

# Clinically important alterations in pharmacogene expression in histologically severe nonalcoholic fatty liver disease

Received: 20 July 2022

Accepted: 7 March 2023

Published online: 17 March 2023

 Check for updates

Nicholas R. Powell<sup>1</sup>, Tiebing Liang<sup>2</sup>, Joseph Ipe<sup>1</sup>, Sha Cao<sup>2</sup>, Todd C. Skaar<sup>1</sup>, Zeruesenay Desta<sup>1</sup>, Hui-Rong Qian<sup>3</sup>, Philip J. Ebert<sup>3</sup>, Yu Chen<sup>3</sup>, Melissa K. Thomas<sup>3</sup> & Naga Chalasani<sup>2</sup>✉

Polypharmacy is common in patients with nonalcoholic fatty liver disease (NAFLD) and previous reports suggest that NAFLD is associated with altered drug disposition. This study aims to determine if patients with NAFLD are at risk for altered drug response by characterizing changes in hepatic mRNA expression of genes mediating drug disposition (pharmacogenes) across the histological NAFLD severity spectrum. We utilize RNA-seq for 93 liver biopsies with histologically staged NAFLD Activity Score (NAS), fibrosis stage, and steatohepatitis (NASH). We identify 37 significant pharmacogene-NAFLD severity associations including *CYP2C19* downregulation. We chose to validate *CYP2C19* due to its actionability in drug prescribing. Meta-analysis of 16 independent studies demonstrate that *CYP2C19* is significantly downregulated to 46% in NASH, to 58% in high NAS, and to 43% in severe fibrosis. Our data demonstrate the downregulation of *CYP2C19* in NAFLD which supports developing personalized medicine approaches for drugs sensitive to metabolism by the CYP2C19 enzyme.

Nonalcoholic Fatty Liver Disease (NAFLD) is one of the most common liver diseases, affecting 25% of the world's population<sup>1</sup>. If untreated, NAFLD progresses to non-alcoholic steatohepatitis (NASH), liver fibrosis, and eventually cirrhosis. While diet, weight-loss, and gastric bypass surgery are effective treatments<sup>2,3</sup>, there are currently no FDA-approved drugs to treat NAFLD. Since NAFLD is highly associated with obesity, diabetes, hypertension, hypercholesterolemia, and cardiovascular disease<sup>4–9</sup>, drugs used to treat cardiometabolic conditions (e.g. clopidogrel) are frequently used in NAFLD patients. However, most of these drugs were not originally studied in NAFLD-specific patient populations. For example, of 3 big trials demonstrating the lifesaving antiplatelet effect of clopidogrel (CURE, CAPRIE, CLARITY-28)<sup>10–12</sup>, none published NAFLD subgroup analyses and one even excluded patients with hepatic insufficiency. Therefore, the approved dosages or other findings from these original drug studies may not be

representative of people with NAFLD. Thus, in this study we aimed to determine if NAFLD patients could be at risk of altered drug response by identifying changes in the hepatic expression of genes that mediate drug disposition (pharmacogenes) across histological NAFLD severity.

Pharmacogenes code for proteins involved in the disposition and response to drugs like anti-hypertensives, lipid-lowering agents, anti-platelets, and agents used to treat diabetes. Many of these drugs are substrates for pharmacogenes like the cytochrome P450 (CYP) metabolic enzymes as well as the ATP-Binding Cassette (ABC) and Solute Carrier (SLC) transporters. Changes in the expression of these genes can impact the response to these drugs via increased or decreased absorption, distribution, metabolism, or excretion (ADME)<sup>13–17</sup>. Knowledge of mRNA expression changes for pharmacogenes has precedence for translating to clinical actionability and adoption into clinical practice, improving successful drug response rates<sup>13–17</sup>.

<sup>1</sup>Indiana University School of Medicine, Department of Medicine, Division of Clinical Pharmacology, Indianapolis, IN, USA. <sup>2</sup>Indiana University School of Medicine, Department of Medicine, Division of Gastroenterology Hepatology, Indianapolis, IN, USA. <sup>3</sup>Eli Lilly and Company, Indianapolis, IN, USA.

✉ e-mail: [nchalasa@iu.edu](mailto:nchalasa@iu.edu)

The clinical importance of pharmacogene expression is also evidenced by 483 FDA drug labels warning of drug-drug interactions and genetic variants that alter the activity of pharmacogenes mediating patient exposure to that drug<sup>18</sup>. Diseases can also alter pharmacogene expression, but have been less studied compared to drug-drug interactions and pharmacogenomics. NAFLD is a disease characterized by the deposition of lipids in liver tissue that is accompanied by significant inflammatory signaling. Inflammation alters pharmacogene expression<sup>19–23</sup> and therefore it is mechanistically plausible that patients with NAFLD would not respond normally to many drugs due to altered pharmacogene-mediated drug disposition.

Existing reports of pharmacogene expression changes in livers of patients with NAFLD is limited to a few studies<sup>24</sup>. Two studies used CYP-targeted real-time PCR<sup>25</sup> or a whole transcriptome array focusing on absorption, distribution, metabolism, or excretion genes<sup>26</sup>, and their findings indicate that CYP enzyme and transporter mRNA expression is altered in steatosis and NASH. Another study found differences in the abundance of certain pharmacogene proteins in NAFLD-cirrhosis, however the sample size was limited ( $n = 9$  cases) and statistically significant conclusions were not made<sup>27</sup>. Other targeted studies found CYP3A4 protein and activity was downregulated and CYP2E1 protein and activity was upregulated in NASH vs. controls<sup>27–29</sup>. Collectively, the findings indicate that certain pharmacogenes may be altered in NASH vs. controls, but pharmacogene expression characterization in other clinically used histological severity markers, like fibrosis and NAFLD activity score (NAS), are still lacking in the literature. Two separate studies reported one pharmacogene, *CYP2C19*, is downregulated in liver fibrosis<sup>30,31</sup> and NAS<sup>31</sup>, however, pharmacogenes were not the focus and the studies did not detail progressive changes in expression across the spectrum of disease. Currently, there are no studies detailing the range of pharmacogene mRNA expression changes that occur over the spectrum of histological disease measures (NAFLD activity score, NASH grade, and fibrosis stage). Therefore, in this study we comprehensively analyzed the changes in pharmacogene expression across the histological severity spectrum in NAFLD. Our study consists of a larger cohort of liver biopsies, utilizes RNA-seq technology, and utilizes linear regression for 3 clinically defined histological measures of NAFLD severity (NAFLD activity score, NASH grade, and fibrosis stage). Addressing this gap in knowledge regarding progressive pharmacogene expression changes in NAS, fibrosis, and NASH has important implications for improving patient care and for NASH drug development.

Focusing on pharmacogenes is advantageous not only by the potential for clinical actionability, but it also increases the likelihood of identifying true positive associations by avoiding transcriptome-wide multiple-testing corrections. Other gene expression studies have been conducted in NAFLD liver samples<sup>25,26,30–48</sup>, but since their focus was transcriptome-wide, they needed to multiply their  $p$ -values by the number of genes tested (often over 20,000) to obtain Bonferroni-corrected significance values. Because of this, many pharmacogene associations remain undiscovered and buried in raw data. Therefore, a second aim of our study was to validate our findings by conducting a meta-analysis in data obtained from 16 studies of transcriptome-wide hepatic gene expression including our own. This 2-step methodologic approach provides a higher level of statistical rigor to demonstrate the magnitude of the validated pharmacogene expression changes in NAFLD. Thus, in this study we determine if NAFLD patients could be at risk of altered drug response due to changes in hepatic expression of genes that mediate drug disposition across histological NAFLD severity.

## Results

We conducted linear regression analyses to identify associations between liver expression of 255 pharmacogenes and NAS, fibrosis stage, and presence of steatohepatitis in 93 individual liver biopsies.

The majority of the pharmacogene mRNA expression values that correlated with the histological phenotypes were unskewed indicating normal distribution. All phenotypes showed good representation across the range of disease severity. Table 1 shows the distribution of these disease phenotypes along with other important clinical and demographic information. In total, we identified 37 pharmacogene-NAFLD severity associations that were statistically significant ( $p \leq 0.05$ ) after Bonferroni correction ( $p$ -value multiplied by 255). We also performed t-tests for disease groups (above or below the disease threshold as described in the methods section) to identify 6 additional genes that were statistically significant after Bonferroni correction. Of the significant differentially expressed genes among the t-tests, 82% were also significant in the linear regressions. We also identified 17 genes that were commonly significant for all 3 disease subgroups after Benjamini-Hochberg multiple-testing correction. *CYP2C19* had the strongest effect size of any pharmacogene.

### NAFLD activity score

The 14 Bonferroni-adjusted statistically significant individual pharmacogene regressions for NAFLD activity score (NAS) are shown in Fig. 1A. The NAS within our cohort ranged from 0–7, within a possible histological range of least severe (0) to most severe (8) disease. The direction of effect was positive (upward) for 9 of the pharmacogenes and negative (downward) for the other 5. The result with the lowest  $p$ -value was a positive correlation for *ABCB4* with an  $R^2$  of 36% (correlation coefficient 0.6). *ABCB8*, *ABCC3*, and *SLC22A12* were also significantly upregulated, while *AOX1* and *SLC16A1* were significantly downregulated with increasing NAS. Table 2 shows the full list and the individual estimates from each regression. To better visualize and compare these results, we plotted the slopes and  $p$ -values for all 255 associations in the volcano plot shown in Fig. 1B. The downregulation of *CYP2C19* and *ABCG2* met the less stringent multiple-comparisons correction (Benjamini-Hochberg). Although not statistically significant, *CYP3A4*, *CYP1A2*, and *CYP2C8* trended downwards. Other major CYP enzymes do not appear to be robustly affected by NAFLD severity. Upregulated genes that met the less stringent threshold are *CES1*, *CES2*, several ABC transporters (*ABCB1*, *ABCC5*, *ABCC4*, *ABCA4*), SLC transporters (*SLC3A1*, *SLC28A1*, *SLC22A11*), *CYP21A2*, and *UGT2B15*. In addition to the regression analysis, we identified significant differences in pharmacogene mRNA expression levels between the most severe disease (NAS 5–8) vs. the less severe disease (NAS 0–4), shown in Fig. 2A. This dichotomous analysis identified the Bonferroni-adjusted significant downregulation of *CYP2C19*.

### Fibrosis stage

We also conducted linear regression analyses to evaluate pharmacogene mRNA expression levels across liver fibrosis stages (ranging from 0–4). There were 24 Bonferroni-adjusted statistically significant associations, with 19 positive, and 5 negative (Fig. 3A). The result with the lowest  $p$ -value was a negative correlation for *GSTZ1* with an  $R^2$  of 33% (correlation coefficient  $-0.57$ ). *CYP2C19* also had a strong negative correlation with an  $R^2$  of 24% (correlation coefficient  $-0.49$ ), and *SLC2A4* was also significantly downregulated. *CYP1B1*, *ABCC1*, *ABCC4*, *SLC3A1*, *SLC6A6*, and *SLC22A17* were significantly upregulated. Table 2 shows the full list and the individual estimates from each regression. To better visualize and compare these results, we plotted the slopes and  $p$ -values for all 255 associations in the volcano plot shown in Fig. 3B. The downregulation of *CYP2C8*, *SLCO1B3*, *SLC22A1*, *SLC10A1*, *CYP4F2*, *CYP2J2*, *AOX1*, and *ABCG2* met the less stringent multiple-comparisons correction (Benjamini-Hochberg). *CYP1A2* and *CYP3A4* trended downward (consistent with the NAS analysis) but were not statistically significant. The upregulation of several minor CYP enzymes (*CYP3A7*, *CYP11A1*, *CYP21A2*, *CYP27B1*, *CYP2R1*), ABC transporters (*ABCB1*, *ABCC10*, *ABCC5*, *ABCB4*, and *ABCC3*), and SLC





**Table 2 | Bonferroni-adjusted significant linear regression correlation estimates between disease sub-groups and pharmacogenes**

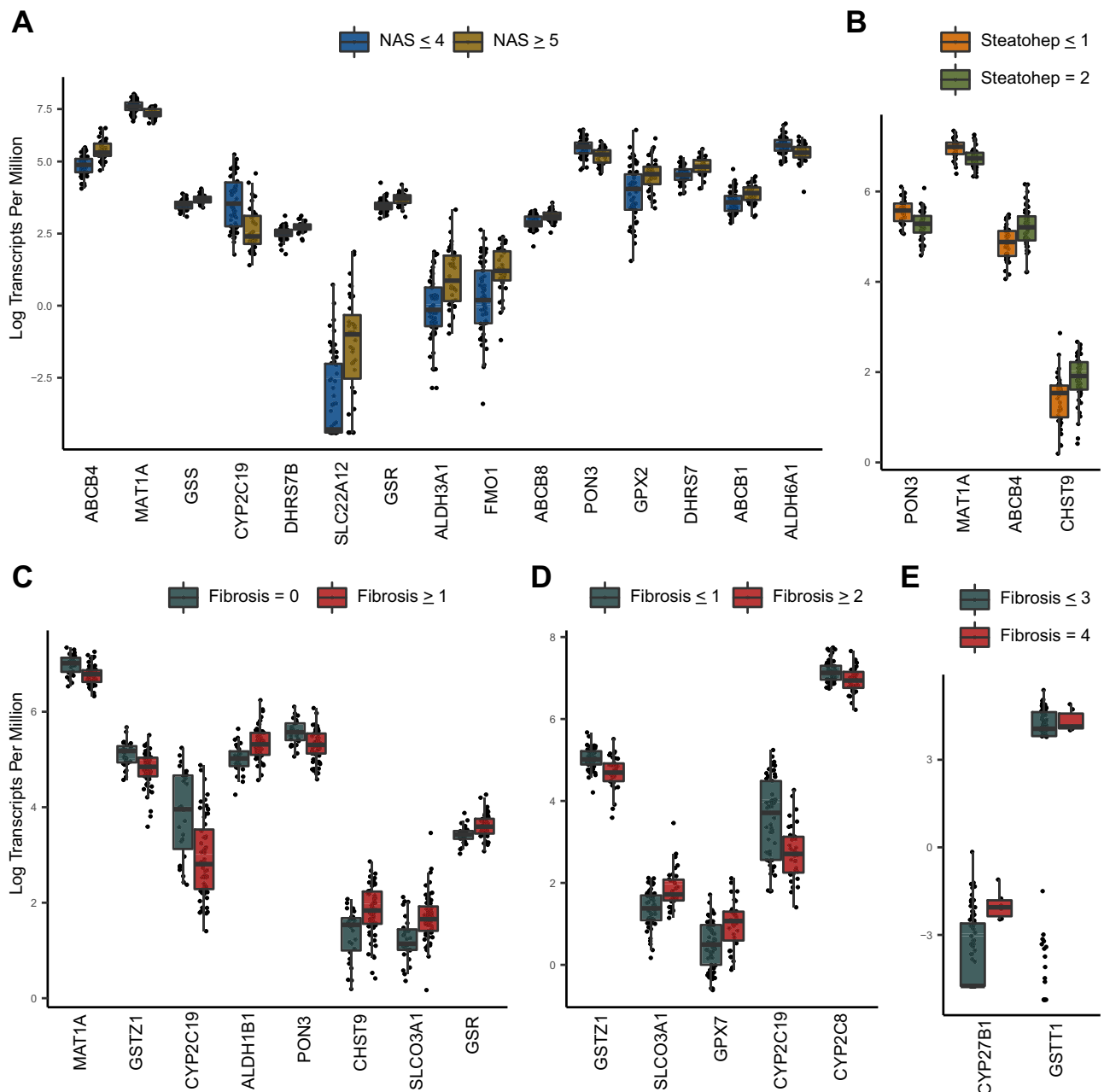
Gene	Slope	Unadjusted p value	R-squared	Adjusted p value (Bonferroni)	Disease sub-group
ABCB4	0.14	1.77E-10	0.36	4.51E-08	NAS
GSS	0.05	6.46E-08	0.28	1.65E-05	NAS
SLC22A12	0.52	7.92E-08	0.27	2.02E-05	NAS
MAT1A	-0.06	2.11E-06	0.22	5.37E-04	NAS
PON3	-0.08	2.94E-06	0.21	7.50E-04	NAS
DHRS7B	0.05	4.64E-06	0.21	1.18E-03	NAS
FMO1	0.28	2.22E-05	0.18	5.67E-03	NAS
ALDH6A1	-0.08	2.92E-05	0.18	7.45E-03	NAS
ABCB8	0.06	4.55E-05	0.17	1.16E-02	NAS
GSR	0.05	5.89E-05	0.16	1.50E-02	NAS
DHRS9	0.17	1.10E-04	0.15	2.80E-02	NAS
SLC16A1	-0.10	1.19E-04	0.15	3.04E-02	NAS
AOX1	-0.06	1.43E-04	0.15	3.66E-02	NAS
ABCC3	0.07	1.85E-04	0.14	4.72E-02	NAS
GSTZ1	-0.17	5.51E-09	0.33	1.40E-06	Fibrosis
SLCO3A1	0.24	7.34E-09	0.33	1.87E-06	Fibrosis
ABCC4	0.27	3.33E-07	0.27	8.50E-05	Fibrosis
MAT1A	-0.09	3.76E-07	0.26	9.60E-05	Fibrosis
GSTP1	0.20	7.95E-07	0.25	2.03E-04	Fibrosis
CYP2C19	-0.37	1.90E-06	0.24	4.85E-04	Fibrosis
CHST9	0.22	2.23E-06	0.23	5.68E-04	Fibrosis
GPX7	0.24	2.96E-06	0.23	7.54E-04	Fibrosis
SLC28A3	0.35	3.28E-06	0.23	8.35E-04	Fibrosis
SOD3	0.36	5.62E-06	0.22	1.43E-03	Fibrosis
CFTR	0.40	9.04E-06	0.21	2.31E-03	Fibrosis
PDE3A	0.24	1.56E-05	0.20	3.98E-03	Fibrosis
SLC2A4	-0.24	1.67E-05	0.20	4.26E-03	Fibrosis
DHRS7B	0.08	2.26E-05	0.19	5.75E-03	Fibrosis
ABCC1	0.22	2.71E-05	0.19	6.92E-03	Fibrosis
ALDH3B1	0.14	2.86E-05	0.19	7.28E-03	Fibrosis
CYP1B1	0.15	4.35E-05	0.18	1.11E-02	Fibrosis
SLC22A17	0.19	6.17E-05	0.17	1.57E-02	Fibrosis
CHST4	0.33	6.28E-05	0.17	1.60E-02	Fibrosis
HSD17B14	-0.26	6.45E-05	0.17	1.65E-02	Fibrosis
SLC6A6	0.23	7.07E-05	0.17	1.80E-02	Fibrosis
CHST10	0.27	1.29E-04	0.16	3.30E-02	Fibrosis
CHST3	0.20	1.35E-04	0.16	3.45E-02	Fibrosis
ALDH1A3	0.23	1.40E-04	0.16	3.58E-02	Fibrosis
MAT1A	-0.14	2.28E-07	0.27	5.81E-05	Steatohepatitis
CHST9	0.31	4.18E-06	0.22	1.07E-03	Steatohepatitis
CYP2C19	-0.50	5.98E-06	0.21	1.53E-03	Steatohepatitis
PON3	-0.16	3.44E-05	0.18	8.77E-03	Steatohepatitis
GSS	0.09	5.04E-05	0.17	1.29E-02	Steatohepatitis
ABCB4	0.21	5.78E-05	0.17	1.47E-02	Steatohepatitis
SLCO3A1	0.24	1.34E-04	0.16	3.42E-02	Steatohepatitis
AOX1	-0.13	1.70E-04	0.15	4.35E-02	Steatohepatitis

NAS Non-Alcoholic Fatty Liver Disease Activity Score.

percent fold change values for the 17 common pharmacogenes (numerical data provided in Table 3). *CYP2C19* showed the largest downregulation that was significant among all three disease measures. This analysis shows that with every 1 unit increase in fibrosis stage, *CYP2C19* transcript abundance decreases to 69% that of the prior stage; reducing *CYP2C19* mRNA expression levels by 77% in individuals with stage 4 fibrosis compared to 0. With every 1 unit increase in NAS,

*CYP2C19* transcript abundance decreases to 83% that of the prior score; reducing *CYP2C19* mRNA expression levels by 73% in individuals with an NAS of 7 compared to 0. As an individual progresses from no-NASH to borderline NASH to definite NASH, *CYP2C19* transcript abundance decreases to 61% that of the prior grade; reducing *CYP2C19* mRNA expression levels by 63% in individuals with NASH compared to those without NASH.





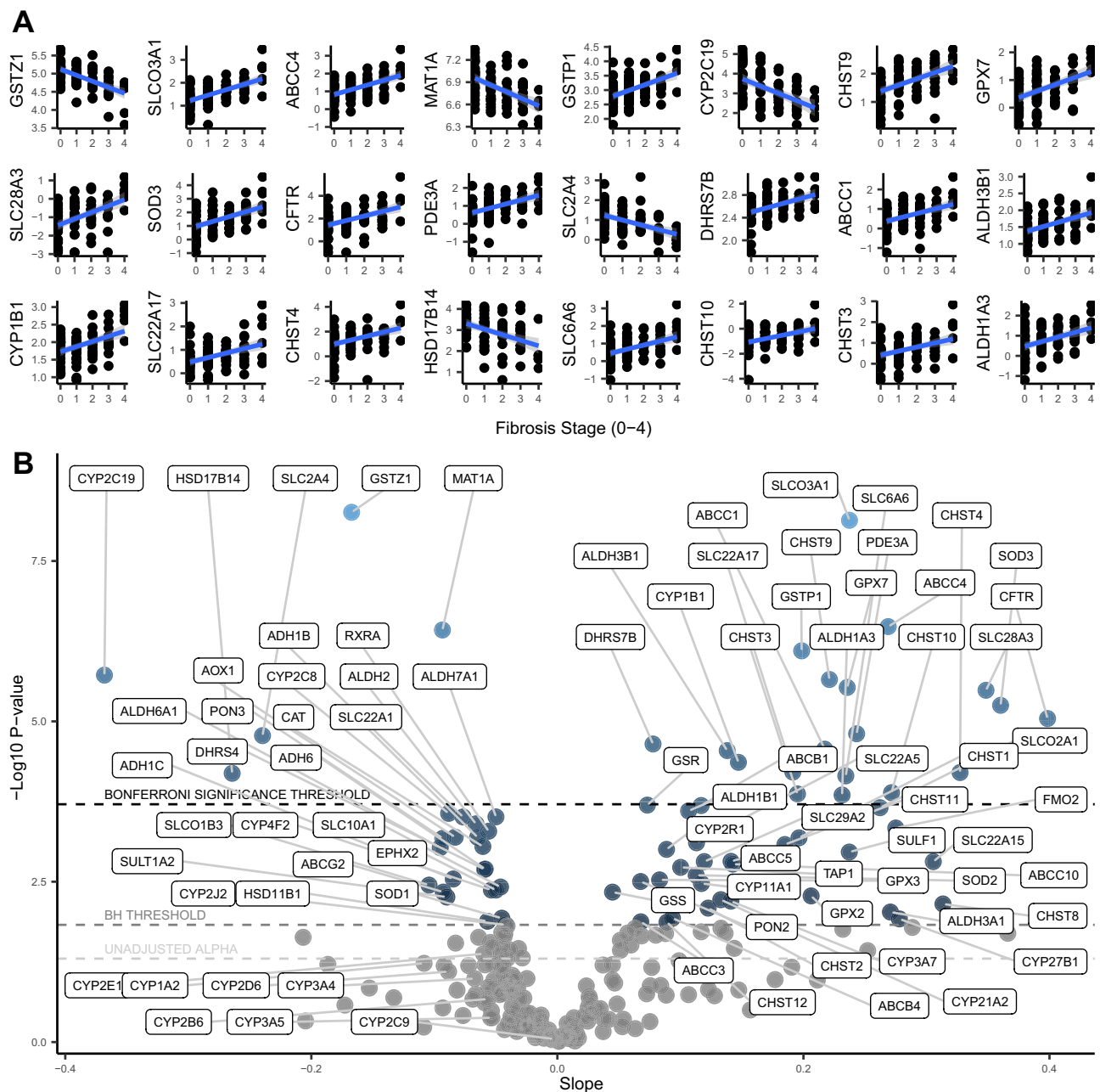
**Fig. 2 | Boxplots showing Bonferroni-adjusted significant t-test results.** Differences are based on two-sided Welch's t-tests. The x-axis is ordered from left to right by  $p$ -value. All  $p$ -values in this figure are below 0.0002. Standard boxplot features (lower quartile, median, upper quartile) are used as defined by ggplot2

version 3.3.6. **A** Expression differences in pharmacogenes between NAS groups ( $n = 93$ ). **B** Expression differences in pharmacogenes between steatohepatitis groups ( $n = 89$ ). **C–E** Expression differences in pharmacogenes between fibrosis groups ( $n = 87$ ). Source data are provided in the supplemental materials.

### Covariate analysis

To determine if the *CYP2C19* downregulation could be better explained by something other than NAFLD, we conducted single linear regressions of *CYP2C19* mRNA abundance with each possible independent variable in our clinical and demographic data (Supplementary Data 12). We found that only the histological markers of NAFLD (fibrosis stage, steatohepatitis, hepatocyte ballooning, and NAS) were statistically significant after Bonferroni multiple corrections (Supplementary Data 13). However, several other factors like AST, ALT, age, metformin use, and diabetes, were possibly associated (uncorrected  $p$ -value  $< 0.05$ ). To determine if these factors, and sex, could have influenced our results, we performed linear regressions between *CYP2C19* mRNA abundance and each NAFLD phenotype with and without correcting for each of these respective covariates. The change in *CYP2C19* slope

before and after covariate correction was measured, and no covariate affected the *CYP2C19* slope by more than 16%, indicating the robustness of the association between the histological NAFLD phenotypes and *CYP2C19* (Supplementary Data 14). This information collectively suggests that metabolic comorbidities like obesity and diabetes, which are highly correlated with NAFLD, are not the primary drivers of the observed *CYP2C19* downregulation. However, it is likely that these comorbidities contribute additional information that, when combined, can improve the characterization of *CYP2C19* downregulation. To provide evidence towards the most important features underlying the *CYP2C19* downregulation, we used a backward elimination approach, narrowing to only remaining coefficients with  $p$ -values  $< 0.05$  in a multiple linear regression model. From this analysis, the following factors explain 39% of the variability in *CYP2C19* mRNA abundance:



**Fig. 3 | Regression results for fibrosis stage ( $n = 87$ ).** **A** Regression plots of the Bonferroni-adjusted significant correlations between fibrosis stage and pharmacogene mRNA expression. Linear trendlines were fit (blue line) with 95% confidence interval region (grey). **B** Volcano plot showing the slopes and  $p$ -values for the

regressions between fibrosis stage and pharmacogene mRNA expression. Horizontal dotted lines correspond to the Bonferroni significance threshold, Benjamini Hochberg (BH) significance threshold, and uncorrected alpha of 0.05, from top to bottom, respectively. Source data are provided in the supplemental materials.

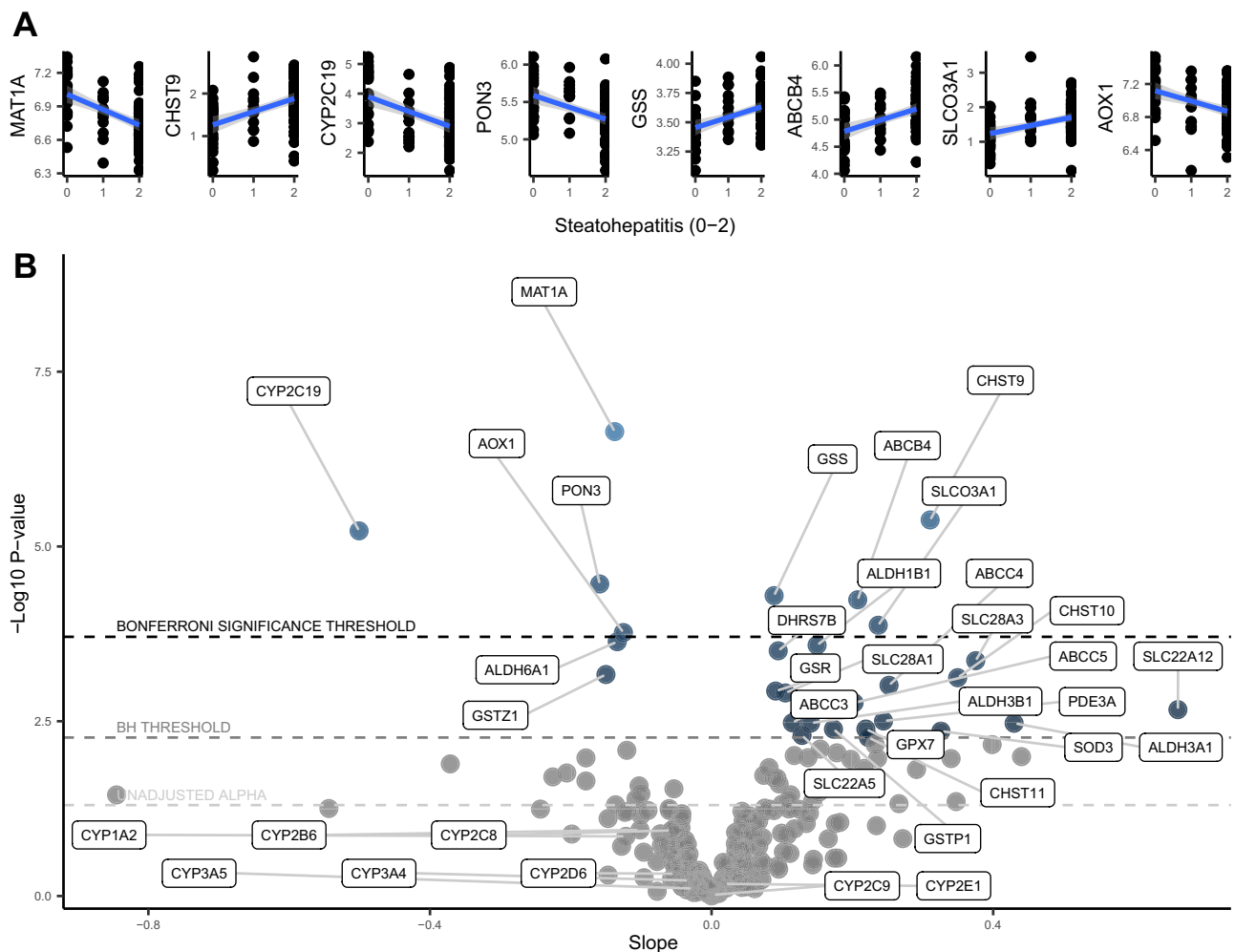
fibrosis stage, age, metformin use, and ALT (Supplementary Data 18). This model is of limited actionability until it is optimized and validated, but it shows that the additional factors increase the variability explained from 24% (fibrosis only) to 39% (after adding age, metformin use, and ALT).

Importantly, NAS and steatohepatitis were dropped from the multiple regression model, suggesting worsening fibrosis as possibly the most important histological NAFLD feature in describing *CYP2C19* downregulation. We tested this further by re-running the regressions of NAS and steatohepatitis in each fibrosis stage separately (results provided in Supplementary Data 3 and 6), showing that these factors are not associated with *CYP2C19* expression outside of the context of worsening fibrosis. An additional multiple regression model was tested by including just fibrosis stage, NAS, and steatohepatitis together.

These results (Supplementary Data 19) show that only fibrosis stage and steatohepatitis diagnosis contribute a meaningful effect size in describing the expression of *CYP2C19*. Further, only fibrosis stage was statistically significant in this analysis.

#### NAFLD-*CYP2C19* meta analysis

Because *CYP2C19* was consistently downregulated across three disease severity phenotypes in our study, and due to its actionability in personalized therapy for *CYP2C19* substrate drugs like clopidogrel, we conducted a meta-analysis using 16 studies that measured *CYP2C19* expression in NAFLD. Figure 6 demonstrates that *CYP2C19* is consistently downregulated in 15 of 16 studies, validating our findings. The meta-analysis model indicates a log<sub>2</sub> fold change of  $-1.13$  in NASH vs. control (Fig. 6A), a value translating to an expression decrease to 46%.



**Fig. 4 | Regression results for steatohepatitis grade ( $n = 89$ ).** **A** Regression plots of the Bonferroni-adjusted significant correlations between fibrosis grade and pharmacogene mRNA expression. Linear trendlines were fit (blue line) with 95% confidence interval region (grey). **B** Volcano plot showing the slopes and  $p$  values for the regressions between steatohepatitis grade and pharmacogene mRNA

expression. Horizontal dotted lines correspond to the Bonferroni significance threshold, Benjamini Hochberg (BH) significance threshold, and uncorrected alpha of 0.05, from top to bottom, respectively. Source data are provided in the supplemental materials.

Fibrosis stage 3–4 vs. 0–1 indicates a log<sub>2</sub> fold change of  $-1.22$  (Fig. 6B), translating to an expression decrease to 43%. NAS 5–8 vs. 0–4 indicates a log<sub>2</sub> fold change of  $-0.79$  (Fig. 6C), translating to an expression decrease to 58%. These meta-analysis values validate our findings.

## Discussion

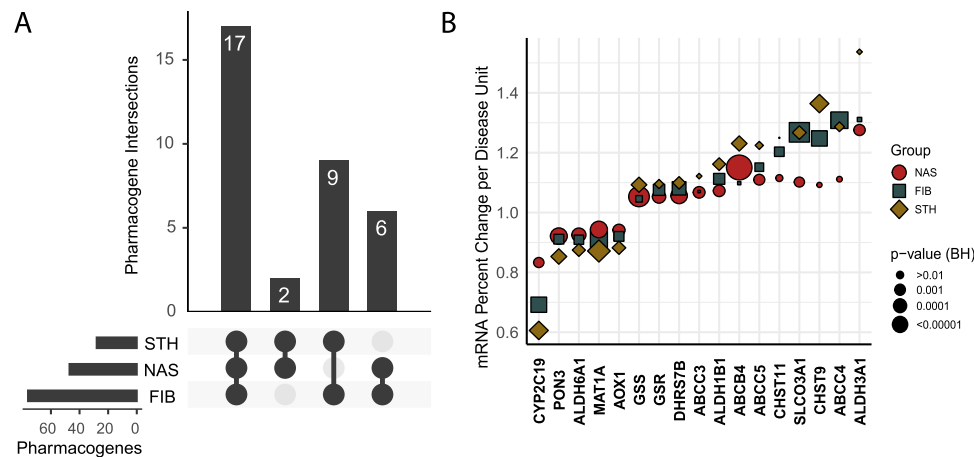
In this study, we screened for altered pharmacogene expression across 93 individual liver biopsies obtained from participants with varying histological stages of NAFLD severity. Our results clearly demonstrate that expression levels of several pharmacogenes are significantly associated with the histological severity of NAFLD. Notably, *CYP2C19* was severely downregulated. By conducting a meta-analysis in 16 similar studies, we validate that *CYP2C19* is strongly downregulated in NASH and advanced fibrosis.

Our study used the current gold standard RNA-seq technology; however, it does have some potential limitations. RNA-seq more readily measures high abundance transcripts, therefore the percent fold changes we report could be affected by this bias. However, since our fold change analyses compare the same gene with similar order-of-magnitude expression levels, this measurement bias is not a major concern. Another potential limitation is the conservative approach taken with the Bonferroni multiple-testing correction. Bonferroni unduly penalizes the statistical significance of results when factors

between tests are not completely independent. Since changes in the expression of one gene could have downstream effects on another genes expression, these changes are not all necessarily independent and therefore the false discovery rate (FDR) Benjamini-Hochberg correction can also be appropriate when interpreting results. A limitation of our meta-analysis is that it was not based on a systematic review for which a protocol was published (rather a narrative review). We also did not conduct sensitivity analyses. Despite these limitations, the meta-analysis accomplished our goal of assessing the repeatability and inferring the magnitude of *CYP2C19* change across a variety of other NAFLD cohorts.

Medications such as rifampin, efavirenz, and ritonavir can induce CYP enzyme expression, including *CYP2C19*<sup>51</sup>. Our analyses would likely be improved by accounting for patients taking these medications, however, this information is not available for our cohort. Despite this limitation, it is unlikely that our findings of *CYP2C19* downregulation could be attributed to medication usage unless for some reason healthier patients tended to take more rifampin, efavirenz, or ritonavir. We do not have reason to suspect this in our patient population nor in the other 15 studies in the meta-analysis. Another limitation of our study is that we did not consider the effect of genotype on mRNA abundance. Expression of functional *CYP2C19* is increased by the *CYP2C19\*17* allele and decreased by *CYP2C19\*2* and *\*3* variants<sup>52–54</sup>.





**Fig. 5 | Comparison of the Benjamini-Hochberg significant pharmacogene expression changes between disease subgroups. A** Upset plot showing the commonality in significantly changed pharmacogenes between disease subgroups. **B** Scatter plot showing mRNA percent change per unit of disease for the 17 pharmacogenes in common between all disease subgroups. For example, the plot shows that with each 1 unit increase in fibrosis stage, *CYP2C19* transcript abundance

decreases to 69% of the expression level at the prior stage. For each 1 unit increase in NAFLD activity score, *CYP2C19* transcript abundance decreases to 83% of the expression level at the prior score. For each 1 unit increase in steatohepatitis grade, *CYP2C19* transcript abundance decreases to 61% of the expression level at the prior grade. FIB fibrosis, NAS NAFLD activity score, STH steatohepatitis, BH Benjamini-Hochberg. Source data are provided in the supplemental materials.

**Table 3 | Regression estimates for the 17 pharmacogenes commonly changed in all 3 disease subgroups, converted to percent fold change per disease unit**

Gene	Fibrosis FC per disease unit	Fibrosis BH <i>p</i> value	NAS FC per disease unit	NAS BH <i>p</i> value	Steatohepatitis FC per disease unit	Steatohepatitis BH <i>p</i> value
CYP2C19	0.69	8.08E-05	0.83	6.73E-03	0.61	5.08E-04
ALDH6A1	0.91	5.64E-03	0.93	9.31E-04	0.87	6.53E-03
MAT1A	0.91	2.40E-05	0.94	1.34E-04	0.87	5.81E-05
PON3	0.91	4.47E-03	0.92	1.50E-04	0.85	2.19E-03
AOX1	0.92	4.47E-03	0.94	2.81E-03	0.88	5.43E-03
GSS	1.05	1.89E-02	1.05	6.74E-06	1.09	2.46E-03
ABCC3	1.07	4.60E-02	1.07	3.37E-03	1.12	3.74E-02
GSR	1.08	2.01E-03	1.05	1.50E-03	1.10	1.86E-02
DHRS7B	1.08	4.11E-04	1.06	1.97E-04	1.10	7.20E-03
ABCB4	1.10	4.21E-02	1.15	4.51E-08	1.23	2.46E-03
ALDH1B1	1.11	2.27E-03	1.07	3.46E-03	1.16	6.58E-03
ABCC5	1.15	8.33E-03	1.11	4.91E-03	1.22	2.44E-02
CHST11	1.20	5.30E-03	1.12	2.13E-02	1.25	4.92E-02
CHST9	1.25	8.11E-05	1.09	3.31E-02	1.36	5.08E-04
SLCO3A1	1.27	9.35E-07	1.10	7.01E-03	1.27	4.88E-03
ABCC4	1.31	2.40E-05	1.11	2.99E-02	1.29	1.65E-02
ALDH3A1	1.31	3.47E-02	1.28	4.18E-03	1.54	3.74E-02

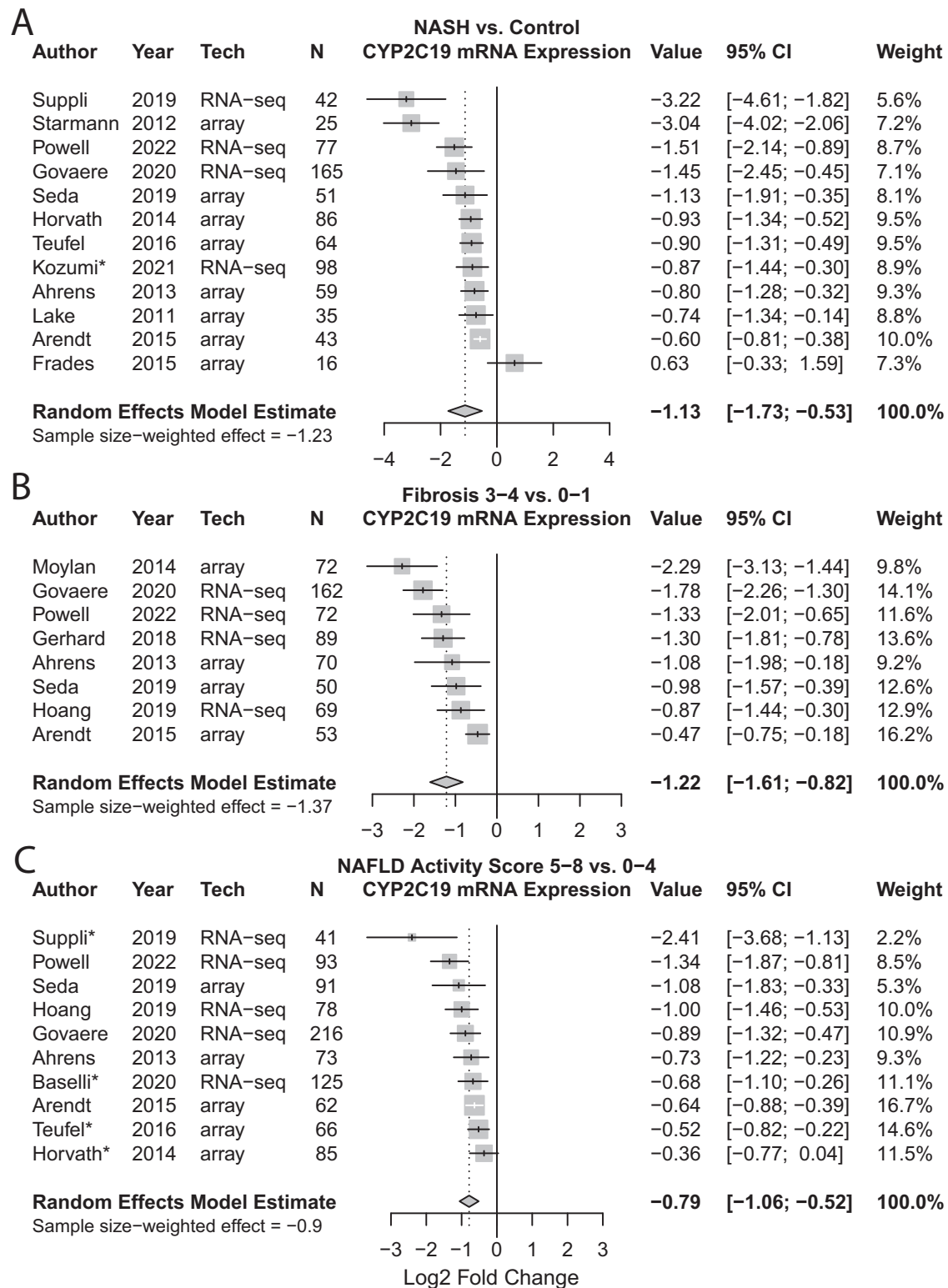
FC fold change, NAS NAFLD activity score, BH Benjamini Hochberg.

The expression changes we found for *CYP2C19* and other pharmacogenes would likely have been more precisely characterized by including known loss-of-function genotypes in the analysis, but since genotypes are unlikely to cluster in severe NAFLD patients, this is not a major limitation. We also recognize there are other ways to classify steatohepatitis. Therefore, we also performed pharmacogene regressions for NASH categorized by the SAF (steatosis, activity, fibrosis) method<sup>55</sup>, and found the slopes were highly correlated with our steatohepatitis results using the CRN method (Pearson correlation coefficient 0.88, Supplementary Data 9).

Another potential limitation of our study is that *CYP2C19* mRNA expression levels may not necessarily correlate with *CYP2C19* enzyme abundance. However, Fisher et al. demonstrated that liver *CYP2C19* mRNA downregulation in NASH ( $p=0.193$ ) corresponded to a significant decrease in liver *CYP2C19* enzyme abundance ( $p=0.01$ )<sup>25</sup>,

indicating *CYP2C19* mRNA is a good surrogate measure for *CYP2C19* enzyme abundance. Additionally, the Human Protein Atlas demonstrates corresponding high levels of mRNA and protein expression of *CYP2C19* in the liver, further supporting mRNA as a good surrogate measure for enzyme abundance<sup>56</sup>. For certain *CYP2C19* genotypes, reduced mRNA expression is the mechanism behind the altered clinical effect of drugs that are *CYP2C19* substrates, providing further evidence<sup>57</sup>. Lastly, NAFLD was recently correlated with 60% lower omeprazole metabolism (a *CYP2C19* substrate)<sup>3</sup>, and a 28% decrease in *CYP2C19* protein<sup>58</sup>, which is broadly consistent with the expected effect of decreased *CYP2C19* mRNA.

Our results have translational significance, as *CYP2C19* is the key enzyme in the bioactivation of the clopidogrel prodrug to its active metabolite, and dysfunction of this activation pathway is known to significantly impair response to clopidogrel<sup>59–61</sup>. Pharmacogenomic



**Fig. 6 | Forest plots demonstrating the meta-analysis effect of NAFLD on CYP2C19 mRNA expression.** Random effects model estimates, weighted based on inverse variance, are provided in bold. Sample size-weighted estimates (effects) are also provided. **A** NASH vs. no-NASH control. Note, studies marked with an asterisk (\*) did not have a healthy control group, therefore subjects labelled “NAFL” were used as the comparator. **B** Fibrosis 3-4 vs. 0-1. **C** NAS 5-8 vs. 0-4. Note, for studies

marked with an asterisk (\*) NAS was not explicitly given therefore the following analyses are included instead: Suppli et al, Teufel et al, and Horvath et al subjects labelled “NAFLD” were compared to healthy controls, Baselli et al subjects labelled “severe NAFLD” were compared to “mild-no NAFLD”. Source data for our study are provided in the supplemental materials and data from other studies is available in GEO.

studies have shown that when correctly transcribed *CYP2C19* is halved due to heterozygosity of the *CYP2C19\*2* allele, the bioactivation of clopidogrel is significantly reduced and the antiplatelet efficacy is also significantly decreased<sup>13,17,53</sup>. Because of this, clinical pharmacogenomic guidelines recommend choosing a different antiplatelet drug for *CYP2C19\*2* carriers<sup>13,17</sup>. Our data demonstrate that NAFLD can have a similar magnitude of effect on *CYP2C19* expression as compared to *CYP2C19\*2*. Even a moderate progression of NAFLD (i.e. NAS from 0 to 4, fibrosis stage from 0 to 2, or steatohepatitis grade 0 to 1) results in a decrease of *CYP2C19* to around 50%. Decreased *CYP2C19* abundance is especially relevant to the NAFLD patient population due to their cardiovascular comorbidities and therefore the increased importance of antiplatelet therapy. Based on our findings of decreased *CYP2C19* expression and the clinical actionability of *CYP2C19\*2*, it is logical to suggest that NAFLD patients are at increased risk for clopidogrel treatment failure and therefore more cardiovascular morbidity. Upon subsequent literature review, we identified a study that demonstrated hepatosteatosis was significantly associated with a lack of clopidogrel anti-platelet effect, further supporting our hypothesis that *CYP2C19* downregulation puts NAFLD patients at increased risk for clopidogrel treatment failure<sup>62</sup>. Further translational research into this area will need to keep in mind that patients with NAFLD can have altered platelet homeostasis, and changing to more potent antiplatelet therapy (i.e., the alternatives to clopidogrel) may contribute to a higher risk of bleeding and will need to be weighed against the risk of low clopidogrel efficacy.

It is especially important to consider special populations like NAFLD because large randomized controlled clinical drug trials are not designed to conclude efficacy and safety for subgroups. Among four big trials studying clopidogrel use (CURE, CAPRIE, CHARISMA, and CLARITY-28)<sup>10–12,63</sup>, none conducted NAFLD subgroup analyses and one of these trials even excluded patients with hepatic insufficiency. Therefore, the approved dosages, efficacy, or safety findings from these pivotal studies may not be representative of people with NAFLD. However, these trials did conduct other subgroup analyses. Both CURE and CHARISMA found diabetes to trend in the direction of less clopidogrel efficacy, though not statistically significant. Additionally, CHARISMA found that obesity, hypertension, and hypercholesterolemia showed similar trends toward lower clopidogrel efficacy, though also not statistically significant. In fact, several studies have shown that patients with diabetes have higher on-clopidogrel platelet reactivity and lower clopidogrel active metabolite compared to controls<sup>64–66</sup>. Diabetes and cardiometabolic diseases are highly associated with NAFLD<sup>5–9</sup> and therefore these data further support the hypothesis that NAFLD is involved in reduced clopidogrel efficacy through the downregulation of *CYP2C19*.

Besides clopidogrel, *CYP2C19* plays a major role in the metabolism of several other drugs including those with narrow therapeutic range (e.g. diazepam, phenytoin, voriconazole, carisoprodol, omeprazole, citalopram, pentamidine, thalidomide, and others)<sup>57</sup>. Thus, our data provide an opportunity to consider personalized treatment of all *CYP2C19* substrates in NAFLD patients. To make this a reality, however, the clinical biomarker will need to be carefully chosen. Our results show that fibrosis stage is the strongest association with *CYP2C19* mRNA downregulation. It is possible that hepatic necroinflammation, as measured by NAS or steatohepatitis, is associated with *CYP2C19* downregulation because it accompanies worsening fibrosis. In contrast, there may still be independent effects of necroinflammation that could not be detangled analytically due to the collinearity with fibrosis. This will need to be further studied for mechanistic conclusions to be reached. We can infer, however, that the mechanism of *CYP2C19* downregulation is not due to transcriptome-wide decreases in expression because there were more genes that were upregulated with worsening fibrosis.

Our results indicate that exposure or pharmacodynamics of drugs that are substrates for other pharmacogenes may be altered in patients

with histologically severe NAFLD. Methionine adenosyltransferase 1A (*MAT1A*) was a robustly downregulated pharmacogene in our data and across other studies<sup>30,33,34,45</sup>. Overexpression of this gene in bladder cancer tumor xenografts has been shown to confer tolerance to gemcitabine<sup>67</sup> suggesting that NAFLD patients could have increased liver toxicity when treated with gemcitabine due to a decreased abundance of *MAT1A*. Aldehyde oxidase 1 (*AOX1*) was another consistently downregulated pharmacogene in our data and across the other studies<sup>26,32–34</sup>. The role of aldehyde oxidase enzyme coded by the *AOX1* gene in human drug metabolism is emerging. Substrates for this phase I metabolic enzyme are many, including clonazepam, nifedipine, and ziprasidone<sup>68,69</sup>. Ziprasidone elimination relies heavily on aldehyde oxidase, therefore it is possible that NAFLD patients would exhibit reduced clearance of ziprasidone due to less *AOX1* expression.

Glutathione s-transferase pi 1 (*GSTP1*) was robustly upregulated in our data and others<sup>26,32,34,45</sup>. *GSTP1* codes for a glutathione s-transferase enzyme that catalyzes the conjugation of polar glutathione groups to enhance systemic elimination of chemotherapeutic agents and toxic metabolites. This gene is the subject of much research due to its variety of roles, one of which is promotion of chemotherapy resistance<sup>70,71</sup>. It is likely that the phenotype of *GSTP1* overexpression in NAFLD patients is multifaceted, but could involve chemotherapeutic resistance in hepatic tumors, and conversely protection from hepatotoxicity for noncancerous liver tissue.

The multidrug resistance-associated protein genes *ABCC3*, *ABCC4*, *ABCC5* were robustly upregulated in our data, agreeing with previous findings at both mRNA and protein levels<sup>24,27,72</sup>. This suggests that chemotherapy-resistance in hepatocellular carcinoma is likely to be a much larger barrier to effective treatment for NAFLD patients. Another hypothesis arises from these data, that chemotherapy resistance in HCC driven by NAFLD will exhibit intrinsic mechanisms of chemotherapy resistance. While a comprehensive review of NAFLD-associated pharmacogenes is out of scope, we highlighted several examples in which NAFLD patients could be at risk for drug failure based on altered pharmacogene expression. These data-driven hypotheses are especially strong when there is strong evidence of the pharmacogenes role in that drugs response, and when findings are repeatable across studies, as is shown in our meta-analysis.

While large, dedicated studies of drug metabolism in every patient disease group (like NAFLD) could create more personalized treatment regimens, such studies are often not practically feasible. Our analyses address this problem by providing effect sizes for each of our regression estimates and detail the percent change occurring with each step in disease grade or stage. These estimates will support the development of NAFLD-specific treatment decisions. Our regression estimates will also allow physiologically-based pharmacokinetic (PBPK) models to be developed to optimize dosing of drugs with established exposure-efficacy relationships. Our study not only characterizes pharmacogene changes in NAFLD, but also provides strong validation evidence, by meta-analysis, that *CYP2C19* is downregulated in NAFLD. Our goal in clinical pharmacology is to customize disease treatments based on specific characteristics of the individual patient. To that end, our study provides evidence that NAFLD patients have unique pharmacogene expression profiles across the disease spectrum, and these data will contribute to developing more individualized treatments for NAFLD patients.

## Methods

This research was conducted ethically and in accordance with research protocols approved by the Indiana University Institutional Review Board (protocol numbers: 1506218127, 1011003025R008).

Our 2-step data analysis approach used (1) pharmacogene-NAFLD regressions in our cohort of patients with NAFLD and (2) meta-analysis specifically focusing on our identified gene(s) with the strongest signal and the most potential for developing clinical interventions.

### Patient selection

This study was undertaken on liver samples obtained from 93 patients with well-characterized NAFLD who underwent percutaneous liver biopsies for their clinical care. Liver specimens were submitted to the pathology laboratory for histological assessment via hematoxylin-eosin (HE) and Masson's trichrome staining and also snap frozen and stored at  $-80\text{ C}$  as part of a research protocol that was approved by the Indiana University Institutional Review Board. All participants signed an informed consent for the liver biopsy procedure by the radiology proceduralists and a separate informed consent for saving a liver specimen for future research purposes.

### Histological assessment

The histological grades of steatosis, inflammation, ballooning, fibrosis stage, NAFLD activity score (NAS), and presence of steatohepatitis (NASH) were assessed across all biopsy samples in a blinded fashion by a single experienced hepatopathologist using the validated NASH Clinical Research Network scoring system<sup>73</sup>.

### RNA Isolation

BioChain Broad Range Total RNA isolation kits were used to isolate RNA (BioChain Institute Inc. Newark, CA). RNA quantity and quality were measured using Nanodrop and Experion RNA StdSens and HighSense analysis kits (BioRad, Hercules, California). RNA samples with Agilent Bioanalyzer RIN  $> 6.0$  were advanced for RNA library generation.

### RNA-seq

RNA libraries were constructed at Covance Genomics Laboratories with Illumina HT Truseq Stranded kits. 93 unique sample libraries, each from one unique patient, proceeded to RNA sequencing. Samples were pooled in groups of 16 with each pool run on the Illumina HiSeq 2500 platform in 4 lanes. Samples were balanced across pools based on distribution of fibrosis stage, NAS score, date of biopsy, age, BMI, date of RNA isolation, RNA quality indicator (RQI), and RNA yield.

### RNA-seq data processing

Raw data QC was performed, including base composition quality, fragment size, mapping read count, 3' bias, read count mapping breakdown, adapter/phiX content, heterologous organism contamination, and sex/ethnicity prediction concordance. Greater than 100 million reads were obtained per sample and reads were mapped to a reference human transcriptome. The ratio between the highest and lowest total mapped reads per sample was 4.2, indicating good read depth consistency across samples. Each sample (or library) was assessed in 4 or 5 replicates (assays). Data correlation of replicates from the same sample were checked. One replicate from all 434 assays was excluded due to low correlation. Replicates were aggregated into one measurement value per sample per gene. There were 93 unique patients in the final cohort with both phenotype information and RNA-seq data.

The read counts for RNA-seq were normalized to the total read counts across all genes for each sample, multiplied by 1,000,000 and  $\log_e$ -adjusted (i.e. transcripts per million, TPM). The TPM method is a widely used normalization procedure that effectively corrects for batch differences in read depth without normalizing away, or dampening, the biological signal<sup>74</sup>. Read counts of zero were assigned the lowest nonzero read count for any gene in each sample to enable inclusion of that data upon log transformation. We used R version 4.0.4 for this data processing.

### Data analyses and statistical tests

The RNA-seq and phenotype data were analyzed using scripts which are provided in the Supplemental documents, and broad descriptions of the analyses are provided here. We used R (version 3.6.0, 4.0.2, and

4.0.4) for all analyses. R code was used to conduct the linear regressions and the t-tests. T-tests were two-sided and based on unequal variances. Bonferroni corrections were applied to  $p$ -values by multiplying by the number of association tests performed, and Benjamini-Hochberg corrections were conducted using the "p-adjust" function in base R. Pharmacogenes were filtered based on a list of 298 pharmacogenes as described by [www.pharmaadme.org](http://www.pharmaadme.org), accessed in 2021. This list is based on input from seven major pharmaceutical companies as to which genes perform or regulate drug metabolism or transport. From this list, 255 pharmacogenes were present in the data. A list of these genes and other raw data can be found in the Supplemental documents. Missing phenotype grades/stages were removed from the respective analysis, resulting in samples sizes of 93 for NAS, 87 for fibrosis, and 89 for steatohepatitis.

For linear regressions, the histological grades or stages of disease were used as independent variables on a continuous numerical scale for testing the association with each pharmacogene. Each linear regression test consisted of one pharmacogene mRNA level as the dependent variable and one NAFLD histological measure as the independent variable. Fibrosis was staged between stage 0 (no fibrosis) through stage 4 (cirrhosis). NAFLD activity score was scored between 0 through 8. Steatohepatitis was graded as grade 0 (no NASH), grade 1 (borderline NASH), and grade 2 (definite NASH). Regression slopes are  $\log_e$ -adjusted differences in expression and therefore represent a  $\log_e$  fold-change (ratio) per unit increase in the NAFLD variable. These values were converted to percent fold change by exponentiating Euler's number ( $-2.71828$ ) to the  $\log_e$  fold change value. For the t-tests, expression was compared between the disease groups as follows: NAFLD activity score was split into low ( $\leq 4$ ) and high ( $\geq 5$ ); fibrosis was split into stage 0 vs. rest, stage 0-1 vs. rest, and stage 4 vs. rest; and steatohepatitis was split into definite vs. borderline/absent. T-tests were performed to provide an orthogonal description of pharmacogene changes using clinically meaningful thresholds of disease severity.

A multiple linear regression model was built to investigate the most important factors in describing *CYP2C19* expression. This approach utilized a backward elimination approach of the clinical and demographic factors in the Supplementary Data 12. Factors were removed from the model until each of the remaining factors had a  $p$ -value of less than 0.05. Variability explained was measured by unadjusted R-squared.

### Meta-analysis

We conducted literature searches using Google Scholar, without restricting results to any time frame, with the following search terms: NAFLD gene expression liver. From these results we manually reviewed titles or abstracts from the first 300 results, narrowing to 16 studies that conducted gene expression profiling in liver samples from patients with NAFLD. We additionally reviewed references from these studies and searched the gene expression omnibus (GEO) identifying 11 more relevant studies for a total of 27 possible studies to include. The following terms were used for the GEO search: Non-alcoholic fatty liver disease, fibrosis, NAFLD, NASH, steatohepatitis, fatty liver. The GEO results were filtered to: Homo sapiens, data sets, series, and publication date: 2000 – 2021. Lastly, we conducted further manual searches of pubmed, (using search terms: NAFLD gene expression liver), finding no further studies, ensuring our literature search exhausted most or all available studies that conducted gene expression analysis in NAFLD liver tissue.

We focused the meta-analysis on *CYP2C19* because it was the most consistently downregulated pharmacogene in our 93 NAFLD patients and knowledge that this gene is downregulated has a large potential for clinical actionability. Of the 27 identified studies, *CYP2C19* expression data was available publicly for 11 studies<sup>30,32,34,45–48,75–78</sup>. We obtained access to data for 3 additional studies thanks to gracious



contributions by the study authors<sup>26,33,79</sup>. We were able to obtain *CYP2C19* differential expression results directly from the supplement of 2 additional studies<sup>80,81</sup>, resulting in a final meta-analysis pool of 17 studies, including our own, with varying phenotypes. One of the supplemental data sets reported FDR-adjusted *p* values using the Benjamini-Hochberg procedure, and we back-calculated using the number of genes and rank order to estimate the non-adjusted *p*-value<sup>80</sup>.

Datasets available as “GEO2R” sets were analyzed in the online GEO portal for Log2 fold-change using the default settings with Limma (version 3.26.8)<sup>82</sup>. Datasets that were available as raw data were analyzed similarly to the GEO2R default settings, using Oligo (version 1.62.1) and Limma packages. For Datasets that showed consistent read depth across samples we used the log counts-per-million normalization in Limma, and for Datasets with more inconsistent read depths, we used the Voom normalization as recommended in the Limma and Voom documentation. All *p*-values were converted to 95% confidence intervals using the metagen<sup>83</sup> R package (version 4.9), followed by forest plot generation.

Phenotype data was not the same for each study, therefore we created groups of analyses that could be compared together. We chose phenotypes that represent the 3 histological measures of NAFLD severity (NAS, NASH, and fibrosis) based on criteria that were available in the most number of studies. These categories are as follows: Fibrosis stages 3-4 vs. stages 0-1, NAS 5-8 vs. 0-4, and NASH vs. absence of NASH. Controls (absence of NASH) were defined as either healthy obese, healthy normal-BMI, and in certain analyses where healthy controls were not available, mild disease was used as the control group. These are further specified in the corresponding figure legends. From the 17 studies, one was not included in our meta-analyses because it only reported steatosis values but not NASH diagnosis, NAS, or fibrosis.

In addition to the strict clinical grouping, the interpretability of the meta-analysis results are enhanced by similar quantification of mRNA in liver samples (using either RNA-seq or an array-based method). Since measurements within each study were compared between disease vs. no disease, the relative numbers (fold changes) eliminate any heterogeneity that could be introduced by theoretical differences between RNA-seq and the array methodology. For each meta-analysis we used a random-effect model, calculating the average Log2 fold-change weighted by the commonly used inverse variance method<sup>83,84</sup>. This approach is rigorous because it considers that NAFLD is not necessarily the only influence on *CYP2C19* expression and is superior to a fixed-effect model which would incorrectly assume that the only influence to between-study heterogeneity comes from population sampling<sup>85</sup>. We confirm that the random-effects model is appropriate because the Cochran’s Q statistic was statistically significantly larger than the degrees of freedom in each analysis, indicating there is between-study heterogeneity (Supplementary Data 21). In addition to results provided by the inverse variance weighting approach, we also provide the sample size-weighted means to show consistency with a more intuitive approach.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The source data contains logTPM pharmacogene RNA-seq values as described in the methods along with the NAFLD histological disease grades/stages/scores. Full RNA-seq gene count data is available under GEO accession number [GSE225740](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE225740). Gene expression data used in the meta-analysis can be found at the data sources provided in Supplementary Data 20 and 22. Source data are provided with this paper.

### Code availability

All code used for this manuscript can be found at GitHub ([https://github.com/Nickpowe/CYP2C19\\_NAFLD\\_code.git](https://github.com/Nickpowe/CYP2C19_NAFLD_code.git)).

### References

1. Younossi, Z. M. et al. Global epidemiology of nonalcoholic fatty liver disease—Meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology* **64**, 73–84 (2016).
2. Chalasani, N. et al. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. *Hepatology* **67**, 328–357 (2018).
3. Kvitne, K. E. et al. Short- and long-term effects of body weight, calorie restriction and gastric bypass on CYP1A2, CYP2C19 and CYP2C9 activity. *Br. J. Clin. Pharmacol.* <https://doi.org/10.1111/bcp.15349> (2022).
4. Targher, G., Byrne, C. D. & Tilg, H. NAFLD and increased risk of cardiovascular disease: clinical associations, pathophysiological mechanisms and pharmacological implications. *Gut* **69**, 1691–1705 (2020).
5. Zhang, Q. Q. & Lu, L. G. Nonalcoholic Fatty Liver Disease: Dyslipidemia, Risk for Cardiovascular Complications, and Treatment Strategy. *J. Clin. Transl. Hepatol.* **3**, 78–84 (2015).
6. Zhao, Y. C. et al. Nonalcoholic Fatty Liver Disease: An Emerging Driver of Hypertension. *Hypertension* **75**, 275–284 (2020).
7. Targher, G., Corey, K. E., Byrne, C. D. & Roden, M. The complex link between NAFLD and type 2 diabetes mellitus - mechanisms and treatments. *Nat. Rev. Gastroenterol. Hepatol.* **18**, 599–612 (2021).
8. Kouvari, M. et al. The presence of NAFLD influences the transition of metabolically healthy to metabolically unhealthy obesity and the ten-year cardiovascular disease risk: a population-based cohort study. *Metabolism*, 154893, <https://doi.org/10.1016/j.metabol.2021.154893> (2021).
9. Fabbrini, E., Sullivan, S. & Klein, S. Obesity and nonalcoholic fatty liver disease: biochemical, metabolic, and clinical implications. *Hepatology* **51**, 679–689 (2010).
10. CAPRIE\_Steering\_Committee. A randomised, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events (CAPRIE). CAPRIE Steering Committee. *Lancet* **348**, 1329–1339 (1996).
11. Sabatine, M. S. et al. Addition of clopidogrel to aspirin and fibrinolytic therapy for myocardial infarction with ST-segment elevation. *N. Engl. J. Med* **352**, 1179–1189 (2005).
12. Yusuf, S. et al. Effects of clopidogrel in addition to aspirin in patients with acute coronary syndromes without ST-segment elevation. *N. Engl. J. Med* **345**, 494–502 (2001).
13. Scott, S. A. et al. Clinical Pharmacogenetics Implementation Consortium guidelines for CYP2C19 genotype and clopidogrel therapy: 2013 update. *Clin. Pharm. Ther.* **94**, 317–323 (2013).
14. Crews, K. R. et al. Clinical Pharmacogenetics Implementation Consortium Guideline for CYP2D6, OPRM1, and COMT Genotypes and Select Opioid Therapy. *Clin. Pharm. Ther.* **110**, 888–896 (2021).
15. Birdwell, K. A. et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guidelines for CYP3A5 Genotype and Tacrolimus Dosing. *Clin. Pharm. Ther.* **98**, 19–24 (2015).
16. Desta, Z. et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2B6 and Efavirenz-Containing Antiretroviral Therapy. *Clin. Pharm. Ther.* **106**, 726–733 (2019).
17. Lee, C. R. et al. Clinical Pharmacogenetics Implementation Consortium Guideline for CYP2C19 Genotype and Clopidogrel Therapy: 2022 Update. *Clin. Pharm. Ther.* **112**, 959–967 (2022).
18. <https://www.fda.gov/drugs/science-and-research-drugs/table-pharmacogenomic-biomarkers-drug-labeling>. Table of Pharmacogenomic Biomarkers in Drug Labeling, Accessed 01/01/2022.
19. Burns, K. E. et al. High CYP2C19 phenotypic variability in gastrointestinal cancer patients. *Cancer Chemother. Pharm.* **77**, 195–204 (2016).



20. Crake, R. L. I. et al. Influence of serum inflammatory cytokines on cytochrome P450 drug metabolising activity during breast cancer chemotherapy: a patient feasibility study. *Sci. Rep.* **11**, 5648 (2021).
21. Goktaş, M. T. et al. Lower CYP2C9 activity in Turkish patients with Behçet's disease compared to healthy subjects: a down-regulation due to inflammation? *Eur. J. Clin. Pharm.* **71**, 1223–1228 (2015).
22. Gravel, S., Chiasson, J. L., Turgeon, J., Grangeon, A. & Michaud, V. Modulation of CYP450 Activities in Patients With Type 2 Diabetes. *Clin. Pharm. Ther.* **106**, 1280–1289 (2019).
23. Lenoir, C. et al. Impact of Acute Inflammation on Cytochromes P450 Activity Assessed by the Geneva Cocktail. *Clin. Pharm. Ther.* **109**, 1668–1676 (2021).
24. Murphy, W. A. et al. Considerations for Physiologically Based Modeling in Liver Disease: From Nonalcoholic Fatty Liver (NAFL) to Nonalcoholic Steatohepatitis (NASH). *Clin. Pharmacol. Ther.* <https://doi.org/10.1002/cpt.2614> (2022).
25. Fisher, C. D. et al. Hepatic cytochrome P450 enzyme alterations in humans with progressive stages of nonalcoholic fatty liver disease. *Drug Metab. Dispos.* **37**, 2087–2094 (2009).
26. Lake, A. D. et al. Analysis of global and absorption, distribution, metabolism, and elimination gene expression in the progressive stages of human nonalcoholic fatty liver disease. *Drug Metab. Dispos.* **39**, 1954–1960 (2011).
27. El-Khateeb, E., Al-Majdoub, Z. M., Rostami-Hodjegan, A., Barber, J. & Achour, B. Proteomic Quantification of Changes in Abundance of Drug-Metabolizing Enzymes and Drug Transporters in Human Liver Cirrhosis: Different Methods, Similar Outcomes. *Drug Metab. Dispos.* **49**, 610–618 (2021).
28. Jamwal, R. et al. Nonalcoholic Fatty Liver Disease and Diabetes Are Associated with Decreased CYP3A4 Protein Expression and Activity in Human Liver. *Mol. Pharm.* **15**, 2621–2632 (2018).
29. Orellana, M. et al. Relationship between in vivo chlorzoxazone hydroxylation, hepatic cytochrome P450 2E1 content and liver injury in obese non-alcoholic fatty liver disease patients. *Hepatol. Res* **34**, 57–63 (2006).
30. Moylan, C. A. et al. Hepatic gene expression profiles differentiate presymptomatic patients with mild versus severe nonalcoholic fatty liver disease. *Hepatology* **59**, 471–482 (2014).
31. Ryaboshapkina, M. & Hammar, M. Human hepatic gene expression signature of non-alcoholic fatty liver disease progression, a meta-analysis. *Sci. Rep.* **7**, 12361 (2017).
32. Govaere, O. et al. Transcriptomic profiling across the nonalcoholic fatty liver disease spectrum reveals gene signatures for steatohepatitis and fibrosis. *Sci. Transl. Med.* **12**, aba4448 (2020).
33. Šeda, O. et al. Hepatic Gene Expression Profiles Differentiate Steatotic and Non-steatotic Grafts in Liver Transplant Recipients. *Front Endocrinol. (Lausanne)* **10**, 270 (2019).
34. Hoang, S. A. et al. Gene Expression Predicts Histological Severity and Reveals Distinct Molecular Profiles of Nonalcoholic Fatty Liver Disease. *Sci. Rep.* **9**, 12541 (2019).
35. Younossi, Z. M. et al. Hepatic gene expression in patients with obesity-related non-alcoholic steatohepatitis. *Liver Int* **25**, 760–771 (2005).
36. Younossi, Z. M. et al. A genomic and proteomic study of the spectrum of nonalcoholic fatty liver disease. *Hepatology* **42**, 665–674 (2005).
37. Rubio, A. et al. Identification of a gene-pathway associated with non-alcoholic steatohepatitis. *J. Hepatol.* **46**, 708–718 (2007).
38. Nakamuta, M. et al. Evaluation of fatty acid metabolism-related gene expression in nonalcoholic fatty liver disease. *Int J. Mol. Med* **16**, 631–635 (2005).
39. Kohjima, M. et al. Re-evaluation of fatty acid metabolism-related gene expression in nonalcoholic fatty liver disease. *Int J. Mol. Med* **20**, 351–358 (2007).
40. Chiappini, F. et al. Exploration of global gene expression in human liver steatosis by high-density oligonucleotide microarray. *Lab Invest* **86**, 154–165 (2006).
41. Baranova, A. et al. Gene expression patterns in hepatic tissue and visceral adipose tissue of patients with non-alcoholic fatty liver disease. *Obes. Surg.* **17**, 1111–1118 (2007).
42. Greco, D. et al. Gene expression in human NAFLD. *Am. J. Physiol. Gastrointest. Liver Physiol.* **294**, G1281–G1287 (2008).
43. Pihlajamäki, J. et al. Thyroid hormone-related regulation of gene expression in human fatty liver. *J. Clin. Endocrinol. Metab.* **94**, 3521–3529 (2009).
44. du Plessis, J. et al. Association of Adipose Tissue Inflammation With Histologic Severity of Nonalcoholic Fatty Liver Disease. *Gastroenterology* **149**, 635–648.e614 (2015).
45. Starman, J. et al. Gene expression profiling unravels cancer-related hepatic molecular signatures in steatohepatitis but not in steatosis. *PLoS One* **7**, e46584 (2012).
46. Arendt, B. M. et al. Altered hepatic gene expression in nonalcoholic fatty liver disease is associated with lower hepatic n-3 and n-6 polyunsaturated fatty acids. *Hepatology* **61**, 1565–1578 (2015).
47. Horvath, S. et al. Obesity accelerates epigenetic aging of human liver. *Proc. Natl Acad. Sci. USA* **111**, 15538–15543 (2014).
48. Ahrens, M. et al. DNA methylation analysis in nonalcoholic fatty liver disease suggests distinct disease-specific and remodeling signatures after bariatric surgery. *Cell Metab.* **18**, 296–302 (2013).
49. Lex, A., Gehlenborg, N., Strobel, H., Vuillemot, R. & Pfister, H. UpSet: Visualization of Intersecting Sets. *IEEE Trans. Vis. Comput Graph* **20**, 1983–1992 (2014).
50. Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**, 2938–2940 (2017).
51. Flockhart, D. A., Thacker, D., McDonald, C. & Desta, Z. *The Flockhart Cytochrome P450 Drug-Drug Interaction Table. Division of Clinical Pharmacology, Indiana University School of Medicine. Updated 2021.* <https://drug-interactions.medicine.iu.edu/>. Accessed Nov 2022.
52. Sim, S. C. et al. A common novel CYP2C19 gene variant causes ultrarapid drug metabolism relevant for the drug response to proton pump inhibitors and antidepressants. *Clin. Pharm. Ther.* **79**, 103–113 (2006).
53. de Morais, S. M. et al. The major genetic defect responsible for the polymorphism of S-mephenytoin metabolism in humans. *J. Biol. Chem.* **269**, 15419–15422 (1994).
54. De Morais, S. M. et al. Identification of a new genetic defect responsible for the polymorphism of (S)-mephenytoin metabolism in Japanese. *Mol. Pharm.* **46**, 594–598 (1994).
55. Bedossa, P. et al. Histopathological algorithm and scoring system for evaluation of liver lesions in morbidly obese patients. *Hepatology* **56**, 1751–1759 (2012).
56. Uhlén, M. et al. Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419. <https://www.proteinatlas.org/> (2015).
57. Desta, Z., Zhao, X., Shin, J. G. & Flockhart, D. A. Clinical significance of the cytochrome P450 2C19 genetic polymorphism. *Clin. Pharmacokinet.* **41**, 913–958 (2002).
58. Kvitne, K. E. et al. Impact of type 2 diabetes on in vivo activities and protein expressions of cytochrome P450 in patients with obesity. *Clin. Transl. Sci.* <https://doi.org/10.1111/cts.13394> (2022).
59. *Plavix Drug Label*, [https://www.accessdata.fda.gov/drugsatfda\\_docs/label/2021/020839s074lbl.pdf](https://www.accessdata.fda.gov/drugsatfda_docs/label/2021/020839s074lbl.pdf) (Plavix Drug Label).
60. Lewis, J. P. et al. Pharmacogenomic polygenic response score predicts ischaemic events and cardiovascular mortality in clopidogrel-treated patients. *Eur. Heart J. Cardiovasc Pharmacother.* **6**, 203–210 (2020).
61. Shuldiner, A. R. et al. Association of cytochrome P450 2C19 genotype with the antiplatelet effect and clinical efficacy of clopidogrel therapy. *Jama* **302**, 849–857 (2009).

62. Velibey, Y. et al. Association between high on-treatment platelet reactivity to clopidogrel and hepatosteatosis in patients undergoing elective stent implantation. *Turk. Kardiyol. Dern. Ars* **46**, 349–357 (2018).
63. Bhatt, D. L. et al. Clopidogrel and aspirin versus aspirin alone for the prevention of atherothrombotic events. *N. Engl. J. Med* **354**, 1706–1717 (2006).
64. Angiolillo, D. J. et al. Platelet function profiles in patients with type 2 diabetes and coronary artery disease on combined aspirin and clopidogrel treatment. *Diabetes* **54**, 2430–2435 (2005).
65. Erlinge, D. et al. Patients with poor responsiveness to thienopyridine treatment or with diabetes have lower levels of circulating active metabolite, but their platelets respond normally to active metabolite added ex vivo. *J. Am. Coll. Cardiol.* **52**, 1968–1977 (2008).
66. Schuette, C. et al. The effect of clopidogrel on platelet activity in patients with and without type-2 diabetes mellitus: a comparative study. *Cardiovasc Diabetol.* **14**, 15 (2015).
67. Martin, K. A. et al. Methionine Adenosyltransferase 1a (MAT1A) Enhances Cell Survival During Chemotherapy Treatment and is Associated with Drug Resistance in Bladder Cancer PDX Mice. *Int. J. Mol. Sci.* **20**, 204983 (2019).
68. Ogiso, T. et al. Substrate selectivity of human aldehyde oxidase 1 in reduction of nitroaromatic drugs. *Arch. Biochem Biophys.* **659**, 85–92 (2018).
69. Beedham, C., Miceli, J. J. & Obach, R. S. Ziprasidone metabolism, aldehyde oxidase, and clinical implications. *J. Clin. Psychopharmacol.* **23**, 229–232 (2003).
70. Miyake, T. et al. GSTP1 expression predicts poor pathological complete response to neoadjuvant chemotherapy in ER-negative breast cancer. *Cancer Sci.* **103**, 913–920 (2012).
71. Dong, X. et al. Glutathione S-transferases P1-mediated interleukin-6 in tumor-associated macrophages augments drug-resistance in MCF-7 breast cancer. *Biochem Pharm.* **182**, 114289 (2020).
72. Hardwick, R. N., Fisher, C. D., Canet, M. J., Scheffer, G. L. & Cherrington, N. J. Variations in ATP-binding cassette transporter regulation during the progression of human nonalcoholic fatty liver disease. *Drug Metab. Dispos.* **39**, 2395–2402 (2011).
73. Kleiner, D. E. et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* **41**, 1313–1321 (2005).
74. Abrams, Z. B., Johnson, T. S., Huang, K., Payne, P. R. O. & Coombes, K. A protocol to evaluate RNA sequencing normalization methods. *BMC Bioinforma.* **20**, 679 (2019).
75. Frades, I. et al. Integrative genomic signatures of hepatocellular carcinoma derived from nonalcoholic Fatty liver disease. *PLoS One* **10**, e0124544 (2015).
76. Suppli, M. P. et al. Hepatic transcriptome signatures in patients with varying degrees of nonalcoholic fatty liver disease compared with healthy normal-weight individuals. *Am. J. Physiol. Gastrointest. Liver Physiol.* **316**, G462–g472 (2019).
77. Wruck, W. et al. Multi-omic profiles of human non-alcoholic fatty liver disease tissue highlight heterogenic phenotypes. *Sci. Data* **2**, 150068 (2015).
78. Kozumi, K. et al. Transcriptomics Identify Thrombospondin-2 as a Biomarker for NASH and Advanced Liver Fibrosis. *Hepatology* **74**, 2452–2466 (2021).
79. Gerhard, G. S. et al. Transcriptomic Profiling of Obesity-Related Nonalcoholic Steatohepatitis Reveals a Core Set of Fibrosis-Specific Genes. *J. Endocr. Soc.* **2**, 710–726 (2018).
80. Teufel, A. et al. Comparison of Gene Expression Patterns Between Mouse Models of Nonalcoholic Fatty Liver Disease and Liver Tissues From Patients. *Gastroenterology* **151**, 513–525.e510 (2016).
81. Baselli, G. A. et al. Liver transcriptomics highlights interleukin-32 as novel NAFLD-related cytokine and candidate biomarker. *Gut* **69**, 1855–1866 (2020).
82. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
83. Viechtbauer, W. Conducting Meta-Analyses in R with the metafor Package. *J. Stat. Softw.* **36**, 1–48 (2010).
84. Van den Noortgate, W., López-López, J. A., Marín-Martínez, F. & Sánchez-Meca, J. Three-level meta-analysis of dependent effect sizes. *Behav. Res Methods* **45**, 576–594 (2013).
85. Tufanaru, C., Munn, Z., Stephenson, M. & Aromataris, E. Fixed or random effects meta-analysis? Common methodological issues in systematic reviews of effectiveness. *Int J. Evid. Based Health.* **13**, 196–207 (2015).

## Acknowledgements

We thank Drs. Johanna K DiStefano, Monika Cahová, Nathan J Cherrington, and all gene expression omnibus (GEO) contributors for benevolently sharing data that enabled the meta-analysis in this manuscript. We thank Julia C. Stumpff, MLIS for her assistance with the literature search for the meta-analysis. We would also like to acknowledge Tamara Graves and Kent E. Williams for their contribution to the RNA isolation and quality measurements. This work was supported, in part, by the following NIH-NIGMS grants: T32GM008425 (NP), R35GM131812 (TS) and by funding from Eli Lilly and Company.

## Author contributions

N.R.P. conducted data analysis, designed research, and wrote manuscript. T.L., J.I., S.C., T.C.S., and Z.D. contributed to the data analyses, designed research, and wrote manuscript. H.R.Q., P.E., Y.C., and M.T. contributed to the RNA-seq experimental design, oversight, performance, data and quality control analyses, and critical review of the manuscript. N.C. oversaw patient selection, histological assessment, verified data analyses, designed research, and wrote manuscript.

## Competing interests

Dr. Chalasani has paid consulting agreements with Abbvie, Zydus, Galectin, Madrigal, Foresite, Altimune, and Boehringer-Ingelheim. He receives research grant support from Exact Sciences, DSM, and Galectin. He has equity ownership in RestUp, LLC, a healthcare placement start-up company. Drs Qian, Ebert, Chen, and Thomas are employees and shareholders of Eli Lilly and Company. Dr. Ipe is an employee of 10 x Genomics and this work was done when he was a faculty member at the Indiana University School of Medicine. Dr. Powell is a shareholder and former employee of Eli Lilly and Company and joined the Indiana University School of Medicine prior to initiating this work. The remaining authors declare no relevant conflicts of interest.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-37209-1>.

**Correspondence** and requests for materials should be addressed to Naga Chalasani.

**Peer review information** *Nature Communications* thanks Ali Canbay, Jörn Schattenberg, Anna Alisi and Annika Hoyer for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023