Carolina Center for Computational Toxicology

Funded by U.S. EPA Cooperative Agreement STAR RD 833825
Progress Report for Project Period 04/01/10 - 03/31/11
Report Date: 04/21/11
Ivan Rusyn, Principal Investigator

Project 2:
Toxico-Genetic Modeling: Population-Wide Predictions from Toxicity Profiling

Preliminary Data and Work Progress

We have continued to make substantial progress in the aims of this Project, including toxicogenetic modeling in population-based *in vitro* models (e.g., mouse hepatocytes, and human lymphoblastoid cells), and developing software and statistical methods for eQTL mapping. The FastMap eQTL analysis software forms the major software development activity, and supports all activities of this Project. FastMap 2.0, with ability to analyze two-genotype (e.g., mouse inbred) data, as well as three-genotype (e.g., human) data, is now complete and in use by our group and the subject of a manuscript soon to be submitted. Innovations in this tool include (i) the use of adaptive permutations; (ii) multiple subset summation trees; (iii) automated reading of HapMap and PLINK genotype formats; (iv) association mapping using linear models or one-way ANOVA; and (v) generation of the output that is both graphical and numerical and is also suitable for plotting using LocusZoom tool. Our *in vitro* population-based toxicity phenotyping data (cell viability and activation of caspase-3/7 endpoints) has been published on 14 EPA-relevant chemicals (pesticide actives, plasticizers, polychlorinated biphenyls, etc.) in a panel of densely genotyped 87 HapMap CEU human lymphoblastoid cell lines. We have also nearly completed analysis of the data from qHTS screening of 240 chemicals (same endpoints) at 12 concentrations, for 80+ HapMap cell lines, through an ambitious collaboration with the Tox21 partner organizations, the National Toxicology Program and the NIH Chemical Genomics Center. New analyses of these data include correlations of publicly available RNA-Seq expression data from these cell lines, which are a useful supplement to genotype-toxicity mapping approaches. These analyses have helped to establish clear inter-individual variation in *in vitro* toxicity response. A similar, but considerably more ambitious experiment to perform qHTS screening of 180- chemicals at 8 concentrations for cell viability endpoint assay in 1100+ lymphoblastoid cell lines from the 1000 Genomes project will begin in May 2011. The data collection should be completed by August 2011. This new data will enable more powerful mapping of susceptibility genes to toxicity response, as well as much more sensitive investigation of the effects of rarer variants.

Results to Date

The goals of this project are to (i) develop toxicogenetic expression quantitative trait loci (eQTL) mapping tools, perform transcription factor network inference and integrative pathway assessment; (ii) perform toxicogenetic modeling of liver toxicity in cultured mouse hepatocytes; (iii) discover chemical-induced regulatory networks using population-based toxicity phenotyping in human cells. We continue to make substantial progress in all of these goals.

Our fast SNP-correlation method, FastMap 2.0, is now operable for human (i.e. three-genotype) data. A number of speedups have been implemented, as described earlier, and the software now has more user-friendly input formats. Several additional developments in eQTL analysis are also providing additional leveraging capabilities for Project 2. Drs. Rusyn, Wright, and Nobel are co-PIs of a newly-funded methods grant to accompany the GTEx initiative

http://commonfund.nih.gov/GTEx/ , which will provide RNA and DNA profiles of ~160 donors in multiple tissues. As part of this work, additional mapping software is being developed with sufficient speed to add an additional (tissue) dimension to eQTL profiling. Dr. Wright is also leading the statistical component of another large eQTL project, mapping lymphocyte eQTL variation using ~1500 twins, which will provide important information about eQTL heritability. In a partnership with multiple investigators, we are developing a new eQTL browser that has more extensive search and display capabilities than competing browsers.

Additional methods work include fast approximate approaches to eQTL permutation testing (Sun and Wright, 2010), and Bayesian iterative adaptive Lasso regression approaches using ensembles of transcripts for eQTL analysis. In order to be prepared for sequence-based expression (RNA-Seq) data, we have been working to establish more powerful approaches to handle transcript count data. A hallmark of such datasets is that they show strong evidence of overdispersion, and must be modeled using beta-binomial or similar distributions. We have developed a new package, BBSeq (Zhou et al., submitted), which performs this modeling, and also attempts to maximize power for small samples by modeling the relationship between the mean and the overdispersion.

We continue to work closely with Project 3 on predictive QSAR modeling of the toxicity data, including Tox21 *in vitro* qHTS data. Project 2 investigators work closely with Project 3 investigators that resulted in several poster presentations and publications. Progress towards the goals of Project 3 is detailed below.

We have completed the analyses of our *in vitro* population-based toxicity phenotyping data (cell viability and apoptosis) on 14 EPA-relevant chemicals (pesticide actives, plasticizers, polychlorinated biphenyls, etc.) in a panel of densely genotyped 87 HapMap CEU human lymphoblastoid cell lines (O'Shea et al., 2011). We have established the degree of inter-individual variability in responses to these agents, examined whether such variability is heritable, and conducted genome-wide association analysis of the individual chemical-assay phenotypes. This work served as an important proof of principle in establishing the utility of *in vitro* toxicity screening in a genetically-defined population model, as well serving as a testing ground for the combination of toxicity phenotypes with baseline expression and genotype data.

Our initial success has led to a much more ambitious collaboration with the Tox21 partner organizations, the National Toxicology Program and the NIH Chemical Genomics Center. In collaboration with NCGC and NTP, we have completed screening of 240 chemicals (in 12 concentrations) in 81 HapMap CEU human lymphoblastoid cell lines for 2 phenotypes (cell viability and apoptosis). This dataset was a major nexus of the computational analyses in Year 3 of the project. We have completed the analyses of this data and the manuscript is in a pre-submission stage. Our goal was to aid in the development of predictive *in vitro* models of chemical-induced toxicity anchored on inter-individual genetic variability. qHTS screening in the genetically-defined population produced robust and reproducible results, which allowed for cross-compound, -assay and -individual comparisons. The generation of high-quality *in vitro* toxicity data on a large library of compounds using qHTS demonstrates the potential of this methodology for assessing the degree of inter-individual variability in toxicity and exploring its genetic determinants.

A further extension of these in vitro studies using a genetically-diverse and –defined in vitro population-based qHTS screening approach is underway. We are in the process of collecting data on 1100+ lymphoblast cell lines from HapMap and 1000 Genomes projects. We will screen 180 chemicals (in 8 concentrations) in qHTS format for cell viability phenotype. These experiments are conducted in collaboration with NTP and NCGC. While the data we collected thus far in a small (80+) population is immediately interpretable as detailed above, the identification of genetic factors that affect responses to xenobiotics is not feasible without expanding the size of the population tested and the experiment underway will address this limitation. By combining the toxicity data and publicly available genetic information, such as that

provided by the HapMap and 1000 Genomes projects, it will be possible to probe the contribution of genetics to adverse phenotypes, and select candidate genes and regulatory networks for further studies to verify their mechanistic relevance with regard to differences in susceptibility to chemical treatment.

To enhance the utility of animal models to detect response biomarkers for genetically diverse populations, we continue experiments with a genetically-defined panel of mouse strains. By taking into the account strain-specific gene expression patterns it is possible to establish genetic polymorphism-dependent and -independent pathways perturbed by the toxicants. In Year 3, this approach has been applied in two case studies where liver effects of tricholoroethylene and ethanol were assessed. Treatment-specific effects on the liver transcriptome in studies with these agents were strongly dependent on the individual's genetic background; still, the molecular effects that are dependent or independent on the individual's genetic background were identified. Since genetic regulation of gene expression is a key contributor to population diversity, these studies provide better understanding of the mechanisms of toxicity that may define susceptibility or resistance.

Additional progress is being made in our experiments that aim at establishing an *in vitro* primary hepatocyte population-based model using a panel of inbred mouse strains. We have completed experiments with 2D cell culture conditions for murine hepatocytes from a large panel of inbred strains. We have tested the reproducibility of the biochemical and molecular function of the cultured cells isolated from several strains, as well as experiments that tested the dose-response to 3 model toxicants in cells from various strains (Martinez et al., 2010). We now perform *in vitro* experiments with mouse hepatocytes using a 3D bioreactor in collaboration with Dr. Linda Griffith at MIT.

Activities for Subsequent Reporting Period

In Year 4 we will concentrate less on novel methods development, and more upon coordinating the existing methods and software and bringing it to bear upon the extensive profiling data developed in Project 2. HapMap 2.0 and additional software developed by our group are relatively mature and are applied to the toxicity profiling studies. Our fast approximations to eQTL significance are used as an adjunct to parametric testing, for situations in which full permutation is not initially feasible. In addition, our previous approaches to fast trans-band testing are relatively mature. Our Bayesian Iterative Adaptive Lasso procedure for eQTL analysis of groups of weakly correlated transcripts is also nearing submission.

We are also moving forward quickly in methods for next-generation sequence analysis. This work is especially relevant for the expansion of our human toxicity profiling to the 1000 Genomes LCLs, which is reaching a high resolution of sequencing coverage for several hundred individuals, and coverage for a large number of individuals that is comparable to SNP array platforms. Effective analyses of such data will raise new issues, including the ability to perform mutation-burden testing in exomes, and otherwise handling rare-variant data in novel ways in comparisons to toxicity susceptibility. Dr. Wright is currently involved in an exome-sequence project for lung disease in cystic fibrosis, and will bring this experience to analyses of the new qHTS data. In addition, the new data will warrant bringing additional expression data into the analyses, including microarray and RNA-Seq data available on the HapMap cell line subset. Accordingly, project personnel have honed their expertise in handling RNA-Seq data, including transcript mapping and resolving different isoforms. We will continue to actively develop methods for differential expression of RNA-Seq data, and investigate approaches to ramp up the analyses for the high-throughput eQTL setting.

For the experimental aims of this project, we will be focusing on both our work with human lymphoblastoid cell lines and primary mouse hepatocytes. In case of the former, we will be analyzing the data from the current experiment on 1100+ cell lines and ~200 chemicals performed in partnership with Tox21. Specifically, we will perform both traditional analyses of

variability and heritability, and a genome-wide association study to not only characterize, but also mechanistically interpret the inter-individual differences in responses to a large number of chemicals across the panel of cells. With regards to the work with primary mouse hepatocytes isolated from a panel of mouse inbred strains, we will be aiming to screen a set of 15-30 model chemicals on 3D mouse liver bioreactors.

Finally, we will continue collaborations with Project 1 on the interpretation of the toxicant-perturbed networks from ToxCast data and other toxicity datasets; and with Project 3 on the QSAR-based analysis of the ToxCast data.

Publications Arising From this Project in Year 3

Papers:

1. Sun, W., and Wright, F.A. (2010) A geometric interpretation of the permutation p-value and its application in eQTL studies. Annals of Applied Statistics 4(2):1014–1033. (*from in press to published*)
2. Adams, T.M., and Nobel, A.B. (2010) Uniform convergence of Vapnik-Chervonenkis classes under ergodic sampling. Annals of Probability 38(4):1345–1367. (*from in press to published*)
3. Gatti, D.M., Barry, W.T., Nobel, A.B., Rusyn, I., and Wright, F.A. (2010) Heading down the wrong pathway: On the influence of correlation within gene sets. BMC Genomics 11:574. (*from submitted to published*)
4. Walter, V., Nobel, A.B., and Wright, F.A. (2011) DiNAMIC: a method to identify recurrent DNA copy number aberrations in tumors. Bioinformatics 27:678-685.
5. Zou, F., Lee, S., Knowles, M.R., Wright, F.A. (2010) Quantification of population structure using correlated SNPs by shrinkage principal components. Hum Hered 70:9-22.
6. Martinez, S.M., Bradford, B.U., Soldatow, V.Y., Kosyk, O., Sandot, A., Witek, R., Kaiser, R., Stewart, T., Amaral, K., Freeman, K., Black, C., LeCluyse, E.L., Ferguson, S.S., and Rusyn, I. (2010) Evaluation of an in vitro toxicogenetic mouse model for hepatotoxicity. Toxicol Appl Pharmacol 249:208-216.
7. Pogribny, I.P., Starlard-Davenport, A., Tryndyak, V., Han, T., Ross, S.A., Rusyn, I., and Beland, F.A., Difference in expression of hepatic microRNAs is associated with strain-specific susceptibility to dietary nonalcoholic steatohepatitis in mice. Lab Invest 90:1437-1446, 2010.
8. Gatti, D.M., Zhao, N., Chesler, E.J., Bradford, B.U., Shabalin, A., Yordanova, R., Lu, L., and Rusyn, I. (2010) Sex-specific gene expression in BXD mouse liver. Physiol Genomics 42:456-468.
9. Rusyn, I., Gatti, D.M., Wiltshire, T., Kleeberger, S.R., and Threadgill, D.W. (2010) Toxicogenetics: population-based testing of drug and chemical safety in mouse models. Pharmacogenomics 11:1127-1136.
10. Rusyn, I., and Daston, G.P. (2010) Computational Toxicology: Realizing the Promise of the Toxicity Testing in the 21st Century. Environ Health Persp 118:1047-1050.
11. Sedykh, A., Zhu, H., Tang, H., Zhang, L., Richard, A., Rusyn I., and Tropsha A. (2011) Use of in vitro HTS-derived concentration-response data as biological descriptors improves the accuracy of QSAR models of in vivo toxicity. Environ Health Persp 119:364-370.
12. Bradford, B.U., Lock, E.F., Kosyk, O., Kim, S.K., Uehara, T., Harbourt, D., DeSimone, M., Threadgill, D.W., Tryndyak, V., Pogribny, I.P., Bleyle, L., Koop, D.R. and Rusyn, I. (2011) Inter-strain differences in the liver effects of trichloroethylene in a multi-strain panel of inbred mice. Toxicol Sci 120:206-217.

13. O'Shea, S.H., Schwarz, J., Kosyk, O., Ross, P.K., Ha, M.J., Wright, F.A., and Rusyn, I. (2011) In vitro screening for population variability in chemical toxicity. Toxicol Sci 119:398-407.
14. Gatti, D.M., Lu, L., Williams, R.W., Sun, W., Wright, F.A., Threadgill, D.W., and Rusyn, I. (2011) MicroRNA expression in the livers of inbred mice. Mutat Res (in press).
15. Low Y., Uehara T., Minowa Y., Yamada H., Ohno Y., Urushidani T., Sedykh A., Muratov E., Fourches D., Zhu H., Rusyn I., Tropsha A. Predicting Drug-induced Hepatotoxicity Using QSAR and Toxicogenomics Approaches. Chem. Res. Tox. Submitted.
16. Zhou, Y., Xia, K., and Wright, F.A. A powerful and flexible approach to the analysis of RNA sequence count data. Bioinformatics. Submitted.

Posters/Abstracts:

1. O'Shea, S., Kosyk, O., Abdo, N., Lock, E., Wright, F., Huang, R., Xia, M., Austin, C., Tice, R., and Rusyn, I. Population-based quantitative high throughput screening (qHTS) for chemical toxicity. Society of Toxicology Annual Meeting, Washington, DC. 2011.
2. Pogribny, I., Koturbash, I., Scherhag, A., Sorrentino, J., Sexton, K., Bodnar, W., James, S., Beland, F., Rusyn, I. Inter-strain differences in susceptibility to 1,3-butadiene-induced DNA damage, epigenetic effects and hepatotoxicity. Society of Toxicology Annual Meeting, Washington, DC. 2011.
3. Low, Y., Uehara, T., Minowa, Y., Yamada, H., Ohno, Y., Urushidani, T., Sedykh, A., Fourches, D., Zhu, H., Rusyn, I., and Tropsha, A. Predictive value of chemical and toxicogenomic descriptors for drug-induced hepatotoxicity. Society of Toxicology Annual Meeting, Washington, DC. 2011.
4. Zhang, L., Zhu, H., Afantitis, A., Melagraki, G., Sarimveis, H., Rusyn, I., and Tropsha, A. QSAR modeling of estrogen receptor binding affinity and virtual screening for potential endocrine disrupting compounds. Society of Toxicology Annual Meeting, Washington, DC. 2011.
5. Campbell, J., Clewell, H., Kim, S., Collins, L., Kosyk, O., and Rusyn, I. Monte Carlo analysis of TCE metabolism across a panel of inbred mouse strains. Society of Toxicology Annual Meeting, Salt Lake City, UT. 2010.
6. Bradford, B.U., Kim, S., Kosyk, O., Grimes, J., O'Connell, T., and Rusyn, I. NMR and microarray based analysis of mouse liver following exposure to trichloroethylene. Society of Toxicology Annual Meeting, Salt Lake City, UT. 2010.