OXFORD

SOT | Society of Toxicology
academic.oup.com/toxsci

TOXICOLOGICAL SCIENCES, 181(1), 2021, 23–34

doi: 10.1093/toxsci/kfab003
Advance Access Publication Date: 23 January 2021
Research Article

Evaluating the Sensitivity and Specificity of Promising Circulating Biomarkers to Diagnose Liver Injury in Humans

Heather P. Llewellyn ,* Vishal S. Vaidya,†,1 Zhenyu Wang,§ Qinghai Peng,* Craig Hyde,§ David Potter,† Jianying Wang,* Qing Zong,* Seda Arat,‡ Matt Martin,‡ Katherine Masek-Hammerman,† Roscoe Warner,¶ Kent Johnson,¶ Gerd A. Kullak-Ublick,‖,‖‖ Guruprasad P. Aithal,‖‖‖ James W. Dear,# and Shashi K. Ramaiah†,1

*Drug Safety Research and Development, Pfizer Inc., San Diego, CA, 92121 USA †Drug Safety Research and Development, Pfizer Inc., Cambridge, MA 02139, USA ‡Drug Safety Research and Development, Pfizer Inc., Groton, CT 06340 USA §Early Clinical Development, Pfizer, Inc., Cambridge, Massachusetts 02139, USA ¶Michigan Medicine, University of Michigan, Ann Arbor, Michigan 48109, USA ‖Mechanistic Safety, CMO & Patient Safety, Global Drug Development, Novartis, Basel 4056, Switzerland ‖‖Department of Clinical Pharmacology and Toxicology, University Hospital Zurich, University of Zurich, Zurich 8091, Switzerland ‖‖‖NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham, Nottingham NG7 2UH, UK; and #Pharmacology, Therapeutics and Toxicology, Centre for Cardiovascular Science, University of Edinburgh, The Queen's Medical Research Institute, Edinburgh EH16 4TJ, UK

[1]To whom correspondence should be addressed at E-mail: vishal.vaidya@pfizer.com and Drug Safety Research & Development, Pfizer, Inc., 300 Technology Square Drive, 3rd Floor, Cambridge, MA 02139, USA. E-mail: shashi.ramaiah@pfizer.com.

Heather P. Llewellyn and Vishal S. Vaidya contributed equally to this study.

The authors certify that all research involving human subjects was done under full compliance with all government policies and the Helsinki Declaration.

ABSTRACT

Early diagnosis of drug-induced liver injury (DILI) continues to be a major hurdle during drug development and postmarketing. 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 with 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 7 promising biomarkers and demonstrated that a 3-biomarker multivariate model can accurately detect liver injury.

Key words: keratin-18; microRNA; glutamate dehydrogenase; diagnosis; liver. .

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 United States 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 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 preclinical 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 APAP 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 7 DILI biomarkers in patients with APAP overdose (APAP, $n = 175$); (2) apply random forest modeling to train, test, and validate a multivariate model with top performing biomarkers to predict ALT; and (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 Material)

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 APAP overdose, who met the inclusion criteria, were asked to provide informed consent to participate in the prospective, APAP overdose cohort study, and their demographics and blood results were recorded. Although the current consensus for defining DILI is an ALT value ≥5× upper limit of normal (ULN) (Aithal *et al.*, 2011), in this study a cut-off of 3 times the upper limit of normal (≥3× 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 ≥3× ULN of study patients compared with controls as a potential signal of DILI during drug development in particular (Senior, 2014). A cut-off of >1 ULN ALT (>50 U/l) was also explored. Serum was collected at 3 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 h (IQT: 4.1, 10.7) after ingestion of APAP. The median collection time for T2 was 12.7 h (IQT: 9.2, 14.1) after T1, and the median for T3 was 22.9 h (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 no. 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 patients were determined by utilizing the EWG definition (≥5× ALT ULN, or ≥2× ALP ULN, or ≥3× ALT ULN and ≥2× TBL ULN) and medical adjudication demonstrating various liver damage etiologies. Healthy subjects were selected as those having normal ranges of ALT (<35 U/l), AST (8–30 U/l), ALP (0.2–1.2 mg/dl), TBL (0.2–1.2 mg/dl), glucose (73–100 mg/dl), blood urea nitrogen (BUN) (8–20 mg/dl), serum creatinine (0.5–1.0 mg/dl for females and 0.7–1.3 mg/dl for males), and creatine kinase (26–180 U/l for females and 38–240 U/l for males). Subjects with clinically demonstrable liver damage typically included those with accidental APAP overdose, ethanol toxicity, drug abuse, transaminitis (elevated transaminases without other evidence of liver injury), metastatic liver disease (diagnosed by biopsy or histopathology after resection), cirrhosis, and liver impairment (Hepatitis B or C, hepatic graft vs host disease). The metastatic group is comprised 12 different sites of origin of the primary cancers. Represented by adrenal, breast, cholangiocarcinoma, colon, endometrial, kidney, liver, melanoma, ancreatic, pleomorphic sarcoma, prostate, and rectal. No single primary cancer site is represented by more than 4

patients of the 27 in this cohort. Some subjects exhibited multiple types of liver injury for example, a subject could be represented in both categories of DILI and acute liver failure. Muscle Injury was diagnosed by either (1) medical adjudication, (2) a muscle biopsy, (3) genetic testing, or (4) clinically determined injuries, which may include, but are not limited to, primary disorders of muscle (dystrophies, myotonic disorders, congenital myopathies, and mitochondrial myopathies) and toxic myopathies (drug, alcohol, and toxicants), as exhibited by, myositis (inflammatory muscle injury), neurogenic atrophy, necrotizing inflammatory muscle injury, chronic severe atrophy, AAF, type II fiber atrophy, nuclear myobags, denervation atrophy, and increased lipids in myofibers. Subjects demonstrating pancreatitis (acute, chronic, hereditary) were diagnosed by either (1) persistent severe epigastric pain, (2) diagnostic armamentarium (endoscopic ultrasound [ES], magnetic resonance cholangiopancreatography [MRCP], computerized tomography [CT], or transabdominal ultrasound), (3) clinically demonstrable deficiencies, or (4) amylase or lipase 3× ULN. Subjects demonstrating gastrointestinal abnormalities were diagnosed by either (1) endoscopy, (2) sigmoidoscopy, or (3) colonoscopy, or (4) clinically demonstrable deficiencies, which could include, but is not limited to, gastroesophageal reflux disease (GERD), esophagitis, irritable bowel syndrome (IBS), celiac disease, Crohn's disease, ulcerative colitis, ulcerative pancolitis, ulcerative proctosigmoiditis, and appendicitis. Subjects having chronic kidney disease (CKD) were diagnosed by either (1) biopsy-proven or (2) clinically demonstrable deficiencies, which could include, but are not limited to, diabetes, high blood pressure, glomerulonephritis, interstitial nephritis, polycystic kidney disease, and malformations, as exhibited by, CKD stage II–V, end stage renal disease (ESRD) and patients on dialysis. Patients were excluded from a cohort if they had any ongoing health problems or immunological flares that could influence liver health, or if they had additional organ injury outside their included cohort. Traditional organ damage biomarkers, such as AST and ALP for liver; lipase and amylase for pancreas, BUN and creatinine for kidney; and creatinine kinase enzyme activity for muscle damage were elevated in their respective clinically diagnosed organ damage 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, and 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 whereas 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 the manufacturers' protocols with a few modifications. The serum biomarker values were calculated using a 6–9 point 5-parameter logarithmic standard curve (Supplementary Figure 1).

MicroRNA-122. Total RNAs from 100 µl plasma/serum were purified by Qiagen's miRNeasy kit (Valencia, California) 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 10× reaction buffer (New England Biolabs, Ipswich, United Kingdom), 10.75 µl of RNA free H_2O, 0.25 µl of Heparinase I (New England Biolabs, Ipswich, United Kingdom), and 1 µl of RNase inhibitor (Promega, Wisconsin). Samples were incubated for 1 h at 30°C followed by 1 min 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 Material.

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 (eg, healthy) predicted to be cases (eg, 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 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 2 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 versus 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, ie, 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 ≥ 150 U/L (≥3× ULN) or ALT > 50 U/l (>1× 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 dataset (T3 data).
- Number of false positives given a sensitivity > 0.95 in the testing dataset; at the same time, defined a DILI threshold to evaluate in the validation dataset (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 dataset to evaluate model performance.

RESULTS

Analysis of Candidate Biomarkers in Cohort of Patients With Acetaminophen Overdose

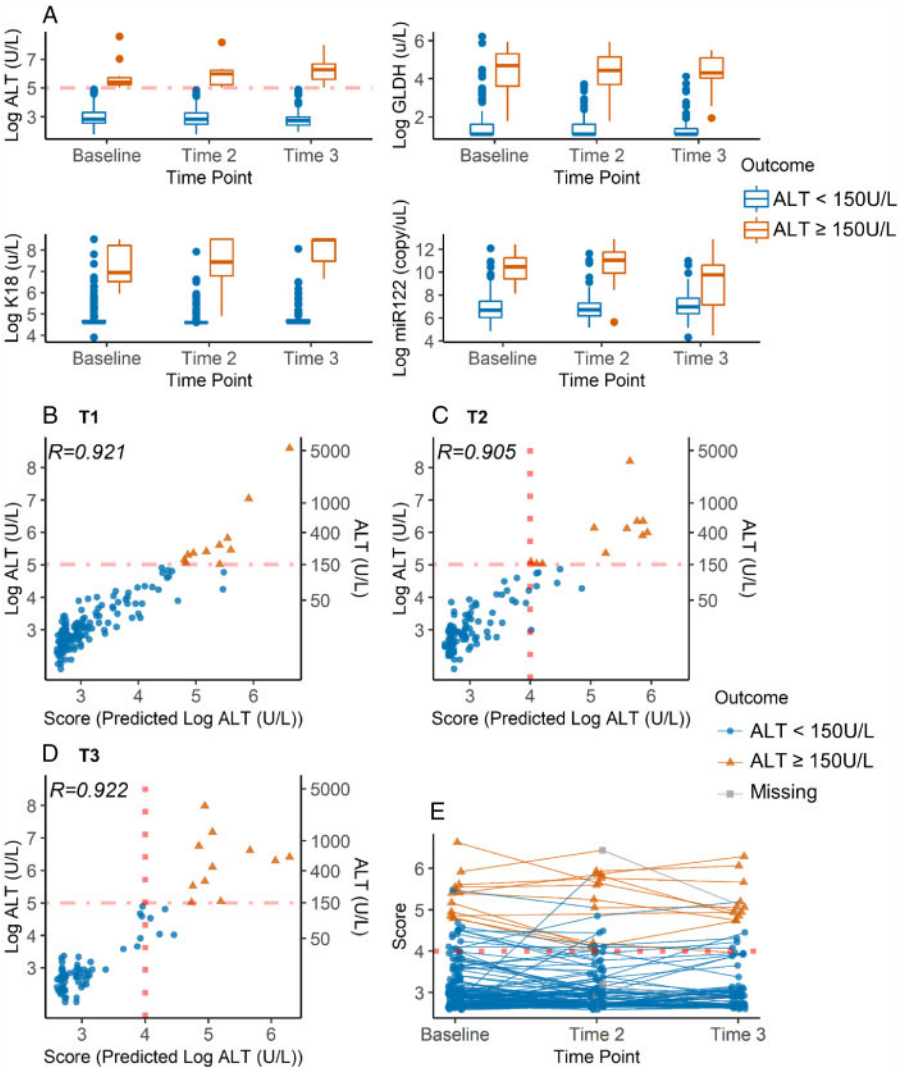
Promising liver injury biomarkers GLDH, K18, ccK18, miR-122, OPN, MCSF, and MCSFR were evaluated for their ability to predict ALT in a cohort of patients with APAP overdose ($n = 175$) (Supplementary Table 2) at 3 timepoints. A random forest model to predict ALT was trained, tested, and validated on 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× 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 ≥ 150U/l compared with APAP exposed patients with ALT < 150U/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 3-biomarker model), the composite score (ie, 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 dataset (Figure 1C). In general, when the values of 2 or 3 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 (Figs. 1E and 1F) or >50 U/l (Supplementary Table 4). Furthermore, all 7 biomarkers were used in the multivariate model and were evaluated with a cutoff of ALT ≥ 150 U/l (Supplementary Figure 3, Table 1) and > 50 U/L (Supplementary Table 1). In this cohort, the specificity and positive predictive value (PPV) of the models were similar between the 3- and 7-biomarker models when using a cutoff of either ALT ≥ 150 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 ROC AUC suggest comparable predictivity between the 2 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 (Supplementary Figure 4) compared with 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 3- and 7-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 3- and 7-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



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

Figure 1. Analysis of a 3-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.

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) ($n = 4$) (Supplementary Methods, Supplementary Table 3, 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, DILI, 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 ($r = -0.043$, $p = .66$; $r = 0.1512$, $p = .125$, respectively) and OPN levels did not correlate with ALT activity ($r = -0.2086$, $p = .0336$).

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, Supplementary Table 4). MCSF achieved an AUC of 0.97, whereas ALT achieved an AUC

Table 1. Assessment of the Random Forest Biomarker Models in the Acetaminophen Overdose and Cross-sectional Cohorts

Dataset	Panel 3[a] R	Panel 3 AUC (CI)	Panel 7[b] 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 versus 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)

[a]Panel 3: GLDH, K18, miR-122.
[b]Panel 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 h, IQR: 4.1, 10.7 after acetaminophen ingestion), T2: Timepoint 2 (11.4 h after T1), T3: Timepoint 3 (21.8 h after T1); GI: gastrointestinal; CI: 95% confidence interval.

Table 2. Comparative Assessment of Candidate Biomarkers at a Fixed Sensitivity for Diagnosis of Liver Injury

Metric	Biomarker Threshold	Sensitivity	Specificity	ROC AUC	PPV	NPV
Acetaminophen overdose cohort[a]						
T1						
GLDH	5.5	1.00	0.77	0.95	0.22	1.00
K18	375.5	1.00	0.94	0.97	0.52	1.00
ccK18	NA	1.00	0.00	0.72	0.06	
MCSF	3.8	1.00	0.04	0.68	0.07	1.00
MCSFR	493.5	1.00	0.35	0.84	0.09	1.00
OPN	3.2	1.00	0.09	0.72	0.07	1.00
miR-122	3412.0	1.00	0.85	0.96	0.31	1.00
Panel-3	4.7	1.00	0.99	0.99	0.85	1.00
Panel-7	4.7	1.00	0.99	0.99	0.85	1.00
T2						
GLDH	5.5	1.00	0.77	0.97	0.29	1.00
K18	135.5	1.00	0.84	0.97	0.38	1.00
ccK18	NA	1.00	0.00	0.87	0.09	
MCSF	19.5	1.00	0.32	0.82	0.12	1.00
MCSFR	416.1	1.00	0.22	0.73	0.11	1.00
OPN	8.6	1.00	0.31	0.78	0.12	1.00
miR-122	280.0	1.00	0.10	0.90	0.10	1.00
Panel 3	4.0	1.00	0.94	0.99	0.61	1.00
Panel 7	4.2	1.00	0.97	0.99	0.79	1.00
T3						
GLDH	6.5	1.00	0.86	0.98	0.52	1.00
K18	716.0	1.00	0.99	0.99	0.92	1.00
ccK18	160.0	1.00	0.81	0.96	0.46	1.00
MCSF	21.0	1.00	0.66	0.90	0.31	1.00
MCSFR	445.6	1.00	0.19	0.70	0.16	1.00
OPN	7.4	1.00	0.14	0.76	0.15	1.00
miR-122	82.0	1.00	0.01	0.72	0.14	1.00
Panel 3	4.6	1.00	1.00	1.00	1.00	1.00
Panel 7	4.4	1.00	0.99	1.00	0.92	1.00
Cross-sectional cohort: healthy versus 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

[a]Sensitivity 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 h, IQT: 4.1, 10.7 after acetaminophen ingestion), T2: Timepoint 2 (12.7 h, IQT: 9.2, 14.1 after T1), T3: Timepoint 3 (22.9 h, IQT: 19.8, 24.2 after T1); PPV: positive predictive value, NPV: negative predictive value.

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) (Supplementary Table 5). However, the K18 AUC for liver versus 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 ($p = 7.2e\text{-}05$) when compared with 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 with 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 with 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 dataset 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 3-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

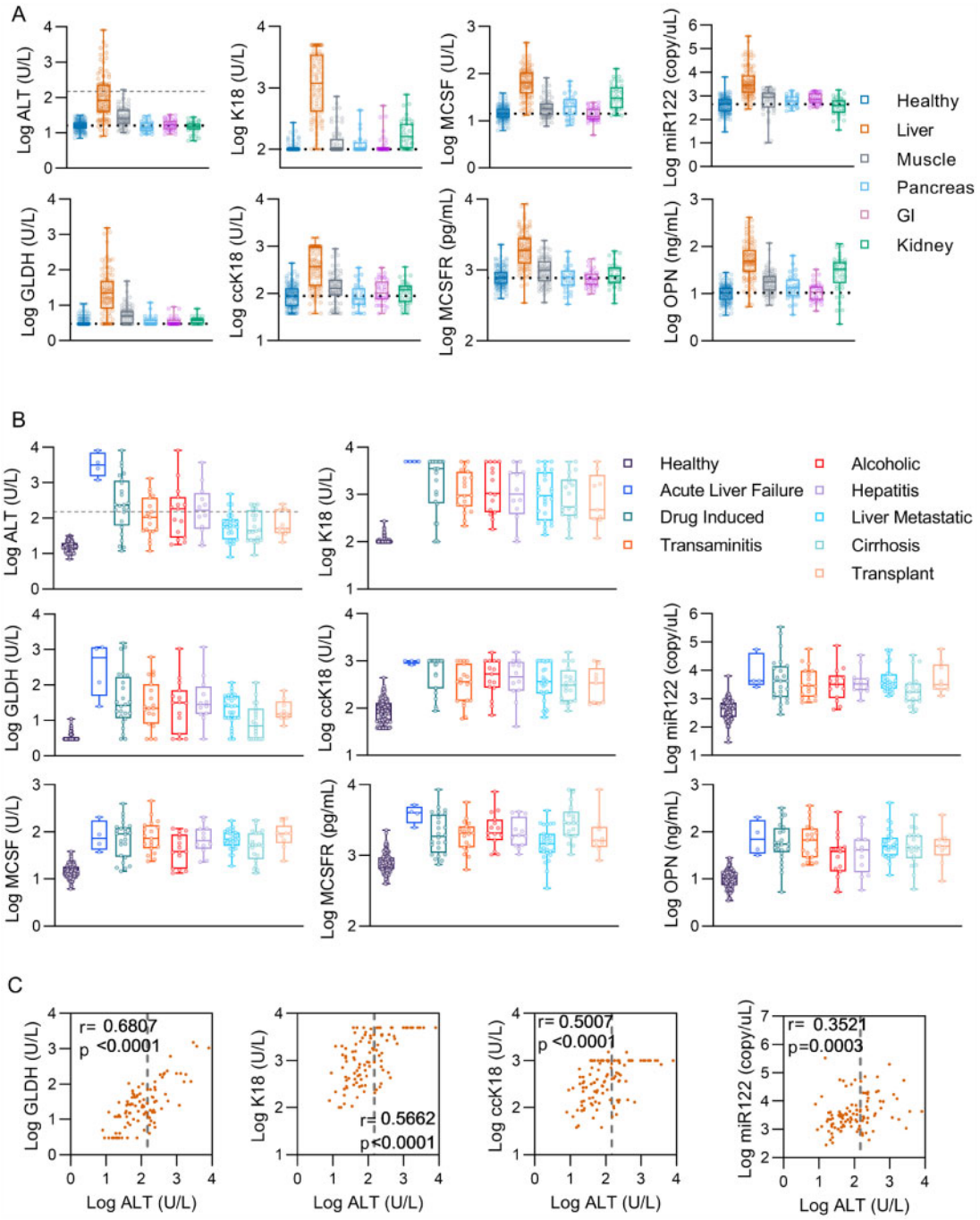
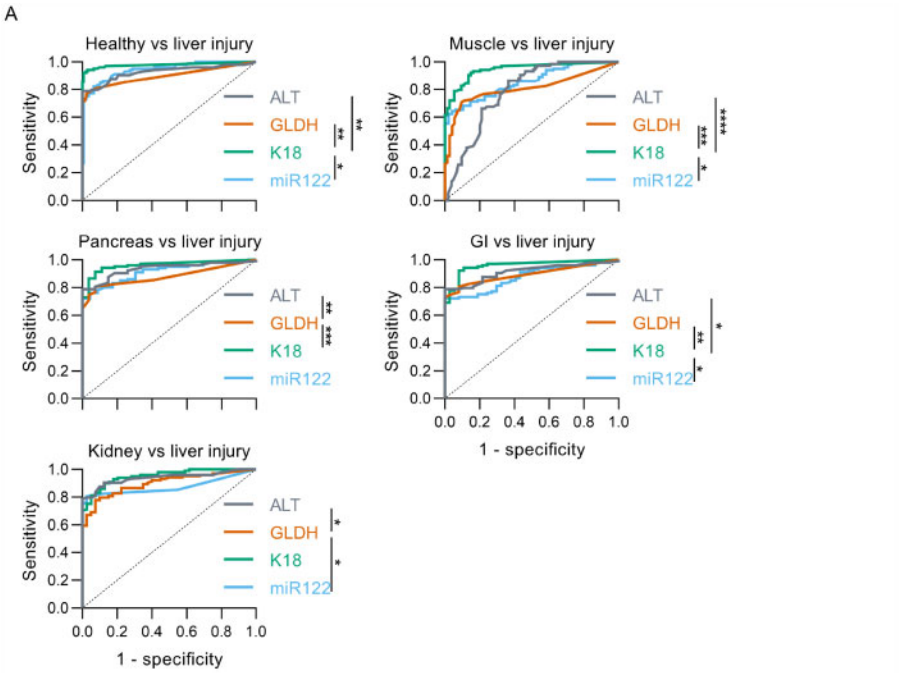


Figure 2. Evaluation of 7-candidate biomarkers for liver injury in comparison to ALT in the cross-sectional cohort. A, Glutamate dehydrogenase (GLDH), cytokeratin 18 (K18), caspase-cleaved K18 (ccK18), macrophage colony-stimulating factor (MCSF), MCSFR (MCSFR), microRNA-122 (miR-122), and osteopontin (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 are repeated from Figure 2A for reference. Values are log 10 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 log 10 normalized.

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 3-biomarker model (Table 2). The 7-biomarker model had a higher specificity for identifying patients with liver damage than the 3-biomarker model or any individual biomarker alone (Supplementary 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 3-biomarker and 7-biomarker panel would correctly identify 101 and 103 patients,



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

Figure 3. Independent validation of the 3-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 versus healthy volunteers. The 3-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. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$, $^{****}p < .0001$.

respectively. If the 3-biomarker model composite score threshold was lowered and set based on an ALT of >50 U/l as defined in Supplementary Figure 3, the number of false negatives decreased (Supplementary 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× 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 (Supplementary Table 4; Table 1). The linear regression 3- biomarker model as defined in the APAP cohort performed slightly better (Supplementary 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 APAP 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

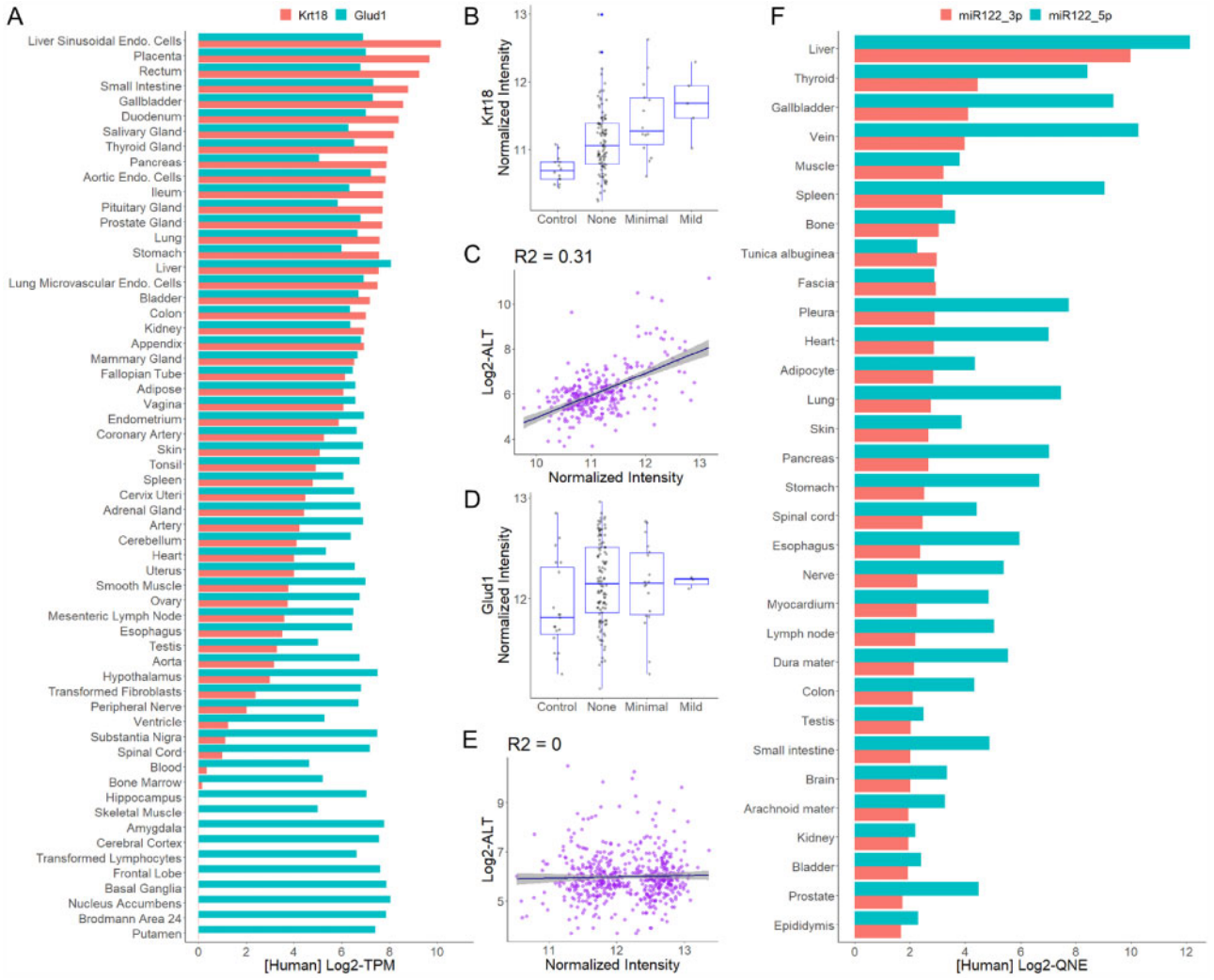


Figure 4. Expression patterns of *KRT18*, *GLUD1*, and miR-122 in human and rat. A, A RNAseq-based database was queried for KRT18 and GLUD1 gene expression in human tissues (see Methods). DrugMatrix, an Affymatrix-based rat toxicogenomics database, was queried for *Krt18* (B, C) and *Glud1* (D, E) expression 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 (gray shading) are shown. Pairwise comparisons were calculated using the Dunn test if the Kruskal-Wallis test was significant. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$, $^{****}p < .0001$. TPM: transcript per million; QNE: quantile normalized expression.

diseased livers ($n = 5$), K18 protein and mRNA were consistently and highly expressed in bile duct epithelium and in peri-portal hepatocytes (Supplementary 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 Supplementary Table 6. In the human, *KRT18* expression is predominantly expressed in the liver compared with other tissues, whereas *GLUD1* is widely expressed, suggesting that *GLUD1* may be less specific for liver than *KRT18* (Figure 4A). In the rat, *Krt18* expression is the highest in bladder, ileum, colon, stomach, and liver (Supplementary Figure 8B). Rat *Krt18* and *Glud1* expression levels in some tissues, including the liver, kidney, and heart correlated with human expression (Supplementary Figs. 8C and 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. Although treated samples had higher expression of *Glud1* compared with 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 (Supplementary 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 ($r^2 = 0.67$) (Supplementary Figs. 8E and 8F). These expression data suggest that K18 and miR-122 maybe be specific biomarkers of liver injury in rats and humans. Although Glud1 mRNA expression doesn't seem to be tissue specific in rats and humans, protein expression and enzyme activity of GLDH across all tissues has not been evaluated and may provide additional information on its 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 (1) 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 (2) a 3- biomarker model with K18, GLDH and miR-122 that was trained, tested, and validated using an APAP overdose cohort, was independently validated in a cross-sectional cohort and able to achieve separation between patients with liver damage and healthy volunteers. The 3-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 postmarketing. 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 APAP 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 3-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 3-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 nonspecifically 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, nonalcoholic fatty liver disease and liver cancer (Ku *et al.*, 2016). Although 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 with 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 with 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; whereas 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 3-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.

SUPPLEMENTARY DATA

Supplementary data are available at *Toxicological Sciences* online.

CODE/DATA AVAILABILITY

Code available upon request.

ACKNOWLEDGMENTS

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

FUNDING

Drug Safety Research and Development Department within Pfizer's Worldwide Research Development and Medical.

AUTHOR CONTRIBUTIONS

Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; H.P.L., V.S.V., Z.W., Q.P., C.H., D.P., J.W., Q.Z., S.A., M.M., K.M.H., R.W., K.J., G.A.K., G.P.A., J.W.D., and S.K.R. Drafting the work or revising it critically for important intellectual content; H.P.L., V.S.V., Z.W., Q.P., C.H., D.P., J.W., Q.Z., S.A., M.M., K.M.H., R.W., K.J., G.A.K., G.P.A., J.W.D., and S.K.R. Final approval of the version to be published; H.P.L., V.S.V., Z.W., Q.P., C.H., D.P., J.W., Q.Z., S.A., M.M., K.M.H., R.W., K.J., G.A.K., G.P.A., J.W.D., and S.K.R. 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. H.P.L., V.S.V., Z.W., Q.P., C.H., D.P., J.W., Q.Z., S.A., M.M., K.M.H., R.W., K.J.., G.A.K., G.P.A., J.W.D., and S.K.R.

DECLARATION OF CONFLICTING INTERESTS

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

REFERENCES

Aithal, G. P., Watkins, P. B., Andrade, R. J., Larrey, D., Molokhia, M., Takikawa, H., Hunt, C. M., Wilke, R. A., Avigan, M., Kaplowitz, N., *et al.* (2011). Case definition and phenotype standardization in drug-induced liver injury. *Clin. Pharmacol. Ther.* **89**, 806–815.

Bala, S., Petrasek, J., Mundkur, S., Catalano, D., Levin, I., Ward, J., Alao, H., Kodys, K., and Szabo, G. (2012). Circulating microrna in exosomes indicate hepatocyte injury and inflammation in alcoholic, drug-induced, and inflammatory liver diseases. *Hepatology* **56**, 1946–1957.

Bernal, W., and Wendon, J. (2014). Acute liver failure. *N. Engl. J. Med.* **370**, 1170–1171.

Bjornsson, E. S., Bergmann, O. M., Bjornsson, H. K., Kvaran, R. B., and Olafsson, S. (2013). Incidence, presentation, and outcomes in patients with drug-induced liver injury in the general population of Iceland. *Gastroenterology* **144**, 1419–1425. 1425.e1411–1413; quiz e1419–1420.

Church, R. J., Kullak-Ublick, G. A., Aubrecht, J., Bonkovsky, H. L., Chalasani, N., Fontana, R. J., Goepfert, J. C., Hackman, F., King, N. M. P., Kirby, S., *et al.* (2019). Candidate biomarkers for the diagnosis and prognosis of drug-induced liver injury: An international collaborative effort. *Hepatology* **69**, 760–773.

Davis, J., and Goadrich, M. 2006. The relationship between precision-recall and roc curves. In: *Proceedings of the 23rd International Conference on Machine Learning*. Association for Computing Machinery, Pittsburgh, PA, USA.

DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* **44**, 837–845.

Fu, S., Wu, D., Jiang, W., Li, J., Long, J., Jia, C., and Zhou, T. (2019). Molecular biomarkers in drug-induced liver injury: Challenges and future perspectives. *Front. Pharmacol.* **10**, 1667.

Giesen, P. L., Peltenburg, H. G., de Zwaan, C., Janson, P. C., Flendrig, J. G., and Hermens, W. T. (1989). Greater than expected alanine aminotransferase activities in plasma and in hearts of patients with acute myocardial infarction. *Clin. Chem.* **35**, 279–283.

Ku, N. O., Strnad, P., Bantel, H., and Omary, M. B. (2016). Keratins: Biomarkers and modulators of apoptotic and necrotic cell death in the liver. *Hepatology* **64**, 966–976.

Kullak-Ublick, G. A., Andrade, R. J., Merz, M., End, P., Benesic, A., Gerbes, A. L., and Aithal, G. P. (2017). Drug-induced liver injury: Recent advances in diagnosis and risk assessment. *Gut* **66**, 1154–1164.

Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., and Tuschl, T. (2002). Identification of tissue-specific microrna from mouse. *Curr. Biol.* **12**, 735–739.

Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A. O., Landthaler, M., *et al.* (2007). A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**, 1401–1414.

Levy, L., Tigert, A., Huszti, E., Saito, T., Mitsakakis, N., Moshkelgosha, S., Joe, B., Boonstra, K. M., Tikkanen, J. M., 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**, 965–973.

Molnar, T., Borocz, K., Berki, T., Szapary, L., Szolics, A., Janszky, J., Illes, Z., and 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**, 719–727.

Nathwani, R. A., Pais, S., Reynolds, T. B., and Kaplowitz, N. (2005). Serum alanine aminotransferase in skeletal muscle diseases. *Hepatology* **41**, 380–382.

Onakpoya, I. J., Heneghan, C. J., and Aronson, J. K. (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, K. R., and Cronin, M. T. (2012). In silico models for drug-induced liver injury–current status. *Expert Opin. Drug Metab. Toxicol*. **8**, 201–217.

R Development Core Team. (2019). *R: A language And environment For statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Roth, S. E., Avigan, M. I., Bourdet, D., Brott, D., Church, R., Dash, A., Keller, D., Sherratt, P., Watkins, P. B., 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**, 333–346.

Senior, J. R. (2014). Evolution of the Food and Drug Administration approach to liver safety assessment for new drugs: Current status and challenges. *Drug Saf*. **37**, 9–17.

Sgro, C., Clinard, F., Ouazir, K., Chanay, H., Allard, C., Guilleminet, C., Lenoir, C., Lemoine, A., and Hillon, P. (2002). Incidence of drug-induced hepatic injuries: A French population-based study. *Hepatology* **36**, 451–455.

Shen, T., Liu, Y., Shang, J., Xie, Q., Li, J., Yan, M., Xu, J., Niu, J., Liu, J., Watkins, P. B., *et al.* (2019). Incidence and etiology of drug-induced liver injury in Mainland China. *Gastroenterology* **156**, 2230–2241.e2211.

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

Smith, A., Calley, J., Mathur, S., Qian, H. R., Wu, H., Farmen, M., Caiment, F., Bushel, P. R., 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, P. J., Dear, J., Platt, V., Simpson, K. J., Craig, D. G. N., Antoine, D. J., French, N. S., Dhaun, N., Webb, D. J., Costello, E. M., *et al.* (2011). Circulating microRNAs as potential markers of human drug-induced liver injury. *Hepatology* **54**, 1767–1776.

Stravitz, R. T., and Lee, W. M. (2019). Acute liver failure. *Lancet* **394**, 869–881.

Tajima, S., Yamamoto, N., and 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, R. J., Auli, M., Nordahl, G., Antoine, D. J., Starkey Lewis, P., Goldring, C. E., Park, B. K., 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**, 461–469.

Wang, K., Zhang, S., Marzolf, B., Troisch, P., Brightman, A., Hu, Z., Hood, L. E., and Galas, D. J. (2009). Circulating microRNAs, potential biomarkers for drug-induced liver injury. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 4402–4407.

Yang, M. C., Liu, H. K., Su, Y. T., Tsai, C. C., and Wu, J. R. (2019). Serum apoptotic marker m30 is positively correlated with early diastolic dysfunction in adolescent obesity. *PLoS One* **14**, e0217429.

Zhang, Y., Jia, Y., Zheng, R., Guo, Y., Wang, Y., Guo, H., Fei, M., and Sun, S. (2010). Plasma microrna-122 as a biomarker for viral-, alcohol-, and chemical-related hepatic diseases. *Clin. Chem*. **56**, 1830–1838.

Zhu, Y., Zhang, H., Sun, Y., Li, Y., Deng, L., Wen, X., Wang, H., and Zhang, C. (2015). Serum enzyme profiles differentiate five types of muscular dystrophy. *Dis. Markers* **2015**, 1–7.+