# Related Reading

**Francis S. Collins, George M. Gray, John R. Bucher. 2008 Transforming Environmental Health Protection. *Science*. 319: 906-907**

http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=2679521&blobtype=pdf&tool=pmcentrez

In 2005, the U.S. Environmental Protection Agency (EPA), with support from the U.S. National Toxicology Program (NTP), funded a project at the National Research Council (NRC) to develop a long-range vision for toxicity testing and a strategic plan for implementing that vision. Both agencies wanted future toxicity testing and assessment paradigms to meet evolving regulatory needs. Challenges include the large numbers of substances that need to be tested and how to incorporate recent advances in molecular toxicology, computational sciences, and information technology; to rely increasingly on human as opposed to animal data; and to offer increased efficiency in design and costs (1–5). In response, the NRC Committee on Toxicity Testing and Assessment of Environmental Agents produced two reports that reviewed current toxicity testing, identified key issues, and developed a vision and implementation strategy to create a major shift in the assessment of chemical hazard and risk (6,7). Although the NRC reports have laid out a solid theoretical rationale, comprehensive and rigorously gathered data (and comparisons with historical animal data) will determine whether the hypothesized improvements will be realized in practice. For this purpose, NTP, EPA, and the National Institutes of Health Chemical Genomics Center (NCGC) (organizations with expertise in experimental toxicology, computational toxicology, and high-throughput technologies, respectively) have established a collaborative research program. PMID: 18276874; PMCID: PMC2679521

**Allan P. Davis, Cynthia G. Murphy, Michael C. Rosenstein, Thomas C. Wiegers and Carolyn J. Mattingly. 2008 The Comparative Toxicogenomics Database facilitates identification and understanding of chemical-gene-disease association: arsenic as a case study. *BMC Medical Genomics*. 1 *(48)*: 1-12**

http://www.biomedcentral.com/content/pdf/1755-8794-1-48.pdf

The etiology of many chronic diseases involves interactions between environmental factors and genes that modulate physiological processes. Understanding interactions between environmental chemicals and genes/proteins may provide insights into the mechanisms of chemical actions, disease susceptibility, toxicity, and therapeutic drug interactions. The Comparative Toxicogenomics Database (CTD; http://ctd.mdibl.org) provides these insights by curating and integrating data describing relationships between chemicals, genes/proteins, and human diseases. To illustrate the scope and application of CTD, we present an analysis of curated data for the chemical arsenic. Arsenic represents

a major global environmental health threat and is associated with many diseases. The mechanisms by which arsenic modulates these diseases are not well understood. METHODS: Curated interactions between arsenic compounds and genes were downloaded using export and batch query tools at CTD. The list of genes was analyzed for molecular interactions, Gene Ontology (GO) terms, KEGG pathway annotations, and inferred disease relationships. RESULTS: CTD contains curated data from the published literature describing 2,738 molecular interactions between 21 different arsenic compounds and 1,456 genes and proteins. Analysis of these genes and proteins provide insight into the biological functions and molecular networks that are affected by exposure to arsenic, including stress response, apoptosis, cell cycle, and specific protein signaling pathways. Integrating arsenic-gene data with gene-disease data yields a list of diseases that may be associated with arsenic exposure and genes that may explain this association. CONCLUSION: CTD data integration and curation strategies yield insight into the actions of environmental chemicals and provide a basis for developing hypotheses about the molecular mechanisms underlying the etiology of environmental diseases. While many reports describe the molecular response to arsenic, CTD integrates these data with additional curated data sets that facilitate construction of chemical-gene-disease networks and provide the groundwork for investigating the molecular basis of arsenic-associated diseases or toxicity. The analysis reported here is extensible to any environmental chemical or therapeutic drug.

**Julia M. Gohlke, Reuben Thomas, Yonqing Zhang, Michael C. Rosenstein, Allan P. Davis, Cynthia Murphy, Kevin G. Becker, Carolyn J. Mattingly and Christopher J. Portier. 2009 Genetic and environmental pathways to complex diseases.** ***BMC Systems Biology*** **3** ***(46)*****: 1-15**

http://www.biomedcentral.com/content/pdf/1752-0509-3-46.pdf

Pathogenesis of complex diseases involves the integration of genetic and environmental factors over time, making it particularly difficult to tease apart relationships between phenotype, genotype, and environmental factors using traditional experimental approaches. RESULTS: Using gene-centered databases, we have developed a network of complex diseases and environmental factors through the identification of key molecular pathways associated with both genetic and environmental contributions. Comparison with known chemical disease relationships and analysis of transcriptional regulation from gene expression datasets for several environmental factors and phenotypes clustered in a metabolic syndrome and neuropsychiatric subnetwork supports our network hypotheses. This analysis identifies natural and synthetic retinoids, antipsychotic medications, Omega 3 fatty acids, and pyrethroid pesticides as potential environmental modulators of metabolic syndrome phenotypes through PPAR and adipocytokine signaling and organophosphate pesticides as potential environmental modulators of neuropsychiatric phenotypes. CONCLUSION: Identification of key regulatory pathways that integrate genetic and environmental modulators define disease associated targets that will allow for

efficient screening of large numbers of environmental factors, screening that could set priorities for further research and guide public health decisions.

**David B Hill, Vinay Swaminathan, Ashley Estes, Jeremy Cribb, E. Timothy O'Brien, C. William Davis and R. Superfine. Force Generation and Dynamics of Individual Cilia under External Loading.** ***In Review***

Motile cilia are unique multimotor systems that display coordination and periodicity while imparting forces to biological fluids. They play important roles in normal physiology, and ciliopathies are implicated in a growing number of human diseases. In the present work we measure the response of individual human airway cilia to calibrated forces transmitted via spot-labeled magnetic microbeads. Cilia respond to applied forces by: (1) a reduction in beat amplitude, up to an 85% reduction by 160-170 pN of force; (2) a decreased tip velocity proportionate to applied force; and (3) no significant change in beat frequency. Tip velocity reduction occurred in each beat direction, independent of the direction of applied force, indicating that the cilium is "driven" in both directions at all times. Through a quasistatic force model, we deduce that axoneme stiffness is dominated by the rigidity of the microtubules, and that cilia can exert 62 ± 18pN of force at the tip via the generation of 5.6 ± 1.6pN/dynein head.

**Ruili Huang, Noel Southall, Ming-Hsuang Cho, Menghang Xia, James Inglese, and Christopher P. Austin. 2008 Characterization of Diversity in Toxicity Mechanism Using in Vitro Cytoxicity Assays in Quantitative High Throughput Screening.** ***Chemical Research in Toxicology*****. 21: 659-667**

http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=2668196&blobtype=pdf

Assessing the potential health risks of environmental chemical compounds is an expensive undertaking that has motivated the development of new alternatives to traditional in vivo toxicological testing. One approach is to stage the evaluation, beginning with less expensive and higher throughput in vitro testing before progressing to more definitive trials. In vitro testing can be used to generate a hypothesis about a compound's mechanism of action, which can then be used to design an appropriate in vivo experiment. Here we begin to address the question of how to design such a battery of in vitro cell-based assays by combining data from two different types of assays, cell viability and caspase activation, with the aim of elucidating the mechanism of action. Because caspase activation is a transient event during apoptosis, it is not possible to design a single end-point assay protocol that would identify all instances of compound-induced caspase activation. Nevertheless, useful information about compound mechanism of action can be obtained from these assays in combination with cell viability data. Unsupervised clustering in combination with Dunn's cluster validity index is a robust method for identifying mechanisms of action without requiring any a priori knowledge

about mechanisms of toxicity. The performance of this clustering method is evaluated by comparing the clustering results against literature annotations of compound mechanisms. PMID: 18281954; PMCID: PMC2668196

**Richard Judson, Ann Richard, David J. Dix, Keith Houck, Matthew Martin, Robert Kavlock, Vicki Dellarco, Tala Henry, Todd Holderman, Philip Sayre, Shirlee Tan, Thomas Carpenter, and Edwin Smith. 2009 The Toxicity Data Landscape for Environmental Chemicals. *Environmental Health Perspectives*. 117*(5)*: 685-695**

http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=2685828&blobtype=pdf&tool=pmcentrez

**Objective:** Thousands of chemicals are in common use, but only a portion of them have undergone significant toxicologic evaluation, leading to the need to prioritize the remainder for targeted testing. To address this issue, the U.S. Environmental Protection Agency (EPA) and other organizations are developing chemical screening and prioritization programs. As part of these efforts, it is important to catalog, from widely dispersed sources, the toxicology information that is available. The main objective of this analysis is to define a list of environmental chemicals that are candidates for the U.S. EPA screening and prioritization process, and to catalog the available toxicology information. **Data sources:** We are developing ACToR (Aggregated Computational Toxicology Resource), which combines information for hundreds of thousands of chemicals from > 200 public sources, including the U.S. EPA, National Institutes of Health, Food and Drug Administration, corresponding agencies in Canada, Europe, and Japan, and academic sources. **Data extraction:** ACToR contains chemical structure information; physical–chemical properties; *in vitro* assay data; tabular *in vivo* data; summary toxicology calls (e.g., a statement that a chemical is considered to be a human carcinogen); and links to online toxicology summaries. Here, we use data from ACToR to assess the toxicity data landscape for environmental chemicals. **Data synthesis:** We show results for a set of 9,912 environmental chemicals being considered for analysis as part of the U.S. EPA ToxCast screening and prioritization program. These include high- and medium-production-volume chemicals, pesticide active and inert ingredients, and drinking water contaminants. **Conclusions:** Approximately two-thirds of these chemicals have at least limited toxicity summaries available. About one-quarter have been assessed in at least one highly curated toxicology evaluation database such as the U.S. EPA Toxicology Reference Database, U.S. EPA Integrated Risk Information System, and the National Toxicology Program. PMID: 19479008; PMCID: PMC2685828

**Matthew T. Martin, Richard S. Judson, David M. Reif, Robert J. Kavlock, and David J. Dix. 2009 Profiling Chemicals Based on Chronic Toxicity Results from the U.S. EPA ToxRef Database. *Environmental Health Perspectives*. 117 *(3)*: 392-399**

http://www.ehponline.org/members/2008/0800074/0800074.pdf

**Background:** Thirty years of pesticide registration toxicity data have been historically stored as hardcopy and scanned documents by the U.S. Environmental Protection Agency (EPA). A significant portion of these data have now been processed into standardized and structured toxicity data within the EPA's Toxicity Reference Database (ToxRefDB) , including chronic, cancer, developmental, and reproductive studies from laboratory animals. These data are now accessible and mineable within ToxRefDB and are serving as a primary source of validation for U.S. EPA's ToxCast research program in predictive toxicology. **Objectives:** We profiled *in vivo* toxicities across 310 chemicals as a model application of ToxRefDB, meeting the need for detailed anchoring end points for development of ToxCast predictive signatures. **Methods:** Using query and structured data-mining approaches, we generated toxicity profiles from ToxRefDB based on long-term rodent bioassays. These chronic/cancer data were analyzed for suitability as anchoring end points based on incidence, target organ, severity, potency, and significance. **Results:** Under conditions of the bioassays, we observed pathologies for 273 of 310 chemicals, with greater preponderance (> 90%) occurring in the liver, kidney, thyroid, lung, testis, and spleen. We observed proliferative lesions for 225 chemicals, and 167 chemicals caused progression to cancer-related pathologies. **Conclusions:** Based on incidence, severity, and potency, we selected 26 primarily tissue-specific pathology end points to uniformly classify the 310 chemicals. The resulting toxicity profile classifications demonstrate the utility of structuring legacy toxicity information and facilitating the computation of these data within ToxRefDB for ToxCast and other applications.

**Carolyn J. Mattingly, Thomas H. Hampton, Kimberly M. Brothers, Nina E. Griffin, and Antonio Planchart. 2009 Perturbation of Defense Pathways by Low-Dose Arsenic Exposure in Zebrafish Embryos. *Environmental Health Perspectives*. 117 *(6)*: 981-987**

http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=2702417&blobtype=pdf

Exposure to arsenic is a critical risk factor in the complex interplay among genetics, the environment, and human disease. Despite the potential for in utero exposure, the mechanism of arsenic action on vertebrate development and disease is unknown. OBJECTIVES: The objective of this study was to identify genes and gene networks perturbed by arsenic during development in order to enhance understanding of the molecular mechanisms of arsenic action. METHODS: We exposed zebrafish embryos at 0.25-1.25 hr postfertilization to 10 or 100 ppb arsenic for 24 or 48 hr. We then used total

RNA to interrogate genome microarrays and to test levels of gene expression changes by quantitative real-time polymerase chain reaction (QPCR). Computational analysis was used to identify gene expression networks perturbed by arsenic during vertebrate development. RESULTS: We identified a set of 99 genes that responded to low levels of arsenic. Nineteen of these genes were predicted to function in a common regulatory network that was significantly associated with immune response and cancer ($p < 10(-41)$). Arsenic-mediated expression changes were validated by QPCR. CONCLUSIONS: In this study we demonstrated that arsenic significantly down-regulates expression levels of multiple genes potentially critical for regulating the establishment of an immune response. The data also provide molecular evidence consistent with phenotypic observations reported in other model systems. Additional mechanistic studies will help explain molecular events regulating early stages of the immune system and long-term consequences of arsenic-mediated perturbation of this system during development. PMID: 19590694 [PubMed - in process]; PMCID: PMC2702417

**S. Mitran. 2007 Metachronal wave formation in a model of pulmonary cilia. *Science Direct: Computers and Structures* 85: 763-764**

A three-dimensional simulation of the formation of metachronal waves in rows of pulmonary cilia is presented. The cilia move in a two-layer fluid model. The fluid layer adjacent to the cilia bases is purely viscous while the tips of the cilia move through a viscoelastic fluid. An overlapping fixed-moving grid formulation is employed to capture the effect of the cilia on the surrounding fluid. In contrast with immersed boundary methods, this technique allows a natural enforcement of boundary conditions without the need for smoothing of singular force distributions. The fluid domains are discretized using a finite volume method. The 9 + 2 internal microtubule structure of an individual cilium is modeled using large-deflection, curved, finite-element beams. The microtubule skeleton is cross-linked to itself and to the cilium membrane through spring elements which model nexin links. The cilium membrane itself is considered to be elastic and subject to fluid stresses computed from the moving grid formulation as well as internal forces transmitted from the microtubule skeleton. A cilium is set into motion by the action of dynein molecules exerting forces between adjacent microtubules. Realistic models of the forces exerted by dynein molecules are extracted from measurements of observed cilia shapes.

**H.S. Wiley, H. Shankaran. 2008 Smad Signaling Dynamics: Insights from a Parsimonious Model. *Sci. Signal.*, 1, pe41.**

The molecular mechanisms that transmit information from cell surface receptors to the nucleus are exceedingly complex; thus, much effort has been expended in developing computational models to understand these processes. A recent study on modeling the

nuclear-cytoplasmic shuttling of Smad2-Smad4 complexes in response to transforming growth factor–β (TGF-β) receptor activation has provided substantial insight into how this signaling network translates the degree of TGF-β receptor activation (input) into the amount of nuclear Smad2-Smad4 complexes (output). The study addressed this question by combining a simple, mechanistic model with targeted experiments, an approach that proved particularly powerful for exploring the fundamental properties of a complex signaling network. The mathematical model revealed that Smad nuclear-cytoplasmic dynamics enables a proportional but time-delayed coupling between the input and the output. As a result, the output can faithfully track gradual changes in the input while the rapid input fluctuations that constitute signaling noise are dampened out.

**H.S. Wiley, S.Y. Shvartsman, and D.A. Lauffenburger. 2003 Computational Modeling of the EGF Receptor System: A Paradigm for Systems Biology. *Trends Cell Biol.* 13, 43-50.**

Computational models have rarely been used as tools by biologists but, when models provide experimentally testable predictions, they can be extremely useful. The epidermal growth factor receptor (EGFR) is probably the best-understood receptor system, and computational models have played a significant part in its elucidation. For many years, models have been used to analyze EGFR dynamics and to interpret mutational studies, and are now being used to understand processes including signal transduction, autocrine loops and developmental patterning. The success of EGFR modeling can be a guide to combining models and experiments productively to understand complex biological processes as integrated systems.

**Hao Zhu, Ivan Rusyn, Ann Richard, and Alexander Tropsha. 2008 Use of Cell Viability Assay Data Improves the Prediction Accuracy of Conventional Quantitative Structure-Activity Relationship Models of Animal Carcinogenicity. *Environmental Health Perspectives*. 116*(4)*: 1-8**

http://www.ehponline.org/members/2008/10573/10573.pdf

**Background:** To develop efficient approaches for rapid evaluation of chemical toxicity and human health risk of environmental compounds, the National Toxicology Program (NTP) in collaboration with the National Center for Chemical Genomics has initiated a project on high-throughput screening (HTS) of environmental chemicals. The first HTS results for a set of 1,408 compounds tested for their effects on cell viability in six different cell lines have recently become available via PubChem. **Objectives:** We have explored these data in terms of their utility for predicting adverse health effects of the environmental agents. **Methods and results:** Initially, the classification *k* nearest neighbor (*k*NN) quantitative structure–activity relationship (QSAR) modeling method was applied to the HTS data only, for a curated data set of 384 compounds. The resulting models had prediction accuracies for training, test (containing 275 compounds together),

and external validation (109 compounds) sets as high as 89%, 71%, and 74%, respectively. We then asked if HTS results could be of value in predicting rodent carcinogenicity. We identified 383 compounds for which data were available from both the Berkeley Carcinogenic Potency Database and NTP–HTS studies. We found that compounds classified by HTS as "actives" in at least one cell line were likely to be rodent carcinogens (sensitivity 77%); however, HTS "inactives" were far less informative (specificity 46%). Using chemical descriptors only, *k*NN QSAR modeling resulted in 62.3% prediction accuracy for rodent carcinogenicity applied to this data set. Importantly, the prediction accuracy of the model was significantly improved (72.7%) when chemical descriptors were augmented by HTS data, which were regarded as biological descriptors. **Conclusions:** Our studies suggest that combining NTP–HTS profiles with conventional chemical descriptors could considerably improve the predictive power of computational approaches in toxicology.

**Hao Zhu, Lin Ye, Ann Richard, Alexander Golbraikh, Fred A. Wright, Ivan Rusyn, and Alexander Tropsha. 2009 A Novel Two-Step Hierachical Quantitative Structure-Activity Relationship Modeling Work Flow for Predicting Acute Toxicity of Chemicals in Rodents.** ***Environmental Health Perspectives*****. 117** ***(8)*****: 1-8**

http://www.ehponline.org/members/2009/0800471/0800471.pdf

**Background:** Accurate prediction of *in vivo* toxicity from *in vitro* testing is a challenging problem. Large public–private consortia have been formed with the goal of improving chemical safety assessment by the means of high-throughput screening.**Objective:** A wealth of available biological data requires new computational approaches to link chemical structure, *in vitro* data, and potential adverse health effects. **Methods and results:** A database containing experimental cytotoxicity values for *in vitro* half-maximal inhibitory concentration ($IC_{50}$) and *in vivo* rodent median lethal dose ($LD_{50}$) for more than 300 chemicals was compiled by Zentralstelle zur Erfassung und Bewertung von Ersatz- und Ergaenzungsmethoden zum Tierversuch (ZEBET ; National Center for Documentation and Evaluation of Alternative Methods to Animal Experiments) . The application of conventional quantitative structure–activity relationship (QSAR) modeling approaches to predict mouse or rat acute $LD_{50}$ values from chemical descriptors of ZEBET compounds yielded no statistically significant models. The analysis of these data showed no significant correlation between $IC_{50}$ and $LD_{50}$. However, a linear $IC_{50}$ versus $LD_{50}$ correlation could be established for a fraction of compounds. To capitalize on this observation, we developed a novel two-step modeling approach as follows. First, all chemicals are partitioned into two groups based on the relationship between $IC_{50}$ and $LD_{50}$ values: One group comprises compounds with linear $IC_{50}$ versus $LD_{50}$ relationships, and another group comprises the remaining compounds. Second, we built conventional binary classification QSAR models to predict the group affiliation based on chemical descriptors only. Third, we developed *k*-nearest neighbor continuous QSAR models for each subclass to predict $LD_{50}$ values from chemical descriptors. All models were

extensively validated using special protocols. **Conclusions:** The novelty of this modeling approach is that it uses the relationships between *in vivo* and *in vitro* data only to inform the initial construction of the hierarchical two-step QSAR models. Models resulting from this approach employ chemical descriptors only for external prediction of acute rodent toxicity.

**Peiying Zuo, Maryse Picher, Seiko F. Okada, Eduardo R. Lazarowski, Brian Button, Richard C. Boucher, and Timothy C. Elston**. **2008 Mathematical Model of Nucleotide Regulation on Airway Epithelia.** ***The Journal of Biological Chemistry*****. 283** ***(39)*****: 1-15**

In the airways, adenine nucleotides support a complex signaling network mediating host defenses. Released by the epithelium into the airway surface liquid (ASL) layer, they regulate mucus clearance through P2 (ATP) receptors, and following surface metabolism through P1 (adenosine; Ado) receptors. The complexity of ASL nucleotide regulation provides an ideal subject for biochemical network modeling. A mathematical model was developed to integrate nucleotide release, the ectoenzymes supporting the dephosphorylation of ATP into Ado, Ado deamination into inosine (Ino), and nucleoside uptake. The model also includes ecto-adenylate kinase activity and feed-forward inhibition of Ado production by ATP and ADP. The parameters were optimized by fitting the model to experimental data for the steady-state and transient concentration profiles generated by adding ATP to polarized primary cultures of human bronchial epithelial (HBE) cells. The model captures major aspects of ATP and Ado regulation, including their >4-fold increase in concentration induced by mechanical stress mimicking normal breathing. The model also confirmed the independence of steady-state nucleotide concentrations on the ASL volume, an important regulator of airway clearance. An interactive approach between simulations and assays revealed that feed-forward inhibition is mediated by selective inhibition of ecto-5'-nucleotidase. Importantly, the model identifies ecto-adenylate kinase as a key regulator of ASL ATP and proposes novel strategies for the treatment of airway diseases characterized by impaired nucleotide-mediated clearance. These new insights into the biochemical processes supporting ASL nucleotide regulation illustrate the potential of this mathematical model for fundamental and clinical research.

**Astrid C Haugen, Ryan Kelley, Jennifer B Collins, Charles J Tucker, Changchun Deng, Cynthia A Afshari, J Martin Brown, Trey Ideker, and Bennett Van Houten. 2004. Integrating phenotypic and expression profiles to map arsenic-response networks.** ***Genome Biology*** **5*(12)*R95.**

**Background**: Arsenic is a nonmutagenic carcinogen affecting millions of people. The cellular impact of this metalloid in Saccharomyces cerevisiae was determined by profiling global gene expression and sensitivity phenotypes. These data were then mapped to a metabolic network composed of all known biochemical reactions in yeast, as

well as the yeast network of 20,985 protein-protein/protein-DNA interactions. **Results**: While the expression data unveiled no significant nodes in the metabolic network, the regulatory network revealed several important nodes as centers of arsenic-induced activity. The highest-scoring proteins included Fhl1, Msn2, Msn4, Yap1, Cad1 (Yap2), Pre1, Hsf1 and Met31. Contrary to the gene-expression analyses, the phenotypic-profiling data mapped to the metabolic network. The two significant metabolic networks unveiled were shikimate, and serine, threonine and glutamate biosynthesis. We also carried out transcriptional profiling of specific deletion strains, confirming that the transcription factors Yap1, Arr1 (Yap8), and Rpn4 strongly mediate the cell's adaptation to arsenic-induced stress but that Cad1 has negligible impact. **Conclusions**: By integrating phenotypic and transcriptional profiling and mapping the data onto the metabolic and regulatory networks, we have shown that arsenic is likely to channel sulfur into glutathione for detoxification, leads to indirect oxidative stress by depleting glutathione pools, and alters protein turnover via arsenation of sulfhydryl groups on proteins. Furthermore, we show that phenotypically sensitive pathways are upstream of differentially expressed ones, indicating that transcriptional and phenotypic profiling implicate distinct, but related, pathways.