Short Communication

Journal of Applied Toxicology

Received: 6 June 2013, Revised: 30 July 2013, Accepted: 31 July 2013 Published online in Wiley Online Library: 11 September 2013

(wileyonlinelibrary.com) DOI 10.1002/jat.2923

LTMap: a web server for assessing the potential liver toxicity by genome-wide transcriptional expression data

Li Xing, Leihong Wu, Yufeng Liu, Ni Ai, Xiaoyan Lu and Xiaohui Fan*

ABSTRACT: Toxicogenomics (TGx) has played a significant role in mechanistic research related with hepatotoxicity as well as liver toxicity prediction. Currently, several large-scale preclinical TGx data sets were made freely accessible to the public, such as Open TG-GATEs. With the availability of a sufficient amount of microarray data, it is important to integrate this information to provide new insights into the risk assessment of potential drug-induced liver toxicity. Here we developed a web server for evaluating the potential liver toxicity based on genome-wide transcriptomics data, namely LTMap. In LTMap, researchers could compare signatures of query compounds against a pregenerated signature database of 20 123 Affymetrix arrays associated with about 170 compounds retrieved from the largest public toxicogenomics data set Open TG-GATEs. Results from this comparison may lead to the unexpected discovery of similar toxicological responses between chemicals. We validated our computational approach for similarity comparison using three example drugs. Our successful applications of LTMap in these case studies demonstrated its utility in revealing the connection of chemicals according to similar toxicological behaviors. Furthermore, a user-friendly web interface is provided by LTMap to browse and search toxicogenomics data (http://tcm.zju.edu.cn/ltmap). Copyright © 2013 John Wiley & Sons, Ltd.

Supporting information may be found in the online version of this article.

Keywords: Toxicogenomics; hepatotoxicity; risk assessment; web server

Introduction

Currently, elongated duration and increasing cost have become significant issues related to safety assessment of new chemical entities as therapeutic agents during the development stage (Collier, 2009). As a critical component of drug safety assessment, liver toxicity of the drug is considered as one important factor for its approval by the regulating agency (Chen *et al*., 2011). Moreover, studies have shown that drug-induced liver injury was associated with over 1000 marketed drugs and contributed to ~40% of failures in clinical trials and market withdrawals (Zhang *et al*., 2012). To advance toxicological research and prevent drug-induced liver injury, it is vital that emerging research tools are incorporated into the study of liver toxicity for drugs. High-throughput technologies such as microarray (i.e., toxicogenomics) have been the active field of research over the past decade (Fan *et al*., 2010, 2011) and there has been great optimism in being able to identify molecular biomarkers for drug safety evaluation efficiently and effectively (Barros, 2005; Leming *et al*., 2003; Lu *et al*., 2013,2013b; Shi *et al*., 2010b).

Recently, several large-scale, high-quality and well-designed hepatic toxicogenomics databases were released, such as the Japanese Toxicogenomics Project (Uehara *et al*., 2010), DrugMatrix database currently hosted by the National Toxicology Program in United States (Ganter *et al*., 2005) and the InnoMed PredTox project of the EU (Suter *et al*., 2011). With the availability of a sufficient amount of microarray data, it is possible to develop a web server to analyze this information and provide new insights into a potential liver toxicity assessment. Inspired by CMap (Lamb *et al*., 2006), which connected drugs, genes and diseases by identifying and interpreting patterns of significant and coordinated changes in gene expression profiles, we described the development of a web server for assessing the potential drug-induced liver toxicity based on genome-wide transcriptional expression data, namely LTMap. Applying the Kolmogorov–Smirnov statistics developed by Lamb *et al*. (2006), the rank ordered mode matching strategy in LTMap would compare signatures of a biological state induced by the query compound with signature databases derived from Open TG-GATEs to discover unexpected similar toxicological changes between drugs. In conclusion, LTMap provides a novel opportunity to recognize connections of compounds that have similar biological effects and understand the chemicals with common mechanisms of action.

Materials and methods

Data content

LTMap is a web server developed to identify liver toxicological patterns from the largest public toxicogenomics data set in terms of the number of microarrays, i.e., Open TG-GATEs (Uehara *et al*., 2010). The information about this data set was summarized in Fig. 1.

*Correspondence to: Dr. Xiaohui Fan, Pharmaceutical Informatics Institute, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China. Email: fanxh@zju.edu.cn

Pharmaceutical Informatics Institute, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China

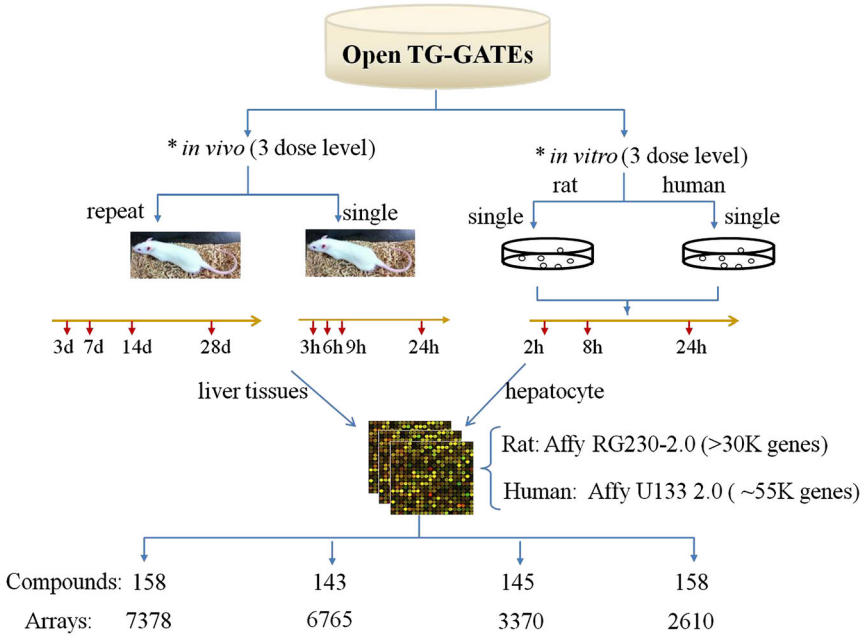

Figure 1. Summary of the data set used for LTMap construction. **In vivo* study: in the repeat-dose study, time points are marked as the drug treatment time, while the time points are the times when test animals were killed after treatment in the single-dose study. *In vitro* study: the time points represent the times when the samples were collected for gene expression analysis after drug treatment.

Raw microarray data (CEL files) are available in Open TG-GATEs (http://toxico.nibio.go.jp/) (Uehara *et al.*, 2010). Arrays were performed on two Affymetrix platforms, i.e., Affymetrix RG230-2.0, Affymetrix human U133 plus 2.0, which contain 31 099 and 54 675 signatures (i.e., probes) respectively.

All Affymetrix CEL (Raw data) files were converted to probe sets with Affymetrix Microarray Suite (MAS) version 5.0 by ArrayTrack (Tong *et al.*, 2003). Probe sets of each reference were ranked based on the fold-change values in descending order. The FC value of each probe is the ratio of the value of one treatment to the average value of its corresponding controls. The web interface of LTMap has been developed using HTML/CSS and Javascript. All data in the database have been stored in SQLite tables. Ruby on Rails has been used as an interface layer between the front and back ends of LTMap

Web server

Figure 2 shows the overview of the LTMap system. There are three key components forming the core functionality of LTMap: data input, analysis and the output.

Input

The approach starts with "query signatures", which includes the upregulated and downregulated probe sets and the number of signatures could fluctuate between 1 and 31 099 or 54 675. LTMap accepts "query signatures" in several different formats. The input list of gene probes could be separated by a comma, space or tab. Such a list could be obtained using standard feature selection methods that statistically rank all the genes and select those above

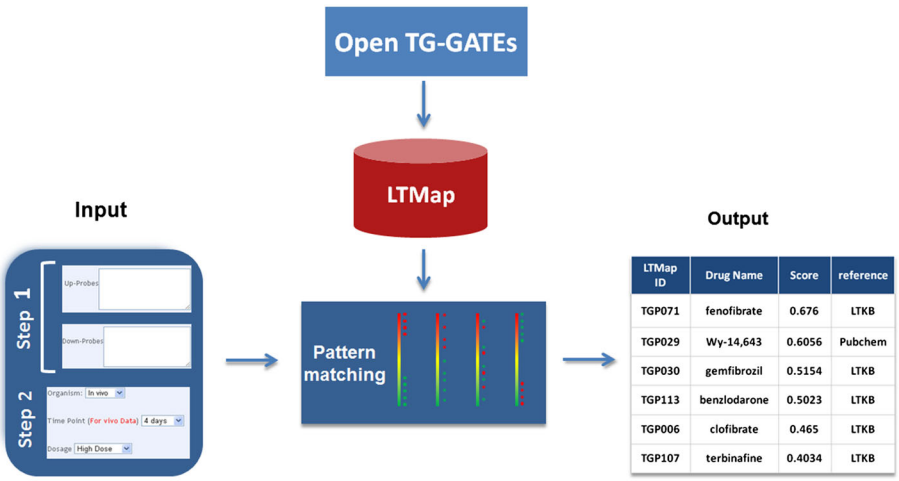

Figure 2. There are three key components forming the core functionality of LTMap: data input, analysis and the output.

Journal of Applied Toxicology

a defined threshold. Examples of this input format are provided on the LTMap website. After submitting probe sets, users can select different time points, dosages and experimental designs as needed. LTMap will automatically search for similar compounds using previously described methodology. Furthermore, DrugBank ID or LTMap ID could be used as query terms to access raw microarray data and more related biological information of specific compounds. The LTMap IDs in the data set were organized in a separate file, which is available for download on the website. In addition to its analysis utilities, LTMap supports the conversion between gene names and probe names in the Affymetrix RG230-2.0 array and human U133 plus 2.0 array.

Analysis

We applied the rank-ordered mode matching strategy based on the Kolmogorov–Smirnov statistic developed by Lamb *et al*. (Kim *et al*., 2005; Lamb *et al*., 2006). The degree of similarities of toxicogenomic profiles between compounds was calculated between the query signatures and each reference array by the following method. Two separate probes sets for up- and downregulated probes were required as the set of signatures for the query compounds. All probes in the array were first sorted according to their fold-changes in term of expression levels. Kolmogorov–Smirnov statistic values were calculated for these two sets separately by searching the position of each tag in the rank-ordered reference probe list, i.e., ks_{up} and ks_{down} respectively. Let us denote n as the number of total probe sets (31 099 or 54 675) of the reference microarray and t as the number of signatures in query signatures that could vary within the range between 1 and 31 099 or 54 675. A vector L represented the position of each probes in the tag list. The positions of all probes in the tag list were sorted in ascending order, i.e., $L(j)$ was in the ordered list, where $j = 1, 2\ldots t$. For each array i, a and b value were calculated according to the following formula.

$$a = \max_{j=1}^{t}\left[\frac{j}{t} - \frac{L(j)}{n}\right]$$

$$b = \max_{j=1}^{t}\left[\frac{L(j)}{n} - \frac{(j-1)}{t}\right]$$

If $a > b$, a is given to ks^i, else $-b$ is given to ks^i. The up scores and down scores are ks^i_{up} and ks^i_{down}, respectively. If the ks^i_{up} and ks^i_{down} have the same sign, the score $(i) = 0$. Otherwise, set score $(i) = (ks^i_{up} - ks^i_{down})/2$.

A positive score means there is a certain level of similarity between two expression profiles mathematically, which suggests a similar biological condition that might be induced by query and compounds in LTMap. A negative score would imply different physiological effects on gene expressions resulted from the treatment of query or data set compounds. If the score is zero, the toxicogenomic profile of reference compounds has no relation with the query signatures. Finally, the score of a compound is calculated by averaging all scores of corresponding reference gene-expression profiles, as each compound was associated with several replicate samples under microarray experimental settings.

Output

After completing the searching process, comprehensive summary information was generated and displayed in a tabular format. The information and parameters that users submitted would be included in the results page and gene probes were annotated with their UniGene IDs, gene names and biology function descriptions. LTMap returns a list of compounds with similarity scores, *P*-values and external links to other databases for related details about pharmacological and toxicological information, such as DrugBank (Knox *et al*., 2011), PubChem (Wang *et al*., 2009), LTKB-DB (Chen *et al*., 2011, 2013). The higher the positive score is, the higher the possibility is that a similar mechanism of the drug action and physiological conditions induced by two drugs. The *P*-value reflects the statistically significance level of connections between query signatures and a certain reference profile. After clicking the right icon buttonbushes, users can visualize detailed information about the gene expression profiles for the selected compounds. The order of the output list could be sorted based on scores or *P*-values. LTMap also provides an online manual to guide the navigation and utility of the website.

Results

Validation of the computational approach

The rationality of the approach was tested to determine whether it is reasonable in liver toxicogenomics study. Internal validation was carried out within the LTMap data set itself. According to the fold-change values, signatures of the top four significantly up- and downregulated probes from each reference profile were first extracted and used as queries through the web server. Our results show that the highest-scoring array is always the array that generated the query signatures, which verified the analysis technique. We took one internal array for example. The array is "WY-14643_High_4_day_Liver_1", which characterizes the biological state of the first SD rat in the group consecutively treated with 100 mg kg^{-1} WY-14643 for 4 days. The highest-scoring compound is WY-14643 itself. This also was put as an example on the website.

Recovering compounds in the LTMap

To examine the ability of the web server recovering those compounds with similar hepatotoxic effects in the LTMap, we used the gene expression profiles obtained from an acute hepatoxic toxicogenomics study, which is part of the MicroArray Quality Control Phase-II(MAQC-II) project (Shi *et al*., 2010a). Thioacetamide is a sulfur-containing compound, which has long been known as a hepatotoxicant. Oxidative stress has been suggested as one important factor involved in thioacetamide-induced hepatocellular necrosis (Stankova *et al*., 2010). Here we chose thioacetamide as a testing compound, using external array with high dose, and 48 h treatment on F344 rats. Probe sets of up- or downregulated by the test compound were the top 20 probes based on the fold-change values in the microarray experiment (see Supplementary information S1). Then these probe sets of thioacetamide were used to query the web server. The expression profile of thioacetamide in the Open TG-GATEs data set with high dose and 3 day treatment ranked at the top. Other top ranking compounds included WY-14643, clofibrozil and benzbromarone. It has been reported that thioacetamide and these top ranking compounds are all related to hepatotoxicity-related abnormal coagulopathy based on the gene analysis results (Hirode *et al*., 2009). In toxicological research, blood coagulation and hepatotoxicity abnormalities are relevant as most coagulation factors

are produced in the liver (Hirode *et al*., 2009). Together with the results from our method, it is possible that thioacetamide may share a novel toxicological mechanism with other peroxisome proliferator-activated receptor (PPAR) agonists for hepatotoxicity.

Identification of potential toxicities of new chemical entities

To exemplify the application of LTMap to identify potential toxicities of new chemical entities, which were not included in LTMap, we took the liver transcriptomics data of Nafenopin retrieved from DrugMatrix (Ganter *et al*., 2005) to compile a query gene signature (see Supplementary information S1), e.g. "nafenopin_338mg/kg_3_day_Liver", where the compound name is first, followed by dose, time point and tissue. Affymetrix array versions RG230-2.0 was used to measure the expression of mRNA from the liver of a male SD rat. Nafenopin was reported as an activator of PPARα (Blomme *et al*., 2009), which is not included in the LTMap data set. For query signatures of nafenopin, by setting search criteria in experimental design, time point and dose, fenofibrate, benzbromarone, gemfibrozil and clofibrozil were identified as the compounds in Open TG-GATEs with strong positive connections with nafenopin. These compounds are all recognized as PPARα agonists (Blomme *et al*., 2009; Kramer *et al*., 2003). Peroxisome proliferators could produce oxidative stress and lead to hepatotoxicity in rats (Blomme *et al*., 2009), therefore representing an important group of hepatic carcinogens in rodents (Cattley, 2004). This study clearly indicated the potential of LTMap to reveal similar toxicogenomic effects induced by different compounds, which would provide clues and opportunities to understand the toxicological mechanism of compounds from some novel perspectives.

Discussion

Microarray technology has been playing a significant role in many areas of biomedical research. Toxicogenomics, one of its applications in toxicology, has achieved significant impact in toxicological research. Here we focused on hepatotoxic side effects of drugs, an important issue for drug safety. It is urgent to enhance our ability to predict, characterize and understand drug-induced hepatotoxicity. In this study, we developed LTMap, a web server to support high-throughput analysis in liver toxicity. LTMap applied the fold-change ranking method established previously by CMap into expression profiles similarity comparison of toxicogenomic data and constructed a core database with prebuilt toxicogenomic expression reference profiles for 170 compounds. The results of our examples demonstrated that this method could work well in revealing the molecular processes related to toxicological effects and is capable of connecting drugs according to their toxicogenomic patterns. In conjunction with experimental validation, LTMap would provide a strategy to facilitate the understanding of toxic mechanisms of compounds. Therefore, LTMap may offer the opportunity to discern toxicity at an early stage of drug development based on the available toxicogenomic information for the known drugs. LTMap is expected to be a powerful tool to help researchers identify and interpret patterns of significant and coordinated changes in gene expression data and may be useful for drug safety and efficacy evaluation.

SUPPORTING INFORMATION

Supporting information may be found in the online version of this article.

Acknowledgements

This work was financially supported by the National S&T Major Project (no. 2012ZX09505001-001), the National Basic Research Program of China (no. 2012CB518405), Program for New Century Excellent Talents in University (NCET-12-0488) and the Natural Science Foundation of China (no. 81173465).

References

Barros SA. 2005. The importance of applying toxicogenomics to increase the efficiency of drug discovery. *Pharmacogenomics* **6**(6): 547–550.

Blomme EAG, Yang Y, Waring JF. 2009. Use of toxicogenomics to understand mechanisms of drug-induced hepatotoxicity during drug discovery and development. *Toxicol. Lett.* **186**: 22–31.

Cattley RC. 2004. Peroxisome proliferators and receptor-mediated hepatic carcinogenesis. *Toxicol. Pathol.* **32**(5): 6–11.

Chen M, Vijay V, Shi Q, Liu Z, Fang H, Tong W. 2011. FDA-approved drug labeling for the study of drug-induced liver injury. *Drug Discov. Today* **16**(15–16): 697–703.

Chen M, Zhang J, Wang Y, Liu Z, Kelly R, Zhou G, Fang H, Borlak J, Tong W. 2013. The liver toxicity knowledge base: a systems approach to a complex end point. *Clin. Pharmacol. Ther.* **93**(5): 409–412.

Collier R. 2009. Rapidly rising clinical trial costs worry researchers. *Can. Med. Assoc. J.* **180**(3): 277–278.

Fan X, Lobenhofer EK, Chen M, Shi W, Huang J, Luo J, Zhang J, Walker SJ, Chu TM, Li L, *et al*. 2010. Consistency of predictive signature genes and classifiers generated using different microarray platforms. *Pharmacogenomics J.* S17–S27.

Fan XH, Shao L, Fang H, Tong WD, Cheng YY. 2011. Cross-platform comparison of microarray-based multiple-class prediction. *PLoS One* **6**(1): e16067.

Ganter B, Tugendreich S, Pearson CI, Ayanoglu E, Baumhueter S, Bostian KA, Brady L, Browne LJ, Calvin JT, Day GJ *et al*. 2005. Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. *J. Biotechnol.* **119**(3): 219–244.

Hirode M, Omura K, Kiyosawa N, Uehara T, Shimuzu T, Ono A, Miyagishima T, Nagao T, Ohno Y, Urushidani T. 2009. Gene expression profiling in rat liver treated with various hepatotoxic-compounds inducing coagulopathy. *J. Toxicol. Sci.* **34**(3): 281–293.

Kim JM, Fisher JW, Yezzi A, Cetin M, Willsky AS. 2005. A nonparametric statistical method for image segmentation using information theory and curve evolution. *Ieee T Image Process* **14**(10): 1486–1502.

Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, Pon A, Banco K, Mak C, Neveu V, *et al*. 2011. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.* **39**(Database issue): D1035–D1041.

Kramer JA, Blomme EA, Bunch RT, Davila JC, Jackson CJ, Jones PF, Kolaja KL, Curtiss SW. 2003. Transcription profiling distinguishes dose-dependent effects in the livers of rats treated with clofibrate. *Toxicol. Pathol.* **31**(4): 417–431.

Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet J-P, Subramanian A, Ross KN, *et al*. 2006. The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**(5795): 1929–1935.

Leming S, Weiming H, Zhenqiang S, Xianping L, Weida T. 2003. Microarrays: Technologies and applications. In *Applied Mycology and Biotechnology*, Dilip KA, George GK (eds). Elsevier; 271–293.

Lu XY, Jin TT, Jin YC, Wu LH, Hu B, Tian Y, Fan XH. 2013. Toxicogenomic analysis of the particle dose- and size-response relationship of silica particles-induced toxicity in mice. *Nanotechnology* **24**(1): 444–455.

Lu XY, Hu B, Shao L, Tian Y, Jin TT, Jin YC, Ji S, Fan XH. 2013b. Integrated analysis of transcriptomics and metabonomics profiles in aflatoxin B1-induced hepatotoxicity in rat. *Food Chem. Toxicol.* **55**: 444–455.

Shi L, Campbell G, Jones WD, Campagne F, Wen Z, Walker SJ, Su Z, Chu TM, Goodsaid FM, Pusztai L, *et al*. 2010a. The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. *Nat. Biotechnol.* **28**(8): 827–838.

Shi Q, Hong H, Senior J, Tong W. 2010b. Biomarkers for drug-induced liver injury. *Expert Rev. Gastroenterol. Hepatol.* **4**(2): 225–234.

Stankova P, Kucera O, Lotkova H, Rousar T, Endlicher R, Cervinkova Z. 2010. The toxic effect of thioacetamide on rat liver in vitro. *Toxicol. In Vitro* **24**(8): 2097–2103.

Suter L, Schroeder S, Meyer K, Gautier JC, Amberg A, Wendt M, Gmuender H, Mally A, Boitier E, Ellinger-Ziegelbauer H, *et al*. 2011. EU framework 6 project: predictive toxicology (PredTox)—overview and outcome. *Toxicol. Appl. Pharmacol.* **252**(2): 73–84.

Tong W, Cao X, Harris S, Sun H, Fang H, Fuscoe J, Harris A, Hong H, Xie Q, Perkins R, *et al*. 2003. ArrayTrack—supporting toxicogenomic research at the U.S. Food and Drug Administration National Center for Toxicological Research. *Environ. Health Perspect.*. **111**(15): 1819–1826.

Uehara T, Ono A, Maruyama T, Kato I, Yamada H, Ohno Y, Urushidani T. 2010. The Japanese toxicogenomics project: application of toxicogenomics. *Mol. Nutr. Food Res.* **54**(2): 218–227.

Wang Y, Xiao J, Suzek TO, Zhang J, Wang J, Bryant SH. 2009. PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res.*. **37**(Web Server issue): W623–W633.

Zhang M, Chen M, Tong W. 2012. Is toxicogenomics a more reliable and sensitive biomarker than conventional indicators from rats to predict drug-induced liver injury in humans? *Chem. Res. Toxicol.* **25**: 122–129.