ELSEVIER

# Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action

Brigitte Ganter [*,1], Stuart Tugendreich [1], Cecelia I. Pearson, Eser Ayanoglu,
Susanne Baumhueter [7], Keith A. Bostian, Lindsay Brady, Leslie J. Browne [8],
John T. Calvin, Gwo-Jen Day, Naiomi Breckenridge, Shane Dunlea,
Barrett P. Eynon, L. Mike Furness [5], Joe Ferng, Mark R. Fielden,
Susan Y. Fujimoto, Li Gong, Christopher Hu, Radha Idury,
Michael S.B. Judo, Kyle L. Kolaja, May D. Lee, Christopher McSorley,
James M. Minor [2], Ramesh V. Nair, Georges Natsoulis, Peter Nguyen,
Simone M. Nicholson, Hang Pham, Alan H. Roter, Dongxu Sun [3], Siqi Tan [6],
Silke Thode [4], Alexander M. Tolley, Antoaneta Vladimirova, Jian Yang,
Zhiming Zhou, Kurt Jarnagin

*Iconix Pharmaceuticals, 325 E. Middlefield Road, Mountain View, CA 94043, USA*

## Abstract

Successful drug discovery requires accurate decision making in order to advance the best candidates from initial lead identification to final approval. Chemogenomics, the use of genomic tools in pharmacology and toxicology, offers a promising enhancement to traditional methods of target identification/validation, lead identification, efficacy evaluation, and toxicity assessment. To realize the value of chemogenomics information, a contextual database is needed to relate the physiological outcomes

* Corresponding author.
  *E-mail address:* bganter@iconixpharm.com (B. Ganter).
[1] They contributed equally to this work.
[2] Present address: Agilent Tech. Inc., 3500 Deer Creek Road, Palo Alto, CA 94304, USA.
[3] Present address: Mpex BioScience Inc., 500 Campanile Drive, San Diego, CA 92182, USA.
[4] Present address: Integrium, 14351 Myford Road, Tustin, CA 92780, USA.
[5] Present address: Nuomics Consult. Ltd., 3 Merlin Drive, Ely, Cambridge CB6 3EA, UK.
[6] Present address: 2077 Tapscott Avenue, El Cerrito, CA 94530, USA.
[7] Present address: 3656 Jefferson Avenue, Redwood City, CA 94062, USA.
[8] Present address: Pharmacopeia Drug Discovery, Inc., P.O. Box 5350, Princeton, NJ 08543-5350, USA.

induced by diverse compounds to the gene expression patterns measured in the same animals. Massively parallel gene expression characterization coupled with traditional assessments of drug candidates provides additional, important mechanistic information, and therefore a means to increase the accuracy of critical decisions. A large-scale chemogenomics database developed from in vivo treated rats provides the context and supporting data to enhance and accelerate accurate interpretation of mechanisms of toxicity and pharmacology of chemicals and drugs. To date, approximately 600 different compounds, including more than 400 FDA approved drugs, 60 drugs approved in Europe and Japan, 25 withdrawn drugs, and 100 toxicants, have been profiled in up to 7 different tissues of rats (representing over 3200 different drug–dose–time–tissue combinations). Accomplishing this task required evaluating and improving a number of in vivo and microarray protocols, including over 80 rigorous quality control steps. The utility of pairing clinical pathology assessments with gene expression data is illustrated using three anti-neoplastic drugs: carmustine, methotrexate, and thioguanine, which had similar effects on the blood compartment, but diverse effects on hepatotoxicity. We will demonstrate that gene expression events monitored in the liver can be used to predict pathological events occurring in that tissue as well as in hematopoietic tissues.

## 1. Introduction

The drug discovery and approval process is long, expensive, and inefficient; less than 1 in 10 promising drug candidates entering phase I clinical trials results in an approved product, thus the decisions made prior to phase I are inaccurate for greater than 90% of drug candidates. There are a number of reasons for this high failure rate, generally stemming from the absence of a full understanding of the biological and toxicological properties and mechanisms of the drug candidate. Comparison of the effects of a candidate compound on rats against an existing database of multiple parameters measured for a large number of marketed drugs, subsequently withdrawn drugs, and known toxicants provides the context to interpret the findings for the candidate and improve the accuracy of development decisions. Analysis of the transcriptional response in the cells of target organs can give an indication of the biochemical or biological mechanism affected by a pharmaceutical compound. Over the last few years, the toxicology community has begun to employ genome-wide gene expression profiling to dissect the mechanisms behind chemical toxicity and to use this information to increase the accuracy and sensitivity of toxicity testing. Specific gene expression profiles have been evaluated in several types of toxicological studies (Burczynski et al., 2000; Debouck and Goodfellow, 1999; Hamadeh et al., 2002; Harris et al., 2001; Nuwaysir et al., 1999; Waring et al., 2001a,b). In this paper, we describe

how this approach, termed "toxicogenomics", or more generally "chemogenomics", has been significantly extended, by expanding both the number of endpoints measured to include clinical chemistry, hematology, histopathology (using a standardized histopathology vocabulary and grading system), and organ weights, and by profiling a large number of compounds in more than one tissue. We demonstrate that coupling gene expression profiling with traditional toxicity measurements enhances the understanding of individual compound effects in rats. Furthermore, although beyond the scope of this paper to discuss in detail, the database contains a great deal of additional information. For example, the gene expression results for each compound are tied to its in vitro pharmacological activity, by first measuring each purified compound in 130 primarily human in vitro molecular pharmacology bioassays that measure selectivity, specificity, and affinity for receptor binding, cytochrome P450 activity and drug target enzymatic activities.

Currently, the database contains the profiles derived from administering approximately 600 different compounds to rats (400 FDA approved drugs, 60 drugs approved in Europe and Japan, 25 withdrawn drugs, 27 standard biochemicals, and 100 molecules of toxicological interest). These have been profiled in up to 7 different tissues, for a total of 3200 different drug–dose–time–tissue combinations. This system is designed to assist drug discovery professionals in the selection of the highest quality leads and drug candi-