Early prediction of incident liver disease using conventional risk factors and gut-microbiome-augmented gradient boosting

Liu, Yang

2022-05-03

Early prediction of incident liver disease using conventional risk factors and gut-microbiome-augmented gradient boosting

Graphical abstract
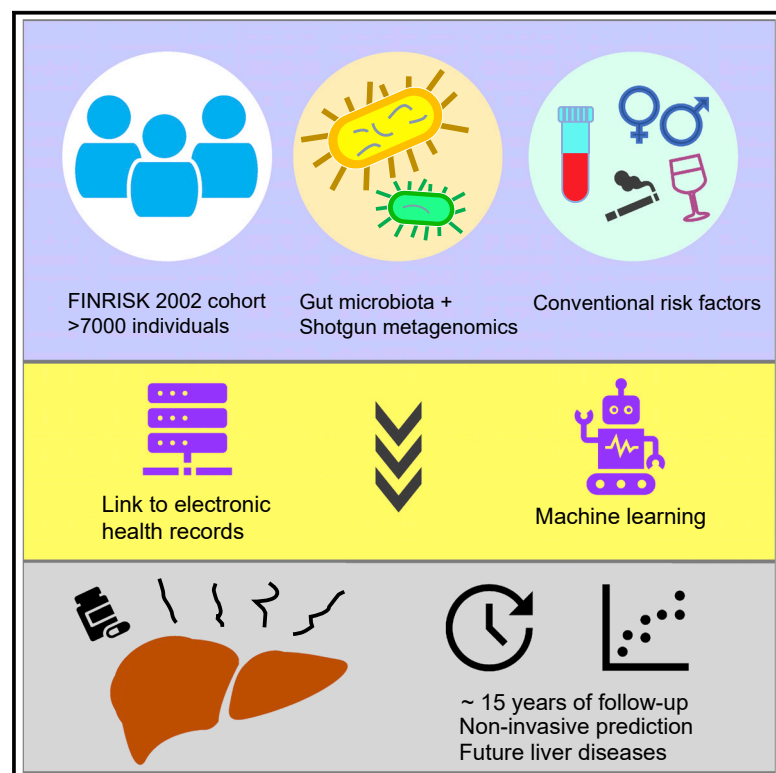
Authors

Yang Liu, Guillaume Méric, Aki S. Havulinna, ..., Veikko Salomaa, Rob Knight, Michael Inouye

Correspondence

yang.liu2@baker.edu.au (Y.L.), minouye@baker.edu.au (M.I.)

In brief

The predictive capacity of the gut microbiome for future liver diseases is investigated using machine learning in a large prospective cohort.

Highlights

- Non-invasive prediction of future liver disease using gut metagenomic sequencing

- Prediction performance of gut microbiome is similar to that of conventional risk factors

- Improved prediction accuracies by combining microbial and conventional predictors

- Machine learning modeling with a large population-based prospective cohort

CellPress

Article

Early prediction of incident liver disease using conventional risk factors and gut-microbiome-augmented gradient boosting

Yang Liu,[1,2,24,*] Guillaume Méric,[1,2,3,4] Aki S. Havulinna,[5,6] Shu Mei Teo,[1,7] Fredrik Åberg,[8] Matti Ruuskanen,[5,9] Jon Sanders,[12] Qiyun Zhu,[12] Anupriya Tripathi,[12,13] Karin Verspoor,[14,15] Susan Cheng,[16] Mohit Jain,[12,17] Pekka Jousilahti,[5] Yoshiki Vázquez-Baeza,[17,18] Rohit Loomba,[19] Leo Lahti,[11] Teemu Niiranen,[5,9,10] Veikko Salomaa,[5] Rob Knight,[12,17,18] and Michael Inouye[1,2,3,7,20,21,22,23,*]

[1]Cambridge Baker Systems Genomics Initiative, Baker Heart and Diabetes Institute, Melbourne, VIC, Australia
[2]Department of Clinical Pathology, Melbourne Medical School, The University of Melbourne, Melbourne, VIC, Australia
[3]Baker Department of Cardiometabolic Health, The University of Melbourne, Melbourne, VIC, Australia
[4]Department of Infectious Diseases, Central Clinical School, Monash University, Melbourne, VIC, Australia
[5]Department of Public Health and Welfare, Finnish Institute for Health and Welfare, Helsinki, Finland
[6]Institute of Molecular Medicine Finland, University of Helsinki, Helsinki, Finland
[7]Cambridge Baker Systems Genomics Initiative, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
[8]Transplantation and Liver Surgery Clinic, Helsinki University Hospital, University of Helsinki, Helsinki, Finland
[9]Department of Internal Medicine, University of Turku, Turku, Finland
[10]Division of Medicine, Turku University Hospital, Turku, Finland
[11]Department of Computing, University of Turku, Turku, Finland
[12]Department of Pediatrics, School of Medicine, University of California, San Diego, La Jolla, CA, USA
[13]Division of Biological Sciences, University of California, San Diego, La Jolla, CA, USA
[14]School of Computing and Information Systems, University of Melbourne, Melbourne, VIC, Australia
[15]School of Computing Technologies, RMIT University, Melbourne, VIC, Australia
[16]Smidt Heart Institute, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA
[17]Center for Microbiome Innovation, University of California, San Diego, La Jolla, CA, USA
[18]Department of Computer Science & Engineering, Jacobs School of Engineering, University of California, San Diego, La Jolla, CA, USA
[19]NAFLD Research Center, Department of Medicine, University of California, San Diego, La Jolla, CA, USA
[20]Health Data Research UK Cambridge, Wellcome Genome Campus, University of Cambridge, Cambridge, UK
[21]British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
[22]British Heart Foundation Centre of Research Excellence, University of Cambridge, Cambridge, UK
[23]The Alan Turing Institute, London, UK
[24]Lead contact
*Correspondence: yang.liu2@baker.edu.au (Y.L.), minouye@baker.edu.au (M.I.)
https://doi.org/10.1016/j.cmet.2022.03.002

SUMMARY

The gut microbiome has shown promise as a predictive biomarker for various diseases. However, the potential of gut microbiota for prospective risk prediction of liver disease has not been assessed. Here, we utilized shallow shotgun metagenomic sequencing of a large population-based cohort (N > 7,000) with ~15 years of follow-up in combination with machine learning to investigate the predictive capacity of gut microbial predictors individually and in conjunction with conventional risk factors for incident liver disease. Separately, conventional and microbial factors showed comparable predictive capacity. However, microbiome augmentation of conventional risk factors using machine learning significantly improved the performance. Similarly, disease-free survival analysis showed significantly improved stratification using microbiome-augmented models. Investigation of predictive microbial signatures revealed previously unknown taxa for liver disease, as well as those previously associated with hepatic function and disease. This study supports the potential clinical validity of gut metagenomic sequencing to complement conventional risk factors for prediction of liver diseases.

INTRODUCTION

Liver disease (LD) causes ~2 million deaths per year worldwide–approximately 3.5% of all deaths–and is increasingly common in

aging populations (World Health Organization, 2019; Asrani et al., 2019). The etiology of LD is complex and includes several inter-related risk factors, such as obesity, age, and excess alcohol consumption (Younossi et al., 2018). Alcohol consumption, in

Table 1. Baseline characteristics of study population

	Female	Male
n = 7,115	n = 55%	n = 45%
Demographics		
Age	49.69 [38.05, 58.78]	51.92 [40.54, 60.70]
Physical parameters		
Body mass index (kg/m²)	25.90 [23.09, 29.47]	26.9 [24.55, 29.58]
Waist-hip ratio	0.84 [0.80, 0.88]	0.97 [0.92, 1.01]
Lifestyles		
Smoking	19%	28%
Pure alcohol consumption (g/week)	18.9 [2.7, 55.8]	75.9 [20.7, 168.3]
Laboratory results		
HDL cholesterol (mmol/L)	1.59 [1.35, 1.89]	1.30 [1.10, 1.53]
LDL cholesterol (mmol/L)	3.19 [2.65, 3.76]	3.46 [2.89, 4.09]
Triglycerides (mmol/L)	1.07 [0.80, 1.45]	1.36 [0.97, 1.97]
Gamma-glutamyl transferase (U/L)	19 [15, 27]	30 [21, 46]
Alanine aminotransferase (U/L)	18 [14, 24]	27 [20, 37]
Aspartate aminotransferase (U/L)[a]	23.5 [20, 28]	28.0 [24, 33]

Median [IQR] for continuous variables; n% for categorical variables.
[a]Available for 6,211 persons.

particular, is a major contributor to LD, accounting for >50% of cirrhosis deaths (Asrani et al., 2019). Consequences of LD can be acute or chronic with highly variable progression rates; however, most patients are not diagnosed until an advanced stage when liver function is overwhelmed (e.g., decompensated cirrhosis) (Younossi et al., 2016; Bellentani, 2017). Currently, liver biopsy remains the gold standard for diagnosis and classification of disease stage, but biopsy is invasive and thus restricted. Although non-invasive tests (NITs) for detecting LD are available, such as ultrasound, elastography, computed tomography, magnetic resonance imaging, and spectroscopy, they are primarily applicable to the detection of advanced severity (Soresi et al., 2014; Cleveland et al., 2018; Moreno et al., 2019). Hence, there is an unmet need for high-fidelity early detection and risk prediction approaches for LD.

The role of the human gut microbiome—the collection of microorganisms residing in the gastrointestinal tract—has been increasingly recognized in various aspects of LD (Hartmann et al., 2019; Tripathi et al., 2018). Interest in the gut microbiome has rapidly grown as sequencing technologies have progressed from 16S rRNA amplicon sequencing to shotgun metagenomics. Recent studies have revealed evidence linking gut microbial composition and the pathogenesis of LD (Adolph et al., 2018; Safari and Gérard, 2019; Zhu et al., 2015), as well as potential therapeutic approaches targeting gut microbial communities (Szabo, 2015; Woodhouse et al., 2018). Importantly, the gut microbiome has shown potential for differentiating cirrhosis and non-cirrhosis controls. Qin et al. have shown that gene- and function-level biomarkers derived from metagenomics could discriminate between liver cirrhosis patients and healthy controls (Qin et al., 2014). Loomba et al. have successfully distinguished advanced fibrosis from mild and moderate nonalcoholic fatty liver disease (NAFLD) using the gut microbiome characterized by whole-genome shotgun sequencing with random forest classifiers (Loomba et al., 2017). Later, Caussy et al. have used random forest classifiers to distinguish NAFLD cirrhosis from non-NAFLD healthy controls based on gut microbial compositions from 16S sequencing (Caussy et al., 2019). However, previous studies have been limited by cross-sectional study design, and there are limited data regarding the association between baseline microbiota and incident LD. This would be an important step in investigating whether the gut microbiome is causally linked to LD or can be used as a stratification tool to identify those at high risk who may benefit from targeted interventions.

Therefore, we designed a prospective study to examine the association and predictive capacity of the gut microbiome for incident LDs, using shallow metagenomic sequencing and supervised machine learning in a large population-based cohort of >7,000 individuals with median ~15 years of electronic health record (EHR) follow-up. Traditional statistical and machine learning approaches are compared on gut metagenomes, and their predictive capacity is evaluated individually and in combination with conventional risk factors, including age, sex, body mass index (BMI), waist-hip ratio (WHR), alcohol consumption, smoking status, triglycerides, high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, and gamma-glutamyl transferase (GGT) levels. The best-performing models are further assessed using survival analysis for time to disease onset. Taken together, our study assesses the potential clinical validity of adding the gut metagenome to conventional risk factors for the prediction of incident LD. We make our predictive models freely available.

RESULTS

Characteristics of study population
This study included 7,115 participants from the population-based FINRISK 2002 cohort, whose participants are representative of Finnish population aged 25–74 years at baseline, with a median follow-up of 14.8 years (STAR Methods) (Borodulin et al., 2018). The detailed baseline characteristics of the study population are provided in Table 1. To investigate the predictive capacity of the baseline gut microbiome and conventional risk factors for incident LDs, we matched phenotype metadata with gut microbial profiles derived from stool samples and linked the baseline data to follow-up diagnoses of any LDs or alcoholic LD (ALD) defined by ICD-10 codes (STAR Methods). After stringent quality control and filtering (STAR Methods), 41 cases of incident ALD and 103 cases of incident LD were considered for prediction analyses. The median time from baseline to onset for ALD and LD was 9.45 (IQR 5.21–11.95) and 9.14 (IQR 5.10–11.98) years, respectively.

Baseline gut microbial composition
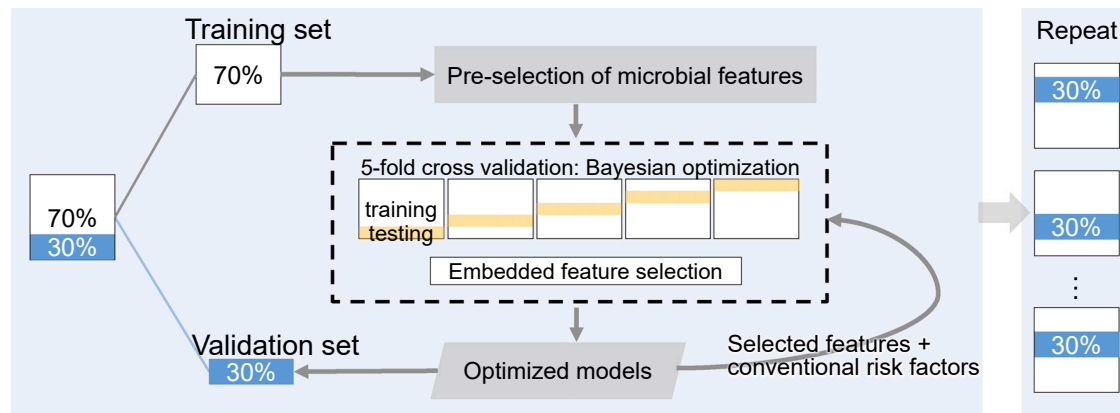Stool samples were sequenced by shallow shotgun metagenomics to a mean depth of approximately 1.056 million reads

Figure 1. Machine learning framework for predicting incident liver disease

per sample. After human sequences, low-quality and adapter reads were removed, and a total of 7.63 billion reads were classified using a Genome Taxonomy Database (GTDB) release 89 index database for taxonomic classification, resulting in 967,000 post-QC and classified reads per sample on average. In total, GTDB classification uniquely identified 151 phyla, 338 classes, 925 orders, 2,254 families, 7,906 genera, and 24,705 species from gut metagenomes. We focused on common bacterial taxa to reduce alignment artifacts and noise; taxa were filtered by relative abundance (>0.01% in at least 1% of samples), which resulted in 46 phyla, 71 classes, 124 orders, 232 families, 617 genera, and 1,224 species for further analysis. Overall, the most abundant taxa were members of phyla *Firmicutes*, *Firmicutes_A* (corresponding to *Firmicutes* in NCBI), *Firmicutes_C* (*Firmicutes*), *Bacteroidota* (*Bacteroidetes*), *Actinobacteriota* (*Actinobacteria*), and *Proteobacteria* (Figure S1).

Developing machine learning models

The workflow for machine learning to predict incident LD is shown in Figure 1. For both ALD and LD, samples were randomly partitioned based on the prediction target into a training set for discovery (70% of samples) and a validation set for evaluation (remaining 30%), and the partitioning itself was randomly performed 10× to assess sampling variation. Within the training set, we developed and tested prediction models through cross-validation, and the optimal models were assessed for final performance in the withheld validation set (STAR Methods). Prediction models were derived from different taxonomic levels separately because taxa at higher ranks are inclusive of their members at lower ranks, and introducing redundant features can lead to impaired prediction performance. The average results of the 10 sample partitions are reported.

To define a subset of informative taxa, we performed pre-selection of microbial features associated with incident LD from the union of three approaches in the training sets (STAR Methods). After pre-selection, there were 10, 16, 42, 123, 355, and 508 microbial taxa on average at phylum, class, order, family, genus, and species levels for incident ALD and 9, 12, 25, 62, 194, and 303 for incident LD. To incorporate microbial diversity measures, Chao1, Pielou, and Shannon indices were included as additional features. These microbial features were then

used to build prediction models in the corresponding training sets.

Gradient boosting classifiers were applied to pre-selected microbial features to develop and optimize prediction models with cross-validation in the training datasets. To assess prediction performance, we also included two robust and common statistical approaches: logistic regression and ridge regression.

Prediction of incident LD

The gradient boosting classifier generally outperformed both multivariable logistic regression and ridge regression, particularly at lower taxonomic levels (Figure 2). The performance across models was measured and compared using the area under the receiver operating characteristic curve (AUROC), which considers the trade-offs between sensitivity and specificity at all possible thresholds (STAR Methods). With the gradient boosting classifier, higher prediction performance was observed at lower taxonomic levels for both incident ALD and LD, suggesting that the strength of association for higher resolution of gut microbial features outweighs their lower abundances at these levels. For LD, we obtained the highest prediction performance at the species level with an average AUROC of 0.733 (IQR 0.709–0.748 across test sets from 10 random splits; Figure 2A). The mean AUROC for LD ranged from 0.622 to 0.725 at phylum and genus levels, respectively. When predicting ALD, we obtained an average AUROC > 0.75 at phylum and class levels and an average AUROC > 0.85 for other taxonomic levels with the highest value of 0.895 (0.886–0.907) at the species level (Figure 2B). For both LD and ALD, the AUROC values of gradient boosting at the species level were significantly higher than those of logistic and ridge regression models (Wilcoxon p < 0.01; Figure S2).

Ridge regression tended to perform better than logistic regression (Figure 2). For LD, ridge regression achieved an average AUROC >0.65 at order, family, genus, and species levels, with the highest average AUROC of 0.675 (0.639–0.707) at the species level; for ALD, an AUROC >0.80 was obtained at family, genus, and species levels, with the highest average AUROC of 0.838 (0.809–0.865) at the species level. Logistic regression yielded its highest average AUROC of 0.651 (0.618–0.673) at the family level and an AUROC <0.60 at other taxonomic levels
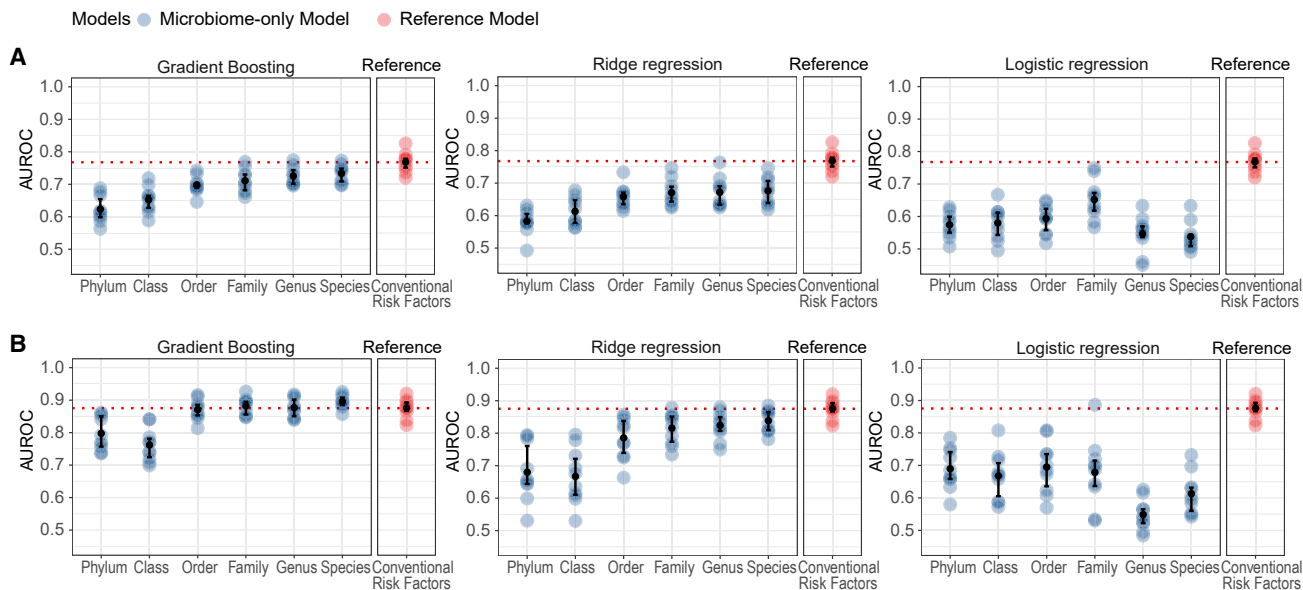
Models ● Microbiome-only Model ● Reference Model

Figure 2. Comparison of approaches for the prediction of incident liver disease using gut microbial features
(A) For the prediction of any liver disease, the gradient boosting classifier outperformed logistic regression and ridge regression across different taxonomic levels.
(B) For the prediction of alcoholic liver disease, similar trends were observed. For comparison, a conventional prediction model is shown in red. Error bars represent mean and IQR (STAR Methods). Horizontal dashed lines mark the mean performance of conventional models.

for predicting LD (Figure 2A); for ALD, the best performance was obtained at the order level with an average AUROC of 0.694 (0.636–0.735; Figure 2B). Although logistic regression is highly efficient and interpretable, it did not perform well in this case where a large number of features are correlated. The L2 regularization of ridge regression better handled inter-correlated microbial features than that of logistic regression. However, both methods underperformed compared with the gradient-boosted decision tree classifier, which is known to better capture nonlinear relationships and is robust to correlated features. The gradient-boosted decision tree classifier was used in subsequent analyses.

Benchmarking reference models using conventional approaches
Conventional risk factors are commonly used for predicting LD risk (Bedogni et al., 2006; Long et al., 2018). We built reference models using a comprehensive set of conventional risk factors, including sex, age, alcohol consumption, smoking status, BMI, WHR, triglycerides, HDL, LDL, and GGT, to compare with the prediction capacity of microbiome-based models (STAR Methods). The conventional risk factor model achieved an average AUROC of 0.768 (0.751–0.779) for incident LD, slightly higher than the highest average AUROC of microbiome-only models achieved at the species level (AUROC 0.733, Figures 2A and S2 [Wilcoxon p < 0.05]). For ALD, the average AUROC of conventional risk factor models reached 0.875 (0.867–0.893), which was lower than the AUROC of the gradient boosting model achieved using species-level microbial features alone (average AUROC 0.895, Figures 2B and S2 [Wilcoxon p < 0.05]). Both conventional models and microbiome-based models had substantial predictive power individually; the next

section evaluates the combination of conventional risk factors and microbial compositions for LD and ALD prediction.

Integrating gut microbiome and conventional risk factors
To investigate the potential of a microbiome-augmented prediction model for LD, we utilized the gradient boosting classifier of microbiome features together with all conventional risk factors related to the disease and followed the same partitioning for training and testing (STAR Methods). To evaluate the performance comprehensively, the optimal models were assessed for both AUROC and area under the precision-recall curve (AUPRC, which considers the trade-offs between precision and recall; STAR Methods). Since greater taxonomic resolution offered better predictive performance, we compared the species-level augmented and the conventional-risk-factors-only models.

Overall, the prediction performance of the microbiome-augmented models achieved greater AUROC and AUPRC compared with that of the conventional prediction models (Wilcoxon p < 0.05). Prediction of LD (Figure 3A) using the species-level augmented model yielded an average AUROC of 0.834 (0.815–0.852), an AUROC increase of 0.066 over the conventional risk factor model (as above, average AUROC 0.768). For ALD, the species-level augmented model yielded an average AUROC of 0.956 (0.945–0.964), an AUROC increase of 0.081 over the conventional risk factor model (as above, average AUROC 0.875) (Figure 3B).

With a baseline AUPRC value of 0.015 for LD, the species-level augmented model achieved an average AUPRC of 0.185 (0.156–0.202), which was higher (Wilcoxon p < 0.05) than the average AUPRC of 0.158 (0.141–0.179) yielded by the conventional risk
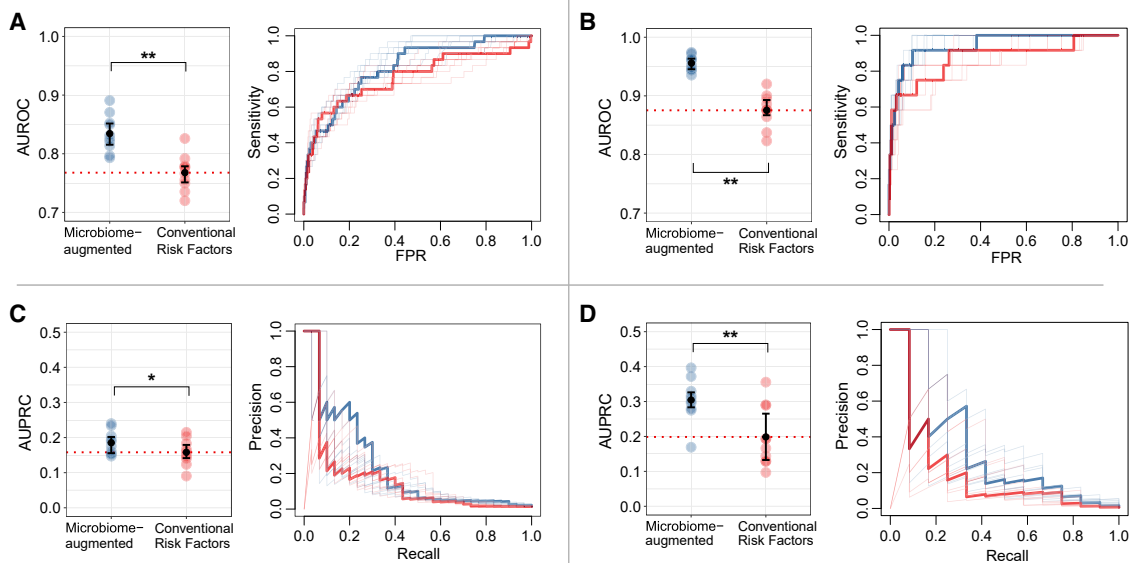
Figure 3. Models of conventional risk factors and gut microbiome data improved the prediction of incident liver disease over conventional prediction models

(A and B) Area under the ROC curve (AUROC) for gradient boosting models using species-level gut microbiome data together with conventional risk factors (blue) or a conventional risk factor model (red), predicting (A) incident any liver disease or (B) alcoholic liver disease.

(C and D) Area under the precision-recall curve (AUPRC) for (C) any liver disease and (D) alcoholic liver disease. Error bars represent mean and IQR (STAR Methods). Difference in performance between the microbiome-augmented model and the conventional risk factor model was tested using one-sided Wilcoxon-signed rank test. *p < 0.05; **p < 0.01. Horizontal dashed lines mark the mean performance of conventional model as a reference. The bolded ROC and precision-recall curves correspond to models with AUROC and AUPRC that are closest to the mean performance reference.

factor model (Figure 3C). For ALD with a baseline AUPRC of 0.006, the species-level augmented model yielded a significantly greater AUPRC than the conventional risk factor model (Wilcoxon p < 0.01; Figure 3D), with average AUPRCs of 0.304 (0.283–0.326) and 0.199 (0.133–0.265), respectively.

Comparison of gut-microbiome-based models and clinical tests for predicting LD

To further evaluate the utility of the gut microbiome for predicting LDs, we assessed the predictive capacity and generalizability of three clinical NITs that aimed to screen liver-related events—fatty liver index (FLI), BARD score, and the dynamic aspartate-to-alanine aminotransferase ratio (dAAR) score—and compared them with gut-microbiome-based predictions. FLI is an algorithm developed for predicting fatty liver (Bedogni et al., 2006) and has been extensively validated in various populations, including a general population not selected for specific LDs (Koehler et al., 2013; Yang et al., 2015; Caballería et al., 2018). FLI is straightforward to use, as it is based on routine measurements of BMI, waist circumference, triglycerides, and GGT (Bedogni et al., 2006). The BARD score combines BMI, aspartate transaminase/alanine aminotransferase (AST/ALT) ratio, and the presence of diabetes mellitus and was developed for predicting advanced liver fibrosis in clinical practice (Harrison et al., 2008). In previous studies, the BARD score has been mostly validated in patients with NAFLD and tended to have limited predictive ability (Younes et al., 2021). The dAAR score is based on age, ALT, and AST and was developed for predicting liver-related outcomes and advanced liver fibrosis/cirrhosis in general populations. The dAAR has shown performance comparable with or

better than other NITs in various populations (Åberg et al., 2021). Although other non-invasive markers for detecting liver-related events, such as FIB4 index and NAFLD fibrosis score, exist, we could not assess these markers as they were not measured at the population level.

We utilized Cox models to investigate the predictive capacity and generalizability of FLI, BARD, and dAAR for incident LDs and compared them with species-level gut-microbiome-based models in the validation sets (STAR Methods). We found that the gut-microbiome-based model and dAAR score performed better than FLI and BARD score in distinguishing incident LD cases from non-cases. For LD, the gut-microbiome-based model achieved a higher predictive capacity than FLI and BARD (both Wilcoxon p < 0.01) and similar performance as dAAR (Wilcoxon p = 0.5). Specifically, the gut-microbiome-based model and dAAR score identified incident cases with average C-statistics of 0.741 (0.724–0.756) and 0.744 (0.721–0.762), respectively, followed by FLI (0.677; 0.636–0.71) and BARD score (0.539; 0.527–0.557). For ALD, the gut-microbiome-based model outperformed (Wilcoxon p < 0.01) each of the dAAR, FLI, and BARD scores, with C-statistics of 0.909 (0.902–0.922), 0.817 (0.781–0.847), 0.725 (0.699–0.751), and 0.561 (0.528–0.581), respectively. For comparison, the conventional risk factor model yielded average C-statistics of 0.771 (0.739–0.803) and 0.894 (0.869–0.922) for LD and ALD, respectively.

A combined model of the gut microbiome and dAAR yielded average C-statistics of 0.786 (0.758–0.801) and 0.883 (0.845–0.908) for LD and ALD, respectively. Our results suggest that the gut microbiome and dAAR have comparable predictive
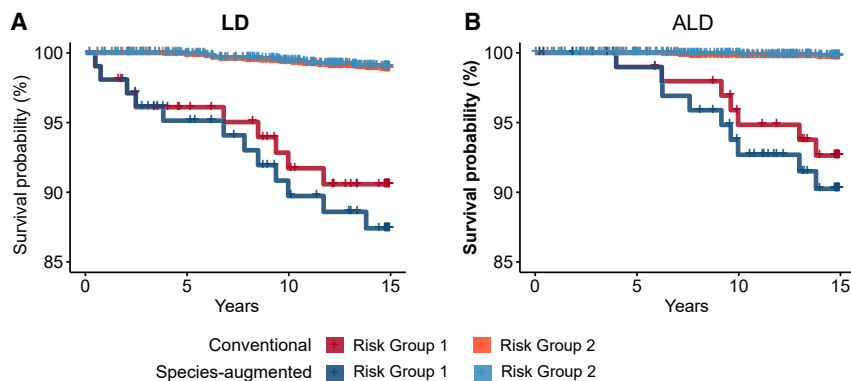
Figure 4. Survival curves of predicted risk groups for incident liver disease

(A and B) Performance in the withheld validation set (30% of samples) of Cox models of conventional risk factors and in combination with species-level microbiome-only scores for (A) any liver disease (# of cases = 30) and (B) alcoholic liver disease (# of cases = 12). Predicted risk groups are the top 5% (risk group 1) versus the bottom 95% (risk group 2).

performance for LD and that the gut microbiome outperforms the existing NITs for ALD. Importantly, the combination of dAAR and the gut microbiome maximized C-indices for LD, indicating that the two predictors are complementary.

Survival analysis using conventional and microbiome-augmented risk models

We next performed survival analysis using time-on-study Cox regression in the validation sets to assess potential clinical validity of microbiome-augmented (species-level) risk models as compared with conventional-risk-factors-only models (STAR Methods). The Cox model of conventional risk factors achieved average C-statistics of 0.813 (0.795–0.826) for LD and 0.922 (0.904–0.940) for ALD, respectively. The microbiome-augmented risk models yielded higher (Wilcoxon $p < 0.01$) average C-statistics of 0.838 (0.819–0.840) for LD and 0.959 (0.954–0.964) for ALD. Consistent with this finding, the microbiome-augmented model fits significantly better (LRT $p < 0.01$) than that using conventional risk factors only. Disease-free survival of those in the highest 5% of microbiome-augmented risk was worse than those for conventional risk factors alone (Figure 4). Among the highest 5% risk predicted for LD, the microbiome-augmented model identified 12 out of 30 cases in the validation set, whereas the conventional risk factor model identified 9 cases. For ALD, the microbiome-augmented and conventional models identified 9 and 7 cases, respectively, out of 12 cases in the validation set.

Application to NAFLD case-control classification

To investigate the generalizability of our findings to NAFLD classification, we applied the prediction models in an independent cohort including 30 NAFLD patients and 97 non-NAFLD controls recruited at the University of California San Diego (UCSD) (Caussy et al., 2019; Oh et al., 2020). The detailed characteristics of the UCSD cohort are given in STAR Methods and Table S1. Here, we tested whether the species-level gut-microbiome-augmented models for any LDs could discriminate NAFLD patients from controls. In the external case-control study, the gut-microbiome-augmented models for LD achieved an average AUROC of 0.705 (IQR 0.697–0.717 in UCSD cohort) and AUPRC of 0.431 (0.422–0.456). Although the predictive models were not trained for NAFLD specifically, the results suggested moderate generalizability of gut-microbiome-augmented models that were trained from a broader spectrum of LDs. For comparison,

we also tested the prediction performance of conventional risk factor models and microbiome-only models trained on the FINRISK cohort. The microbiome-only models for LD achieved an average AUROC of 0.61 (0.57–0.651) and AUPRC of 0.315 (0.277–0.355), whereas the conventional risk factor models had an average AUROC of 0.497 (0.454–0.549) and AUPRC of 0.256 (0.237–0.273), indicating that the gut-microbiome-augmented model was the most generalizable.

Furthermore, we examined the ability of conventional risk factors to distinguish NAFLD cases from controls and whether gut microbiome could improve the discrimination ability. Using logistic regression models of individual risk factors, we found that BMI, WHR, TRIG, and HDL each achieved AUROC scores greater than 0.750 (Figures S3A–S3I). The combination of all conventional risk factors achieved an AUROC of 0.890 (Figure S3J). We next utilized ridge logistic regression to evaluate the discrimination ability of species-level gut microbiome. Notably, species-level gut microbiome alone identified NAFLD cases with an AUC score of 0.832 (Figure S3K), which was higher than any individual conventional risk factor. Subsequently, we combined the gut-microbiome-based model and conventional risk factors using logistic regression, which improved the discrimination performance to an AUROC of 0.936 (Figure S3L).

Composition of gut microbiome signatures

To better understand which bacterial taxa contribute to ALD and LD prediction, we considered those that most frequently contributed to the optimal gradient boosting classifiers at each taxonomic level (Table S2).

Notably, Pielou evenness and Chao1 and Shannon diversity, which were found to be negatively associated with both ALD and LD, were all selected as predictive contributors at phylum, class, order, and family levels. This was consistent with previous findings that the richness and diversity of gut microbiome communities are positively correlated with human health (Acharya and Bajaj, 2017; Bäckhed et al., 2012).

The microbial signatures mainly comprised taxa from phylum *Actinobacteriota* (*Actinobacteria* in NCBI taxonomy), *Bacteroidota* (*Bacteroidetes*), *Firmicutes* and *Firmicutes_A* (*Firmicutes*), and *Proteobacteria* (*Proteobacteria*) (Figures 5, S4, and S5). Overall, most of the selected microbial taxa were significantly (FDR < 0.05) and positively associated with LD. Many bacterial taxa have been previously reported to be related to LD and its progression. The families *Chitinophagaceae* (mainly contributed by *Chitinophaga*) (Puri et al., 2018), *Streptococcaceae* (mainly *Streptococcus* spp.) (Puri et al., 2018; Dubinkina et al., 2017; Sarin et al., 2019), and *Enterobacteriaceae* (mainly *Klebsiella*
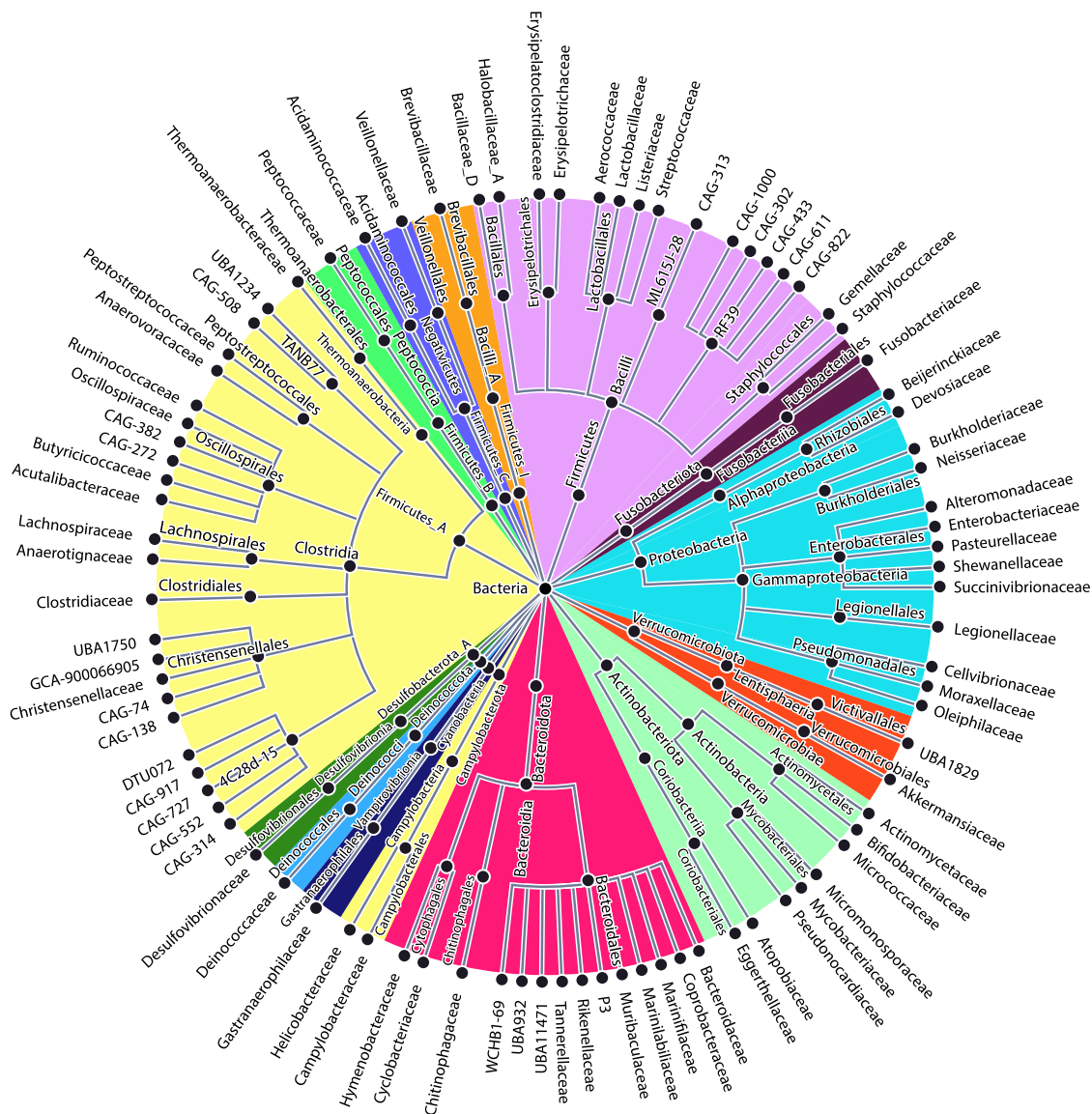
Figure 5. Predictive microbial taxa for liver disease

A bacterial taxonomy tree (phylum to family level) whose members at lower ranks showed predictive signal for incident liver disease. For full taxonomy, see Figure S4.

and *Klebsiella_A*) (Dubinkina et al., 2017); and genera *Actinomyces* (mainly *A. graevenitzii*) (Fan et al., 2018; Hwang et al., 2011), *Rikenella* (Xue et al., 2017), *Blautia* (Dubinkina et al., 2017; Leclercq et al., 2014), *Dorea* (Leclercq et al., 2014; Waters and Ley, 2019), *Neisseria* (Fan et al., 2018; Chen et al., 2016), etc., have been frequently reported to be enriched in patients with alcoholism and ALD; the families *Streptococcaceae* (mainly *Streptococcus* spp.), *Erysipelotrichaceae*, and *Enterobacteriaceae* (mainly *Escherichia*); and genera *Actinomyces* (Caussy et al., 2019), *Lactobacillus_C* and *Lactobacillus_H* as former *Lactobacillus* (Nobili et al., 2018; Shao et al., 2018; Jiang et al., 2015), *Veillonella* (Shao et al., 2018; Chen et al., 2016), *Prevotella* spp. (Chen et al., 2016; Jiang et al., 2015; Zhu et al., 2015; Shen et al., 2017; Boursier et al., 2016), etc., have been found to be

positively associated with a broad range of LDs, including acute-on-chronic liver failure, non-alcoholic fatty LD, and cirrhosis. Several members of *Actinomyces* spp. (Ávila et al., 2015; Könönen and Wade, 2015), *Escherichia* spp. (Yamamoto et al., 2017; Commander et al., 2017; Chen et al., 2005), *Klebsiella* spp. (Paasch et al., 2017; Kamal et al., 2017), *Desulfovibrio* spp. (Koyano et al., 2015), etc., have been identified as pathogens for liver abscess and sepsis. Consistent with previous studies, *UBA11524* (former *Faecalibacterium*) (Caussy et al., 2019; Leclercq et al., 2014; Dubinkina et al., 2017; Zhu et al., 2015; Yun et al., 2019), *Coprococcus* (Dubinkina et al., 2017), and *Akkermansia* (Lowe et al., 2017; Dubinkina et al., 2017; Grander et al., 2018; Wu et al., 2017) were negatively associated (FDR < 0.05) with LD. Notably, genus *Akkermansia*, of which

A. muciniphila was previously suggested as having potential protective effect on liver function and gut microbiota ecology (Wu et al., 2017; Dao et al., 2016; Grander et al., 2018; Kim et al., 2020), uniquely contributed to every higher rank within phylum *Verrucomicrobia* for prediction.

Among the prediction signatures, many bacterial taxa have been found in association with the development of liver damage. Intestinal barrier dysfunction, marked by increased intestinal permeability, plays a key role in the pathogenesis of LD and is directly associated with cirrhosis (Albillos et al., 2020). At the genus level, *Ruminococcus*, *Dorea*, *Faecalibacterium*, and *Blautia* were found to be responsible for increased intestinal permeability (Leclercq et al., 2014), which can induce translocation of microbes and microbial metabolites and subsequently worsen hepatic inflammation (Albillos et al., 2020). Conversely, *Bifidobacterium* was found to be negatively correlated with intestinal permeability (Leclercq et al., 2014). Gut microbial lipopolysaccharide (LPS) is one of the most potent LPSs that trigger a cascade of proinflammatory response and promote the progression of fatty liver (Albillos et al., 2020). Besides, LPS-producing bacteria are linked to obesity (Zhao, 2013), a major risk factor for NAFLD (Younossi et al., 2018). Although members of phylum *Bacteroidota* (*Bacteroidetes*) are the largest group of LPS producers, such as *Bacteroides* and *Prevotella* spp., family *Enterobacteriaceae* of phylum *Proteobacteria* and family *Desulfovibrionaceae* of phylum *Desulfobacterota_A* (*Proteobacteria*) exhibit an immense amount of endotoxin activity (Zhao, 2013). A recent study has shown that endotoxin producers that overgrow in patients with fatty liver, including strain members of *Escherichia* and *Klebsiella*, can induce NAFLD in mice models and suggested a potential causative role for them in NAFLD (Fei et al., 2020). The altered gut microbiota composition in cirrhosis is partially attributed to reduced primary bile acids and increased secondary bile acids in the gut lumen that are resulted from liver insufficiency (Albillos et al., 2020). The reduction in total bile acids in the gut contributes to an overgrowth of pathobiont microbes, including members of *Enterobacteriaceae* and *Enterobacteriaceae* (Albillos et al., 2020). The elevation of secondary bile acids is largely associated with an abundance of bacterial producers of secondary bile acid, such as members of *Clostridium* and *Eubacterium* (Albillos et al., 2020; Wahlström et al., 2016). Bile salt hydrolase activity is associated with resistance of hepatocytes to bile toxicity and is broadly present in gut microbes, including *Bacteroides*, *Bifidobacterium*, *Clostridium*, and *Lactobacillus* (Wahlström et al., 2016).

DISCUSSION

In this study, we investigated the potential analytic and clinical validity of the gut microbiome to improve the prediction of future LD. Using baseline gut metagenomic sequencing and 15 years of EHR follow-up, we developed a framework to predict incident LD and ALD using machine learning approaches. We demonstrated that the gut microbiome and conventional risk factor models exhibited similar prediction performances separately, but importantly, the microbiome-augmented conventional risk factor models markedly improved the prediction. Furthermore, microbiome-augmented models achieved higher prediction accuracies for incident LD and ALD than clinical NITs FLI and BARD

score. These results indicate that the combination of conventional risk factors with gut microbiota may have potential clinical utility in early risk stratification for LD.

Few studies so far have investigated the prediction of incident LD events using gut microbiota. Currently, clinical risk prediction models for LD events are commonly derived from demographic, lifestyle, and biochemical factors that resulted from routine blood tests. Although these prediction rules have reasonable accuracy in clinical practice, they tend to be influenced by extrahepatic conditions and have reduced accuracy for early-stage disease (Vilar-Gomez and Chalasani, 2018; Carbone et al., 2016). Furthermore, there is a lack of guidance for primary care and necessity of referral based on the test results, as a large number of patients with abnormal test results are asymptomatic during LD progression (Williams et al., 2018; Standing et al., 2018; Härmälä et al., 2019). Thus, there is an urgent need for new tools that improve early detection of high-risk individuals.

Our findings are consistent with those of previous studies on the relationship of bacterial taxa with hepatic function, disease, and progression, and we identified several taxa with potential probiotic effects. However, the precise role of gut microbiota is poorly understood, and our results support the need for species-level resolution or indeed greater levels of resolution offered by even deeper metagenomic sequencing. For example, the abundance of the *Bifidobacterium* genus has been reported to be associated with alcoholism and liver injury in various ways (Leclercq et al., 2014; Xu et al., 2012): at the species level, *B. dentium* has been found to be enriched in advanced LD (Dubinkina et al., 2017); conversely, *B. pseudocatenulatum* and *B. bifidum* have been recognized as potential probiotics that may attenuate liver damage (Fang et al., 2017; Nobili et al., 2018; Gómez-Hurtado et al., 2019). This indicates the importance of lower-level taxa resolution in interpreting how bacteria contribute to the disease pathology.

We demonstrated the generalizability of our findings in an independent cohort of an American NAFLD case-control study. We showed that the models trained for any LD were able to distinguish NAFLD cases from controls with moderate classification performance. The attenuated performance of models in this external cohort can be possibly explained by culturally and geographically distinct populations (He et al., 2018), a different study design, and shallower sequencing depth. Moreover, we demonstrated that the gut microbiome had higher discrimination performance than individual conventional risk factors in distinguishing NAFLD cases from controls and combination of all factors markedly improved the performance.

Notwithstanding the challenging necessity for validation of novel biomarkers as well as the development of standards for interpretation as prerequisites for clinical implementation, our study provides an evidence base and corresponding risk prediction models for the translation of metagenomic sequencing in risk prediction of LD.

Limitations of the study
Our study has several limitations. Due to the necessity of a prospective early detection study to consider a large number of apparently healthy individuals, we were limited in the number of incident disease cases, and therefore, we were not well-powered to investigate subtypes and stages of LD, which might lead

to greater clinical significance. The imbalanced data problem, where the distribution of cases and controls is not equal, is almost inevitable in disease prediction studies. It also imposes challenges to applications of a large variety of machine learning classification algorithms, as the classifiers tend to bias toward the more common class. In our study, the imbalanced classification is handled with nested cross-validation design (Vabalas et al., 2019) and stochastic gradient boosting (Friedman, 2002) for producing robust and unbiased estimates of performance. The practicality for shallow metagenomic sequencing for a large prospective cohort also meant that we were not able to evaluate the added information of deep sequencing to risk prediction. In addition, due to limited feasibility of LD screening tests at the population level, we were not able to perform more comprehensive comparisons of the gut microbiome and NITs such as FIB-4. The prevention measures available to individuals at high risk of LD are also somewhat limited. These include weight reduction and alcohol and smoking cessation and may extend to caution with pharmaceutical prescriptions. Finally, our cohort is of European ancestry, and these prediction models may have attenuated performance in non-European ancestries.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cmet.2022.03.002.

AUTHOR CONTRIBUTIONS

Study design, Y.L., S.M.T., V.S., and M.I.; data analyses and interpretation, Y.L., G.M., A.S.H., M.R., and M.I.; investigation, Y.L., G.M., A.S.H., F.Å., M.R., J.S., Q.Z., Y.V.-B., R.L., V.S., and M.I.; metagenomic sequencing analyses, J.S., Q.Z., A.T., Y.V.-B., and R.K.; supervision, S.M.T., K.V., S.C., M.J., P.J., R.L., L.L., T.N., V.S., R.K., and M.I.; all authors reviewed the manuscript and approved the final manuscript.

DECLARATION OF INTERESTS

V.S. has received honoraria for consulting from Novo Nordisk and Sanofi and travel support from Novo Nordisk. He also has ongoing research collaboration with Bayer Ltd (all unrelated to the present study). R.L. serves as a consultant or advisory board member for Anylam/Regeneron, Arrowhead Pharmaceuticals, Astra Zeneca, Bird Rock Bio, Boehringer Ingelheim, Bristol-Myer Squibb, Celgene, Cirius, CohBar, Conatus, Eli Lilly, Galmed, Gemphire, Gilead, Glympse Bio, GNI, GRI Bio, Inipharm, Intercept, Ionis, Janssen, Inc., Merck, Metacrine, Inc., NGM Biopharmaceuticals, Novartis, Novo Nordisk, Pfizer, Prometheus, Promethera, Sanofi, Siemens, and Viking Therapeutics. In addition, his institution has received grant support from Allergan, Boehringer Ingelheim, Bristol-Myers Squibb, Cirius, Eli Lilly and Company, Galectin Therapeutics, Galmed Pharmaceuticals, GE, Genfit, Gilead, Intercept, Grail, Janssen, Madrigal Pharmaceuticals, Merck, NGM Biopharmaceuticals, NuSirt, Pfizer, pH Pharma, Prometheus, and Siemens. He is also co-founder of Liponexus, Inc.

REFERENCES

Åberg, F., Danford, C.J., Thiele, M., Talbäck, M., Rasmussen, D.N., Jiang, Z.G., Hammar, N., Nasr, P., Ekstedt, M., But, A., et al. (2021). A dynamic aspartate-to-alanine aminotransferase ratio provides valid predictions of incident severe liver disease. Hepatol. Commun. 5, 1021–1035.

Acharya, C., and Bajaj, J.S. (2017). Gut microbiota and complications of liver disease. Gastroenterol. Clin. North Am. 46, 155–169.

Adolph, T.E., Grander, C., Moschen, A.R., and Tilg, H. (2018). Liver-microbiome axis in health and disease. Trends Immunol. 39, 712–723.

Albillos, A., De Gottardi, A., and Rescigno, M. (2020). The gut-liver axis in liver disease: pathophysiological basis for therapy. J. Hepatol. 72, 558–577.

Asrani, S.K., Devarbhavi, H., Eaton, J., and Kamath, P.S. (2019). Burden of liver diseases in the world. J. Hepatol. 70, 151–171.

Ávila, F., Santos, V., Massinha, P., Pereira, J.R., Quintanilha, R., Figueiredo, A., Lázaro, A., Carrelho, S., Coelho, J.S., Barroso, E., and Duarte, M.A. (2015). Hepatic actinomycosis. GE Port. J. Gastroenterol. *22*, 19–23.

Bäckhed, F., Fraser, C.M., Ringel, Y., Sanders, M.E., Sartor, R.B., Sherman, P.M., Versalovic, J., Young, V., and Finlay, B.B. (2012). Defining a healthy human gut microbiome: current concepts, future directions, and clinical applications. Cell Host Microbe *12*, 611–622.

Bedogni, G., Bellentani, S., Miglioli, L., Masutti, F., Passalacqua, M., Castiglione, A., and Tiribelli, C. (2006). The fatty liver index: a simple and accurate predictor of hepatic steatosis in the general population. BMC Gastroenterol. *6*, 33.

Bellentani, S. (2017). The epidemiology of non-alcoholic fatty liver disease. Liver Int. *37* (suppl 1), 81–84.

Borodulin, K., Tolonen, H., Jousilahti, P., Jula, A., Juolevi, A., Koskinen, S., Kuulasmaa, K., Laatikainen, T., Männistö, S., Peltonen, M., et al. (2018). Cohort profile: the national FINRISK study. Int. J. Epidemiol. *47*, 696–696i.

Borodulin, K., Vartiainen, E., Peltonen, M., Jousilahti, P., Juolevi, A., Laatikainen, T., Männistö, S., Salomaa, V., Sundvall, J., and Puska, P. (2015). Forty-year trends in cardiovascular risk factors in Finland. Eur. J. Public Health *25*, 539–546.

Boursier, J., Mueller, O., Barret, M., Machado, M., Fizanne, L., Araujo-Perez, F., Guy, C.D., Seed, P.C., Rawls, J.F., David, L.A., et al. (2016). The severity of nonalcoholic fatty liver disease is associated with gut dysbiosis and shift in the metabolic function of the gut microbiota. Hepatology *63*, 764–775.

Caballería, L., Pera, G., Arteaga, I., Rodríguez, L., Alumà, A., Morillas, R.M., De La Ossa, N., Díaz, A., Expósito, C., Miranda, D., et al. (2018). High prevalence of liver fibrosis among European adults with unknown liver disease: a population-based study. Clin. Gastroenterol. Hepatol. *16*, 1138–1145.e5.

Carbone, M., Sharp, S.J., Flack, S., Paximadas, D., Spiess, K., Adgey, C., Griffiths, L., Lim, R., Trembling, P., Williamson, K., et al. (2016). The UK-PBC risk scores: derivation and validation of a scoring system for long-term prediction of end-stage liver disease in primary biliary cholangitis. Hepatology *63*, 930–950.

Caussy, C., Tripathi, A., Humphrey, G., Bassirian, S., Singh, S., Faulkner, C., Bettencourt, R., Rizo, E., Richards, L., Xu, Z.Z., et al. (2019). A gut microbiome signature for cirrhosis due to nonalcoholic fatty liver disease. Nat. Commun. *10*, 1406.

Chalasani, N., Younossi, Z., Lavine, J.E., Diehl, A.M., Brunt, E.M., Cusi, K., Charlton, M., and Sanyal, A.J. (2012). The diagnosis and management of non-alcoholic fatty liver disease: practice Guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association. Hepatology *55*, 2005–2023.

Chen, S.C., Yen, C.H., Lai, K.C., Tsao, S.M., Cheng, K.S., Chen, C.C., Lee, M.C., and Chou, M.C. (2005). Pyogenic liver abscesses with *Escherichia coli*: etiology, clinical course, outcome, and prognostic factors. Wien. klin. Wochenschr. *117*, 809–815.

Chen, T., and Guestrin, C. (2016). XGBoost: a scalable tree boosting system. In KDD '16: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (ACM).

Chen, Y., Ji, F., Guo, J., Shi, D., Fang, D., and Li, L. (2016). Dysbiosis of small intestinal microbiota in liver cirrhosis and its association with etiology. Sci. Rep. *6*, 34055.

Cleveland, E., Bandy, A., and Vanwagner, L.B. (2018). Diagnostic challenges of nonalcoholic fatty liver disease/nonalcoholic steatohepatitis. Clin. Liver Dis. (Hoboken) *11*, 98–104.

Commander, S.J., Goss, M., Shi, Y., Finegold, M., Heczey, A., Kukreja, K., Masand, P., and Vasudevan, S. (2017). Liver abscesses secondary to *Escherichia coli* infection mimicking multifocal hepatoblastoma: a case report. J. Pediatr. Surg. Case Rep. *18*, 42–44.

Dao, M.C., Everard, A., Aron-Wisnewsky, J., Sokolovska, N., Prifti, E., Verger, E.O., Kayser, B.D., Levenez, F., Chilloux, J., Hoyles, L., et al. (2016). *Akkermansia muciniphila* and improved metabolic health during a dietary intervention in obesity: relationship with gut microbiome richness and ecology. Gut *65*, 426–436.

Dubinkina, V.B., Tyakht, A.V., Odintsova, V.Y., Yarygin, K.S., Kovarsky, B.A., Pavlenko, A.V., Ischenko, D.S., Popenko, A.S., Alexeev, D.G., Taraskina, A.Y., et al. (2017). Links of gut microbiota composition with alcohol dependence syndrome and alcoholic liver disease. Microbiome *5*, 141.

Fan, X., Peters, B.A., Jacobs, E.J., Gapstur, S.M., Purdue, M.P., Freedman, N.D., Alekseyenko, A.V., Wu, J., Yang, L., Pei, Z., et al. (2018). Drinking alcohol is associated with variation in the human oral microbiome in a large study of American adults. Microbiome *6*, 59.

Fang, D., Shi, D., Lv, L., Gu, S., Wu, W., Chen, Y., Guo, J., Li, A., Hu, X., Guo, F., et al. (2017). Bifidobacterium pseudocatenulatum LI09 and Bifidobacterium catenulatum LI10 attenuate D-galactosamine-induced liver injury by modifying the gut microbiota. Sci. Rep. *7*, 8770.

Fei, N., Bruneau, A., Zhang, X., Wang, R., Wang, J., Rabot, S., Gérard, P., and Zhao, L. (2020). Endotoxin producers overgrowing in human gut microbiota as the causative agents for nonalcoholic fatty liver disease. mBio *11*, e03263-19.

Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. Ann. Statist. *29*, 1189–1232.

Friedman, J.H. (2002). Stochastic gradient boosting. Comp. Stat. Data Anal. *38*, 367–378.

Gómez-Hurtado, I., Zapater, P., Portune, K., Juanola, O., Fernández-Iglesias, A., González-Navajas, J.M., Gracia-Sancho, J., Sanz, Y., and Francés, R. (2019). Improved hemodynamic and liver function in portal hypertensive cirrhotic rats after administration of B. pseudocatenulatum CECT 7765. Eur. J. Nutr. *58*, 1647–1658.

Grander, C., Adolph, T.E., Wieser, V., Lowe, P., Wrzosek, L., Gyongyosi, B., Ward, D.V., Grabherr, F., Gerner, R.R., Pfister, A., et al. (2018). Recovery of ethanol-induced *Akkermansia muciniphila* depletion ameliorates alcoholic liver disease. Gut *67*, 891–901.

Härmälä, S., O'Brien, A., Parisinos, C.A., Direk, K., Shallcross, L., and Hayward, A. (2019). Development and validation of a prediction model to estimate the risk of liver cirrhosis in primary care patients with abnormal liver blood test results: protocol for an electronic health record study in Clinical Practice Research Datalink. Diagn. Progn. Res. *3*, 10.

Harrison, S.A., Oliver, D., Arnold, H.L., Gogia, S., and Neuschwander-Tetri, B.A. (2008). Development and validation of a simple NAFLD clinical scoring system for identifying patients without advanced disease. Gut *57*, 1441–1447.

Hartmann, P., Chu, H., Duan, Y., and Schnabl, B. (2019). Gut microbiota in liver disease: too much is harmful, nothing at all is not helpful either. Am. J. Physiol. Gastrointest. Liver Physiol. *316*, G563–G573.

Havulinna, A.S., Sysi-Aho, M., Hilvo, M., Kauhanen, D., Hurme, R., Ekroos, K., Salomaa, V., and Laaksonen, R. (2016). Circulating ceramides predict cardiovascular outcomes in the population-based FINRISK 2002 cohort. Arterioscler. Thromb. Vasc. Biol. *36*, 2424–2430.

He, Y., Wu, W., Zheng, H.M., Li, P., Mcdonald, D., Sheng, H.F., Chen, M.X., Chen, Z.H., Ji, G.Y., Zheng, Z.D., et al. (2018). Regional variation limits applications of healthy gut microbiome reference ranges and disease models. Nat. Med. *24*, 1532–1535.

Hwang, S.S., Park, S.D., Jang, I.H., Uh, Y., Yoon, K.J., and Kim, H.Y. (2011). Actinomyces graevenitzii bacteremia in a patient with alcoholic liver cirrhosis. Anaerobe *17*, 87–89.

Jiang, W., Wu, N., Wang, X., Chi, Y., Zhang, Y., Qiu, X., Hu, Y., Li, J., and Liu, Y. (2015). Dysbiosis gut microbiota associated with inflammation and impaired mucosal immune function in intestine of humans with non-alcoholic fatty liver disease. Sci. Rep. *5*, 8096.

Juutilainen, A., Kastarinen, H., Antikainen, R., Peltonen, M., Salomaa, V., Tuomilehto, J., Jousilahti, P., Sundvall, J., Laatikainen, T., and Kastarinen, M. (2012). Trends in estimated kidney function: the FINRISK surveys. Eur. J. Epidemiol. *27*, 305–313.

Kamal, F., Williams, G., Akbar, H., Khan, M.A., and Kadaria, D. (2017). *Klebsiella* pneumoniae Liver abscess: a case report and review of literature. Cureus *9*, e970.

Kim, D., Song, L., Breitwieser, F.P., and Salzberg, S.L. (2016). Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res. *26*, 1721–1729.

Kim, S., Lee, Y., Kim, Y., Seo, Y., Lee, H., Ha, J., Lee, J., Choi, Y., Oh, H., and Yoon, Y. (2020). *Akkermansia muciniphila*. Prevents fatty liver, decreases serum triglycerides, and maintains gut homeostasis. Appl. Environ. Microbiol. *86*, 03004–03019.

Koehler, E.M., Schouten, J.N., Hansen, B.E., Hofman, A., Stricker, B.H., and Janssen, H.L. (2013). External validation of the fatty liver index for identifying nonalcoholic fatty liver disease in a population-based study. Clin. Gastroenterol. Hepatol. *11*, 1201–1204.

Könönen, E., and Wade, W.G. (2015). Actinomyces and related organisms in human infections. Clin. Microbiol. Rev. *28*, 419–442.

Koyano, S., Tatsuno, K., Okazaki, M., Ohkusu, K., Sasaki, T., Saito, R., Okugawa, S., and Moriya, K. (2015). A case of liver abscess with Desulfovibrio desulfuricans bacteremia. Case Rep. Infect. Dis. *2015*, 354168.

Leclercq, S., Matamoros, S., Cani, P.D., Neyrinck, A.M., Jamar, F., Stärkel, P., Windey, K., Tremaroli, V., Bäckhed, F., Verbeke, K., et al. (2014). Intestinal permeability, gut-bacterial dysbiosis, and behavioral markers of alcohol-dependence severity. Proc. Natl. Acad. Sci. USA *111*, E4485–E4493.

Long, M.T., Pedley, A., Massaro, J.M., Hoffmann, U., Ma, J., Loomba, R., Chung, R.T., and Benjamin, E.J. (2018). A simple clinical model predicts incident hepatic steatosis in a community-based cohort: the Framingham Heart Study. Liver Int. *38*, 1495–1503.

Loomba, R., Seguritan, V., Li, W., Long, T., Klitgord, N., Bhatt, A., Dulai, P.S., Caussy, C., Bettencourt, R., Highlander, S.K., et al. (2017). Gut microbiome-based metagenomic signature for non-invasive detection of advanced fibrosis in human nonalcoholic fatty liver disease. Cell Metab. *25*, 1054–1062.e5.

Lowe, P.P., Gyongyosi, B., Satishchandran, A., Iracheta-Vellve, A., Ambade, A., Kodys, K., Catalano, D., Ward, D.V., and Szabo, G. (2017). Alcohol-related changes in the intestinal microbiome influence neutrophil infiltration, inflammation and steatosis in early alcoholic hepatitis in mice. PLoS One *12*, e0174544.

Moreno, C., Mueller, S., and Szabo, G. (2019). Non-invasive diagnosis and biomarkers in alcohol-related liver disease. J. Hepatol. *70*, 273–283.

Nobili, V., Putignani, L., Mosca, A., Del Chierico, F.D., Vernocchi, P., Alisi, A., Stronati, L., Cucchiara, S., Toscano, M., and Drago, L. (2018). Bifidobacteria and lactobacilli in the gut microbiome of children with non-alcoholic fatty liver disease: which strains act as health players? Arch. Med. Sci. *14*, 81–87.

Oh, T.G., Kim, S.M., Caussy, C., Fu, T., Guo, J., Bassirian, S., Singh, S., Madamba, E.V., Bettencourt, R., Richards, L., et al. (2020). A universal gut-microbiome-derived signature predicts cirrhosis. Cell Metab. *32*, 878–888.

Paasch, C., Wilczek, S., and Strik, M.W. (2017). Liver abscess and sepsis caused by Clostridium perfringens and *Klebsiella* oxytoca. Int. J. Surg. Case Rep. *41*, 180–183.

Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.A., and Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat. Biotechnol. *36*, 996–1004.

Puri, P., Liangpunsakul, S., Christensen, J.E., Shah, V.H., Kamath, P.S., Gores, G.J., Walker, S., Comerford, M., Katz, B., Borst, A., et al. (2018). The circulating microbiome signature and inferred functional metagenomics in alcoholic hepatitis. Hepatology *67*, 1284–1302.

Qin, N., Yang, F., Li, A., Prifti, E., Chen, Y., Shao, L., Guo, J., Le Chatelier, E., Yao, J., Wu, L., et al. (2014). Alterations of the human gut microbiome in liver cirrhosis. Nature *513*, 59–64.

Safari, Z., and Gérard, P. (2019). The links between the gut microbiome and non-alcoholic fatty liver disease (NAFLD). Cell. Mol. Life Sci. *76*, 1541–1558.

Salosensaari, A., Laitinen, V., Havulinna, A.S., Meric, G., Cheng, S., Perola, M., Valsta, L., Alfthan, G., Inouye, M., Watrous, J.D., et al. (2020). Taxonomic signatures of long-term mortality risk in human gut microbiota. medRxiv, 2019.12.

Sarin, S.K., Pande, A., and Schnabl, B. (2019). Microbiome as a therapeutic target in alcohol-related liver disease. J. Hepatol. *70*, 260–272.

Shao, L., Ling, Z., Chen, D., Liu, Y., Yang, F., and Li, L. (2018). Disorganized gut microbiome contributed to liver cirrhosis progression: a meta-omics-based study. Front. Microbiol. *9*, 3166.

Shen, F., Zheng, R.D., Sun, X.Q., Ding, W.J., Wang, X.Y., and Fan, J.G. (2017). Gut microbiota dysbiosis in patients with non-alcoholic fatty liver disease. Hepatobiliary Pancreat. Dis. Int. *16*, 375–381.

Soresi, M., Giannitrapani, L., Cervello, M., Licata, A., and Montalto, G. (2014). Non invasive tools for the diagnosis of liver cirrhosis. World J. Gastroenterol. *20*, 18131–18150.

Standing, H.C., Jarvis, H., Orr, J., Exley, C., Hudson, M., Kaner, E., and Hanratty, B. (2018). GPs' experiences and perceptions of early detection of liver disease: a qualitative study in primary care. Br. J. Gen. Pract. *68*, e743–e749.

Szabo, G. (2015). Gut-liver axis in alcoholic liver disease. Gastroenterology *148*, 30–36.

Tripathi, A., Debelius, J., Brenner, D.A., Karin, M., Loomba, R., Schnabl, B., and Knight, R. (2018). The gut-liver axis and the intersection with the microbiome. Nat. Rev. Gastroenterol. Hepatol. *15*, 397–411.

Vabalas, A., Gowen, E., Poliakoff, E., and Casson, A.J. (2019). Machine learning algorithm validation with a limited sample size. PLoS One *14*, e0224365.

Vilar-Gomez, E., and Chalasani, N. (2018). Non-invasive assessment of non-alcoholic fatty liver disease: clinical prediction rules and blood-based biomarkers. J. Hepatol. *68*, 305–315.

Wahlström, A., Sayin, S.I., Marschall, H.U., and Bäckhed, F. (2016). Intestinal crosstalk between bile acids and microbiota and its impact on host metabolism. Cell Metab. *24*, 41–50.

Waters, J.L., and Ley, R.E. (2019). The human gut bacteria Christensenellaceae are widespread, heritable, and associated with health. BMC Biol. *17*, 83.

Wick, R., and Méric, G. (2019). Metagenomics index correction. Preprint at bioRxiv. https://doi.org/10.1101/712166.

Williams, R., Alexander, G., Armstrong, I., Baker, A., Bhala, N., Camps-Walsh, G., Cramp, M.E., De Lusignan, S., Day, N., Dhawan, A., et al. (2018). Disease burden and costs from excess alcohol consumption, obesity, and viral hepatitis: fourth report of the Lancet Standing Commission on Liver Disease in the UK. Lancet *391*, 1097–1107.

Woodhouse, C.A., Patel, V.C., Singanayagam, A., and Shawcross, D.L. (2018). Review article: the gut microbiome as a therapeutic target in the pathogenesis and treatment of chronic liver disease. Aliment. Pharmacol. Ther. *47*, 192–202.

2019). World Health Statistics Overview 2019: Monitoring Health for the SDGs, Sustainable Development Goals (World Health Organization).

Wu, W., Lv, L., Shi, D., Ye, J., Fang, D., Guo, F., Li, Y., He, X., and Li, L. (2017). Protective effect of *Akkermansia muciniphila* against immune-mediated liver injury in a mouse model. Front. Microbiol. *8*, 1804.

Xu, M., Wang, B., Fu, Y., Chen, Y., Yang, F., Lu, H., Chen, Y., Xu, J., and Li, L. (2012). Changes of fecal Bifidobacterium species in adult patients with hepatitis B virus-induced chronic liver disease. Microb. Ecol. *63*, 304–313.

Xue, M., Liu, Y., Lyu, R., Ge, N., Liu, M., Ma, Y., and Liang, H. (2017). Protective effect of aplysin on liver tissue and the gut microbiota in alcohol-fed rats. PLoS One *12*, e0178684.

Yamamoto, D., Hernandes, R.T., Liberatore, A.M.A., Abe, C.M., Souza, R.B.D., Romão, F.T., Sperandio, V., Koh, I.H., and Gomes, T.A.T. (2017). *Escherichia albertii*, a novel human enteropathogen, colonizes rat enterocytes and translocates to extra-intestinal sites. PLoS One *12*, e0171385.

Yang, B.L., Wu, W.C., Fang, K.C., Wang, Y.C., Huo, T.I., Huang, Y.H., Yang, H.I., Su, C.W., Lin, H.C., Lee, F.Y., et al. (2015). External validation of fatty liver index for identifying ultrasonographic fatty liver in a large-scale cross-sectional study in Taiwan. PLoS One *10*, e0120443.

Younes, R., Caviglia, G.P., Govaere, O., Rosso, C., Armandi, A., Sanavia, T., Pennisi, G., Liguori, A., Francione, P., Gallego-Durán, R., et al. (2021). Long-term outcomes and predictive ability of non-invasive scoring systems in patients with non-alcoholic fatty liver disease. J. Hepatol. *75*, 786–794.

Younossi, Z., Anstee, Q.M., Marietti, M., Hardy, T., Henry, L., Eslam, M., George, J., and Bugianesi, E. (2018). Global burden of NAFLD and

NASH: trends, predictions, risk factors and prevention. Nat. Rev. Gastroenterol. Hepatol. *15*, 11–20.

Younossi, Z.M., Koenig, A.B., Abdelatif, D., Fazel, Y., Henry, L., and Wymer, M. (2016). Global epidemiology of nonalcoholic fatty liver disease-meta-analytic assessment of prevalence, incidence, and outcomes. Hepatology *64*, 73–84.

Yun, Y., Kim, H.N., Lee, E.J., Ryu, S., Chang, Y., Shin, H., Kim, H.L., Kim, T.H., Yoo, K., and Kim, H.Y. (2019). Fecal and blood microbiota profiles and pres-ence of nonalcoholic fatty liver disease in obese versus lean subjects. PLoS One *14*, e0213692.

Zhao, L. (2013). The gut microbiota and obesity: from correlation to causality. Nat. Rev. Microbiol. *11*, 639–647.

Zhu, L., Baker, R.D., and Baker, S.S. (2015). Gut microbiome and nonalcoholic fatty liver diseases. Pediatr. Res. *77*, 245–251.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human stool samples	FINRISK 2002 Cohort	https://thl.fi/en/web/thl-biobank
Human stool samples	UCSD NAFLD Cohort (PI: Rohit Loomba)	UCSD IRB #140084 UCSD IRB #111282
Deposited data		
Raw metagenomic data	FINRISK 2002 Cohort	EGAS00001005020
Raw metagenomic data	UCSD NAFLDCohort	EGAS00001004600
FINRISK 2002 individual level data	FINRISK 2002 Cohort	https://thl.fi/en/web/thl-biobank
GTDB release 89	Genome Taxonomy Database	https://gtdb.ecogenomic.org/
Software and algorithms		
Centrifuge v.1.0.4		https://github.com/infphilo/centrifuge
XGBoost		https://xgboost.ai/
mlrMBO		https://github.com/mlr-org/mlrMBO
Caret		https://cran.r-project.org/web/packages/caret/index.html
Glmnet		https://cran.r-project.org/web/packages/glmnet/index.html
Survival		https://cran.r-project.org/web/packages/survival/index.html
Other		
Illumina HiSeq4000	Illumina	

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Yang Liu (yang.liu2@baker.edu.au).

Materials availability
This study did not result in the generation of novel reagents. Data in this study are available with written application.

Data and code availability
The FINRISK data for the present study are available with a written application to the THL Biobank as instructed on the website of the Biobank: https://thl.fi/en/web/thl-biobank/for-researchers. A separate permission is needed from FINDATA (https://www.findata.fi/en/) for use of the EHR data. The accession number for the metagenomics data for the validation cohort reported in this paper is available at the European Genome-Phenome Archive (EGAD00001006364). Predictive models are available at https://doi.org/10.26188/12554573.v1.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Study population
The FINRISK population surveys have been performed every 5 years since 1972 to monitor trends in cardiovascular disease risk factors in the Finnish population (Borodulin et al., 2015, 2018). The FINRISK 2002 study was based on a stratified random sample of the population aged 25–74 years from six specific geographical areas of Finland (Salosensaari et al., 2020). The sampling was stratified by sex, region and 10-year age group so that each stratum had 250 participants. The overall participation rate was 65.5% (n = 8798). The participants filled out a questionnaire at home, then participated in a clinical examination carried out by specifically trained nurses and gave a blood sample. They also received a sampling kit and instructions to donate a stool sample at home and mail it to the Finnish Institute for Health and Welfare in an overnight mail. The follow up of the cohort took place by record linkage of the study

data with the Finnish national electronic health registers (Hospital Discharge Register and Causes of Death Register), which provide in practice 100% coverage of relevant health events in Finnish residents. For the present analyses the follow-up extended until Dec 31st, 2016. The study protocol of FINRISK 2002 was approved by the Coordinating Ethical Committee of the Helsinki and Uusimaa Hospital District (Ref. 558/E3/2001). All participants signed an informed consent. The study was conducted according to the World Medical Association's Declaration of Helsinki on ethical principles.

Disease case definitions
The liver disease investigated in this study consists of two groups, alcoholic liver disease (ALD) and a broader range of any liver disease (LD) according to the ICD-10 codes (Finnish modification). A sample was considered as an incident case of any liver disease if the follow-up register-based diagnostic classification was under the ICD-10 codes K70 - K77; the alcoholic liver disease was defined by the ICD-10 code K70. Table S3 shows the numbers of subjects of main categories for any liver disease.

Inclusion and exclusion criteria
The inclusion criteria of FINRISK 2002 cohort have been previously described (Borodulin et al., 2018). Samples with gut microbiome profiles, phenotype metadata and follow-up all available were included in our analysis (n=7115). The exclusion criteria of our analysis were: (1) samples with gut metagenomic sequencing yielding <400K reads; (2) presence of baseline prevalent diagnosis of target disease for prediction; (3) baseline pregnancy during the survey year. Altogether, 41 cases of incident ALD in 7005 samples and 103 cases of incident LD in 6965 samples were considered for modelling analyses.

Case-control study for validation
Participants were drawn from the Twins and Family cohort recruited at the University of California, San Diego (UCSD) NAFLD Research Center between December 2011 and December 2017 (Table S1). All participants underwent exhaustive clinical research visit and provided stool samples which were collected and immediately stored at -80°C. The study design, clinical assessments and metagenome profiling have been detailed previously (Caussy et al., 2019; Oh et al., 2020). Definition for NAFLD was in concordance with the American Association for the Study of Liver Study (AASLD) Practice Guidelines (Chalasani et al., 2012). All participants provided written informed consent.

METHOD DETAILS

Phenotype metadata
The phenotype data in this study comprised of demographic characteristics, life habits, disease history and medications, laboratory test results and follow-up EHRs. Baseline phenotype variables used as conventional risk factors included age, sex, body mass index (BMI), waist-hip ratio (WHR), smoking status, alcohol consumption, triglyceride (TRIG), gamma-glutamyl transferase (GGT), high-density lipoprotein (HDL) and low-density lipoprotein (LDL) cholesterol measurements. BMI was computed as the weight in kilograms divided by the square of height in meters measure with light clothing(Borodulin et al., 2018). Smoking status described whether a participant was a current daily smoker at the time of the survey. Alcohol consumption, based on self-reported questionnaire, was measured as the average weekly pure alcohol use in grams during the past 12 months. TRIG, GGT, HDL and LDL- cholesterol were measured from blood samples collected from participants advised to fast for at least 4 hours prior to collection and avoid heavy meals earlier during the day (Borodulin et al., 2018, Juutilainen et al., 2012, Havulinna et al., 2016). Baseline alanine aminotransferase (ALT) and aspartate aminotransferase (AST), which were measured in a sub-sample of subjects, were included for developing non-invasive test scores of liver diseases. The median follow-up was 14.84 years and the end point was the date of death or last follow-up. Incident disease was coded as a binary variable indicating disease case (1) or non-case (0) with matched time from baseline to event or end of follow-up also utilised for analyses.

Metagenomic sequencing and taxonomic profiling
Stool samples were collected by participants and mailed overnight to Finnish Institute for Health and Welfare for storing at -20°C; the samples were sequenced at the University of California San Diego in 2017. The gut microbiome was characterized by shallow shotgun metagenomics sequencing with Illumina HiSeq 4000 Systems. We successfully performed stool shotgun sequencing in n = 7231 individuals. The detailed procedures for DNA extraction, library preparation and sequence processing have been previously described (Salosensaari et al., 2020). Adapter and host sequences were removed. To preserve the quality of data while retaining most of the disease cases, samples with sequencing depth less than 400,000 were excluded from our analysis. The metagenomes were classified using default parameters in Centrifuge 1.0.4 (Kim et al., 2016), and using an index database based on taxonomic definitions from the Genome Taxonomy Database (GTDB) release 89 (Parks et al., 2018) (Wick and Méric, 2019).

Characterization of the gut microbiome
The gut microbial composition was represented as relative abundance of taxa. For each metagenome at phylum, class, order, family, genus and species levels, the relative abundance of a taxon was computed as the proportion of reads assigned to the clade rooted at this taxon among total classified reads of this metagenome. The relative abundance of a bacteria that had no reads assigned in a metagenome was considered as zero in the corresponding profile. We focused on common and relatively abundant taxa of a

within-sample relative abundance greater than 0.01% in more than 1% of samples. The centered log-ratio (CLR) transformation was carried out on abundance data by taking the log of taxa abundance divided by geometric mean of abundance in each metagenome profile. Abundance of zero was replaced with a value representing 1/10 of the minimum abundance in a metagenome before transformation. In this study, all analyses except for microbial diversity calculation were based on CLR transformed data.

QUANTIFICATION AND STATISTICAL ANALYSIS

General framework of prediction modelling

Prediction models were developed for any liver disease and alcoholic liver disease at phylum, class, order, family, genus and species levels separately. For each incident disease to be predicted, samples were randomly shuffled and partitioned into a training cohort for discovery and a validation cohort for evaluation at a 7:3 ratio according to the target disease variable such that the distribution of disease cases and healthy controls in training and testing datasets were consistent. Within the training set, we first performed pre-selection of features (detailed in next section) and then developed models using pre-selected features through 5-fold cross validation stratified according to the prediction target, which further created random splits of internal training and testing sets at a 8:2 ratios five times with testing sets being mutually exclusive. The models were optimized based on cross-validated results. The optimal models were then trained on the full training set and finally assessed on the withheld validation set that was excluded from the training and optimization process to avoid data leakage from the training set. Considering the variation of attribute distributions that can occur during random data partitioning, we repeated the whole process described above 10 times and reported the average and IQR results (given in parentheses). The detailed procedures were elucidated in the rest of this section.

Pre-selection of microbial taxa

To select a set of informative microbial taxa that were individually associated with incident liver disease, we analyzed the relationship between microbial abundance and incident disease using (1) logistic regression adjusted for age and gender, (2) Cox regression for time to disease occurrence adjusted for age and gender, and (3) Spearman correlation. This feature selection step was performed only within the training datasets accounting for 70% of samples. A microbial taxon was included in further analyses if statistical significance ($P<0.05$) was found by any of the above three methods. After adjusting for age and gender, on average 8 phyla, 14 classes, 35 orders, 103 families, 299 genus and 406 species were associated with incident ALD at statistical significance using logistic regression; 8 phyla, 14 classes, 36 orders, 106 families, 306 genera and 416 species were found significant using cox regression. The Spearman correlation found 7 phyla, 12 classes, 36 orders, 112 families, 314 genera and 428 species, on average, significantly correlated with alcoholic liver disease. For LD, the average numbers of significantly associated taxa at each taxonomic level were 7 phyla, 10 classes, 19 orders, 49 families, 157 genera and 245 species with logistic regression; 7 phyla, 10 classes, 20 orders, 52 families, 164 genera and 255 species with cox regression; 5 phyla, 8 classes, 19 orders, 51 families, 148 genera and 218 species with Spearman correlation. As the selected taxa were not always agreed by all three approaches, taxa selected by any approach in the training cohort were included for developing prediction models with the corresponding data partition. Of the 10 differently sampled training sets, the average numbers of microbial features at phylum, class, order, family, genus and species levels were 10, 16, 42, 123, 355, 508 for predicting incident ALD, and 9, 12, 25, 62, 194, 303 for predicting LD, respectively.

Microbial and conventional features

Conventional risk factors include baseline age, gender, BMI (kg/m^2), WHR, alcohol consumption (g), smoking status, TRIG (mmol/l), GGT (U/L), HDL and LDL cholesterol (mmol/l). Microbial features comprised taxa abundance along with microbial diversity metrics at phylum, class, order, family, genus and species levels. To characterize microbial diversity in samples, Chao1 index, Pielou's evenness index and Shannon diversity index were calculated using raw abundance data without filtering. Chao1 index estimates the total species richness for a given community considering the presence of rare species. Pielou's evenness index measures how evenly the species are distributed in a given sample. Shannon's index takes into account both species richness and evenness.

Model development

The machine learning approach extreme gradient boosting was applied to predict the incidence of liver disease from baseline phenotype and microbial data using *Xgboost* library in R. Xgboost is a distributed and optimized implementation of gradient boosting decision trees, an ensemble method of sequential and additive training of trees with regularizations (Chen and Guestrin, 2016). In the boosting process, each decision tree is a base learner that learns the errors from the previous tree. At each iteration, randomized sampling approach is implemented to increase robustness of the base learner (Friedman, 2002). The trees are trained iteratively in a sequence to minimize the errors and are finally aggregated to make a strong learner (Friedman, 2001). The prediction procedure was a twofold process which involved developing models using microbial features alone and in combination with conventional risk factors. In the first step the gradient boosting classifiers were trained on microbial features consisting of taxa abundance and diversity metrics at different taxonomic levels separately. In the second step, microbial features selected by the embedded feature selection of gradient boosting classifiers in the first step, together with conventional risk factors, were deployed to predict incident disease. The models were trained with Bayesian optimization (*mlrMBO* in R) through 5-fold cross validation in the training dataset. The optimal models selected based on cross-validated results were evaluated in the withheld evaluation dataset as the final performance for predicting incident disease. The highly ranked and frequently selected (by more than half of the models) microbial features were

considered as predictive signatures for further interpretation. Since logistic regression was one of the most widely used statistical tools for building clinical prediction models, we compared its prediction performance with gradient boosting classifiers using the same training and evaluation sets. In addition, we performed Ridge regression, which was more suited to correlated microbiome features by adding an L2 penalty term to the loss function, following consistent data partitioning strategies. The Ridge regression was optimized by a fine grid search of parameters with cross-validation of the same divisions of folds as the gradient boosting classifier.

Reference models with conventional methods

Currently, prediction models for liver disease are commonly built by regression methods of conventional risk factors. Therefore, reference models were built using logistic regression of commonly used liver disease predictors including age, gender, BMI (kg/m^2), WHR, alcoholic consumption (g), smoking status, TRIG (mmol/l), GGT (U/L), HDL and LDL cholesterol (mmol/l), as a benchmark procedure.

Model evaluation

The prediction performance of all models was evaluated in the corresponding withheld validation dataset (30% of samples) that were not used for discovery. The area under the receiver operating characteristic curve (AUROC) was used to compare the performance across models of different methods and features. Differences in prediction performance were tested with one-sided Wilcoxon-signed rank test with continuity correction. The AUROC is a widely applied metric that considers the trade-offs between sensitivity and specificity at all possible thresholds for comparing the performance across various classifiers with a baseline value of 0.5 for a random classifier. Area under the precision-recall curve (AUPRC) was provided as a complementary assessment, particularly when constructing risk models combining microbiome and conventional risk factors. AUPRC considers the trade-offs between precision (or positive predictive value) and recall (or sensitivity) with a baseline that equals the proportion of positive disease cases in all samples. Since AUPRC is more sensitive to higher ranks of the positive class, it is preferred for highly imbalanced datasets where, for example, case numbers are small relative to controls. As the entire model development process was repeated 10 times, following the 10 randomly sampled partitions of training and validation datasets, each data partitioning led to a set of optimal models developed in the corresponding training dataset. The final performance of optimal models developed from discovery data was evaluated in the corresponding validation data that were set apart in the beginning. The average results of data partitions were reported. To further assess the final prediction result, we considered the species-level microbiome models using gradient boosting classifiers, which outperformed microbiome-only models based on other taxonomic levels for both LD and ALD. In the withheld validation datasets of various partitions, Cox regression models of conventional predictors and in combination with predicted scores of microbiome-only models were built using the time difference between baseline and follow-up disease occurrence or the end of follow-up. The Cox models were evaluated by the concordance statistic (c-statistic). The fit of the model was assessed by likelihood ratio test.