Assess Genomic Biomarkers of Toxicity in Drug Development

Wei Zhang, Scott Emrich and Erliang Zeng*
Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN, USA
∗ E-mail: ezeng@nd.edu

1 Introduction

Typical risk assessment strategies use animal models and *in vitro* experiments as surrogates for human studies in the early stages of drug development. Toxicity assessment is then conducted using conventional indicators such as pathology and clinical chemistry data. Although these methods are widely used, around 40% of drug-induced liver injury (DILI) cases are not detected in the preclinical studies using these conventional indicators, and agreement between studies on animal models and human clinical trials is often poor. To overcome these issues, advances in modern "-omics" including high-throughput microarrays and next-generation sequencing technologies have allowed using genomic biomarkers in risk assessment. The underlying hypothesis is that genomic biomarkers will be more sensitive than conventional markers in detecting toxicity signals.

In this paper, we predicted DILI based on microarray data sets provided by the Japanese toxicogenomics project [1], a CAMDA 2013 challenge. We first explored the possibility of replacing the animal model with in vitro assay coupled with toxicogenomics. Previous studies addressed this problem using the agreement of differentially expressed gene lists from *in vivo* and *in vitro* data, and found poor agreement between the two [2]. Pessiot *et al.* then proposed to evaluate the *in vivo-in vitro* agreement using Gene Set Enrichment Analysis (GSEA) on collapsed probesets [2], which has shown success in improving such agreement. Here we took an alternative approach. Instead of comparing features (for example, differentially expressed genes) resulting from *in vivo* and *in vitro* experiments, we evaluated biological consequences such as pathological measurements and observed DILI by comparing the power of gene features to predict these consequences. The underlying hypothesis was that the processes that cause pathology and DILI effects are complicated and may involve many factors; and although *in vivo* and *in vitro* data sets may share many common characteristics, they could also capture different biological information. Because drug toxicity could result from perturbations of biological metabolic pathways, these effects could happen at any level and could be induced by several key players. We also explored the possibility of predicting the DILI potential in humans using the *in vitro* data from rat primary hepatocytes or human primary hepatocytes. Our method focused on the analysis of resulting pathological data and therefore could provide a more fair comparison using downstream effects as our key. Because the pathological data is available for *in vivo* assays only, we assumed that drug will cause similar pathological

results in *in vitro* experiments as it does for *in vivo* assays. For DILI data, we assumed a drug has similar liver injury effect in humans and rats for predicting liver injury using toxicogenomics data from animals.

2   Materials and Methods

We explored the possibility of predicting the DILI potential in humans using Japanese toxicogenomics project data, which provided close to 20,000 pre-processed Affymetrix microarrays used to measure the effects of 131 drugs on the liver [1]. These included rat *in vivo* data with two experiment designs (single and repeated) and *in vitro* data of both rat and human (using rat and human hepatocytes). Various drug dose levels and sacrifice time after treatments were applied in the experiment design. We used FARMS-summarized and collapsed gene expression values as described previously [3] in modeling and analysis for this paper.

Among the 131 drugs, 101 were associated with one of the following categories: "Most DILI concern", "less DILI concern" and "no DILI concern". We considered DILI prediction as a binary classification problem. Unlike previous work [3] that used two classes of "Most DILI" against "Less DILI" or "No DILI", we used the control microarray data as one class and the microarray data from " Most DILI concern" and "Less DILI concern" drugs as the other class because we found it is difficult to differentiate between these two labels. Each microarray data is represented by 12,088 genes (rats) or 18,988 genes (humans).

CAMDA challenge also provided a total of 5569 summaries of the rat liver pathology reports as previously described [1], which makes supervised training possible to predict pathology. For each pathology finding, we ignore severity and create a dataset for that finding and use binary classification model to classify it. As in previous work [4], only the five most frequent pathology findings, for which the largest data sets were available, were evaluated: hypertrophy, necrosis, cellular infiltration, microgranuloma and cellular change.

As to the classification model used we applied Random Forest (RF) [5], which is an ensemble approach based on the aggregation of a set of decision trees, where each tree is grown from a bootstrap sample (sampling with replacement) of the original data. The average over all of the predictions from the individual trees was considered as the final predicted value. In addition to achieving competing prediction accuracy compared to the state-of-the-art machine learning methods, Random Forest also could cope with high dimensional data and has good model interpretability while incorporating variable selection inside the learning process.

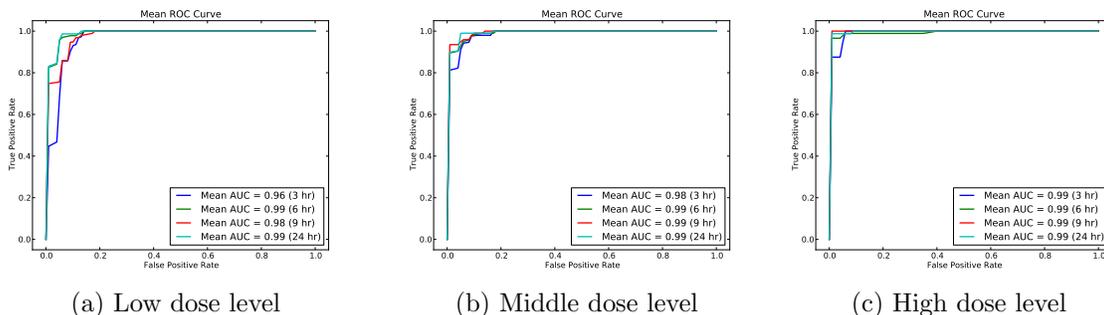Source code and additional results from the analysis are available at: `https://bitbucket.org/davidzhang/camda2013`.

Figure 1: ROC curves in classifying DILI using Rat *in vitro* data

3 Results

We first explored influences of different experiment designs (dose and sacrifice time) on the prediction of DILI potential. Figure 1 demonstrates ROC curves for the classification of DILI categories using rat *in vitro* data sets of different combinations of dose level (low, middle and high) and time point (3, 6, 9 and 24 hour after treatment). The ROC curves are averaged results using 5-fold cross validation. Although various combinations of dose and time points have different sets of differentially expressed genes, they all have good and similar discriminative power in classifying samples as damaged and non-damaged. This observation suggests that time information and dose level are not critical factors in assessing drug toxicity in these data, which is consistent with the prior findings of Pessiot *et al.* [2].

3.1 Comparisons between rat *in vivo* and *in vitro* studies

To evaluate the possibility of replacing an animal *in vivo* study with *in vitro* assay, we built a classifier on available rat *in vitro* data and then used it to predict rat *in vivo* data. Figure 2(a) demonstrates the ROC curve of DILI classification using 1,000 trees in the Random Forest. The result is promising (AUC=0.83). On the other hand, the RF model built on *in vivo* data can perfectly predict DILI potential of *in vitro* assay (AUC=1.00, results not shown). In addition to providing prediction, RF also calculates variable importance (VIM) for each feature when constructing the model. We then compared two gene lists: one list include genes in *in vivo* study with VIMs not larger than 0 (referred to as *in vivo* gene list), and the other list contains genes from *in vitro* assay whose VIMs are not larger than 0 (referred to as *in vitro* gene list). These two gene lists were obtained according to RF VIMs based on RF models constructed on *in vivo* and *in vitro* data sets separately. Figure 4 shows the *Venn* diagram of the comparison. Among 5,845 total important genes, only 1575 (26.88%) genes are shared by two lists, indicating poor agreement when comparison is performed between gene features. Nevertheless, using

(a) Rat *in vitro* classify rat *in vivo*

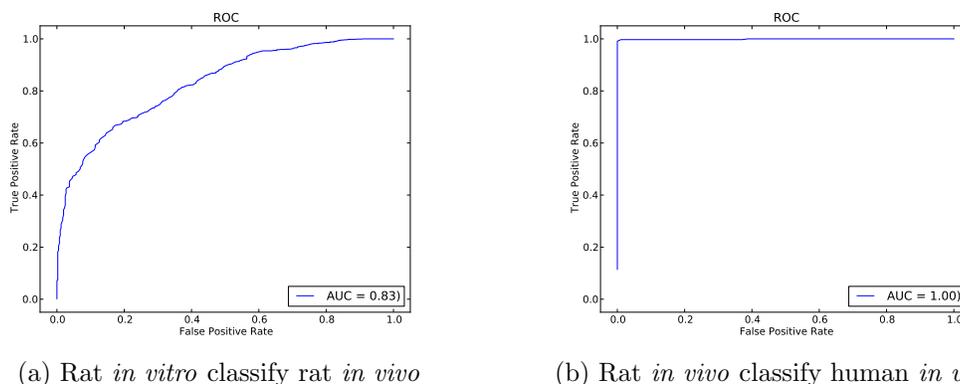(b) Rat *in vivo* classify human *in vitro*

Figure 2: ROC curves of rat *in vivo* DILI classification using RF model built on rat *in vitro* data (a) and of human *in vitro* DILI classification using RF model built on rat *in vivo* data (b)

DILI classification as an interpretation of important genes from the two lists is a better approach to compare rat *in vivo* and *in vitro* data sets. Since DILI is derived eventually from pathology and clinical chemistry data, thus we expect classifier using RF model for predicting pathological data also preform good. Figure 3 demonstrates five leading pathologies classification using RF model built on rat *in vivo* data (Figure 3(a)) and rat *in vitro* data (Figure 3(b)) .

## 3.2   Comparisons between human *in vitro* and rat *in vitro* data

The results as shown in Figure 2(a) demonstrate that replace animal model with *in vitro* assay is possible. Furthermore, we attempted to predict DILI in humans using rat toxicogenomics data. We created an ortholog gene mapping between human genes and rat genes according to their corresponding probe-set common gene names and obtained a list of 9,947 pairs of ortholog genes. A RF classification model was constructed using rat *in vitro* data and such model was then used to predict DILI potential of human *in vitro* gene expression data. The result as shown in Figure 2(b) demonstrates that we could accurately classify human DILI (AUC=1.0) using model built on rat *in vitro* data, which implies it is possible to predict the liver injury in humans using toxicogenomics data from animal *in vitro* assays.

References

1. Uehara T, Ono A, Maruyama T, Kato I, Yamada H, et al. (2010) The japanese toxicogenomics project: application of toxicogenomics. Molecular nutrition & food
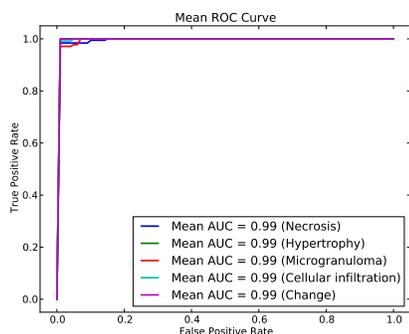
(a) Using RF model built on rat *in vivo* data
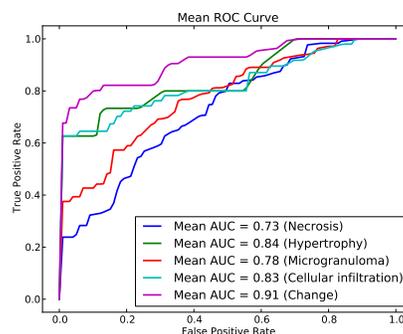
(b) Using RF model built on rat *in vitro* data

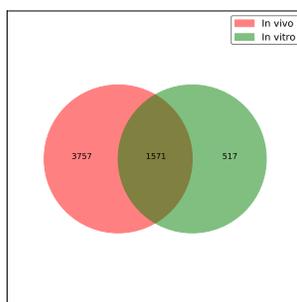Figure 3: ROC curves for classifying five pathologies

Figure 4: Venn diagram of common differentially expressed genes in *in vivo* and *in vitro* microarray

research 54: 218–227.

2. Pessiotm JF, Wong PS, Maruyama T, Morioka R, Aburatani S, et al. (2013) The impact of collapsing data on microarray analysis and dili prediction. Systems Biomedicine 1: 1–7.

3. Pessiot JF, Wong PS, Maruyama T, Morioka R, Aburatani S, et al. (2013) The impact of collapsing data on microarray analysis and dili prediction. Systems Biomedicine 1: 1–7.

4. Shigetta R, Bowles M (2013) Statistical models for predicting liver toxicity from genomic data. Systems Biomedicine 1: 1–6.

5. Breiman L (2001) Random forests. Machine learning 45: 5–32.