators, aryl hydrocarbon receptor agonists, noncoplanar polychlorinated biphenyls, inflammatory agents, and hypoxia-inducing agents. Analysis of 1,200 transcripts using

both a correlation-based approach and a probabilistic approach resulted in a classification accuracy between 50 and 70%. However, a probabilistic approach based on Bayesian statistics was used to identify a diagnostic set of 12 transcripts that provided an estimated 100% predictive accuracy based on "leave-one-out cross-validation." Expansion of this approach to additional chemicals of regulatory concern could serve as an important screening step in a new era of toxicologic testing.

However, despite such encouraging findings, the full potential of microarray-based approaches has yet to be realized. Many obstacles must be overcome before gene expression profiling data can be applied meaningfully to such areas as clinical diagnosis and risk assessment. Some of these obstacles involve the technology itself, although perhaps the most significant problems lie in data analysis.

The consensus was that the future of toxicogenomics is very much dependent on the development of advanced database systems and query tools for depositing, storing, and mining data. The Jackson Laboratory (Bar Harbor, Maine, USA) already knows the importance and power of databases, and J.T. Eppig (The Jackson Laboratory) provided an overview of their important and internationally recognized Mouse Genome Informatics (MGI) database (Table 1). The genetics of the mouse are extremely well known; so much so that, as the title of Eppig's talk suggested, we can use the mouse genome to learn about ourselves. To fully realize the power of the mouse as a genetic model depends on making integrated genetic, genomic, and phenotypic information available to promote knowledge discovery, and the goal of the Jackson Laboratory is to facilitate use of the mouse as a genetic model for human biology. As such, the MGI database contains information on topics such as gene characterization, maps (genetic, cytogenetic, physical, and comparative), comparative genomic data, various phenotypic and genetic variants (mutants, polymorphisms, and strains), and access to an e-mail user group. One of the most important features of the MGI database is data are not just deposited but fully integrated. It has further increased its utility through the implementation of comprehensive gene ontology (GO) descriptions. D.P. Hill, also from the Jackson Laboratory, discussed how the Jackson Laboratory has contributed to the GO Consortium (Table 1). The problem with most data sets is their interpretation is a human endeavor, and people say things differently, use different words for

Table 1. Websites discussed during the Ninth Mount Desert Island Biological Laboratory Environmental Health Sciences Symposium.[a]

Gene Ontology (GO) Consortium
(*http://www.geneontology.org*)
[accessed 4 September 2002].
The goal of the Gene Ontology Consortium is to produce a dynamic controlled vocabulary that can be applied to all organisms, even as knowledge of gene and protein roles in cells is accumulating and changing.

Comparative Mouse Genomics Centers Consortium (CMGCC)
(*http://www.niehs.nih.gov/cmgcc/home.htm*)
[accessed 4 September 2002]
CMGCC was initiated by the Environmental Genome Project to develop transgenic and knockout mouse models based on human DNA sequence variants in environmentally responsive genes. These mouse models are tools to improve understanding of the biologic significance of human DNA polymorphisms.

Cytoscape
(*http://www.cytoscape.org*)
[accessed 4 September 2002]
Cytoscape is a bioinformatics software platform for visualizing molecular interaction networks and integrating these interactions with gene expression profiles and other state data. Cytoscape provides an algorithm that filters through a network to identify the particular cellular signaling and regulatory pathways that control the changes in gene expression observed by microarray.

Environmental Genome Project (EGP)
(*http://www.niehs.nih.gov/envgenom/home.htm*)
[accessed 4 September 2002]
The mission of the EGP is to improve understanding of human genetic susceptibility to environmental exposures. The rationale of the EGP is certain genes have a greater than average influence over human susceptibility to environmental agents. If we identify and characterize the polymorphism in those genes, we will increase our understanding of human disease susceptibility. This knowledge can be used to protect susceptible individuals from disease and to reduce adverse exposure and environmentally induced disease.

Mount Desert Island Biological Laboratory (MDIBL)
(*http://www.mdibl.org/*)
[accessed 4 September 2002]
The mission of MDIBL is to promote research and education in the biology of marine organisms, to foster understanding and preservation of the environment, and to advance human health. The MDIBL is home to the Center for Membrane Toxicity Studies, an NIEHS-supported Marine and Freshwater Biomedical Sciences Center.

Mouse Genome Informatics (MGI) database
(*www.informatics.jax.org*)
[accessed 4 September 2002]
MGD includes information on mouse genetic markers, molecular segments (probes, primers and yeast artificial chromosomes), phenotypes, comparative mapping data, graphic displays of linkage, cytogenetic, and physical maps, experimental mapping data, and strain distribution patterns for recombinant inbred strains and cross-haplotypes.

National Center for Toxicogenomics (NCT)
(*http://www.niehs.nih.gov/nct/home.htm*)
[accessed 4 September 2002]
NIEHS has established the NCT to coordinate an international research effort to develop the field of toxicogenomics. The NCT will provide a unified strategy and a public database and will develop the informatics infrastructure to promote the development of the field of toxicogenomics. NIEHS will pay special attention to toxicogenomics as applied to the prevention of environmentally related diseases.

[a]Summaries taken from websites.

the same thing, and generate different conceptualizations.

The objective of the GO Consortium is to develop a shaped language for annotation of molecular characteristics across organisms to achieve a mutual understanding of the definitions and meaning of word use. In this way, databases conforming to GO annotation rules will be able to communicate with one another and will be compatible for unified mining by software such as Cytoscape (Table 1). L.D. Samson (MIT, Cambridge, Massachusetts, USA) described this bioinformatics software platform, designed by T. Ideker (MIT), as a tool for "visualizing molecular interaction networks and integrating these interactions with gene expression profiles and other state data." Its key feature is an algorithm that finds active pathways, i.e., subnetworks of genes that jointly show significant differential expression over a set of experimental conditions observed by microarray experiment. The identification of signaling and regulatory pathways will play an important role in risk assessments, as the regulation of a single gene means very little. Genes such as *p53*, for example, are involved in multiple independent cellular processes that may lead to the life or death of a cell, depending on a number of other factors. Thus, interpreting toxicogenomic data will require complex computer analysis using algorithms not yet developed that apply toxicogenomic data to gene expression networks.

For the William B. Kinter memorial lecture, the Deputy Director of NIEHS, Dr. Samuel H. Wilson, gave an overview of the contributions of the Environmental Genome Project (EGP) (Table 1) to the field of environmental genomics. The EGP was launched in 1998 by NIEHS to improve understanding of human genetic susceptibility to environmental exposures. It explores the hypothesis that risk for common diseases is a function of disease-predisposing alleles in environmental-responsive genes combined with specific exposures. The knowledge obtained through the EGP will be used to improve risk assessment and provide a science-based framework for development of environmental policy. The ultimate goal of the EGP, and of policies based on EGP research, is to prevent adverse effects from environmental exposure and to protect individuals who are at risk for these adverse events. The initial objectives of the EGP were to discover gene variants (polymorphisms) in environmentally responsive genes and characterize their structure–function relationship to better understand genetic susceptibility. The program has made significant progress

toward achieving this target: more than 2,000 single nucleotide polymorphisms have been found on 123 genes. With such information in hand, the next steps are to study gene–environment interactions in mouse and cell models. To this end the EGP has established the Comparative Mouse Genomics Centers Consortium (CMGCC) (Table 1) to use transgenic mouse models to discover the functional significance of human DNA polymorphism in environmentally responsive genes.

Mice are not the only models that can tell us about ourselves. There is also keen interest in carrying out comparative genomics with alternative models, including many marine and freshwater species. This interest has been fueled by an increasing realization that the Earth's aquatic environments harbor a rich diversity of organisms, many of which provide invaluable models for human diseases. Although over 80% of the Earth's earth's living organisms exist only in such ecosystems, we know relatively little about these creatures; not surprisingly, only a few have been used as experimental models in biomedical research. Foremost among these are the fish, and the many advantages of fish models have been summarized in reviews by Powers (1989) and, more recently, by Ballatori and Villalobos (2002). Fish are represented by groups of morphologically diverse animals that inhabit a wide variety of ecosystems. Specialized adaptations allow them to thrive in environments that differ significantly in terms of oxygen tension, temperature, salinity, atmospheric pressure, light intensity, predator density, and chemical concentrations, including relatively high concentrations of man-made chemicals (Powers 1989). Unraveling the fundamental mechanisms responsible for these adaptations in ion regulation, respiration, chemical signaling, and reproduction is likely to provide important insight into normal human physiology and into the etiology of environmentally related diseases. Moreover, these specialized traits and adaptations can be exploited to address specific research questions (Ballatori and Villalobos 2002).

The oceans also play a wider role in human health. More oxygen is provided to the Earth's atmosphere by oceanic photosynthesis than by all terrestrial plants combined. Oceans are essential components of the water cycle. They are the main source of food for a significant portion of the earth's population. They receive and break down much of our waste. They are considered the richest potential source of new drugs to treat human disease. In recognition of these and other significant roles of oceans in

human health, NIEHS is planning to establish centers of excellence in "Oceans and Human Health." This global initiative, first drafted after a 1999 meeting at the Bermuda Biological Station for Research (2002), has overall goals of the early detection of potential marine-based contaminants, prevention of associated human illness, and the development of products to enhance human well-being. Dr. Wilson suggested this initiative could be expanded to include the type of genomic studies being carried out on marine animals at laboratories such as the MDIBL (Table 1).

One of the most important uses of marine animals is derivation of genomic information for use in comparative genomic studies. For example, J.L. Boyer (Yale University School of Medicine, New Haven, Connecticut, USA) has used the little skate (*Raja erinacea)* for studying the bile salt export pump (BSEP; Cai et al. 2001) and an organic anion transporting polypeptide (Cai et al. 2002). The BSEP regulates bile salt–dependent bile flow and is important in the elimination of many endogenous and exogenous steroids. Mutations in the human BSEP gene result in a form of liver disease called progressive familial intrahepatic cholestasis, or PFIC type II. Boyer's laboratory has found that all mutations in BSEP in patients with PFIC type II disease are in conserved regions of human, rat, mouse, and skate genes, showing how comparative genomics can be used to identify functionally important regions of DNA sequences and provide insights into human disease.

Comparative genomics is also beginning to exert a major influence in the field of toxicogenomics. A reasonable effort is being directed toward examining the way environmental toxicants regulate gene expression and how this might be turned into a way of monitoring the extent and degree of impact of such toxicants on environmental systems. Thus, for example, Hemmer et al. (2001) recently developed a sheepshead minnow bioassay to detect hepatic vitellogenin mRNA in male sheepshead minnows, suggesting a mechanism by which these small coastal fishes can be used as a sentinel species to detect contamination with estrogenic compounds. In the same way, meeting participants learned how Drs. William S. Baldwin and Lisa J. Bain from the University of Texas at El Paso are studying winter flounder (*Pseuopleuronectes americanus*) and mummichog (*Fundulus heteroclitus*), respectively, to discover genes regulated by chromium exposure. These studies are being carried out using

differential display and suppression polymerase chain reaction subtractive hybridization—techniques for identifying differentially expressed genes. In contrast to DNA arrays, which currently consume most of the interest and headlines in the genomics community, these techniques are of greater use to people working with species such as these marine fish, whose genomes have not been well characterized. Unlike arrays, these techniques do not require gene sequence data to identify genes differentially regulated by the model exposure. Other than the Japanese puffer fish *Fugu rubripes*, very little sequence data are available for marine fish. Thus, techniques such as differential display and subtractive hybridization are ideal for pursuing genes involved in models where limited genetic information is available. In their preliminary studies, Baldwin and Bain discovered a number of genes that appear to be regulated by chromium exposure; however, they also reported it can sometimes be difficult to discover the identity or function of newly discovered gene sequences. Several expressed sequence tags were identified that were altered in chromium-treated animals but appear to have no similarities with sequences deposited in GenBank (*http://www.ncbi.nlm.nih.gov/Genbank/GenbankSearch.html*).

The increasing volume of work being conducted on marine species and the increasing importance of comparative genomics have prompted the MDIBL and its NIEHS-supported Marine and Freshwater Biomedical Science Centers to begin development of the Comparative Toxicogenomics Database (CTD). This project, supported by NIEHS as part of its program in toxicogenomics, is led by two investigators at the MDIBL, C. Mattingly and G. Colby, and is directed by J.L. Boyer. The aim is to develop a comprehensive database to compare gene sequences and functions between humans, laboratory animals such as rats and mice, and aquatic organisms. It will be the first community-based, publicly accessible database devoted to genes of human toxicologic significance and will include much useful information: genomic data (e.g., ontology, DNA and protein sequences, and tissue expression data), literature references, toxicants associated with specific genes, a repository for reagents and associated contact information, comparative analysis tools, and integration with related databases such as PubMed (*http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed*), GenBank, OMIM (Online Mendelian Inheritance in Man) (*http://www.ncbi.nlm.nih.gov/omim*),

and TOXNET (*http://toxnet.nlm.nih.gov*). It is anticipated the CTD will provide novel insights into the dynamics of gene–environment interaction and human health. CTD is slated to go online in 2006, but Mattingly and Colby predict a limited version of the database may be available for collaborators and registered users in as little as 2 years' time. In the meantime, they encourage interested parties to contact them (ctd@mdibl.org) for information or comments on developing the database to fully meet the needs of the toxicogenomics community.

Marine species are useful for comparative genomic analyses and also provide excellent models for basic experimental studies. For example, isolated brain capillaries are often used to study blood–brain barrier function and its regulation. A major shortcoming of capillaries isolated from mammalian species is they are short lived. However, D.S. Miller (NIEHS) and G. Fricker (University of Heidelberg, Heidelberg, Germany) found brain capillaries isolated from killifish or dogfish shark (*Squalus acanthias*) are suitable and relatively long-lived models for mammalian systems (Miller et al. 2002). Another example is the rectal gland of the spiny dogfish (*S. acanthias*), which has long been used as a model for the mechanism of secondary active chloride transport in epithelial organs throughout the vertebrate kingdom. The rectal gland consists of a homogenous population of tubules composed predominantly of a single cell type whose major function is sodium and chloride secretion. The content of Na^+,K^+-ATPase and the $Na^+,K^+,2Cl^-$ cotransporter exceeds that of mammalian kidney by almost two orders of magnitude. Techniques have been established to study various aspects of chloride transport in intact dogfish shark, as well as *in vitro,* using the isolated perfused rectal gland, primary rectal gland cell cultures, and isolated membrane vesicles (Forrest 1996).

The hepatic uptake of organic solutes is a fundamental property of all vertebrates, and N. Ballatori (University of Rochester School of Medicine, Rochester, New York) has been using another cartilaginous fish, the small skate (*R. erinacea*), as a means to search for novel organic solute transporters. By using *Xenopus laevis* oocytes to screen a skate liver cDNA library for taurochlorate and estrone sulfate transport activity, his group recently identified a novel type of organic solute and steroid transporter (Wang et al. 2001). In contrast to other such transporters, transport activity requires the coexpression of two distinct gene products (Ostα and Ostβ). His group

is now investigating the mechanism by which these two gene products interact to generate transport activity and are searching for possible human orthologs of these skate genes.

Summary

The preliminary draft sequence of the human genome was recently completed. This first draft has been recognized as a landmark achievement and has produced a tremendous amount of information on the human genome, including information on toxicologically relevant genes. Nevertheless, it is one of the greatest truths of our age that information does not equate to knowledge, and there is still a lot to learn about the human genome, including the precise number of encoded genes, their biologic functions, and their relevance to toxicology. It has been opined that one of the best ways to learn about the human genome is through comparative genomics. This approach, which aims to increase our understanding of human and animal physiology through the comparison of gene sequence and function from unrelated species, will become increasingly important as more genomes are sequenced. The structural and functional characteristics of many genes and proteins are often remarkably conserved across kingdoms and phyla. Thus, comparisons of genomic and amino acid sequences can provide new and useful insights into gene and protein structure and function. The main message from this meeting was that toxicogenomic and physiologic studies into human biology are greatly enhanced by the use of marine and other nontraditional animal models. In addition, the vast amount of information generated by contemporary genomic studies offers more information than most researchers are interested in or can deal with. The key to storage, analysis, interpretation, and widespread dissemination of such data appears to lie in publicly accessible integrated databases containing fully annotated and standardized data.

REFERENCES

Ballatori N, Villalobos AR. 2002. Defining the molecular and cellular basis of toxicity using comparative models. Toxicol Appl Pharmacol 183:207–220.

The Bermuda Station for Research. Available: *http://www.bbsr.edu/index.html* [accessed 4 September 2002]

Cai SY, Wang L, Ballatori N, Boyer JL. 2001. Bile salt export pump is highly conserved during vertebrate evolution and its expression is inhibited by PFIC type II mutations. Am J Physiol Gastrointest Liver Physiol 281:G316–G322.

Cai SY, Wang W, Soroka CJ, Ballatori N, Boyer JL.

2002. An evolutionarily ancient Oatp: insights into conserved functional domains of these proteins. Am J Physiol Gastrointest Liver Physiol 282:G702–G710.

Forrest JN Jr. 1996. Cellular and molecular biology of chloride secretion in the shark rectal gland: regulation by adenosine receptors. Kid Int 49:1557–1562.

Hamadeh HK, Amin RP, Paules RS, Afshari CA. 2002a. An overview of toxicogenomics. Curr Issues Mol Biol 4:45–56.

Hamadeh HK, Bushel PR, Jayadev S, Martin K, DiSorbo O, Sieber S, et al. 2002b. Gene expression analysis reveals chemical-specific profiles. Toxicol Sci 67:219–231.

Hamadeh HK, Bushel PR, Jayadev S, DiSorbo O, Bennett L, Li L, et al. 2002c. Prediction of compound signature using high density gene expression profiling. Toxicol Sci 67:232–240.

Hemmer MJ, Hemmer BL, Bowman CJ, Kroll KJ, Folmar LC, Marcovich D, et al. 2001. Effects of p-nonylphenol, methoxychlor, and endosulfan on vitellogenin induction and expression in sheepshead minnow (*Cyprinodon variegatus*). Environ Toxicol Chem 20:336–343.

Miller DS, Graeff C, Droulle L, Fricker S, Fricker G. 2002. Xenobiotic efflux pumps in isolated fish brain capillaries. Am J Physiol 282:R191–R198.

Olden K, Guthrie J, Newton S. 2001. A bold new direction for environmental health research. Am J Public Health 91:1964–1967.

Powers DA. 1989. Fish as model systems. Science 246:352–358.

Selkirk J, Tenant RW. Gordon Research Conference on Toxicogenomics. Available:*http://www.grc.uri.edu/programs/2003/toxico.htm* [accessed 22 July 2002].

Simmons PT, Portier CJ. 2002. Toxicogenomics: the new frontier in risk analysis. Carcinogenesis 23:903–905.

Tennant RW. 2002. The National Center for Toxicogenomics: using new technologies to inform mechanistic toxicology. Environ Health Perspect 110:A8–A10.

Thomas RS, Rank DR, Penn SG, Zastrow GM, Hayes KR, Pande K, et al. 2001. Identification of toxicologically predictive gene sets using cDNA microarrays. Mol Pharmacol 60:1189–1194.

Wang W, Seward DJ, Li L, Boyer JL, Ballatori N. 2001. Expression cloning of two genes that together mediate organic solute and steroid transport in the liver of a marine vertebrate. Proc Natl Acad Sci U S A 98:9431–9436.