



Carolina Center for Computational Toxicology

Ivan Rusyn, Karen Yeatts, Shawn Gomez, Timothy Elston, Fred Wright, Alexander Tropsha
The University of North Carolina at Chapel Hill

This research is funded by
U.S. EPA - Science To Achieve
Results (STAR) Program
Grant # 833825

Science Questions

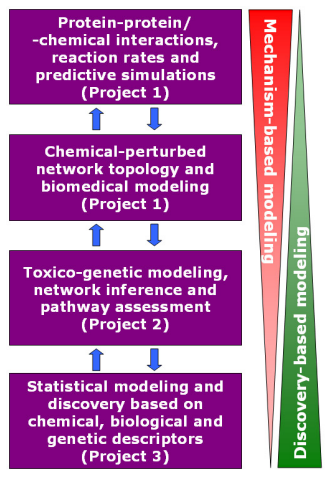
The Carolina Center for Computational Toxicology is engaged in a broad interdisciplinary effort to devise novel tools, methods and knowledge to assist the regulatory agencies and the greater environmental health sciences community in protecting the environment and human health. Using publicly available data, the Center applies knowledge and expertise of the individual investigators and teams to develop complex predictive modeling solutions that range from mechanistic, interpretative data modeling to discovery and decision support efforts. Furthermore, each of the Center's three Research Projects is engaged in active collaboration with other projects within both the Center and the US EPA.

The overall *hypothesis* being addressed in our work is that current knowledge on the modes of action of toxic agents can be efficiently combined with advances in *in silico* mathematical modeling, high throughput screening, toxicogenomics and -genetics to develop novel mechanistic models and approaches that strengthen the ability of predictive scientific data to guide policy decisions, enhance risk assessment, and improve hazard identification.

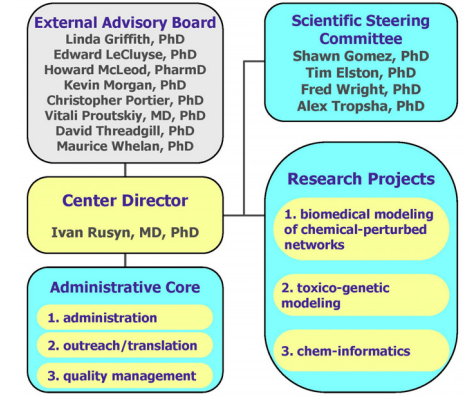
Research Goals

- Major Inter-Disciplinary Objectives are:
- to develop innovative methods and tools for dramatically increasing throughput of chemical screening and precision of hazard identification;
 - to improve existing methodologies for chemical hazard testing *in silico* and *in vitro* with novel strategies that take into account the genetic diversity of the population;
 - to improve linkages in the source-to-outcome paradigm and quantitative risk assessment by translating and applying the scientific findings of the Center into easy-to-use high-throughput computational models of chemical-perturbed signaling networks.

Environmental and biomedical modeling areas in the Carolina Center for Computational Toxicology



The organizational chart of the Carolina Center for Computational Toxicology



Administrative Core Unit

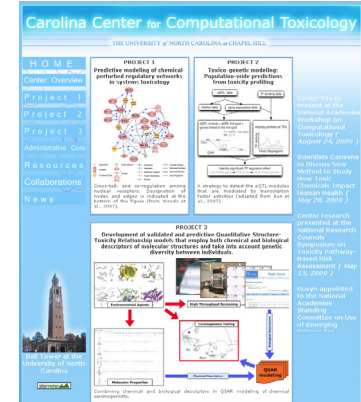
- Administration Function:**
- Project and budget management
 - Communications
 - Reporting to EPA and UNC
 - Organization of the annual EAB meetings

- Integration Function**
- Promoting interactions within the Center
 - Promoting interactions with EPA/NCCT and other partners
 - Facilitating scientific interactions between Projects

- Public Outreach/Translation Function**
- Created Center website: <http://comptox.unc.edu>
 - Implementing bioinformatics and chemo-informatics tools into GUI-enabled software
 - Conduct joint research meetings with EPA/NCCT
 - Make presentations at the state, national and international scientific meetings

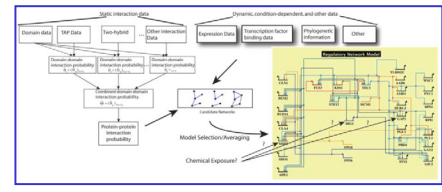
- Quality Management Function**
- Center-wide quality management plan developed and approved by the EPA
 - Quality assurance project plans developed and annual audits performed for Year 1 of the grant
 - Remedial actions will be completed by November 01, 2009

Center website: comptox.unc.edu

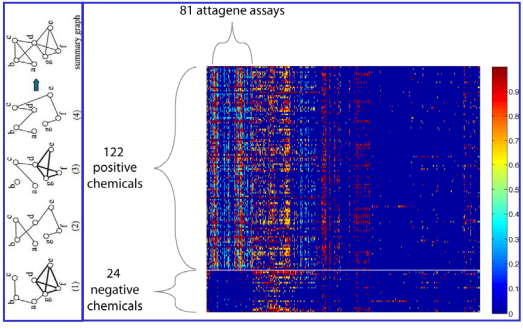


PROJECT 1 Predictive modeling of chemical-perturbed regulatory networks in systems toxicology

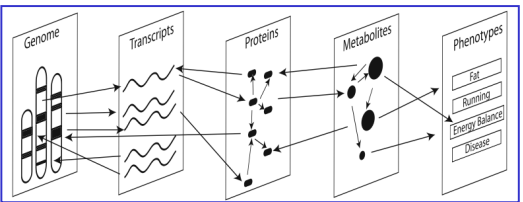
Project 1 has initiated the analysis of the Phase I ToxCast data that has started to become available as of February of 2009. Initial work has focused on establishing computational approaches for establishing links between chemical activity and biological function, with a primary objective being to establish chemical activity within the context of functional biological networks and pathways.



In addition, we expect that such contextual information will help establish new predictor variables that could potentially be used in modeling efforts in all projects. Similarly, the breadth of the assays represented in the ToxCast data presents an additional challenge and we have begun analysis across ToxCast assays, again with a focus on attempting to discern similarities and/or differences in assays with regard to biological context. We are initially attempting the use of graph-based methodologies for the integration of these different data types and the finding of frequent co-occurrence functional modules.

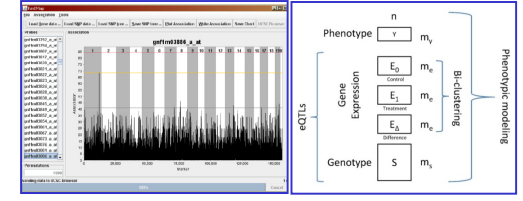


In combination with additional external data sets describing known functional networks, such approaches can potentially reduce the dimension of the biological variables to a set that is both statistically predictive as well as biologically relevant. The application of a recently developed Bayesian data integration approach is also being undertaken, with the goal of linking assays into an integrated representation of biological relationships and associated perturbations.

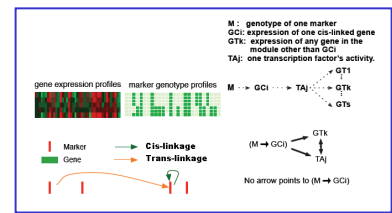


PROJECT 2 Toxico-genetic modeling: Population-wide predictions from toxicity profiling

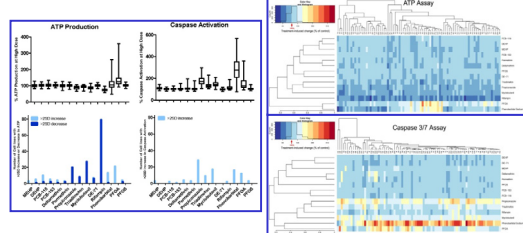
We have made substantial progress in the aims of this Research Project, including software and statistical methods development for eQTL mapping, toxicogenetic modeling in mouse hepatocytes, and toxicity phenotyping in human cells. The FastMap eQTL analysis software forms the major software development activity, and is being used to support all activities of this Research Project. We envision FastMap as the platform of choice upon which to build additional future eQTL analysis capabilities. We have also made substantial progress in enabling p-value based inference in eQTL settings that does not require permutation. This work requires careful handling of the correlation structure among transcripts using matrix decomposition, and may enable very fast and valid inference using a quick "snapshot" of the dataset. More sophisticated and CPU-intensive analyses of the same data can then be performed using FastMap. We have also developed a new approach to estimating permutation p-values that is essentially time-constant, and does not increase with decreasing p-values.



Our continued work on analyzing transcription factor binding will help integrate data from eQTL studies with data about transcription regulation, eventually providing a much more complete view of transcriptional regulation in toxicity response.

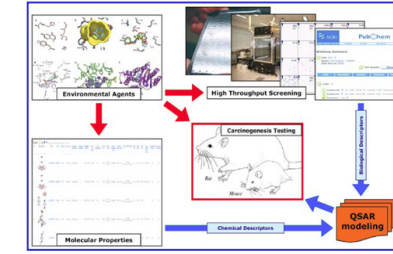


Finally, we have finished collection of phenotyping (cell viability and apoptosis) data on 14 EPA-relevant chemicals in 87 HapMap human lymphoblastoid cell lines and are close to establishing appropriate cell culture conditions for murine hepatocytes from a large panel of inbred strains.

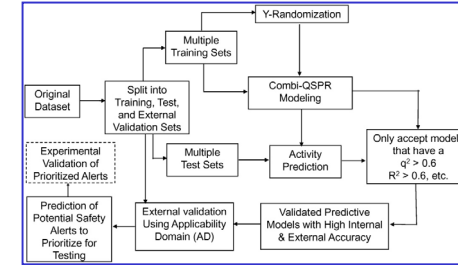


PROJECT 3 Development of validated and predictive Quantitative Structure-Toxicity Relationship models that employ both chemical and biological descriptors of molecular structures and take into account genetic diversity between individuals

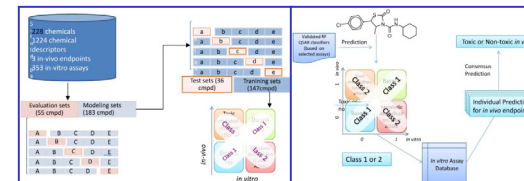
In the first year, the main effort of Project 3 was devoted to further development of Quantitative Structure-Activity Relationship (QSAR) methodology and its applications to chemical toxicity datasets such as ToxCast. The gist of our approach is the joint use of both chemical descriptors of molecules and the results of their biological testing *in vitro* as hybrid *chemical biological descriptors* in our predictive QSAR modeling workflow.



Armed with new methodologies, we have recently initiated the analysis of ToxCast data that were released early February 2009. In one of these studies we have employed cheminformatics approaches to analyze the EPA ToxCast chemical libraries to develop predictive toxicity models and prioritize compounds for *in vivo* toxicity testing.



In another study we focused on a fraction of ToxCast assays that reported the largest numbers of compounds "active" in both *in vitro* and *in vivo* assays. We have employed ToxCast cell-viability and gene-expression assays as biological descriptors in QSAR modeling of animal toxicity endpoints. Specifically, we have analyzed the relationship between the short-term bioassay results and different chronic *in vivo* toxicity responses for the ToxCast Phase I compounds.



Conclusions

The Center not only advances the field of computational toxicology through the development of new methods, software, and applied models but also advocates the application of the knowledge into practice through collaborative efforts with EPA and other governmental stakeholders. The emphasis of our work is on the usability of the outputs by the risk assessment community and the investigative toxicologists thus facilitating the transition of the field of computational toxicology from a hypothesis-driven toward a predictive science.

Impact and Outcomes

- The Center produced 12 publications (see list below) and 18 poster presentations in Year 1;
- The Center interacted actively with US EPA (v-Liver®, ToxCast®, ToxRefDB®, ACToR and DSSTox), NIEHS/NTP and NCGC.

Future Directions

- Project 1:**
- Continue in depth analysis of ToxCast Phase I data;
 - Develop methods for the integration of data types;
 - Investigate the applicability of the metabolism model as a tool for the prediction of the effects of chemical perturbation of metabolic pathways;
 - Integration of the eQTL analyses/approaches with the network-focused methodologies (with Proj. 2);
 - Establish the network context for QSAR (with Proj. 3).

- Project 2:**
- Continue development of FastMap software;
 - Construct transcription regulation networks in the Bayesian framework by combining eQTLs, nucleosome occupancy, and transcriptional regulation data;
 - Complete characterization of the mouse hepatocyte cultures and perform experiments with key toxicants;
 - Complete GWAS analyses of the HapMap lymphoblast cell viability and apoptosis data and correlate the toxicity endpoints with basal gene expression profiles.

- Project 3:**
- Complete the analysis of ToxCast data;
 - Continue to explore other datasets that provide both *in vivo* and *in vitro* data for chemicals;
 - Build models that could be used by EPA to prioritize the selection of ToxCast Phase 2 compounds.

Publications

- Choi K, and Gomez SM. (2009) BMC Bioinformatics (In revision)
- Staab J et al. (2009) BMC Bioinformatics (In revision)
- Gatti DM et al. (2009) Bioinformatics 4:482-489
- Sun W et al. (2009) PLoS One 4:e4721
- Zhu H et al. (2009) Environ Health Persp 117:1257-1264
- Gatti, DM et al. (2009) Mamm Genome 20:437-454
- Harrill AH et al. (2009) Tox Sci 110:235-243
- Sun W, and Wright FA (2009) Ann Appl Stat (In revision)
- Sun W et al. (2009) BMC Bioinformatics (In revision)
- Zhu H et al. (2008) Environ. Health Persp 116: 506-513
- Zhu H et al. (2009) Chem Res Tox (In revision)
- Artemenko AG et al. (2009) Chem Res Tox (In revision)