Evaluating the sensitivity and specificity of promising circulating biomarkers to diagnose liver injury in humans

Llewellyn, Heather P ; Vaidya, Vishal S ; Wang, Zhenyu ; Peng, Qinghai ; Hyde, Craig ; Potter, David ; Wang, Jianying ; Zong, Qing ; Arat, Seda ; Martin, Matt ; Masek-Hammerman, Katherine ; Warner, Roscoe ; Johnson, Kent ; Kullak-Ublick, Gerd A ; Aithal, Guruprasad P ; Dear, James W ; Ramaiah, Shashi K

Abstract: Early diagnosis of drug-induced liver injury (DILI) continues to be a major hurdle during drug development and post marketing. The objective of this study was to evaluate the diagnostic performance of promising biomarkers of liver injury—glutamate dehydrogenase (GLDH), cytokeratin-18 (K18), caspase-cleaved K18 (ccK18), osteopontin (OPN), macrophage colony-stimulating factor (MCSF), MCSF receptor (MCSFR), and microRNA-122 (miR-122) in comparison to the traditional biomarker alanine aminotransferase (ALT). Biomarkers were evaluated individually and as a multivariate model in a cohort of acetaminophen overdose (n = 175) subjects and were further tested in cohorts of healthy adults (n = 135), patients with liver damage from various causes (n = 104), and patients with damage to the muscle (n = 74), kidney (n = 40), gastrointestinal tract (n = 37) and pancreas (n = 34). In the acetaminophen cohort, a multivariate model with GLDH, K18 and miR-122 was able to detect DILI more accurately than individual biomarkers alone. Furthermore, the three-biomarker model could accurately predict patients with liver injury compared to healthy volunteers or patients with damage to muscle, pancreas, gastrointestinal tract and kidney. Expression of K18, GLDH and miR-122 was evaluated using a database of transcriptomic profiles across multiple tissues/organs in humans and rats. K18 mRNA (Krt18) and MiR-122 were highly expressed in liver whereas GLDH mRNA (Glud1) was widely expressed. We performed a comprehensive, comparative performance assessment of seven promising biomarkers and demonstrated that a three-biomarker multivariate model can accurately detect liver injury.

DOI: <https://doi.org/10.1093/toxsci/kfab003>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-200202>

Journal Article

Accepted Version

Originally published at:

Llewellyn, Heather P; Vaidya, Vishal S; Wang, Zhenyu; Peng, Qinghai; Hyde, Craig; Potter, David; Wang, Jianying; Zong, Qing; Arat, Seda; Martin, Matt; Masek-Hammerman, Katherine; Warner, Roscoe; Johnson, Kent; Kullak-Ublick, Gerd A; Aithal, Guruprasad P; Dear, James W; Ramaiah, Shashi K (2021). Evaluating the sensitivity and specificity of promising circulating biomarkers to diagnose liver injury in humans. *Toxicological Sciences*, 181(1):23-34.

DOI: <https://doi.org/10.1093/toxsci/kfab003>

**Evaluating the sensitivity and specificity of promising circulating biomarkers to
diagnose liver injury in humans**

Heather P. Llewellyn^{1*}, Vishal S. Vaidya^{1*#}, Zhenyu Wang², Qinghai Peng¹, Craig
Hyde², David Potter¹, Jianying Wang¹, Qing Zong¹, Seda Arat¹, Matt Martin¹, Katherine
Masek-Hammerman¹, Roscoe Warner³, Kent Johnson³, Gerd A. Kullak-Ublick^{4, 5},
Guruprasad P. Aithal⁶, James W. Dear⁷ and Shashi K. Ramaiah^{1#}

¹Drug Safety Research and Development, ²Early Clinical Development, Pfizer, Inc;
³Michigan Medicine, University of Michigan, Ann Arbor, MI; ⁴Mechanistic Safety, CMO &
Patient Safety, Global Drug Development, Novartis, Basel, Switzerland, ⁵Department of
Clinical Pharmacology and Toxicology, University Hospital Zurich, University of Zurich,
Zurich, Switzerland; ⁶NIHR Nottingham Biomedical research Centre, Nottingham
University Hospitals NHS Trust and the University of Nottingham, Nottingham, UK;
⁷Pharmacology, Therapeutics and Toxicology, Centre for Cardiovascular Science,
University of Edinburgh, The Queen's Medical Research Institute, Edinburgh, UK

*Co first author

#Co corresponding author

Running Title: Biomarkers of Liver Injury

Key Words: keratin-18, microRNA, glutamate dehydrogenase, diagnosis, liver

24 **Contact information**

25 Vishal S. Vaidya, PhD, Tel: 617.894.0927, Email: vishal.vaidya@pfizer.com or Shashi
26 K. Ramaiah, DVM, PhD, Tel: 636.222,0593, Email: Shashi.Ramaiah@pfizer.com
27 Drug Safety Research & Development, Pfizer, Inc, 300 Technology Square Drive, 3rd
28 Floor, Cambridge, MA 02139

29

30 **List of abbreviations**

31 Drug-induced liver injury (DILI)
32 Glutamate dehydrogenase (GLDH),
33 Cytokeratin-18 (K18),
34 Caspase-cleaved K18 (ccK18),
35 Osteopontin (OPN),
36 Macrophage colony-stimulating factor (MCSF),
37 MCSF receptor (MCSFR)
38 MicroRNA 122 (miR-122)
39 Alanine aminotransferase (ALT)
40 Acetaminophen (APAP)
41 Gastrointestinal (GI)
42 Food and Drug Administration (FDA)
43 Upper limit of normal (ULN)
44 Receiver operator characteristic (ROC)
45 Area under the curve (AUC)
46 Precision-Recall curve (PRC)

- 47 Aspartate amino transferase (AST)
- 48 Alkaline phosphatase (ALP)
- 49 Blood urea nitrogen (BUN)
- 50 Total bilirubin (TBL)
- 51 Acute liver failure (ALF)

Abstract

Early diagnosis of drug-induced liver injury (DILI) continues to be a major hurdle during drug development and post marketing. The objective of this study was to evaluate the diagnostic performance of promising biomarkers of liver injury - glutamate dehydrogenase (GLDH), cytokeratin-18 (K18), caspase-cleaved K18 (ccK18), osteopontin (OPN), macrophage colony-stimulating factor (MCSF), MCSF receptor (MCSFR), and microRNA-122 (miR-122) in comparison to the traditional biomarker alanine aminotransferase (ALT). Biomarkers were evaluated individually and as a multivariate model in a cohort of acetaminophen overdose (n=175) subjects and were further tested in cohorts of healthy adults (n=135), patients with liver damage from various causes (n=104), and patients with damage to the muscle (n=74), kidney (n=40), gastrointestinal tract (n=37) and pancreas (n=34). In the acetaminophen cohort, a multivariate model with GLDH, K18 and miR-122 was able to detect DILI more accurately than individual biomarkers alone. Furthermore, the three-biomarker model could accurately predict patients with liver injury compared to healthy volunteers or patients with damage to muscle, pancreas, gastrointestinal tract and kidney. Expression of K18, GLDH and miR-122 was evaluated using a database of transcriptomic profiles across multiple tissues/organs in humans and rats. K18 mRNA (*Krt18*) and MiR-122 were highly expressed in liver whereas GLDH mRNA (*Glud1*) was widely expressed. We performed a comprehensive, comparative performance assessment of seven promising biomarkers and demonstrated that a three-biomarker multivariate model can accurately detect liver injury.

Introduction

Drug-induced liver injury (DILI) is a major concern for patients, clinicians, regulatory agencies and drug makers, as it is the leading cause of acute liver failure among patients referred for liver transplantation (Bernal and Wendon 2014; Przybylak and Cronin 2012). The annual incidence of DILI is about 14-24 per 100,000 people (Bjornsson et al. 2013; Sgro et al. 2002; Shen et al. 2019). An overdose of acetaminophen (APAP/paracetamol) is the most common cause of DILI and acute liver failure in the US and Europe (Stravitz and Lee 2019). DILI is also a leading cause of compound attrition during drug development, and drug withdrawals and restrictions after drug approval and marketing (Kullak-Ublick et al. 2017) (Onakpoya et al. 2016). Although idiosyncratic and intrinsic DILI have different pathophysiologies, many biomarkers likely overlap in their ability to detect DILI. A large effort is currently under way in academia, industry and via public-private partnerships to identify early, sensitive and specific translational biomarkers for diagnosis and prognosis of DILI in humans. Furthermore, the Food and Drug Administration (FDA) has a renewed interest to expand guidance on biomarker research to determine hepatotoxic liability of drugs and avenues for biomarker regulatory qualification opportunities.

The current DILI biomarkers are a combination of serum alanine aminotransferase (ALT) and aspartate aminotransferase (AST) which are general indicators of hepatocellular injury, serum alkaline phosphatase (ALP) which is partially predictive of cholestatic liver injury, and total bilirubin (TBL) concentration which is frequently used to predict global liver function (Church et al. 2019; Shi et al. 2010). It is widely accepted that current diagnosis of DILI relies on biomarkers which lack sufficient

specificity and sensitivity for detecting liver injury and therefore, there is a need for development of better biomarkers (Shi et al. 2010), especially those that can be used both in preclinical and clinical studies for drug development.

Promising biomarkers for diagnosis of DILI, that have also been supported by the FDA, include total cytokeratin 18 (K18), caspase cleaved K18 (ccK18), macrophage colony-stimulating factor (MCSF), MCSF receptor (MCSFR), osteopontin (OPN), glutamate dehydrogenase (GLDH) and microRNA-122 (miR-122)(Church et al. 2019; Roth et al. 2020). Although these biomarkers have been evaluated in pre-clinical and clinical studies, a comprehensive study to quantitatively evaluate the performance characteristics of all 7 candidate biomarkers individually and in combination has not been performed. Therefore, the objective of this study was to evaluate the sensitivity and specificity of these promising safety biomarkers individually and in combination for detecting liver injury using acetaminophen overdose and cross-sectional cohorts of patients with liver damage due to diverse etiologies. Specifically, our aims were to 1) compare the diagnostic performance of the seven DILI biomarkers in patients with acetaminophen overdose (APAP, n=175); 2) apply random forest modeling to train, test and validate a multivariate model with top performing biomarkers to predict ALT; 3) independently confirm the performance of biomarkers individually and as a multivariate model in a cross-sectional study involving patients with clinically established liver damage (n=104) as well as patients with other organ damage (n=185) and healthy volunteers (n=135).

Brief Experimental Procedures (Details provided in Supplementary Materials)

Study Populations

Acetaminophen overdose study participants: Ethical approval for this study was provided by London - South East Research Ethics Committee (18/LO/0894) (ClinicalTrials.gov identifier: NCT03497104). Patients presenting to Royal Infirmary of Edinburgh, UK (RIE) following acetaminophen overdose, who met the inclusion criteria, were asked to provide informed consent to participate in the prospective, acetaminophen overdose cohort study and their demographics and blood results were recorded. Although the current consensus for defining DILI is an ALT value $\geq 5x$ upper limit of normal (ULN) (Aithal et al. 2011), in this study a cutoff of three times the upper limit of normal ($\geq 3x$ ULN) ALT (150 U/L) was used as this is consistent with prior studies (Starkey Lewis et al. 2011) and because the FDA has defined an ALT $\geq 3x$ ULN of study patients compared to controls as a potential signal of DILI during drug development in particular (Senior 2014). A cutoff of > 1 ULN ALT (> 50 U/L) was also explored. Serum was collected at three timepoints, baseline (T1, $n = 175$), T2 ($n = 127$) and T3 ($n = 81$). T1 was collected when the patient was admitted to the hospital, 4.6 hours (IQT: 4.1, 10.7) after ingestion of acetaminophen. The median collection time for T2 was 12.7 hours (IQT: 9.2, 14.1) after T1, and the median for T3 was 22.9 hours (IQT: 19.8, 24.2) after T1.

Cross-sectional cohort study participants: Patient samples were collected from the University of Michigan health care system with informed consent (IRB approval # HUM-44422). Patient cohorts were selected based on their individual disease states, their serum chemistry values and medical adjudication of their clinical files. Liver damage

141 patients were determined by utilizing the EWG definition ($\geq 5\times$ ALT ULN, or $\geq 2\times$ ALP ULN,
142 or $\geq 3\times$ ALT ULN and $\geq 2\times$ total bilirubin ULN) and medical adjudication demonstrating
143 various liver damage etiologies. Healthy subjects were selected as those having normal
144 ranges of ALT (< 35 U/L), AST ($8 - 30$ U/L), ALP ($0.2 - 1.2$ mg/dL), total bilirubin ($0.2 -$
145 1.2 mg/dL), glucose ($73 - 100$ mg/dL), blood urea nitrogen ($8 - 20$ mg/dL), serum
146 creatinine ($0.5 - 1.0$ mg/dL for females and $0.7 - 1.3$ mg/dL for males), and creatine
147 kinase ($26 - 180$ U/L for females and $38 - 240$ U/L for males). Subjects with clinically
148 demonstrable liver damage typically included those with accidental acetaminophen
149 overdose, ethanol toxicity, drug abuse, transaminitis (elevated transaminases without
150 other evidence of liver injury), metastatic liver disease (diagnosed by biopsy or
151 histopathology after resection), cirrhosis, and liver impairment (Hepatitis B or C, hepatic
152 graft vs host disease). The metastatic group is comprised of 12 different sites of origin of
153 the primary cancers. Represented by adrenal, breast, cholangiocarcinoma, colon,
154 endometrial, kidney, liver, melanoma, ancreatic, pleomorphic sarcoma, prostate and
155 rectal. No single primary cancer site is represented by more than 4 patients of the 27 in
156 this cohort. Some subjects exhibited multiple types of liver injury for example, a subject
157 could be represented in both categories of drug induced liver injury and acute liver failure.
158 Muscle Injury was diagnosed by either i) medical adjudication, ii) a muscle biopsy, iii)
159 genetic testing or iv) clinically determined injuries, which may include, but are not limited
160 to, Primary Disorders of Muscle (Dystrophies, Myotonic Disorders, Congenital
161 Myopathies and Mitochondrial Myopathies) and Toxic Myopathies (Drug, Alcohol and
162 Toxicants), as exhibited by, myositis (inflammatory muscle injury), neurogenic atrophy,
163 necrotizing inflammatory muscle injury, chronic severe atrophy, AAF, type II fiber atrophy,

164 nuclear myobags, denervation atrophy, and increased lipids in myofibers. Subjects
165 demonstrating Pancreatitis (Acute, Chronic, Hereditary) were diagnosed by either i)
166 Persistent Severe Epigastric Pain, ii) Diagnostic Armamentarium [Endoscopic Ultrasound
167 (ES), Magnetic Resonance Cholangiopancreatography (MRCP), Computerized
168 Tomography (CT) or Transabdominal ultrasound] iii) Clinically Demonstrable Deficiencies
169 or iv) Amylase or Lipase 3x ULN. Subjects demonstrating Gastrointestinal abnormalities
170 were diagnosed by either i) Endoscopy, ii) Sigmoidoscopy or iii) Colonoscopy, or iv)
171 Clinically Demonstrable Deficiencies, which could include, but is not limited to,
172 Gastroesophageal Reflux Disease (GERD), Esophagitis, Irritable Bowel Syndrome (IBS),
173 Celiac Disease, Crohn's Disease, Ulcerative Colitis, Ulcerative Pancolitis, Ulcerative
174 Proctosigmoiditis and Appendicitis. Subjects having Chronic Kidney Disease (CKD) were
175 diagnosed by either i) Biopsy-Proven or ii) Clinically Demonstrable Deficiencies, which
176 could include, but are not limited to, Diabetes, High Blood Pressure, Glomerulonephritis,
177 Interstitial Nephritis, Polycystic Kidney Disease and Malformations, as exhibited by, CKD
178 stage II – V, End Stage Renal Disease (ESRD) and patients on Dialysis. Patients were
179 excluded from a cohort if they had any ongoing health problems or immunological flares
180 that could influence liver health, or if they had additional organ injury outside their included
181 cohort. Traditional organ damage biomarkers such as aspartate amino transferase (AST)
182 and alkaline phosphatase (ALP) for liver; lipase and amylase for pancreas, blood urea
183 nitrogen (BUN) and creatinine for kidney; and creatinine kinase enzyme activity for
184 muscle damage were elevated in their respective clinically diagnosed organ damage
185 cohorts (Supplementary Figure 5). All human serum was collected in serum separator

tubes, aliquoted, frozen at -80°C and sent to Pfizer's Drug Safety Research and Development's Biomarker Laboratories for biomarker analysis.

Biomarker Measurements

Clinical chemistry parameters ALT, GLDH, AST, ALP, TBIL, Lipase, AMYL, GLUC, BUN, CREA, CK were evaluated using a Siemens Advia 1800 chemistry analyzer.

Protein biomarkers: K18 and ccK18 were measured by SpectraMax 500 from Molecular Devices using CK_M65 EpiDeath® ELISA kit and CK_M30 Apoptosense® ELISA kit respectively (Manufacturer: PEVIVA AB, Bromma, Sweden; Distributor: DiaPharma, West Chester Township, Ohio, catalog numbers 10040 and 10010). M65 assay can detect full length, nonapoptotic and apoptotic fragments of K18 while M30 assay detects only caspase-cleaved fragments of K18 (Ku et al. 2016). MCSF and OPN were measured by electro-chemiluminescent using Meso Scale Discovery (MSD) Kits (catalog number K151XRK-1 and K151HJC-2) and light intensity signal was detected by Meso Sector S600, Model 1201. MCSFR was measured by fluorescent labeled microbeads using Luminex Magnetic MultiPlex Human MCSFR kit (R&D Systems Inc., Minneapolis, Minnesota, catalog number LXSAHM-01) and the fluorescent signal was detected by Bio-Plex 200, Model Luminex XYP. Biomarker assays were performed according to manufacturers' protocols with a few modifications. The serum biomarker values were calculated using a 6 to 9 point five-parameter logarithmic standard curve (Supplementary Figure 1).

MicroRNA-122: Total RNAs from 100 µl plasma/serum were purified by Qiagen's miRNeasy kit (Valencia, CA, USA) according to the manufacturer's protocol and a total final 20 µl of the purified RNAs was eluted. To remove possible heparin contamination, 6 µl of extracted RNA was added to a master mix consisting of 2 µl of 10x reaction buffer (New England Biolabs, Ipswich, United Kingdom), 10.75 µl of RNA free H₂O, 0.25 µl of Heparinase I (New England Biolabs, Ipswich, United Kingdom) and 1 µl of RNase inhibitor (Promega, WI, United States). Samples were incubated for 1 hour at 30°C followed by 1 minute at 99°C. Samples were stored at -80°C. Five µl of the purified miRNA was subjected to ddPCR quantification. Three step reactions were employed in the quantification of miRNAs. First, a poly(A) tail was added to the miRNAs using a poly(A) enzyme from New England Lab. Next, polyadenylated miRNAs were transcribed to cDNA by reverse transcriptase (MultiScribe™, Applied Biosystems) with poly(T) oligos containing an adapter primer sequence. The cDNA was then quantified with specific forward (5'-GCTGGAGTGTGACAATGGTGTT-3') and universal reverse (5'-TTTCGGCTGCCATGTACGTTTTTTTTTTVN-3') primers using Eva-green in droplet digital PCR (ddPCR). All primers were acquired through Integrated DNA Technologies (IDT, Coralville, Iowa). Circulating miR-122 was assayed in singleton by QX200™ Droplet Digital™ PCR System from Bio-Rad using Evagreen-based detection method. The performance of miR-122 in ddPCR was evaluated to determine assay sensitivity, range of the assay, reproducibility, dilutional linearity, and freeze-thaw stability.

Performance characteristics are described in the supplemental materials.

Statistical Analysis

Area under the curve analysis

Global predictivity across potential cutoffs for any single biomarker was assessed using the AUC (Area Under the Curve) of the Receiver Operator Characteristic (ROC) curve. The ROC curve plots the False Positive Rate horizontally versus the True Positive rate vertically, which represents, respectively, the fraction of actual control samples (e.g., healthy) predicted to be cases (e.g., liver injury), and the fraction of actual case samples predicted to be cases. The curve is generated by visiting every distinguishable cutoff, which corresponds to a cutoff between every pair of adjacent unique sorted values in the observed biomarker dataset. An AUC of 1.0 represents perfectly separable cases and controls, while an AUC of 0.5 represents predictability no better than random guessing.

For each biomarker, we assessed the distinguishability of liver injury as cases versus healthy subjects as controls by calculating the AUC for that biomarker as calculated from the auc function using the pROC package in R (R Core Team, 2019) (R Development Core Team 2019). The significance levels of the AUC values were evaluated using the roc test() function in that same package, using the default DeLong method (DeLong et al. 1988) for comparing AUCs from two datasets. Here, the p-value of a single AUC was evaluated by using roc test() to compare it to the AUC of the null set for the same biomarker values, where the null set was generated by randomly permuting the case and control labels of the biomarker values. To evaluate biomarker specificity for liver injury, AUCs were also evaluated via the same method using other organ injury cohorts as controls against the liver injury cohort as cases. Using roc test() as above, we were also able to assess the statistical significance of AUC differences

between different biomarkers, as well as for comparisons between different control cohorts vs. liver injury for the same biomarker.

Multivariate modeling

To evaluate the predictivity of a panel of the candidate biomarkers (GLDH, K18, miR-122, OPN, ccK18, MCSFR and MCSF) to predict the measured ALT activity value, multivariate models were built using the baseline (T1) APAP overdose patient data. First, the natural logarithm of ALT was used as the dependent variable and candidate biomarkers were used as predictors. Random forest and linear regression models were then built to assess the predictivity of the biomarker panel, i.e. composite score. Importance values were generated from the random forest modeling. All biomarker values were generally comparable between timepoints and the difference in biomarker kinetics were not expected to influence the modeling.

Biomarker selection was based on their importance value > 20 (scaled maximum score is 100). Next, thresholds of predicted log(ALT) were used to categorize subjects into DILI or non-DILI given the condition that sensitivity > 0.95 at T2. DILI was defined as $ALT \geq 150$ U/L ($\geq 3 \times$ ULN) or $ALT > 50$ U/L ($> 1 \times$ ULN). The threshold of 50 U/L was used as the ULN as this is the locally defined ULN at RIE. After the model was built with baseline data (training set) and a threshold was chosen at T2, the model was validated using T3 data.

Random forest was chosen as an optimal model based on the following considerations:

- Correlation coefficient between score (predictive log(ALT)) and measured log(ALT) in both testing set (T2 data) and validation data set (T3 data).

- Number of false positives given a sensitivity > 0.95 in the testing data set; at the same time, defined a DILI threshold to evaluate in the validation data set (T3 data).

Since models were built at baseline, with thresholds decided based on timepoint 2, and validation conducted on data from T3 with the same set of patients, models were also tested in the cross-sectional cohort as an independent data set to evaluate model performance.

Results

Analysis of candidate biomarkers in cohort of patients with acetaminophen overdose

Promising liver injury biomarkers glutamate dehydrogenase (GLDH), cytokeratin 18 (K18), caspase-cleaved K18 (ccK18), microRNA-122 (miR-122), osteopontin (OPN), macrophage colony-stimulating factor (MCSF) and MCSF receptor (MCSFR) were evaluated for their ability to predict ALT in a cohort of patients with acetaminophen (APAP) overdose (n=175) (Supplementary Table 2) at three timepoints. A random forest model to predict ALT was trained, tested and validated on this APAP overdose cohort using GLDH, K18 and miR-122 as they had a high importance value as determined by the random forest model (100, 88.05, 54.57 respectively) relative to OPN, ccK18, MCSFR and MCSF (16, 15.21, 8.27 and 0 respectively). Consistent with prior APAP cohort studies (13) and because ALT ≥ 3 x ULN may be a potential signal of DILI during drug development in particular (14), we first evaluated the predictability of the model using an ALT cutoff of ≥ 150 U/L. GLDH, K18 and miR-122 concentrations were elevated at all timepoints in APAP overdose subjects with ALT ≥ 150 U/L compared to APAP exposed patients with ALT < 150 U/L, with few exceptions (Figure 1A). Using baseline (T1) data to train the model with GLDH, K18 and miR-122 (also referred to as the three-biomarker model), the composite score (i.e. predicted log ALT) produced by this model was highly correlated (R = 0.921) with measured ALT activity (Figure 1B). The model was then tested at the second timepoint (T2) (Figure 1C) and validated at the third timepoint (T3) (Figure 1D). The composite score highly correlated with measured log ALT activity at T2 and T3 and the correlation coefficients (0.905 and

0.922, respectively) were comparable to those from the training data (T1), suggesting generalizability of the model. With the objective of maximizing sensitivity (fixed at ≥ 0.95), the composite score threshold was set at the lowest composite score (4) in subjects with ALT ≥ 150 U/L in the testing data set (Figure 1C). In general, when the values of two or three of the biomarkers were high, the patient tended to have a high composite score (Supplemental figure 2). The composite scores at each timepoint demonstrated high specificity, with few false positives with an ALT cutoff of ≥ 150 U/L (Figure 1E, F) or > 50 U/L (Supplemental table 4). Furthermore, all seven-biomarkers were used in the multivariate model and were evaluated with a cutoff of ALT ≥ 150 U/L (Supplemental Figure 3, Table 1) and > 50 U/L (Table 1, Supplementary table). In this cohort, the specificity and positive predictive value (PPV) of the models were similar between the three- and seven- biomarker models when using a cutoff of either ALT ≥ 150 U/L or > 50 U/L.

In addition to random forest, we also evaluated a linear regression approach to develop a multivariate model for predicting ALT. Values of the receiver operator characteristic area under the curve (ROC AUC) suggest comparable predictivity between the two approaches at T1, T2, T3 (random forest ROC AUC = 0.99, 0.99, 1.00 and linear regression ROC AUC = 0.98, 0.98, 0.99 respectively). However, in cases of significant class imbalance (the total number of a class of data is far less than the total number of another class of data), it is recognized the ROC AUC values can sometimes be overoptimistic (Davis and Goadrich 2006). With that in mind, we also computed the Precision-Recall curve (PRC) AUC values for both approaches. Where ROC curves summarize the tradeoff between sensitivity and specificity, P-R curves summarize the

tradeoff between sensitivity (“recall”) and positive predictive value (“precision”). The results at T1 (random forest PRC AUC = 0.86 and linear regression PRC AUC = 0.61) suggest an advantage to the random forest approach. PRC AUC are similar at T2 and T3 between the random forest (PRC AUC = 0.91, 1.00) and linear regression (PRC AUC = 0.90, 0.97) approaches. While the composite score using a linear regression model correlated with ALT ($R=0.83$, $R=0.91$, $R=0.94$ for T1, T2 and T3 respectively), there were more false positives (Supplemental Figure 4) compared to the random forest model. Therefore, we focused on the results from the random forest model only.

To compare the performance of individual biomarkers to the models, sensitivity was set to ≥ 0.95 and specificity was compared (Table 2) within each timepoint or injury damage cohort. The threshold was determined by maximizing the specificity given sensitivity ≥ 0.95 within each timepoint or injury damage cohort. Consistent with the above findings, the three and seven biomarker model had similar specificities and in general, were higher than the individual biomarkers. K18 had a higher specificity than any other biomarkers at each timepoint and slightly lower specificity than the three and seven-biomarker models, suggesting that K18 might be a sufficient standalone liver injury biomarker.

Performance characteristics of biomarkers in a cross-sectional cohort of patients with liver injury

The performance characteristics of the 7 candidate liver injury biomarkers and multivariate model in comparison with the traditional biomarker, ALT, was further tested in an independent cross-sectional study with healthy volunteers ($n=135$) and patients

with damage to liver (n=104), muscle (n=74), pancreas (n=34), GI (n=37), and kidney (n=40). Liver damage patients included transaminitis (n=54), metastatic liver disease (n=27), drug induced (n=24), cirrhosis (n=20), alcoholic (n=15), hepatitis (n=12), liver transplant (n=9) and acute liver failure (ALF) (n=4) (Supplemental methods, Supplemental Table 3, Supplemental Figure 5). All 7 measured candidate liver injury biomarkers as well as ALT were elevated in patients with liver damage relative to other organ damages (Figure 2A). Of these biomarkers, GLDH, K18 and miR-122 had a greater fold-increase (7.3-, 12.0- and 6.3-fold, respectively) in liver damage over healthy volunteers than ALT (5.3-fold). Candidate biomarkers were also stratified by the type of liver damage (Figure 2B). Biomarkers tended to be highest in patients with ALF, drug-induced liver injury, and transaminitis (elevated transaminases without other evidence of liver injury). As previously reported (Church et al. 2019), GLDH activity showed a positive correlation with ALT activity (Figure 2C). K18, ccK18 and miR-122 levels were also positively correlated with ALT activity (Figure 2C) suggesting that these biomarkers positively associate with ALT. MCSF and MCSFR levels did not correlate with ALT activity ($r=-0.043$, $p=0.66$; $r=0.1512$, $p=0.125$ respectively) and OPN levels did not correlate with ALT activity ($r=-0.2086$, $p=0.0336$).

For each biomarker, we assessed the distinguishability of liver damage (cases) versus healthy subjects (controls) by calculating the area under the receiver operator characteristic curve (AUC) for each biomarker. K18 achieved near complete separation between patients with liver damage and healthy subjects with an AUC of 0.98 (Table 2, Figure 3A, Supplemental Table 4). MCSF achieved an AUC of 0.97, whereas ALT achieved an AUC of 0.93, and GLDH, ccK18, MCSFR, OPN and miR-122 demonstrated

AUCs of 0.87 – 0.92. K18 also distinguished patients with liver damage from those with GI tract, pancreatic, muscle and kidney damage (AUC = 0.959, 0.963, 0.937, and 0.90, respectively) (Supplemental Table 5). However, the K18 AUC for liver vs kidney damage subjects was only 0.90. By comparison, ALT had similar AUC values in healthy compared to GI tract, pancreas and kidney, but was significantly reduced ($p = 7.2e-05$) when compared to the muscle damage patients. We also assessed the statistical significance of AUC differences between different biomarkers using the same comparison cohorts. When comparing AUCs, K18 was superior in terms of sensitivity and specificity over ALT and GLDH in diagnosing liver damage compared to healthy volunteers, GI tract and muscle damage patients (Figure 3A). K18 outperformed ALT for liver damage in all cohorts except kidney injury where they were similar. GLDH only outperformed ALT for muscle injury. cck18 did not outperform ALT in any cohort. MCSF outperformed ALT for healthy, GI tract, and muscle but not for pancreas and kidney. MCSFR only outperformed ALT for GI tract. Overall, in this cross-sectional analysis, GLDH, K18, and miR-122 were more sensitive and specific compared to other biomarkers in a liver damage patient cohort.

The cross-sectional cohort of patients with liver damage, other organ damage and healthy volunteers was used as an independent validation data set for the multivariate models. The models were constructed to predict ALT with the APAP overdose cohort and therefore, we used the same composite score thresholds defined in the APAP cohort for validation in the cross-sectional cohort. The three-biomarker model was able to achieve near perfect separation between patients with liver injury and healthy volunteers (Table 1) and composite scores were highly correlated with the

measured log ALT (Figure 3B). The model exhibited strong predictability as reflected by the ROC AUC (Figure 3C, Table 1) when comparing liver damage to healthy or other organ damage cohorts. When setting the sensitivity ≥ 0.95 and comparing the individual biomarkers to the models, K18 had a similar specificity to the three-biomarker model (Table 2). The seven-biomarker model had a higher specificity for identifying patients with liver damage than the three-biomarker model or any individual biomarker alone (Supplemental figure 6). Of ALT, K18, GLDH and miR-122, K18 has the highest specificity in the cross-sectional data, consistent with findings in Figure 3A. In the case of setting the specificity to 0.95, in this cohort of 104 patients with liver damage, ALT, GLDH, K18 and miR-122 would correctly identify 83, 82, 98 and 80 patients, respectively. CcK18, MCSF, MCSFR and OPN would correctly identify 73, 94, 75 and 89 patients respectively. The three biomarker and seven biomarker panel would correctly identify 101 and 103 patients, respectively. If the three-biomarker model composite score threshold was lowered and set based on an ALT of $> 50\text{U/L}$ as defined in supplemental figure 3, the number of false negatives decreased (Supplemental Table 4). Patients with liver damage in the cross-sectional cohort contained multiple different types of liver disease and some had low ALT measurements (diagnosed using $>2\text{x ULN}$ ALP), which may be why the lower threshold performed better. The predictability of the model was also enhanced when all 7 biomarkers were included (Supplemental Table 4; Table 1). The linear regression three biomarker model as defined in the APAP cohort performed slightly better (Supplemental Figure 7) than the random forest (Figure 3) when independently validated in the cross-sectional liver damage cohort.

In summary, the three-biomarker model with GLDH, K18 and miR-122 was trained, tested and validated in the acetaminophen overdose cohort, demonstrated high predictability of ALT and accurately identified liver damage subjects in an independent validation cohort.

Expression patterns of K18, GLDH and miR-122 in humans and rat

K18 protein and gene expression was evaluated in healthy and injured human livers. Using immunohistochemistry and in-situ hybridization we found that in both normal (n = 5) and diseased livers (n=5), K18 protein and mRNA were consistently and highly expressed in bile duct epithelium and in peri-portal hepatocytes (Supplemental figure 8A). Expression in midzonal and centrilobular hepatocytes was also observed, however this was more variable both within and across samples.

To evaluate the physiological gene expression profiles of *KRT18* (gene for K18) and *GLUD1* (gene for GLDH) across different tissues in human and rat, we queried (1) GTEX (Genotype-Tissue Expression) and HPA (Human Protein Atlas), public human gene/protein expression databases and (2) Pfizer Zoomap, an internal tissue atlas for preclinical species. Rat gene expression data in each tissue can be found in supplemental table 6. In the human, *KRT18* expression is predominantly expressed in the liver compared to other tissues, whereas *GLUD1* is widely expressed, suggesting that *GLUD1* may be less specific for liver than *KRT18* (Fig 4A). In the rat, *Krt18* expression is the highest in bladder, ileum, colon, stomach and liver (Supplemental figure 8B). Rat *Krt18* and *Glud1* expression levels in some tissues, including the liver, kidney and heart correlated with human expression (Supplemental figure 8C, 8D).

To further assess the utility of expression profiles of *Krt18* and *Glud1* in rat hepatotoxicity, we queried DrugMatrix, a public rat toxicogenomics database that includes tissue gene expression and pathological evaluations. *Krt18* expression had minimal variability in control samples with an associated pathology score of 0 (Figure 4B). In samples treated with compound, *Krt18* expression increased with more severe pathology scores, suggesting that *Krt18* gene expression is actively regulated during liver injury and that upregulation of *Krt18* may start to occur prior to any overt pathology or occur as a secondary effect of hepatocyte regeneration in the context of injury. Additionally, ALT activity correlated with *Krt18* expression (Figure 4C), which is consistent with the patient data (Figure 1B). We also filtered for *Glud1* and found 153 samples with a reported pathology term including liver necrosis and/or apoptosis. While treated samples had higher expression of *Glud1* compared to controls, there was no correlation of pathology with *Glud1* expression levels (Figure 4D). *Glud1* gene expression was not correlated with measured ALT activity (Figure 4E).

Tissue Atlas, a human miRNA tissue expression database, was interrogated for miR-122 expression. MiR-122-3p and miR-122-5p were highly expressed in the liver (Figure 4F) with a tissue specificity index > 0.91 and tissue expression correlated with each other ($r^2 = 0.91$). Rat miR-122 expression was evaluated in the RATEmiR database. Rat miR-122-3p and miR-122-5p were liver tissue specific (Supplemental figure 8G) with a tissue specificity index = 1, and tissue expression was highly correlated with each other ($r^2 = 1$). Furthermore, rat miR-122-3p expression was highly correlated with human miR-122-3p ($r^2 = 0.96$) and miR-122-5p is correlated with human miR-122-5p ($r^2 = 0.67$) (Supplemental figure 8E, 8F). These expression data suggest

468 that K18 and miR-122 maybe be specific biomarkers of liver injury in rats and humans.
469 Although Glud1 mRNA expression doesn't seem to be tissue specific in rats and
470 humans, protein expression and enzyme activity of GLDH across all tissues has not
471 been evaluated and may provide additional information on it's utility and specificity.

Discussion

The present study evaluated the diagnostic performance of seven promising biomarkers of liver injury in humans. We provide evidence to suggest that (i) K18 was superior in terms of sensitivity and specificity over ALT and GLDH in diagnosing liver damage compared to healthy volunteers, GI tract and muscle damage patients; and (ii) a three biomarker model with K18, GLDH and miR-122 that was trained, tested, and validated using an acetaminophen overdose cohort, was independently validated in a cross-sectional cohort and able to achieve separation between patients with liver damage and healthy volunteers. The three-biomarker model also demonstrated strong diagnostic potential when comparing liver damage patients and patients with damage to the muscle, pancreas, GI tract and kidney. Early detection, accurate diagnosis and determining outcomes of DILI continue to be major hurdles during drug development and post marketing. Significant biomarker gaps exist in the current methods to diagnose, provide mechanistic information and determine prognosis of DILI in clinical trials. These results not only provide a comprehensive assessment of individual biomarker performance in acetaminophen and liver damage cohorts due to different etiologies, but also highlight the utility of K18, GLDH and miR-122 in a multivariate model to provide greater sensitivity and specificity than each biomarker alone in detecting liver injury.

Elevations in ALT activity can occur in other settings such as muscle movement(Fu et al. 2019) and myocardial(Giesen et al. 1989) and skeletal muscle injury(Nathwani et al. 2005). Data from this study demonstrate that the three-biomarker model (GLDH, K18 and miR-122) clearly separated patients with muscle injury from

patients with liver damage thereby offering significant advantages over measuring ALT. This finding suggests that the three-biomarker model could be deployed as monitoring biomarker panel for liver injury in clinical trials involving patients with muscular dystrophies(Zhu et al. 2015) where ALT is non-specifically elevated due to muscle damage and a specific biomarker to monitor liver health is desired.

In this study, all candidate biomarkers were elevated in liver damage patients relative to healthy volunteers, muscle, pancreas, GI tract and kidney patients. Furthermore, candidate biomarkers were elevated in each type of liver damage, including DILI. K18 had superior sensitivity and specificity over ALT, GLDH and miR-122 in liver compared to healthy, muscle and GI tract damage patients. K18 has been proposed as a biomarker for a range of liver conditions including acute liver failure and chronic liver diseases such as viral hepatitis, non-alcoholic fatty liver disease and liver cancer(Ku et al. 2016). While an advantage of K18 as a biomarker is that it is an early marker of apoptosis/necrosis; a disadvantage is that it is also a biomarker for dysfunction in tissues other than the liver including the lung (Fu et al. 2019; Levy et al. 2019; Molnar et al. 2019; Tajima et al. 2019) (Yang et al. 2019). Thus, a panel of 3 or 7 biomarkers may be advantageous over a single biomarker. An advantage of GLDH as a biomarker is that it is an early marker of liver-specific mitochondrial damage and has low inter- and intra-individual variability compared to other liver injury biomarkers(Tajima et al. 2019). GLDH has also been shown to be more readily detectable than ALT in a rat model of APAP-DILI (Thulin et al. 2017). However, GLDH has been shown to have a shorter half-life than ALT (Tajima et al. 2019) and by itself did not offer any advantage over ALT in detecting liver damage compared to healthy controls. miR-122 is

advantageous as an early marker of liver-specific damage but its use has been limited due to the higher inter-and intra-individual variability(Levy et al. 2019) and a potentially short half-life(Thulin et al. 2017).

Traditional liver injury biomarkers are passively released from necrotic hepatocytes and lack mechanistic understanding of underlying liver injury. Our data and others(Ku et al. 2016) demonstrate that hepatocytes and cholangiocytes specifically express K18. With a direct hepatotoxic insult, in early apoptosis K18 is cleaved and released into circulation as ccK18; while full-length K18 is released with necrosis. Therefore, levels may reflect different cell death processes in the liver (Fu et al. 2019). Although ccK18 performed well with an AUC of 0.873, the relatively reduced sensitivity and specificity can be investigated in subsequent studies to understand if this is associated with kinetics of ccK18 release, severity of injury and/or underlying pathologic mechanism of liver injury using longitudinal cohort of patients with DILI. Gene expression was increased with the degree of liver apoptosis and necrosis, suggesting K18 has an active role in liver damage. We and others show that miR-122 is specifically expressed in the livers of humans(Landgraf et al. 2007) and rats(Smith et al. 2016). MiR-122 accounts for 70% of hepatic miRNAs(Lagos-Quintana et al. 2002) and is superior to ALT in detecting liver injury in muscle injury patients(Zhang et al. 2010). MiR-122 is an early marker of liver injury(Wang et al. 2009) and found in protein rich fraction of plasma and specifically packaged into exosomes(Bala et al. 2012) during liver injury. GLDH, MCSF, MCSFR and OPN may also be used as mechanistic biomarkers. GLDH, a mitochondrial protein, reflects loss of mitochondrial integrity. MCSFR, a receptor for MCSF, is shed from activated macrophages during DILI(Church

et al. 2019) and is a biomarker of inflammation. Notably, we observe high levels of MCSFR in patients with cirrhosis relative to the other candidate biomarkers. OPN may also be a marker of liver inflammation and necrosis(Roth et al. 2020).

In summary, our results identify a three-biomarker model with K18, GLDH and miR-122 for sensitive and specific detection of APAP DILI and liver damage due to other causes. Whether these biomarkers either alone or in combination outperform traditional markers such as ALT as safety biomarkers for diagnosis and prediction of DILI remains to be tested in larger multicentered longitudinal cohort.

550 **Acknowledgements**

551 We would like to thank Scott Auerbach, NIEHS, Toxicoinformatics Group Leader and
552 Dan Svoboda, Manager for DrugMatrix for providing access and supporting the
553 DrugMatrix data query and to Lila Ramaiah for providing critical input on the manuscript.
554 We acknowledge the leadership and engagement of external stake holders and
555 consortia such as IMI supported SAFE-T and TransBioLine as well as C-Path supported
556 PSTC. They are critical drivers for biomarker science innovations and regulatory
557 qualifications impacting drug development and patient care.

558

559 **Funding Statement**

560 Funding was made available from the Drug Safety Research and Development
561 department within Pfizer's Worldwide Research Development and Medical.

562

References

- Aithal GP, Watkins PB, Andrade RJ, Larrey D, Molokhia M, Takikawa H, Hunt CM, Wilke RA, Avigan M, Kaplowitz N et al. 2011. Case definition and phenotype standardization in drug-induced liver injury. *Clin Pharmacol Ther.* 89(6):806-815.
- Bala S, Petrasek J, Mundkur S, Catalano D, Levin I, Ward J, Alao H, Kodys K, Szabo G. 2012. Circulating micrnas in exosomes indicate hepatocyte injury and inflammation in alcoholic, drug-induced, and inflammatory liver diseases. *Hepatology.* 56(5):1946-1957.
- Bernal W, Wendon J. 2014. Acute liver failure. *N Engl J Med.* 370(12):1170-1171.
- Bjornsson ES, Bergmann OM, Bjornsson HK, Kvaran RB, Olafsson S. 2013. Incidence, presentation, and outcomes in patients with drug-induced liver injury in the general population of iceland. *Gastroenterology.* 144(7):1419-1425, 1425 e1411-1413; quiz e1419-1420.
- Church RJ, Kullak-Ublick GA, Aubrecht J, Bonkovsky HL, Chalasani N, Fontana RJ, Goepfert JC, Hackman F, King NMP, Kirby S et al. 2019. Candidate biomarkers for the diagnosis and prognosis of drug-induced liver injury: An international collaborative effort. *Hepatology.* 69(2):760-773.
- Davis J, Goadrich M. 2006. The relationship between precision-recall and roc curves. Paper presented at: Proceedings of the 23rd international conference on Machine learning. Association for Computing Machinery; Pittsburgh, Pennsylvania, USA.
- DeLong ER, DeLong DM, Clarke-Pearson DL. 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics.* 44(3):837-845.
- Fu S, Wu D, Jiang W, Li J, Long J, Jia C, Zhou T. 2019. Molecular biomarkers in drug-induced liver injury: Challenges and future perspectives. *Front Pharmacol.* 10:1667.
- Giesen PL, Peltenburg HG, de Zwaan C, Janson PC, Flendrig JG, Hermens WT. 1989. Greater than expected alanine aminotransferase activities in plasma and in hearts of patients with acute myocardial infarction. *Clinical Chemistry.* 35(2):279-283.
- Ku NO, Strnad P, Bantel H, Omary MB. 2016. Keratins: Biomarkers and modulators of apoptotic and necrotic cell death in the liver. *Hepatology.* 64(3):966-976.
- Kullak-Ublick GA, Andrade RJ, Merz M, End P, Benesic A, Gerbes AL, Aithal GP. 2017. Drug-induced liver injury: Recent advances in diagnosis and risk assessment. *Gut.* 66(6):1154-1164.
- Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T. 2002. Identification of tissue-specific micrnas from mouse. *Curr Biol.* 12(9):735-739.
- Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M et al. 2007. A mammalian micrna expression atlas based on small rna library sequencing. *Cell.* 129(7):1401-1414.
- Levy L, Tigert A, Huszti E, Saito T, Mitsakakis N, Moshkelgosha S, Joe B, Boonstra KM, Tikkanen JM, Keshavjee S et al. 2019. Epithelial cell death markers in bronchoalveolar lavage correlate with chronic lung allograft dysfunction subtypes and survival in lung transplant recipients-a single-center retrospective cohort study. *Transpl Int.* 32(9):965-973.
- Molnar T, Borocz K, Berki T, Szapary L, Szolics A, Janszky J, Illes Z, Csecsei P. 2019. Subacute elevation of plasma level of caspase-cleaved cytokeratin-18 is associated with hemorrhagic transformation and functional outcome in ischemic stroke. *J Stroke Cerebrovasc Dis.* 28(3):719-727.
- Nathwani RA, Pais S, Reynolds TB, Kaplowitz N. 2005. Serum alanine aminotransferase in skeletal muscle diseases. *Hepatology.* 41(2):380-382.
- Onakpoya IJ, Heneghan CJ, Aronson JK. 2016. Post-marketing withdrawal of 462 medicinal products because of adverse drug reactions: A systematic review of the world literature. *BMC Med.* 14:10.

- Przybylak KR, Cronin MT. 2012. In silico models for drug-induced liver injury--current status. *Expert Opin Drug Metab Toxicol.* 8(2):201-217.
- R Development Core Team. 2019. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- Roth SE, Avigan MI, Bourdet D, Brott D, Church R, Dash A, Keller D, Sherratt P, Watkins PB, Westcott-Baker L et al. 2020. Next-generation dili biomarkers: Prioritization of biomarkers for qualification and best practices for biospecimen collection in drug development. *Clin Pharmacol Ther.* 107(2):333-346.
- Senior JR. 2014. Evolution of the food and drug administration approach to liver safety assessment for new drugs: Current status and challenges. *Drug Saf.* 37 Suppl 1:S9-17.
- Sgro C, Clinard F, Ouazir K, Chanay H, Allard C, Guilleminet C, Lenoir C, Lemoine A, Hillon P. 2002. Incidence of drug-induced hepatic injuries: A french population-based study. *Hepatology.* 36(2):451-455.
- Shen T, Liu Y, Shang J, Xie Q, Li J, Yan M, Xu J, Niu J, Liu J, Watkins PB et al. 2019. Incidence and etiology of drug-induced liver injury in mainland china. *Gastroenterology.* 156(8):2230-2241 e2211.
- Shi Q, Hong H, Senior J, Tong W. 2010. Biomarkers for drug-induced liver injury. *Expert Rev Gastroenterol Hepatol.* 4(2):225-234.
- Smith A, Calley J, Mathur S, Qian HR, Wu H, Farmen M, Caiment F, Bushel PR, Li J, Fisher C et al. 2016. The rat microrna body atlas; evaluation of the microrna content of rat organs through deep sequencing and characterization of pancreas enriched mirnas as biomarkers of pancreatic toxicity in the rat and dog. *BMC Genomics.* 17:694.
- Starkey Lewis PJ, Dear J, Platt V, Simpson KJ, Craig DGN, Antoine DJ, French NS, Dhaun N, Webb DJ, Costello EM et al. 2011. Circulating micrnas as potential markers of human drug-induced liver injury. *Hepatology.* 54(5):1767-1776.
- Stravitz RT, Lee WM. 2019. Acute liver failure. *Lancet.* 394(10201):869-881.
- Tajima S, Yamamoto N, Masuda S. 2019. Clinical prospects of biomarkers for the early detection and/or prediction of organ injury associated with pharmacotherapy. *Biochem Pharmacol.* 170:113664.
- Thulin P, Hornby RJ, Auli M, Nordahl G, Antoine DJ, Starkey Lewis P, Goldring CE, Park BK, Prats N, Glinghammar B et al. 2017. A longitudinal assessment of mir-122 and gldh as biomarkers of drug-induced liver injury in the rat. *Biomarkers.* 22(5):461-469.
- Wang K, Zhang S, Marzolf B, Troisch P, Brightman A, Hu Z, Hood LE, Galas DJ. 2009. Circulating micrnas, potential biomarkers for drug-induced liver injury. *Proc Natl Acad Sci U S A.* 106(11):4402-4407.
- Yang MC, Liu HK, Su YT, Tsai CC, Wu JR. 2019. Serum apoptotic marker m30 is positively correlated with early diastolic dysfunction in adolescent obesity. *PLoS One.* 14(5):e0217429.
- Zhang Y, Jia Y, Zheng R, Guo Y, Wang Y, Guo H, Fei M, Sun S. 2010. Plasma microrna-122 as a biomarker for viral-, alcohol-, and chemical-related hepatic diseases. *Clin Chem.* 56(12):1830-1838.
- Zhu Y, Zhang H, Sun Y, Li Y, Deng L, Wen X, Wang H, Zhang C. 2015. Serum enzyme profiles differentiate five types of muscular dystrophy. *Dis Markers.* 2015:543282.

Tables

Table 1. Assessment of the random forest biomarker models in the acetaminophen overdose and cross-sectional cohorts

Data set	Panel -3 ¹ R	Panel -3 AUC (CI)	Panel -7 ² R	Panel -7 AUC (CI)
Acetaminophen overdose cohort				
ALT ≥ 150 U/L cutoff				
Training (T1)	0.921	0.99 (0.98, 1)	0.964	0.99 (0.99, 1)
Testing (T2)	0.905	0.99 (0.97, 1)	0.912	0.99 (0.98, 1)
Validation (T3)	0.922	1 (1, 1)	0.922	1 (0.99, 1)
ALT > 50 U/L cutoff				
Training (T1)	0.921	0.98 (0.96, 1)	0.964	0.99 (0.97, 1)
Testing (T2)	0.905	0.98 (0.96, 1)	0.912	0.97 (0.93, 1)
Validation (T3)	0.922	1 (1, 1)	0.922	0.99 (0.98, 1)
Cross-sectional cohort: comparator group vs liver				
Healthy	0.856	0.99 (0.98, 1)	0.815	1 (1, 1)
Muscle	0.76	0.92 (0.88, 0.97)	0.688	0.96 (0.93, 0.99)
Pancreas	0.799	0.97 (0.95, 1)	0.731	0.99 (0.98, 1)
GI tract	0.811	0.98 (0.96, 1)	0.748	0.99 (0.98, 1)
Kidney	0.814	0.93 (0.90, 0.97)	0.751	0.97 (0.94, 0.99)

¹Panel-3: GLDH, K18, miR-122; ²Panel-7: GLDH, K18, miR-122, ccK18, MCSF,

MCSFR, OPN; R: Pearson's correlation coefficient to measured ALT activity; AUC: area under the curve; T1: Timepoint 1 (collected at hospital admission, median: 4.6 hours, IQR: 4.1, 10.7 after acetaminophen ingestion), T2: Timepoint 2 (11.4 hours after T1), T3: Timepoint 3 (21.8 hours after T1); GI: gastrointestinal; CI: 95% confidence interval.

Metric	Biomarker Threshold	Sensitivity	Specificity	ROC AUC	PPV	NPV
Cross-sectional cohort: healthy vs liver						
GLDH	2.3	0.88	0.58	0.90	0.62	0.87
K18	137.0	0.95	0.93	0.98	0.92	0.96
ccK18	70.5	0.94	0.36	0.87	0.53	0.89
MCSF	17.1	0.95	0.72	0.97	0.72	0.95
MCSFR	828.5	0.95	0.64	0.92	0.67	0.95
OPN	11.8	0.95	0.59	0.95	0.64	0.94
miR-122	614.0	0.95	0.71	0.94	0.71	0.95
Panel-3	3.1	0.95	0.96	0.99	0.94	0.96
Panel-7	3.6	0.95	1.00	1.00	1.00	0.96
ALT	15.5	0.95	0.46	0.93	0.58	0.93

¹Sensitivity was fixed at ≥ 0.95 where possible and thresholds for the APAP

cohorts were determined with ALT ≥ 150 U/L within each timepoint. Panel-3:

GLDH, K18, miR-122; Panel-7: GLDH, K18, miR-122, ccK18, MCSF, MCSFR, OPN; R:

Pearson's correlation coefficient to measured ALT activity; ROC AUC: receiver operator

curve area under the curve; T1: Timepoint 1 (collected at hospital admission, median:

4.6 hours, IQT: 4.1, 10.7 after acetaminophen ingestion), T2: Timepoint 2 (12.7 hours,

IQT: 9.2, 14.1 after T1), T3: Timepoint 3 (22.9 hours, IQT: 19.8, 24.2 after T1); PPV:

positive predictive value, NPV: negative predictive value.

Figure legends

Figure 1. Analysis of a three biomarker-based multivariate model for detection of liver injury using longitudinal cohort of patients with acetaminophen overdose.

Three liver injury biomarkers (GLDH, K18 and miR-122) were used to build a predictive model for log ALT. **(A)** ALT, GLDH, K18 and miR-122 levels at each timepoint.

Correlation between composite score and measured log ALT activity at **(B)** baseline (time 1, training), **(C)** time 2 (testing) and **(D)** time 3 (validation). **(E)** Score of each patient overtime. **(F)** Summary results from setting the threshold at an ALT cutoff of ≥ 150 U/L. Values are shown as raw and natural logarithm ALT. Pearson's R coefficient is shown based on the measured log ALT and score.

Figure 2. Evaluation of seven candidate biomarkers for liver injury in comparison to ALT in the cross-sectional cohort. (A)

Glutamate dehydrogenase (GLDH), cytokeratin 18 (K18), caspase-cleaved K18 (ccK18), macrophage colony-stimulating factor (MCSF), MCSFR (MCSFR), microRNA-122 (miR-122) and osteopontin (OPN) in comparison with the traditional biomarker, alanine aminotransferase (ALT) were measured in healthy volunteers (n=135) and patients with damage to liver (n=104), muscle (n=74), pancreas (n=34), GI (n=37) or kidney (n=40). **(B)** Candidate biomarkers in the liver damage cohort in subjects with acute liver failure (n=4), drug induced (n=24), transaminitis (n=54), alcoholic (n=15), hepatitis (n=12), liver metastatic (n=27), cirrhosis (n=20) and transplant (n=9). Healthy data is repeated from Fig 2A for reference. Values are log₁₀ normalized. Some subjects exhibited multiple types of liver damage. **(C)** Spearman's r correlation coefficient between ALT activity and candidate biomarkers

GLDH, K18, ccK18 and miR-122 in patients with liver damage. Values are log10 normalized.

Figure 3. Independent validation of the three biomarker-based multivariate model

for detection of liver damage in the cross-sectional cohort. (A) Area under the receiver operator characteristic curve (AUC-ROC) for candidate biomarkers in patients with liver damage vs healthy volunteers. The three-liver injury biomarker (GLDH, K18 and miR-122) model was evaluated in the cross-sectional cohort. Correlation between composite score and measured log ALT activity in **(B)** liver damage patients and healthy volunteers and **(C)** all organ damages. **(D)** Summary results from setting the threshold at 4 (as defined in Figure 1C). Values are shown as raw and natural logarithm ALT. *P < 0.05, **P < 0.01, ***P < 0.001, **** P< 0.0001.

Figure 4. Expression patterns of *KRT18*, *GLUD1* and miR-122 in human and rat. (A)

A RNAseq-based database was queried for KRT18 and GLUD1 gene expression in human tissues (see methods). DrugMatrix, an Affymatrix-based rat toxicogenomics database, was queried for *Krt18* (B, C) and *Glud1* (D, E) expression across several tissues. **(B)** *Krt18* in control samples (n = 77) and samples treated with a compound (n = 115), classified by histopathology score for liver necrosis and or apoptosis of none (n = 96), minimal (n = 14) and mild (n = 5). The Kruskal-Wallis p-value is 5.76e-5. **(C)** Correlation between ALT activity and rat *Krt18* gene expression. **(D)** *Glud1* expression in control samples (n = 19) and samples treated with a compound (n = 134), classified by histopathology score for liver necrosis and or apoptosis of none (n = 113), minimal (n =

18) and mild ($n = 3$). The Kruskal-Wallis p-value is 0.119. **(E)** Correlation between ALT activity and rat *Glud1* gene expression. **(F)** Tissue Atlas was queried for 3p and 5p miR-122 in human tissues. The linear regression line (blue) and confidence interval (grey shading) are shown. Pairwise comparisons were calculated using the Dunn test if the Kruskal–Wallis test was significant. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. TPM: transcript per million; QNE: quantile normalized expression.

Author Contributions

Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; HPL, VSV, ZW, QP, CH, DP, JW, QZ, SA, MM, KMH, RW, KJ, GAK, GPA, JWD, SKR

Drafting the work or revising it critically for important intellectual content; HPL, VSV, ZW, QP, CH, DP, JW, QZ, SA, MM, KMH, RW, KJ, GAK, GPA, JWD, SKR

Final approval of the version to be published; HPL, VSV, ZW, QP, CH, DP, JW, QZ, SA, MM, KMH, RW, KJ, GAK, GPA, JWD, SKR

Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. HPL, VSV, ZW, QP, CH, DP, JW, QZ, SA, MM, KMH, RW, KJ, GAK, GPA, JWD, SKR

Code/Data availability

Code available upon request

F

DataSet	TimePoint	Threshold	Sensitivity	Specificity	PPV	NPV	TP	FN	TN	FP
Testing	Time 2	4	1	0.940	0.611	1	11	0	109	7
Validation	Time 3	4	1	0.943	0.733	1	11	0	66	4

A

B

C

D

Comparison	Threshold	Sensitivity	Specificity	PPV	NPV	TP	FN	TN	FP
Liver vs Healthy	4	0.723	1.000	1.000	0.827	73	28	134	0
Liver vs Muscle	4	0.723	0.957	0.973	0.616	73	28	45	2
Liver vs Pancreas	4	0.723	1.000	1.000	0.451	73	28	23	0
Liver vs GI	4	0.723	1.000	1.000	0.569	73	28	37	0
Liver vs Kidney	4	0.723	1.000	1.000	0.582	73	28	39	0

Evaluating the sensitivity and specificity of promising circulating biomarkers to diagnose liver injury in humans

Heather P. Llewellyn, Vishal S. Vaidya, et.al.

SUPPLEMENTAL MATERIALS

1. Supplemental Methods

2. Supplementary Figures

- **Supplemental Figure 1.** Standard curves for each biomarker measurements on respective assay platforms using commercially available kits.
- **Supplemental Figure 2.** Contribution of each of the 3 biomarkers to the random forest model.
- **Supplemental Figure 3.** Analysis of a multivariate model using seven biomarkers for patients with acetaminophen overdose and a cutoff of ALT ≥ 150 U/L.
- **Supplemental Figure 4.** Analysis of a linear regression multivariate model using three biomarkers for patients with acetaminophen overdose evaluated at a cutoff of ALT ≥ 150 U/L.
- **Supplemental Figure 5.** Conventional biomarkers in the cross-sectional cohort.
- **Supplemental Figure 6.** Independent validation of the seven biomarker model in the cross-sectional cohort of patients with liver damage.
- **Supplemental Figure 7.** Independent validation of the linear regression model in the cross-sectional cohort with three biomarkers.
- **Supplemental Figure 8.** Expression patterns of K18, Glud1 and miR-122.

3. Supplementary Tables

- **Supplemental Table 1.** Bioanalytical characteristics of biomarker assays.
- **Supplemental Table 2.** Basic and clinical demographics of longitudinal cohort with acetaminophen overdose.
- **Supplemental Table 3.** Basic and clinical demographics of cross-sectional cohort of healthy volunteers and patients with injury to liver, muscle, pancreas, gastrointestinal tract or kidney.
- **Supplemental Table 4.** Assessment of the multivariate biomarker models in the acetaminophen and cross-sectional cohorts with a cutoff of ALT > 50 U/L
- **Supplemental Table 5.** Performance characteristics of candidate biomarkers in the cross-sectional cohort.
- **Supplemental Table 6.** Krt18 and Glud1 gene expression in rat tissues

Supplementary methods

Study Population

Longitudinal cohort of acetaminophen overdose study participants: The study inclusion criteria were i) age 16 years and over, ii) hospital attendance with acetaminophen overdose alone or as part of a mixed overdose and iii) patient able to give informed consent. The exclusion criteria were i) patient detained under the Mental Health Act (Scotland) or ii) inability to provide informed consent or iii) unreliable history of overdose or iv) prisoners. Patients who presented to RIE with a subsequent overdose were eligible to be enrolled into the study more than once. Data collected included: age, gender, past medical history, date and time of overdose, amount of acetaminophen taken, other drugs taken, date and times of all treatment including blood tests, results of all blood tests. The patients' blood samples were collected, processed to isolate serum, and aliquoted before being frozen in a -20°C freezer. The frozen serum samples were then transferred into a -80°C freezer at the co-located Queen's Medical Research Institute (University of Edinburgh) prior to being sent to Pfizer, Inc.

Biomarker Measurements

The performance characteristics for the biomarker assays including CK_M30 (ccK18), CK_M65 (K18), MCSF, MCSFR and Osteopontin were evaluated by measuring sensitivity, assay range, reproducibility (intra- and inter- assay imprecision, dilutional linearity, recovery and sample freeze-thaw stability (Supplementary Table 1). The sensitivity or the lowest limit of detection (LLOD) was determined by analyzing 10

replications of the respective standard sets; the lowest standard concentration for which the coefficient of variation (CV) was less than a target of 20% was determined to be the LLOD. The analytical recovery in control was determined by low, medium and high concentration of the validation control pools, which were prepared in human serum samples acquired from a patient demonstrating liver injury based on high serum liver enzyme values and a normal healthy donor without liver injury. Intra-assay imprecision was established by analyzing 10 replicates of low, medium and high concentration of validation aliquots in a single assay run. Inter-assay imprecision was established by analyzing aliquots of low, medium, and high concentration validation aliquots over a minimum of 3 separate assay runs. Dilutional linearity was evaluated to demonstrate the reliability of sample dilution. Sample dilution was required for MCSFR and Osteopontin to fit the concentrations of respective antigens in the linear range of the standard curve. The low and high concentration validation controls, which were obtained from healthy and DILI patient serum, were diluted 1:2, 1:4, 1:8 using sample diluent to be measured by the respective biomarker assays. Freeze-thaw stability was evaluated as recovered analyte concentration relative to the sample undergoing initial freeze/thaw cycle. Three freeze-thaw cycles for low, medium and high concentration of the validation control aliquots were analyzed by the respective biomarker assays.

In-Situ Hybridization and Immunohistochemistry

Formalin-fixed, paraffin embedded human liver samples acquired by autopsy from healthy donors or patients diagnosed with liver injury secondary to alcohol toxicity (purchased from ProteoGenex, Inglewood CA) were assessed for K18 expression by

immunohistochemistry (IHC) and in situ hybridization (ISH). For ISH, serial sections of each sample were stained on a Leica Bond-Rx (Leica Biosystems, Buffalo Grove, IL) automated platform using a positive control probe (human PPIB), a negative control probe (dapB), or the *Krt18* target probe purchased from ACD Bio (Newark, CA). For IHC, liver samples were stained for K18 using a monoclonal (DC-10) mouse primary antibody (Novocastra, Leica Biosystems, Buffalo Grove, IL) on a Leica Bond-Rx automated platform at a dilution of 1:750 (0.104ug/mL) with a heat induced antigen retrieval with ER1 for 20 minutes and no serum blocking.

Evaluation of candidate biomarker expression using an *in vivo* toxicogenomic database

To determine the gene expression profile of *Krt18* and *Glut1* following liver toxicity *in vivo*, a public rat toxicogenomics data set, DrugMatrix (of the National Toxicology Program; <https://ntp.niehs.nih.gov/drugmatrix/index.html>), which tracks tissue gene expression using Affymetrix 230 2.0 microarrays in combination with pathological evaluations, was mined. Investigators contributing to this database performed *in vivo* studies following guidelines from the ToxFX Analysis Suite User Manual. Briefly, a minimum of 3 animals per group were required and compounds were tested at a minimum the fully effective/therapeutic dose and maximum tolerated/toxic dose. Compound exposure time was chosen to be sufficient to elicit informative gene expression changes, 1 to 14 days and at least 2 timepoints were included. Prior to the analysis described below, microarray data sets were normalized using the RMA algorithm 42, 43 implemented in GeneSpring GX 12.6 (Agilent Technologies, Palo Alto,

CA, USA). The DrugMatrix gene expression data are available through the GEO (Series ID: GSE57811).

For the gene expression profile of rat *Krt18* and *Glud1* in DrugMatrix, we focused on a subset of control and treatment groups with liver histopathology results (i.e., Tissue = “Liver” and histopathology results containing “Necrosis” and “Apoptotic” terms) as well as ALT measurements. Within DrugMatrix, pathological severity scores range from 0 to 4 based on the percentage of the tissue damage, and as described previously^{15,16}. If a compound had multiple treatment groups, the maximum pathology score was used for analysis. Additionally, control groups were broken out from compound treated groups for global and pairwise comparisons. The Kruskal-Wallis test was employed as a global test to assess if there were significant differences among different groups¹⁷. If the Kruskal-Wallis p-value was significant (< 0.05), the Dunn multiple comparison test was performed with the Benjamini-Hochberg correction method for pairwise comparisons¹⁸.

Entry/value	Classification	% of Tissue Damage
0	None	0
1	Minimal	0 - 30
2	Mild	30 - 60
3	Moderate	60 - 80
4	Severe	> 80

Evaluation of candidate gene biomarker expression using GTEX, HPA and Pfizer ZooMap databases

To determine the gene expression profile of *KRT18* and *GLUD1* in human and *Krt18* and *Glud1* in rat, we queried Zoomap, Pfizer’s internal human and preclinical transcriptomic profile database. Human data was pulled from GTEX (Genotype-Tissue

Expression, www.gtexportal.org) as read count and transcript per million (TPM) values. Additional human gene expression data was pulled from HPA (Human Protein Atlas, www.proteinatlas.org) as raw fastq files and processed in Salmon-based RNA-Seq pipeline at Pfizer. The combined human dataset consisted of 63 tissues. Preclinical species data were generated internally from 54 tissues and only rat data was used for this study. Gene expression data is provided in supplemental table 6. The tissue samples were sequenced using TruSeq and aligned using the same Salmon-based RNA-Seq pipeline in Pfizer. Human *KRT18* and rat *Krt18* gene expression correlation analyses were based on the 38 common tissues between human and rat. The human *KRT18* and rat *Krt18* gene expression data are log2-transformed TPM values.

Evaluation of candidate miRNA biomarker expression using RATEmiRs database

To determine the microRNA expression profile of miR-122 in rat, we queried RATEmiRs database, which consists of the microRNA-Seq data from 23 rat tissues. The tissue expression data was processed by 3 different institutions. According to our correlation analysis, all 3 datasets for miR-122-3p and miR-122-5p were highly correlated ($R^2 > 0.99$) and we further analyzed the dataset processed by Maastricht University. The rat miRNA data is represented in Log2-transformed TMM (Trimmed mean of M-values).

Evaluation of candidate miRNA biomarker expression using Human Tissue Atlas database

To determine the microRNA expression profile of miR-122 in human, we queried Tissue Atlas database, which consists of the microRNA-Seq data from 31 human tissues. The human miRNA data is represented in 3 ways: raw expression, quantile normalized expression (QNE) and variance stabilizing normalized expression. Log2-transformed QNE data was used for visualization.

There are 8 common tissues between human and rat miRNA expression data, which was used for correlation analyses. There are 4 different brain sections (brainstem, cerebellum, cerebrum, hippocampus) in RATEmiRs and we selected the maximum expression value of miR-122-3p and miR-122-5p among all the brain sections to have one human brain expression data, which then could be comparable to rat brain expression data.

Supplementary Figures

Supplemental Figure 1. Standard curves for each biomarker measurement. Serum biomarkers CK_M65 (K18), CK_M30 (ccK18), osteopontin, MCSF and MCSFR, values were calculated on their respective platform using a 6 to 9 point five-parametric logarithmic standard curve.

Supplemental Figure 2. Contribution of each biomarker in the three-biomarker multivariate model. In this heat map, warmer the color is, higher the score a patient gets based on the individual's three biomarker values. miR-122 values are summarized by quartiles – lower left plot (first quartile), lower right plot (second quartile), upper left plot (third quartile) and upper right plot (fourth quartile).

F	DataSet	TimePoint	Threshold	Sensitivity	Specificity	PPV	NPV	TP	FN	TN	FP
	Testing	Time 2	4.2	1	0.974	0.786	1	11	0	113	3
	Validation	Time 3	4.2	1	0.943	0.733	1	11	0	66	4

Supplemental Figure 3. Analysis of a multivariate model using seven biomarkers for patients with acetaminophen overdose and a cutoff of ALT $\geq$ 150 U/L. Seven liver injury biomarkers (GLDH, K18, ccK18, MCSF, MCSFR, OPN and miR-122) were used to build a predictive model for log ALT. (A) ccK18, MCSF, MCSFR, OPN levels at each timepoint. (B) Correlation between predicted log ALT (score) and measured log ALT activity at (A) baseline (time 1, training), (C) time 2 (testing) and (D) time 3 (validation). (E) Score of each patient overtime. (F) Summary results from setting the threshold at 4.2. Values are shown as raw and natural logarithm ALT. Pearson's correlation coefficient (R) is shown based on log ALT and score.

Supplemental Figure 4. Analysis of a linear regression multivariate model using three biomarkers for patients with acetaminophen overdose evaluated at a cutoff of ALT ≥ 150 U/L. Three liver injury biomarkers (GLDH, K18 and miR-122) were used to build a predictive model for log ALT. (A) Correlation between predicted log ALT (score) and measured log ALT activity at (A) baseline (time 1, training), (B) time 2 (testing) and (C) time 3 (validation). (D) Score of each patient overtime. (E) Summary results from setting the threshold at 3.6. Values are shown as raw and natural log ALT. Pearson's correlation coefficient (R) is shown based on log ALT and score.

Supplemental Figure 5. Conventional biomarkers in the cross-sectional cohort.

Liver injury biomarkers (AST, ALP, TBIL), pancreatic injury biomarkers (lipase, AMYL, GLUC), kidney injury biomarkers (BUN, creatinine) and muscle injury biomarkers (CK) were measured in healthy volunteers and patients with liver, muscle, pancreas, GI and kidney damage. Line is at the median of the healthy cohort. Values are log₁₀ normalized. AST: aspartate aminotransferase, ALP: alkaline phosphatase, TBIL: total bilirubin, AMYL: amylase, GLUC: glucose, BUN: blood urea nitrogen, CREA: creatinine.

Supplemental Figure 6. Independent validation of the seven biomarker model in

the cross-sectional cohort of patients with liver damage. The seven liver injury

biomarker (GLDH, K18, ccK18, MCSF, MCSFR, OPN and miR-122) model was

evaluated in the cross-sectional cohort with thresholds set at 4.2 as defined in

supplemental figure 3. Correlation between the composite score and measured

log(ALT) activity in (A) liver injury patients and healthy volunteers and (B) all organ

injuries. (C) Summary results from setting the threshold at 4.2. Values are shown as raw

and natural log ALT.

Supplemental Figure 7. Independent validation of the linear regression model in the cross-sectional cohort with three biomarkers. The three liver injury biomarkers (GLDH, K18 and miR-122) model was evaluated in the cross-sectional cohort with the parameters defined in supplemental figure 6. (A) Correlation between predicted log ALT (score) and measured log ALT activity in (A) liver injury patients and healthy volunteers and (B) all organ injuries. (C) Summary results from setting the composite score threshold at 3.6. Values are shown as raw and natural logarithms ALT.

Supplemental Figure 8. Expression patterns of K18, Glud1 and miR-122. (A) Representative immunohistochemistry and in situ hybridization images of K18 protein and KRT18 gene expression in the liver from a healthy volunteer and liver injury patient. An internal RNAseq-based database was queried for KRT18 and GLUD1 gene expression in rat and human tissues. (B) *Krt18* and *Glud1* expression in rat tissues. Correlation of rat with human (C) KRT18 and (D) GLUD1 expression (log2-TPM) in 38 different tissues from the internal database. The linear regression line (purple) and confidence interval (grey shading) are shown. RATEmiRs was queried for miR-122 in rat tissues and Tissue Atlas was queried for miR-122 in human tissues. Correlation between human and rat (E) 3p and (F) 5p miR-122. (G) miR-122-3p and miR-122-5p expression in rat tissues.

Supplementary Tables

Supplemental table 1. Bioanalytical characteristics of biomarker assays.

Parameters	CK_M30	CK_M65	MCSF	MCSFR	Osteopontin	MiR-122
Assay principle	Enzyme-based sandwich ELISA	Enzyme-based sandwich ELISA	Electrical chemiluminescent -based MSD assay	Microbead-based MultiPlex Luminex assay	Electrical chemiluminescent -based MSD assay	Reverse-transcription droplet digital PCR based assay
LLD	38 U/L	100 U/L	0.57 pg/mL	25.78 ng/mL	0.8 ng/mL	2 copies/uL
Assay range	38-1000 U/L	100-5000 U/L	0.57-2320 pg/mL	0.52-125.27 ng/mL	0.04-200 ng/mL	1-100,000 copies/uL
Intra assay	<10%	<20%	<5%	<5%	<10%	<20%
Inter assay	<10%	<15%	<10%	<10%	<15%	<20%
Recovery	90-110%	90-110%	95-100%	95-100%	90-115%	ND
Linearity	Linear over dilutions: 1:2, 1:4, 1:8	Linear over dilutions: 1:2, 1:4, 1:8	Linear over dilutions: 1:2, 1:4, 1:8	Linear over dilutions: 1:2	Linear over dilutions: 1:2, 1:4, 1:8	linear over cDNA dilutions 1:2, 1:4, 1:8, 1:16
Freeze-thaw Stability	83.8-109.8% for all 3 freeze-thaw cycles	83.5-106.0% for 1 st and 2 nd freeze-thaw cycles	84.4-101.1% for all 3 freeze-thaw cycles	97.7-103.5% for all 3 freeze-thaw cycles	87.2-122.1% for all 3 freeze-thaw cycles	40-140% for all 3 freeze-thaw cycles

Supplemental Table 2. Basic and clinical demographics of the longitudinal cohort with acetaminophen overdose.

Variable	APAP exposure cohort
n	175
Age, median (IQR)	30 (21, 45)
Gender (M:F)	48:127
Ethnicity (number, % of total)	
White Scottish	113, 64.57%
White British	43, 24.57%
White English	6, 3.03%
White Other	6, 3.03%
Asian or British Asian	2, 1.01%
Mixed	5, 2.86%
Type of overdose (number, % of total)	
Acute Overdose <8 hours from ingestion	95, 54.29%
Acute Overdose >8 hours from ingestion	27, 15.43%
Supra-therapeutic overdose	28, 16%
Staggered intentional overdose	23, 13.14%
Amount of acetaminophen ingested (g) median (IQR)	12 (8, 20)
Required NAC treatment (number, % of total)	123, 70.69%
Required NAC treatment beyond initial dose of 300mg/kg (number, % of total)	
First presentation acetaminophen concentration (mg/L) median (IQR)	54.5 (12.75, 122)
ALT > ULN at first presentation (number, % of total)	30, 17.14%
DILI at first presentation (number, % of total)	11, 6.29%
DILI at any time (number, % of total)	16, 9.14%

Supplemental Table 3. Basic and clinical demographics of cross-sectional cohort of healthy volunteers and patients with damage to the liver, muscle, pancreas, gastrointestinal tract or kidney.

Variable	Healthy	Liver	Muscle	Pancreatic	GI¹	Kidney
n	135	104	74	34	37	40
Age (median)	38	52	56	50	31	65
Sex (F, %)	50.4	51.0	60.8	47.1	59.5	35.0
Ethnicity (number, % of total)						
Caucasian	57, 42%	72, 69%	56, 76%	28, 82%	32, 86%	29, 73%
Black	4, 2%	16, 15%	10, 14%	1, 3%	1, 3%	9, 23%
Asian	68, 50%	4, 4%	3, 4%	1, 3%	0, 0%	2, 5%
Mixed	4, 2%	7, 7%	3, 4%	3, 9%	2, 5%	0, 0%
Unknown	2, 1%	5, 5%	2, 3%	1, 3%	2, 5%	0, 0%
Type of liver damage (number, % of total)						
Cirrhosis	0, 0%	20, 19%	0, 0%	0, 0%	0, 0%	0, 0%
Transaminitis	0, 0%	54, 52%	0, 0%	0, 0%	0, 0%	0, 0%
Liver Transplant	0, 0%	9, 9%	0, 0%	0, 0%	0, 0%	0, 0%
Acute Liver Failure	0, 0%	4, 4%	0, 0%	0, 0%	0, 0%	0, 0%
Drug Induced	0, 0%	24, 23%	0, 0%	0, 0%	0, 0%	0, 0%
Metastatic Liver Disease	0, 0%	27, 25%	0, 0%	0, 0%	0, 0%	0, 0%
Hepatitis	0, 0%	12, 12%	0, 0%	0, 0%	0, 0%	0, 0%
Alcoholic	0, 0%	15, 14%	0, 0%	0, 0%	0, 0%	0, 0%
None	135, 100%	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%
Co-Morbidities (number, % of total)						
Pancreatitis Acute	0, 0%	5, 5%	0, 0%	11, 32%	0, 0%	0, 0%
Pancreatitis Chronic	0, 0%	0, 0%	0, 0%	18, 53%	0, 0%	0, 0%
Pancreatitis Necrotizing	0, 0%	0, 0%	0, 0%	3, 9%	0, 0%	0, 0%
Pancreatitis Idiopathic	0, 0%	0, 0%	0, 0%	5, 15%	0, 0%	0, 0%
Myotosis	0, 0%	0, 0%	19, 26%	0, 0%	0, 0%	0, 0%
Myopathy	0, 0%	0, 0%	32, 43%	0, 0%	0, 0%	0, 0%
Necrotizing Myopathy	0, 0%	0, 0%	3, 4%	0, 0%	0, 0%	0, 0%
Muscular Dystrophy	0, 0%	0, 0%	7, 9%	0, 0%	0, 0%	0, 0%
Myotonic Dystrophy	0, 0%	0, 0%	6, 8%	0, 0%	0, 0%	0, 0%
Dermatomyositis	0, 0%	0, 0%	28, 38%	0, 0%	0, 0%	0, 0%

Variable	Healthy	Liver	Muscle	Pancreatic	GI¹	Kidney
Diabetes Mellitus	0, 0%	13, 13%	5, 7%	10, 29%	2, 5%	16, 40%
Cholelithiasis	0, 0%	8, 8%	0, 0%	1, 3%	1, 3%	0, 0%
Hypertension	1, 1%	29, 28%	17, 23%	16, 47%	4, 11%	29, 73%
Obesity	1, 1%	3, 3%	0, 0%	0, 0%	0, 0%	0, 0%
Neuropathy	0, 0%	6, 6%	2, 3%	4, 12%	0, 0%	3, 8%
Cancer	0, 0%	32, 31%	0, 0%	0, 0%	0, 0%	0, 0%
Irritable Bowel Syndrome	0, 0%	0, 0%	1, 1%	0, 0%	3, 8%	0, 0%
Ulcerative Colitis	0, 0%	1, 1%	0, 0%	0, 0%	22, 59%	0, 0%
Crohn's	0, 0%	0, 0%	0, 0%	0, 0%	10, 27%	0, 0%
Celiac Disease	0, 0%	0, 0%	0, 0%	0, 0%	2, 5%	0, 0%
GERD²	0, 0%	18, 17%	7, 9%	10, 29%	4, 11%	11, 28%
Immunosuppression	1, 1%	6, 6%	1, 1%	1, 3%	15, 41%	0, 0%
CAD CHF³	0, 0%	1, 1%	2, 3%	3, 9%	0, 0%	22, 55%
Chronic Kidney Disease	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	40, 100%
Hyperlipidemia	0, 0%	1, 1%	9, 12%	15, 44%	2, 5%	12, 30%
Acute Kidney Injury	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	11, 28%
ESRD⁴	0, 0%	0, 0%	0, 0%	0, 0%	0, 0%	4, 10%

¹Gastrointestinal, ²Gastroesophageal reflux disease, ³Coronary Artery Disease, Congestive Heart Failure,

⁴End Stage Renal Disease

Supplemental table 4. Assessment of the multivariate biomarker models in the acetaminophen and cross-sectional cohorts with a cutoff of ALT > 50 U/L

Data set	Sensitivity	Specificity	PPV	NPV
Acetaminophen overdose cohort				
Panel-3				
Testing (T2)	0.96	0.94	0.79	0.99
Validation (T3)	1.00	0.95	0.86	1.00
Panel-7				
Testing (T2)	0.96	0.92	0.74	0.99
Validation (T3)	1.00	0.92	0.78	1.00
Cross-sectional cohort: comparator group vs liver				
Panel-3				
Healthy	0.90	0.99	0.99	0.93
Muscle	0.90	0.79	0.90	0.79
Pancreas	0.90	0.96	0.99	0.69
GI	0.90	0.97	0.99	0.78
Kidney	0.90	0.72	0.89	0.74
Panel-7				
Healthy	0.98	0.99	0.99	0.99
Muscle	0.98	0.72	0.88	0.94
Pancreas	0.98	0.87	0.97	0.91
GI tract	0.98	0.92	0.97	0.94
Kidney	0.98	0.56	0.85	0.92

A score threshold of 3.4 was set based on a cutoff of ALT > 50 U/L in the testing (T2) dataset in the acetaminophen overdose cohort to achieve a sensitivity of ≥ 0.95 . Panel-3: GLDH, K18, miR-122, Panel-7: GLDH, K18, miR-122, ccK18, MCSF, MCSFR, OPN, GI: gastrointestinal

Supplemental table 5. Performance characteristics of candidate biomarkers in cross-sectional cohort.

Comparator Group vs. Liver Injury					
Metric	Healthy	GI	Pancreas	Muscle	Kidney
ALT					
AUC	0.927	0.915	0.932	0.786	0.940
(lower 95% CI)	0.889	0.870	0.891	0.721	0.903
(upper 95% CI)	0.965	0.960	0.972	0.852	0.977
p-value	1.39E-22	1.20E-09	3.25E-08	5.23E-08	1.61E-15
GLDH					
AUC	0.898	0.863	0.848	0.802	0.875
(lower 95% CI)	0.853	0.801	0.783	0.734	0.819
(upper 95% CI)	0.943	0.926	0.913	0.869	0.931
p-value	4.33E-15	6.39E-05	4.22E-05	5.63E-07	2.71E-07
K18					
AUC	0.980	0.959	0.963	0.937	0.897
(lower 95% CI)	0.961	0.929	0.935	0.899	0.846
(upper 95% CI)	0.999	0.990	0.992	0.975	0.947
p-value	1.65E-41	1.62E-13	4.20E-16	5.38E-25	7.10E-09
ccK18					
AUC	0.873	0.829	0.874	0.778	0.837
(lower 95% CI)	0.825	0.763	0.816	0.710	0.774
(upper 95% CI)	0.921	0.895	0.932	0.845	0.900
p-value	6.74E-13	1.94E-07	1.59E-07	8.20E-05	9.30E-09
MCSF					
AUC	0.968	0.970	0.889	0.913	0.764
(lower 95% CI)	0.947	0.947	0.832	0.873	0.681
(upper 95% CI)	0.988	0.992	0.945	0.954	0.847
p-value	1.35E-27	2.65E-16	9.75E-12	6.45E-18	3.43E-04
MCSFR					
AUC	0.919	0.941	0.913	0.833	0.880
(lower 95% CI)	0.882	0.903	0.864	0.775	0.824
(upper 95% CI)	0.956	0.979	0.962	0.891	0.936
p-value	2.03E-21	1.09E-14	5.98E-07	1.30E-08	2.07E-10

Comparator Group vs. Liver Injury					
Metric	Healthy	GI	Pancreas	Muscle	Kidney
OPN					
AUC	0.945	0.934	0.868	0.851	0.679
(lower 95% CI)	0.911	0.894	0.805	0.794	0.586
(upper 95% CI)	0.980	0.975	0.930	0.909	0.772
p-value	1.66E-20	7.01E-08	3.95E-08	4.54E-09	4.36E-02
MiR-122					
AUC	0.942	0.874	0.913	0.859	0.952
(lower 95% CI)	0.913	0.818	0.862	0.803	0.921
(upper 95% CI)	0.972	0.931	0.963	0.915	0.984
p-value	1.30E-22	7.03E-07	5.59E-06	2.37E-10	3.98E-15

AUC: area under the curve, GI: gastrointestinal, CI: confidence interval; AUC values in healthy column are also presented in Table 2.

Supplemental table 6. Krt18 and Glud1 gene expression (transcript per million) in rat tissues

Ensembl gene ID	ENSRNOG00000047393	ENSRNOG00000057367
Gene name	<i>Krt18</i>	<i>Glud1</i>
Adipose	29.88528	248.9478
Adrenal Gland	0.809365	109.3341
Aorta	3.672417	164.0643
Aortic Endothelial Cells	0.042041	255.682
Bladder	1224.304	284.7448
Bone Marrow	0.053724	38.79093
Brown Adipose	0.347206	128.1751
Cecum	425.2671	163.2745
Cerebellum	0.460357	265.6096
Cerebral Cortex	0.911499	429.1953
Cervix Uteri	101.3481	74.38472
Colon	547.7608	190.2412
Cornea	61.30009	146.4273
Dorsal Root Ganglion	0.27439	301.1755
Epididymis	271.5411	193.0114
Esophagus	0.27648	119.6904
Fallopian Tube	90.68141	180.3572
Heart	4.29917	136.0678
Hippocampus	2.640748	470.4592
Hypothalamus	22.34186	751.0555
Ileum	602.1224	337.8785
Jejunum	468.5455	328.3326
Kidney	97.5769	431.2272
Lens	1.597902	18.74528
Liver	434.7969	623.0281
Liver Sinusoidal Endothelial Cells	5.391445	535.9823
Lung	160.5427	223.823
Lung Microvascular Endothelial Cells	4.88206	590.9548
Mammary Gland	271.0016	177.5404
Mesenteric Lymph Node	87.33451	129.053
Ovary	7.665875	169.8032
Peripheral Nerve	0.365716	221.6998
Pituitary Gland	163.3907	75.47128
Prostate Gland	0.334858	48.37447
Salivary Gland	320.2824	16.63248

Seminal Vesicle	152.7679	57.22239
Skeletal Muscle	0.195903	139.6985
Skin	0.529863	158.1688
Spinal Cord	1.401544	532.5201
Spleen	4.376594	92.45521
Stomach	496.8318	41.49174
Striatum	8.33308	536.9551
Testis	0.170369	44.34947
Thymus Gland	197.2049	139.0348
Thyroid Gland	191.6185	119.1097
Tongue	0.389779	154.4846
Trachea	297.8835	150.4735
Ureter	253.62	227.6216
Uterus	315.2315	175.7814
Vagina	66.42612	86.80762
Ventricle	0.094285	126.8217