International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:6, No:4, 2012

Protein Profiling in Alanine Aminotransferase Induced Patient cohort using Acetaminophen

Gry M, Bergström J, Lengquist J, Lindberg J, Drobin K, Schwenk J, Nilsson P, and Schuppe-Koistinen I.

Abstract**—**Sensitive and predictive DILI (Drug Induced Liver Injury) biomarkers are needed in drug R&D to improve early detection of hepatotoxicity. The discovery of DILI biomarkers that demonstrate the predictive power to identify individuals at risk to DILI would represent a major advance in the development of personalized healthcare approaches. In this healthy volunteer acetaminophen study (4g/day for 7 days, with 3 monitored non-treatment days before and 4 after), 450 serum samples from 32 subjects were analyzed using protein profiling by antibody suspension bead arrays. Multiparallel protein profiles were generated using a DILI target protein array with 300 antibodies, where the antibodies were selected based on previous literature findings of putative DILI biomarkers and a screening process using pre dose samples from the same cohort. Of the 32 subjects, 16 were found to develop an elevated ALT value (2Xbaseline, responders). Using the plasma profiling approach together with multivariate statistical analysis some novel findings linked to lipid metabolism were found and more important, endogenous protein profiles in baseline samples (prior to treatment) with predictive power for ALT elevations were identified.

Keywords**—**DILI, Plasma profiling, PLSDA, Randomforest.

I. Introduction

DRUG induced liver injury (DILI) is the leading cause of drug attrition due to either preclinical toxicity or toxicity in man in clinical trials. Moreover, liver injury is among the most common causes of acute liver failure in the United States, accounting for approximately 13% of all cases [1]. The liver is the most exposed organ related to drug toxicity due to its metabolizing capacity of exogenous compounds to reactive intermediates. These metabolites can cause progressive hepaocyte damage, fulminant hepatic failure and in severe cases even death if liver transplantation is not performed.

DILI is often categorized into two different types; type 1 and type 2, where type 1 is a dose- dependant state which usually is possible to predict, while type 2 is an ideosyncratic condition which refers to the combination of genetic and non-genetic factors that make some patients more susceptible of developing injuries [2].

From a regulatory perspective; one case of DILI in an entire clinical trial population is considered ominous and may reject the drug candidate. Indications of drug induced liver injuries are thus a major bottleneck in drug development, and inflict huge costs for the pharmaceutical industry.

Gry M, Bergström J, Lengquist J, Lindberg J, and Schuppe-Koistinen I are with the molecular toxicology, AstraZeneca R&D, Södertälje, Sweden.

Drobin K, Schwenk J and Nilsson P are with Science for life laboratory, Stockholm, KTH –Royal Institue of Technology, Box 1031, Solna, Sweden.

To further complicate the issues of drug development and DILI, is idiosyncratic DILI incidence is ranging from 5-20 / 1000 exposed for the more DILI-inducing drugs, to 1-10 / 1 million exposed to drugs which has very few recordings of DILI onset. It is therefore not surprising that DILI is not often considered a potential diagnosis when liver dysfunction is identified, and this is a major hurdle when clinical studies are planned [3].

Presently, there is no specific test for DILI, nor any means of singling out a DILI-inducing drug among many received, so the physician must empirically decide if a certain treatment should be terminated or not. In order to clinically define and diagnose DILI onset, the main clinical bio-marker used today is alanine transferase (ALT). Additional markers are aspartate-transaminase (AST) and alkaline phosphatase (ALP), where ALT and ALP data often are combined in order to estimate the type of liver damage. The concentration of the amino transferases (ALT, AST and ALP) in plasma, in combination with the medical history of the patient and the experience of the physician, are the factors that are utilized to diagnose a DILI onset [1]. However, amino-tranferases suffer from limitations regarding specificity and insensitivity to discriminate DILI from other forms of causes of amino-tranferase elevations, such as muscle tissue breakdown and alcohol intake. In addition, it is believed that severe liver damage occurs only in a subset of the patients experiencing increased levels of trans-aminases. In general, even though an elevation in amino-tranferases, some patients (typically 90 %) show no or mild signs of liver dysfunctions (adaptation), and only a small proportion (<1%) fail to adapt and develop DILI are seen [4].

In order to predict drug disposition and liver toxicity, several models have been used [5]. Acetaminophen is known to result in severe, dose-dependent liver injury when administered in large quantities, and is the leading cause of DILI. The drug has been used for many years, and the metabolic breakdown of the drug is well documented as well as the potential risks of the drug metabolites. It has been shown that when the drug is used in therapeutical quantities, usually 4 mg per day, some might experience an increase in amino-tranferases, although no liver injury is detected.

The aim of this study is to find alternative bio-markers for DILI with better specificity and sensitivity than the existing aminotransferases. Further, the new biomarkers should give mechanistic insights, information and reliable signals about the liver injury.To insert images in *Word*, position the cursor at the insertion point and either use Insert | Picture | From File or copy the image to the Windows clipboard and then Edit | Paste Special | Picture (with “Float over text” unchecked).

II. Material and Methods

Urine and blood samples from a cohort consisting of 32 healthy individuals were collected. The patients were administered acetaminophen, 1 mg four times daily, for seven days during a 14 day trial. The first 3 days were an acclimation period and dosing started on day four and ended at day 10. Day 11 to 14 composed a recovery phase. Clinical parameters, such as ALT and AST were measured and recorded throughout the study [6]. The maximum ALT level during the trial was used to stratify the samples into two different strata; non-responders, and responders. The peak ALT-value was compared to the mean value from the three acclimation days and the cutoff was peak ALT value <1.5 * mean baseline values for non-responders and peak ALT value >2 times the baseline values for the responders.

A. Data Generation

Protein expression data was collected using a bead based solution assay [7]. 300 antibodies were related to DILI and liver specific expression based on literature knowledge, RNA-transcriptional data, protein expression data using the human protein atlas, and protein expression data from a previous experiment using pre dose samples from day one (data not shown). In short, antibodies were coupled to beads with a unique identity which can be identified and quantified using a flourometer. The samples were labeled using biotine and once the antibodies had been pooled with the samples and washed; streptavidin was used to quantify the relative amount of antibodies that has an interaction to the target proteins. The amount of expression is estimated by using the median intensity per antibody for all beads that are recorded by the flourometer. In all, 300 protein expression values per sample were measured and collected.

B. Data Analysis

The intensity data were normalized using probability quotient normalization [8] and potential biases were evaluated using principal component analysis [9].

Since the data have longitudinal characteristics, a general mixed model was applied. The model was used on two different sets of data, one where all samples were included and a second where the recovery phase was excluded.

In addition, the data was analyzed using three classifying procedures (partial least squares discriminant analysis (pls-da)[10], random forest [11] and receiver operator statistics (ROC)). These three methods were applied cross-sectionally using data stratified into four different time spans; pre dose (day 1-3), early dose (day 4-6), late dose (day 7-9) and post dose (day 11-14) where measurements for each patient were averaged within the time span.

The models were applied using the lme4-package and the R-software (cran.r-project.org). Pls-da and random forest were applied using the randomForest- and the caret (classification and regression training) –packages [12]. Using the caret functions, the maximum number of components in pls-da was set to 18 and number of trees in randomForest was set to 2000. Prior the classification procedure the variables were preprocessed, (scaled and centered) using the preprocess and predict functions in the caret package.

Using the abovementioned methods, p-values and the coefficients from the generalized linear models as well as the variable importance values from the classifiers were calculated. For further evaluation of data and selection of proteins, the p-value cutoff for the mixed models was set to 0.05 (following multiple adjustments) and the variable importance cutoff was selected based on a random sampling procedure. In this procedure all samples were randomly assigned and used as input to the classifiers and the variable importance was calculated. This was iterated 1000 times and the distribution of variable importance's was compared with the variable importance value from the model using the correct sample stratification. If the variable importance was within the 5 % quintile the protein was selected for further analysis.

The proteins that were selected from the analysis procedure were further investigated by using Ingenuity Pathway Analysis using default parameters [13].

III. Results

Analyzing 300 proteins, 10 of them were found significantly changed across time and between groups using the mixed models across all days and 5 for the reduced data. Two examples of significant variables are shown in Fig. 1.

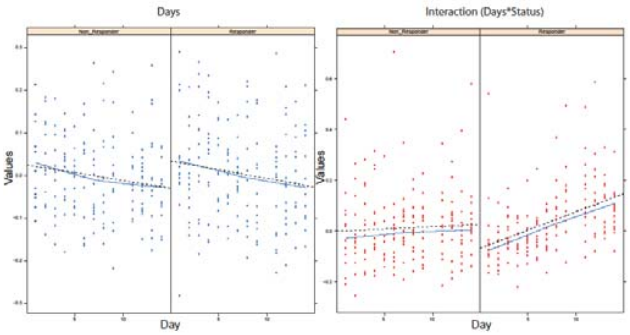

Fig. 1 Two examples of mixed models. The plot to the left shows data from a variable significant across days and the right plot depicts a variable significant for the interaction between individual status and time. The lines are linear regression and a lowess fit (solid line)

The number of variables that were in the 95 percentile for each specific variable importance distribution using the plsda classifier distributions was 27 across the different time strata. The correlation between the different classification model and the variable importance is depicted in Fig. 2.

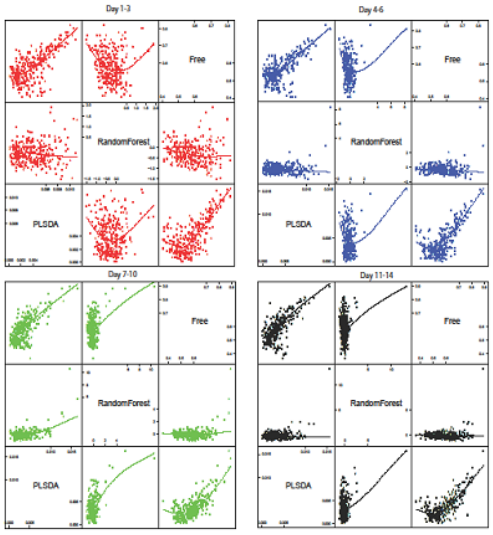

Fig. 2 The correlation between different classifiers

The overlap between days for the plsda classifying model is depicted in a venndiagram in Fig. 3.

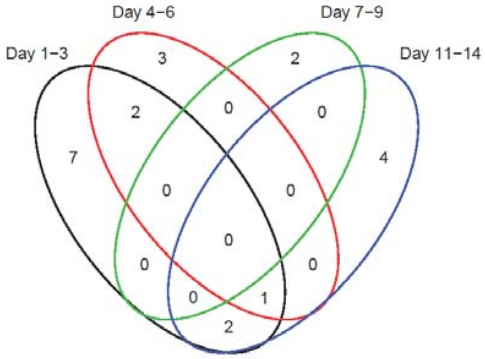

Fig. 3 All variables found as significant using the plsda classifier. The variables were selected based on iterative comparisons of variables originating from the model with correct class partitioning and the models with randomly sampled data. The 95:th percentile was used as cutoff value

The union of all the significant variables was analyzed regarding biological context in Ingenuity Pathway Analysis (REF) were the top ranked canonical pathways and bio functions were related to conversion and homeostasis of lipids. There were also significant pathways related to amino acid perturbations.

IV. Discussion

Using cross-sectional t-tests, none of the antibodies used as proxy for protein expression showed any significant change between the two groups at any time point. Although the difference in ALT-levels between the two groups is small and that the dose given is at therapeutic amounts, a subset of proteins was expected to differ between the groups. By applying power analysis, it was shown that the difference in protein expression had to be quite substantial due to both the relatively small number of patients and the inherent variability of the antibodies. By applying methods which utilize all the data (generalized mixed models) and multivariate methods we still managed to distinguish a few interesting proteins among the 300 investigated. In order to increase the power of the analysis more samples have to be included since the inherent variability of the antibodies is a fixed atribute. To estimate the variability make more technical replicates could be included if possible. These estimations could be utilized to select which antibodies to include in an additional experiment and to exclude antibodies with high variance. By reducing the total number of different antibodies, antibodies targeted towards the same protein could be included and the estimations of protein abundance in plasma should be more accurate.

References

[1] Chalasani N, Fontana RJ, Bonkovsky HL, Watkins PB, Davern T, Serrano J, Yang H, Rochon J; Drug Induced Liver Injury Network (DILIN), *"Causes, clinical features, and outcomes from prospective study of drug-induced liver injury in the United States."* Gastroenterology, 2008, Dec;135(6):1924-34. Epub 2008 Sep 17.

[2] Toru Usui, Masashi Mise, Takanori Hashizume, Masashi Yabuki and Setsumo Komuro;*"Evaluation of the Potential for Drug-Induced Liver Injury Based on in Vitro Covalent Binding to Human Liver Proteins" Drug Metabolism and Disposition*, The American Society for Pharmacology and Experimental Therapeutics, DMD 37:2383-2392, 2009, Vol. 37, No. 12 28860/3529781.

[3] Russman,S, Kullak-Ublick G A and Grattagliano I;*Current Concepts of Mechanisms in Drug-Induced Hepatotoxicity. Curr Med Chem, 2009 August; 16(23):3041-3053.*

[4] Rawlins & Thompson, Textbook of adverse drug reactions, Oxford University Press 1991:18-45. Gruchalla, Lancet 2000,356:1505-11.

[5] Patel SJ, Milwid JM, King KR, Bohr S, Iracheta-Velle A, Li M, Vitalo A, Parekkadan B, Jindal R and Yarmush ML. *"Gap junction inhibition prevents drug-induced liver toxicity and fulminant hepatic failure"*. Nat Biotechnology 2012 Jan 15;30(2):179-83.doi: 10.1038/nbt.2089.

[6] O'Connel TM, Watkins PB. *"The application of metabonomics to predict drug-induced liver injury"*. Clin. Pharmacol. Ther. 2010 Sep;88(3):394-9. Epub 2010 Jul 28.

[7] Neiman M, Hedberg J, Dönnes P, Schuppe-Koistinen I, Hanschke S, Schindler R, Uhlén M, Schwenk JM and Nilsson P. *"Plasma profiling reveals human fibulin-1 as candidate marker for renal impairment". J Proteome Res. 2011 Nov 4;10(11):4925-34. Epub 2011 Sep 29.*

[8] Dieterle F, Ross A, Schlotterbeck G and Senn H. *"Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics."*. Anal. Chem. 2006 Jul 1;78(13):4281-90.

[9] Yao F, Coquery J, Le Cao KA. *"Independent Principal Component Analysis for biologically meaningful dimension reduction of large biological data sets"*. BMC Bioinformatics. 2012 Feb 3;13(1):24.

[10] Westerhuis JA, van Velzen EJ, Hoefsloot HC and Smilde AK.*"Multivariate paired data analysis: multilevel PLSDA versus OPLSDA"*. Metabolomics. 2010 Mar;6(1):119-128. Epub 2009 Oct 28.J.

[11] Golugula A, Lee G, Medabhushi A. *"Evaluating feature selection strategies for high dimensional, small sample size datasets"*. Conf Proc IEEE Eng Med Biol Soc. 2011 Aug;2011:949-52. PMID:22254468.

[12] Kuhn M, Contributions from Wing J, Weston S, Williams A, Keefer C and Engelhardt A. *"Classification and regression training"*. URL:http://caret-r.forge.r-project.org/. Published:2012-02-06.

[13] IPA. The data was analyzed using the Ingenuity Pathway Analysis (www.ingenuity.com).