A blood-based biomarker panel (NIS4) for non-invasive diagnosis of non-alcoholic steatohepatitis and liver fibrosis : a prospective derivation and global validation study

NIS4 for non-invasive diagnosis of nonalcoholic steatohepatitis and liver fibrosis

Stephen A Harrison, MD[1*], Vlad Ratziu, MD[2*], Jérôme Boursier, MD[3], Sven Francque, MD[4], Pierre Bedossa, MD[5], Zouher Majd, PhD[6], Geneviève Cordonnier, PhD[6], Fouad Ben Sudrik, MSc[6], Raphael Darteil, PhD[6], Roman Liebe, PhD[6], John Brozek, MSc[6], Alice Roudot, PharmD[6] , Bart Staels, PhD[7], Dean Hum, PhD[6], Sophie Jeannin Megnien, MD[6¥], Suneil Hosmane, PhD[6], Noémie Dam, PhD[6], Pierre Chaumat, MSc[6], Rémy Hanf, PhD[6¶], Quentin M Anstee, FRCP[8§], Arun J Sanyal, MD[9§]

[*]Co-first author; [¥]Current affiliation: Summit Clinical Research, San Antonio, Texas, United States; [¶]corresponding author; [§]co-senior author

Affiliations:

[1]Pinnacle Clinical Research, San Antonio, TX, United States; Radcliffe Department of Medicine, University of Oxford, Oxford, United Kingdom

[2]Sorbonne Université, Institute for Cardiometabolism and Nutrition, Hôpital Pitié-Salpêtrière, Paris, France

[3]Hepatology Department, Angers University Hospital and HIFIH Laboratory, Angers University, Angers, France

[4]Department of Gastroenterology and Hepatology, Antwerp University Hospital; Translational Sciences in Inflammation and Immunology & InflaMed Consortium of Excellence, University of Antwerp, Antwerp, Belgium

[5]Liverpat, Paris, France; Institute of Translational and Clinical Research, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom

[6]GENFIT, Loos, France

[7]Univ. Lille, INSERM, CHU Lille, Institut Pasteur de Lille, Lille, France

[8]Institute of Translational and Clinical Research, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom; Liver Unit, Newcastle upon Tyne Hospitals NHS Trust, Freeman Hospital, Newcastle upon Tyne, United Kingdom

[9]Division of Gastroenterology, Hepatology and Nutrition, Virginia Commonwealth University School of Medicine, Richmond, Virginia, United States

Corresponding author:

Rémy Hanf

GENFIT, Parc Eurasante, 885 Avenue Eugène Avinée, 59210 Loos, France

Email: remy.hanf@genfit.com

Phone: +33(0)3 2016 4000

RESEARCH IN CONTEXT

Evidence before this study

Non-alcoholic steatohepatitis (NASH) is currently the most common chronic liver disease and carries an increasing toll on hepatic morbidity and mortality. The clinical outcomes of NASH constitute a rapidly growing public health concern due to the increasing prevalence of obesity and type 2 diabetes, two conditions closely associated with this disease. Early identification of patients with NASH and fibrosis (fibrosis stage [F] $\geq$2) is essential, as they are at higher risk of disease progression. Currently, conclusive information on NASH activity and fibrosis stage can only be obtained through liver biopsy, which is associated with limitations that significantly contribute to the underdiagnosis of this disease and represents a barrier to treatment access.

Added value of this study

We have developed a blood-based diagnostic test (NIS4) that quantitatively measures four independent biomarkers to produce a score that identifies, among patients with metabolic risk factors, those with "at risk NASH" (defined as NAFLD [nonalcoholic fatty liver disease] activity score [NAS] $\geq$4 and F $\geq$2), who are at higher risk of disease progression. To validate its diagnostic performance, NIS4 was tested in independent patient cohorts selected by similar criteria to that of the future intended use population. The NIS4 test showed a robust diagnostic performance across multiple clinically relevant subpopulations, and was neither influenced by nor dependent on patients' age, gender, body mass index, transaminase levels, or metabolic comorbidities as compared to other diagnostic approaches.

Implications of all the available evidence

NIS4 is the first molecular diagnostic test developed to specifically identify patients with at risk NASH, who are at high risk of progressive liver disease, without recourse to liver biopsy. This non-invasive test is expected to increase the rate of diagnosis of patients with potentially deleterious outcomes and thereby benefit those in need for specific management, including regular monitoring and

pharmacotherapy. The availability of a simple to perform, blood-based test will ultimately help increase disease awareness for NASH.

SUMMARY

Background: Non-invasive tests that allow identification of patients with NASH at high risk of disease progression are lacking. We aimed to establish a blood-based test to help identify patients with at risk NASH (NAS ≥4 and F ≥2), currently considered candidates for pharmacological intervention to prevent disease progression.

Methods: A hypothesis-free, stepwise regression analysis identified a biomarker algorithm that minimised panel size and assay complexity, while maintaining high diagnostic accuracy for the identification of at risk NASH in patients with metabolic risk factors. Blood samples, clinical data, and paired biopsies from three independent NAFLD cohorts were used to develop and validate the test.

Findings:
Modelling was developed on a full dataset matrix (N=239) and 27 non-colinear discriminating variables were introduced in a logistic regression model. The NIS4 algorithm was identified with the lowest Akaike Information Criterion value and comprised miR-34a-5p, A2M, YKL-40, and HbA1c (AUROC=0.80). NIS4 was retrospectively validated in a multicentre prospectively recruited cohort (N=475; AUROC=0.83) and a monocentric retrospective cohort (N=228; AUROC=0.77). In a META-ANALYSIS cohort (N=714), patients with NIS4 <0.36 were classified as not having at risk NASH with 85% sensitivity, 60% specificity, and a NPV of 80%. Conversely, patients with NIS4 >0.62 were classified as having at risk NASH with 85% specificity, 60% sensitivity, and a PPV of 81%. The NIS4 score significantly outperformed other blood-based tests in the identification of at risk NASH, including ELF (AUROC=0.75) and FIB-4 (AUROC=0.75). The overall diagnostic performance of NIS4 was neither influenced by nor dependent on age, gender, BMI, transaminase levels, or metabolic comorbidities.

Interpretation: NIS4 is a novel, robust, non-invasive test with the potential to serve as a clinical diagnostic in populations with metabolic risk factors for the identification of at risk NASH.

Funding: GENFIT

INTRODUCTION

Nonalcoholic fatty liver disease (NAFLD)—characterised by excessive liver fat accumulation—has a high prevalence worldwide, affecting up to 30% of the general population, and is the leading cause of chronic liver disease. (National Institutes of Health, Dept. of Health & Human Services. Genetics Home Reference. Reviewed Nov 2016; published Nov 2019. Available at https://ghr.nlm.nih.gov/condition/non-alcoholic-fatty-liver-disease. Accessed Nov 30, 2019; Younossi Z, et al. *Nat Rev Gastroenterol Hepatol*. 2018;15:11-20; Wu S, et al. *Sci Rep*. 2016;6:33386) Nonalcoholic steatohepatitis (NASH), the most severe form of NAFLD, is characterised by the presence of hepatocyte ballooning and inflammation and can progress silently toward cirrhosis, precluding the opportunity for clinicians to diagnose and intervene therapeutically. (Brodosi L, et al. *Ann Hepatol*. 2016;15(5):673-678; Chalasani N, et al. *Hepatology*. 2018;67(1):328-357; Jiang-Hua Zhou J-H, et al. World J Gastroenterol. 2019;25(11):1307-1326) NASH is a growing cause of cirrhosis and hepatocellular carcinoma globally. (Estes et al. *J Hepatol.* 2018;69(4):896-904) Furthermore, NASH is projected to become the leading cause of liver transplantation in the US; it is already the leading cause among women and the second leading cause overall. (Sumida Y, et al. Hepatol Res. 2019 Sep 8. doi: 10.1111/hepr.13425. [Epub ahead of print]; Noureddin M, et al. *Am J Gastroenterol.* 2018;113(11):1649-1659) The clinical burden of NASH also comprises a cardiovascular disease risk greater than that of the general population, even after adjustment for traditional cardiovascular risk factors and metabolic syndrome. (Patil R, et al. *World J Gastrointest Pathophysiol*. 2017;8(2):5158) Sumida Y, et al. Hepatol Res. 2019 Sep 8. doi: 10.1111/hepr.13425. [Epub ahead of print]) In fact, hepatic and cardiovascular mortality account for an overall reduced survival compared with the age and sex-matched general population. (Adams LA, et al. *Gastroenterology.* 2005;129(1):113-21; Söderberg C, et al. *Hepatology*. 2010 ;51(2):595-602; Ekstedt M, et al. *Hepatology*. 2006 Oct;44(4):865-73). Given this clinical scenario, there is a pressing need to identify patients at risk of disease progression who should be considered for therapeutic intervention.

The main histological determinants of the risk for long-term liver outcomes are NASH activity and fibrosis stage (F). NASH activity is assessed by the NAFLD activity score (NAS), a composite index derived from the sum of the scores for macrovesicular steatosis, hepatocyte ballooning, and

lobular inflammation. (Kleiner DE, et al. *Hepatology* 2005;41:1313-21) In studies with paired liver biopsies, steatohepatitis was associated with liver-related outcomes, and higher NAS scores at baseline were correlated with a high probability of fibrosis stage increase after ≥1 year, suggesting that NASH drives disease progression. Fibrosis stage reflects the extent of disease progression toward cirrhosis—F2 (significant fibrosis) or higher increases the risk of liver-related clinical outcomes. (Kleiner DE, et al. *JAMA Netw Open*. 2019;2(10):e1912565; Sanyal AJ, et al. *Hepatology*. 2019;70(6):1913-1927; Angulo P, et al. *Gastroenterology*. 2015;149(2):389-97) Given that the overall disease state is defined by both NASH activity and fibrosis stage, this is the rationale for inclusion of patients with NASH (NAS ≥4) and F ≥2 (referred to as "at risk NASH" in this report) in pharmacological intervention clinical trials. (Sanyal AJ, et al. *Hepatology*. 2011;54(1):344-53)

Liver biopsy is the clinical reference standard for the diagnosis of NASH among patients with clinical risk factors for this disease, such as metabolic disorders (with or without abnormal liver biochemistries) in the absence of alternative causes for steatosis. (EASL-EASD-EASO. *J Hepatol*. 2016;64:1388-402; Ratziu V, et al. *J Hepatol*. 2015;62(1 Suppl):S65-75) The implementation of this diagnostic approach, however, is limited in routine clinical practice by its invasiveness, cost, attendant risks, variability in interpretation, and the restricted number of professionals able to perform and interpret the test. (Castera L, et al. *Gastroenterology*. 2019;156(5):1264-1281; Russo MW, et al. *Hepatology*. 2017;65(1):336-340.) These limitations may preclude the liver biopsy from being a primary diagnostic in such a prevalent disease. Understandably, this will represent a major barrier for patient diagnosis, management, and future treatment access in routine clinical practice.

Currently, there are limited non-invasive diagnostics specifically designed to identify at risk NASH. Many commonly used tests (e.g. , Fibrosis-4 [FIB-4] score, Enhanced Liver Fibrosis [ELF] score, FibroTest™, FibroScan™ (Vallet-Pichard A, et al. *Hepatology* 2007;46:32-36; Parkes J, et al. *J Viral Hepat*. 2011;18(1):23-31; Myers RP, et al. *AIDS*. 2003;17:721-725; Sandrin L, et al. *Ultrasound Med Biol*. 2003;29(12):1705-13)) were originally developed to identify advanced fibrosis (or cirrhosis) in mixed liver etiologies (e.g., hepatitis C virus, hepatitis B virus) and may be influenced by clinical parameters common to NASH. Recent studies have highlighted test performances (area under the receiver operating characteristic curve [AUROC] <0.80) of OWLiver® and FibroTest™ among other non-invasive diagnostics in the

identification of NASH and/or F ≥2, particularly in individuals with type 2 diabetes. (Bril F, et al. *Diabetes Obes Metab.* 2018;20(7):1702-1709; Bril F, et al. *J Investig Med.* 2019;67(2):303-311; Bril F, et al. *Diabetes Care.* 2019 Oct 11. pii: dc191071. doi: 10.2337/dc19-1071. [Epub ahead of print]) Even widely utilised non-invasive techniques such as transient elastography have shown to be influenced by a number of clinical features, including the presence of type 2 diabetes, dyslipidaemia, high waist circumference, elevated aspartate aminotransferase levels, as well as elevated systolic blood pressure at the time of examination. (Bazerbachi F, et al. *Clin Gastroenterol Hepatol*. 2019;17(1):54-64.e1) Nevertheless, transient elastography remains a commonly used technique outside of primary care. (Pandyarajan V, et al. *Gastroenterol Hepatol (N Y).* 2019; 15(7): 357-365) Limitations or confounders of non-invasive tests are of critical importance to future prescribers of NASH diagnostics in order to help assess the right test or tests to use in their patients.

We here report the development and validation of NIS4, a blood-based diagnostic multivariate index test specifically designed to identify at risk NASH. NIS4 quantitatively measures four independent and NASH-associated biomarkers to produce a score that identifies at risk NASH among patients with metabolic risk factors. To anticipate its diagnostic performance, the test was fully validated in independent patient cohorts selected by criteria similar to that of the future intended use population. In addition, NIS4 was analysed across multiple clinical subgroups to assess test robustness and relative clinical performance as compared with commonly used blood-based diagnostic tests in the same patient populations.

METHODS

The study was performed in three independent cohorts of patients from different geographical regions. All patients provided informed consent, and the corresponding studies have been approved by the institutional review boards at each site.

Patient cohorts

GOLDEN-505 was a multicentre international cohort comprising 239 non-cirrhotic patients with biopsy-confirmed NASH and NAS ≥3 (with contribution from all components of the NAS score) who were consecutively included in a Phase 2b trial of elafibranor (NCT01694849). (Ratziu V, et al.

Gastroenterology. 2016;150:1147-59 e5) Clinical data, blood samples, and paired liver biopsy results were collected from all patients at inclusion. This cohort is further characterised in the Supplementary Methods section.

RESOLVE-IT-DIAG was a multicentre international cohort comprising the first 475 patients with suspected NASH who presented with ≥1 metabolic risk factor (type 2 diabetes, dyslipidaemia, hypertension, obesity) and were consecutively screened for potential inclusion into an ongoing Phase 3 NASH clinical trial (NCT02704403). During the screening window, clinical data, blood samples, and paired liver biopsy results were collected from all patients. This cohort is further characterised in the Supplementary Methods section.

The ANGERS cohort was a single-centre (Angers University Hospital, France) retrospective cohort of 228 patients with suspected NAFLD and clinical risk factors for NASH and/or suspicion of significant fibrosis according to abnormal elastography results and/or abnormal liver biochemistry.

Liver biopsy and histological scoring

All liver biopsies were scored based on the NASH Clinical Research Network (CRN) classification system. For the GOLDEN-505 and RESOLVE-IT-DIAG cohorts, histological scoring was centralised and performed by a single trained pathologist at Hôpital Beaujon (Paris, France). For the ANGERS cohort, liver biopsies were performed at the Hepatology Department, Angers University Hospital, read by a local pathologist expert in chronic liver diseases, and scored according to the NASH-CRN classification system. In all cases, pathologists were blinded to clinical and biochemistry data.

Clinical and laboratory assessments

Demographic and clinical data were obtained from patients' medical records. Biochemical analyses were performed in blood samples collected at a median time vs liver biopsy of +1 day (Q1=-56 days, Q3=+40 days) for the GOLDEN-505 cohort, of +13 days (Q1=-39 days, Q3=+26 days) for the RESOLVE-IT-DIAG cohort, and at the time of liver biopsy for the ANGERS cohort. Clinical and laboratory assessments conducted are further characterised in the Supplementary Methods. All

biochemical analyses were performed utilising commercially available kits. Alpha-2 macroglobulin (A2M; Siemens, BN II or Prospect instrument, anti-serum OSAMG15C0502), haemoglobin A1c (HbA1c; Bio-Rad, Variant II or Variant Turbo, Kit # 270-2455), and YKL-40 (also known as chitinase 3-like protein 1 [CHI3L1]; Bio-Techne, Kit # DC3L10) were quantified as directed by the manufacturers' product inserts. Circulating levels of miR-34a-5p (all cohorts) and other microRNAs (miRNAs; GOLDEN-505 and RESOLVE-IT-DIAG cohorts) were selected on the basis of biological plausibility (Szabo G, Csak T. *Dig Dis Sci.* 2016;61(5):1314-24) and from prior Next Generation Sequencing experiments identifying a number of miRNA which were consistently over-expressed in serum of patients with at risk NASH. (Francque S, et al. *J of Hepatology*. 2017.Vol 66, Issue 1, Supplement; S110-S111) A subset of these miRNAs was then quantified through quantitative reverse transcription polymerase chain reaction (RT-qPCR) as described for miRNA-34a-5p in Supplementary Methods. Among upregulated miRNAs in patients with at risk NASH, miR-34a-5p was identified as the most significantly and consistently overexpressed miRNA. (Francque S, et al. *J of Hepatology*. 2017.Vol 66, Issue 1, Supplement; S110-S111) YKL-40 and miR-34a-5p assays' analytical performances were validated at GENFIT Laboratories before use in this study in accordance with Clinical and Laboratory Standards Institute (CLSI) guidelines (EP05A3E, EP06AE). Both YKL-40 and miR-34a-5p assays have been cross-validated by a College of American Pathologists/Clinical Laboratory Improvements Amendments (CAP/CLIA)-certified central laboratory (COVANCE Central Labs).

Modelling approach

A hypothesis-free, stepwise regression analysis was performed to identify a biomarker panel that minimised assay complexity, while maintaining high diagnostic accuracy (AUROC ≥0.80) to identify at risk NASH among patients with metabolic risk factors.

The GOLDEN-505 training cohort was used to develop a minimal biomarker panel to identify at risk NASH among patients with metabolic risk factors biopsied for suspected NAFLD. Briefly, the liver biopsy data were classified to 1 if histology met the definition of at risk NASH (i.e., NAS ≥4 and F ≥2), and 0 otherwise. The modelling was developed on a full dataset matrix comprising 239 patients with no missing values, using a total of 108 different variables (Supplementary Table S1). Univariate

analyses (*t*-test or Wilcoxon test) identified 47 discriminating variables, and a colinearity assessment was performed. In case of colinearity (Pearson coefficient ≥0.6), the more significant variable was retained. A total of 27 discriminating variables were finally introduced in a logistic regression model of backward stepwise selection using Akaike Information Criterion (AIC). The optimal model with the lowest AIC included: miR-34a-5p, A2M, YKL-40, and HbA1c.

C-statistics

The diagnostic performance of individual biomarkers and panels was assessed using area under the receiver operating characteristic curve (AUROC). In each cohort tested, an internal validation was performed by assessing performance in 1000 bootstrap samples and calculation of 95% confidence intervals (CI) that excluded the zero value. Differences in AUROC were assessed using the DeLong test. Accuracy, sensitivity, specificity, positive predictive values (PPV), and negative predictive values (NPV) are provided with 95% CI calculated with the asymptotic formula based on the normal approximation to the binomial distribution.

Subpopulation analyses were performed by comparing AUROC values of each score. In each subpopulation, and for each score, an internal validation was performed by assessing performance in 1000 bootstrap samples, leading to 95% CIs.

NIS4 validation

The diagnostic performance of the GOLDEN-505 cohort-derived algorithm was first validated in two independent cohorts: the RESOLVE-IT-DIAG cohort (N=475) and the ANGERS cohort (N=228). The diagnostic performance of the GOLDEN-505 cohort-derived algorithm was then assessed in a META-ANALYSIS cohort combining GOLDEN-505 and RESOLVE-IT-DIAG cohorts (N=714). The META-ANALYSIS cohort, considered to be a closer representation of the future intended use population of NIS4 (i.e., broader spectrum of patients), was used to refine algorithm coefficients to better define clinically useful cutoffs to rule out or rule in patients with at risk NASH. The diagnostic performance of the refined algorithm was then independently validated in the

ANGERS cohort. In each cohort, an internal validation was conducted by running the diagnostic algorithm in 1000 bootstrap samples and calculating the 95% CI that excluded the zero value.

Comparison of NIS4 vs other blood-based tests

The META-ANALYSIS cohort combining patients from the GOLDEN-505 and RESOLVE-IT-DIAG cohorts (N=714) was used to conduct a head-to-head comparison of the NIS4 score to the FIB-4 score, (Shah AG, et al. *Clin Gastroenterol Hepatol*. 2009;7(10):1104-) NAFLD fibrosis score (NFS), (Angulo P, et al. *Hepatology*. 2007;45(4):846-54) ELF score, (Guha IN, et al. *Hepatology*. 2008;47(2):455-60), BARD, (Harrison SA, et al. *Gut*. 2008; 57(10):1441-7) Aspartate Aminotransferase (AST) to Platelet Ratio Index (APRI), (Wai CT, et al. *Hepatology*. 2003; 38(2): 518-26) FibroTest™, (Ratziu V, et al. *BMC Gastroenterol*. 2006;6:6) and FibroMeter™, (Cales P, et al. *J Hepatol*. 2009;50(1):165-73) in which all scores were calculated from the same patient samples; the META-ANALYSIS cohort was also used to assess clinical performance in a target population and sensitivity analyses in clinically relevant subpopulations with NASH.

RESULTS

Demographic and baseline characteristics of cohorts

The demographic and baseline characteristics of cohorts are shown in Table 1, and the patients' distribution across the histological spectrum of NAFLD is presented in Supplementary Table S2. The mean age and mean BMI across cohorts ranged from 52 to 55 years and 31.31 to 34.07 kg/m^2, respectively. The prevalence of at risk NASH across cohorts ranged from 40% to 55% (GOLDEN-505 cohort: 43.5%; RESOLVE-IT-DIAG cohort: 54.7%; META-ANALYSIS cohort: 51%; ANGERS cohort: 39.5%). Approximately half of patients across cohorts had a NAS score of 4-5 (GOLDEN-505 cohort: 52%, RESOLVE-IT-DIAG cohort: 48%; META-ANALYSIS cohort: 49%; ANGERS cohort: 42%). The distribution of fibrosis stage was similar across cohorts, with fewer patients with cirrhosis compared with other stages. Patients without fibrosis (F0) or with early fibrosis (F1) represented 41% to 51% of patients across cohorts.

Univariate analysis of laboratory parameters associated with at risk NASH

In the GOLDEN-505 cohort, levels of biomarkers of liver injury (alanine aminotransferase [ALT], AST, gamma-glutamyl transferase), insulin resistance (fasting glucose and insulin), and glycaemia (HbA1c, fructosamine) were higher in patients with at risk NASH than in those with less severe disease activity. Levels of biomarkers reflective of apoptosis (cytokeratin 18-M30 and cytokeratin 18-M65) and fibrosis (YKL-40, A2M, procollagen type III N-terminal peptide, and tissue inhibitor matrix metalloproteinase 1) were also higher in patients with at risk NASH. Levels of miR-34a-5p and miR-122-5p were also significantly higher in this patient population. Similar findings were obtained in the RESOLVE-IT-DIAG cohort. The comparison of major biochemical markers, clinical parameters, and histology in patients with vs without at risk NASH in the GOLDEN-505 and RESOLVE-IT-DIAG cohorts is shown in Supplementary Table S3. A biomarker comparative analysis showed miR-34a-5p to be the single most discriminatory biomarker to identify at risk NASH (Figure 1 and Table 2).

Development and validation of a diagnostic signature for at risk NASH

As described in the Methods section, a hypothesis-free, stepwise regression analysis was conducted in the GOLDEN-505 cohort to identify a biomarker algorithm that minimised assay size/complexity, while maintaining high diagnostic accuracy (AUROC $\geq$0.80) to identify patients with at risk NASH (NAS $\geq$4 and F $\geq$2) among individuals with metabolic risk factors. The NIS4 model was selected as the one linked to the lowest AIC value among the different models obtained through the stepwise-backward process and contained miR-34a-5p, A2M, YKL-40, and HbA1c, which collectively showed high discriminatory accuracy to identify at risk NASH. The AUROC (general metric to assess overall diagnostic performance) for NIS4 in the training set was 0.80 (95% CI 0.74–0.85; Figure 2 and Table 3). The AUROC for the GOLDEN-505 cohort-derived NIS4 algorithm was similar across the validation cohorts (Figure 2). Similarly, the accuracy metrics (sensitivity, specificity, PPV, and NPV at a balanced cutoff of 0.43 were comparable in all cohorts tested (Table 3 and Supplementary Table S4). Given the similar diagnostic performances of NIS4 in the GOLDEN-505 and RESOLVE-IT-DIAG cohorts, these two datasets were combined into a META-ANALYSIS cohort (N=714) to increase sample size/statistical power to assess clinical performance and anticipate real-world clinical utility.

In the META-ANALYSIS cohort, optimization of NIS4 had marginal impact on overall diagnostic performance and accuracy metrics, but slightly shifted the balanced cutoff to 0.50. At low (85% sensitivity), balanced, and high (85% specificity) cutoffs, the refined NIS4 algorithm showed comparable accuracy metrics in all cohorts. Notably, the ANGERS cohort served as an independent validation cohort, confirming that refinement of the NIS4 algorithm had no major performance impact upon comparison with the GOLDEN-505 cohort-derived algorithm. In the META-ANALYSIS cohort, the refined NIS4 algorithm performed significantly better (AUROC=0.83) than the individual components of the score: 0.77 for miR-34a-5p, 0.70 for A2M, 0.70 for YKL-40 and 0.65 for HbA1c (Figure 1, Table 2). Significant correlations between the NIS4 score and both fibrosis stage and NASH severity were observed (Supplementary Figure S1). Unless otherwise specified, performance analyses were based on the refined NIS4 algorithm. Mean PPV and NPV metrics for NIS4 as a function of test cut-off configuration and disease prevalence (5% to 70%) were also generated from the META-ANALYSIS cohort to illustrate test performances within different clinical contexts (Supplementary Table S5).

Comparison of NIS4 vs other blood-based scores for the identification of at risk NASH

In the META-ANALYSIS cohort, NIS4 (AUROC=0.83) significantly outperformed other blood marker-based NASH/fibrosis diagnostics (*P* <0.001), including FIB-4 (AUROC=0.75), NFS (AUROC=0.68), ELF (AUROC=0.74), APRI (AUROC=0.75), FibroMeter™ (AUROC=0.73) and FibroTest™ (AUROC=0.69) for the identification of at risk NASH in patients with metabolic risk factors (Figure 3 and Table 4). In addition, while NIS4 was not developed to specifically identify the subpopulation of at risk NASH with advanced fibrosis (F ≥3), it achieved high diagnostic performance compared with other blood-based tests (Table 4). All diagnostic performance comparisons were conducted in the same patients.

In the META-ANALYSIS cohort, 37% (265/714) of patients had a NIS4 <0.36 and were classified as not having at risk NASH with 85% sensitivity, 60% specificity, and a NPV of 80%. Of patients with NIS4 <0.36, 29% (76/265) had a NAS <4 and 75% (199/265) had minimal to no fibrosis (F0/F1). Among individuals with NIS4 <0.36, 20% (54/265) were misclassified and—of these—72% (39/54) had NAS ≥4 and F2, 28% (15/54) had NAS ≥4 and F3, and 0% had cirrhosis. Conversely, 38% (272/714) of patients had a NIS4 >0.62 and were classified as having at risk NASH with 85% specificity, 60% sensitivity, and a PPV of 81%. Among patients with NIS4 >0.62, 93% (252/272) had a NAS ≥4, 51% (139/272) had a NAS ≥6, 86% had stage ≥2 fibrosis, and 54% had stage ≥3 fibrosis.

Among individuals with NIS4 >0.62, 19% (52/272) were misclassified and—of these—6% (3/52) had NAS ≥4 and F0 (no fibrosis), 56% (29/52) had NAS ≥4 and F1, 11% had NAS <4 and F1, 27% (14/52) had NAS <4 and F2-F3, and 0% had cirrhosis.

A NIS4 balanced cutoff of 0.50 optimised for both sensitivity and specificity. Overall, ~75% of patients within the META-ANALYSIS cohort were accurately classified. In the ANGERS cohort 50% (114/227) of patients had NIS4 <0.36 while 34% (78/227) of patients had a NIS4 >0.62. Similar to the META-ANALYSIS cohort, the majority of patients within the ANGERS cohort (~73%) were also properly classified with NIS4. Additional NIS4 accuracy metrics are provided in Table 3 and the histological distribution of patients as a function of NIS4 range is provided in Figure 4 and Supplementary Table S6.

Finally, sensitivity analyses were performed in the META-ANALYSIS cohort to evaluate the performance of NIS4 in specific subpopulations of clinical relevance in NASH. The overall diagnostic performance of NIS4 was the highest in the analysed cohorts, and was neither dependent on (i.e., included as variables in the NIS4 algorithm) nor statistically impacted by patient age within the range studied, gender, BMI, or metabolic comorbidities (Figure 5, Supplementary Figure S2, and Supplementary Table S7). *PNPLA3* status did not appear to impact clinical performance of the blood-based diagnostic test evaluated. The clinical performance of simple scores such as NFS appeared to be impacted by type 2 diabetic status, age, and gender, whereas the FIB-4 test maintained moderate test performance compared with NIS4 in all categories tested and APRI demonstrated higher test performance in female patients than in male patients. Among higher complexity multi-analyte blood-based tests, ELF was numerically impacted by gender, type 2 diabetic status, and presence of a normal ALT. FibroMeter™ was also numerically impacted by type 2 diabetic status, and FibroTest™ appeared to have lower test performance in patients with normal ALT. Further studies with larger patient numbers are needed to confirm these findings.

DISCUSSION

We have developed a blood-based diagnostic test (NIS4) that quantitatively measures four independent and NASH-associated blood-based biomarkers to produce a score intended to help identify at risk NASH. This report describes the development and validation of NIS4 among patients

with suspected NAFLD based on the presence of metabolic risk factors (e.g., type 2 diabetes, obesity, dyslipidaemia, hypertension), and without other causes of chronic liver disease or of steatosis. Importantly, patients with and without increased serum aminotransferase were included. A sizeable portion of the patient population used to develop and validate the test was generated from multiple clinical trial centers across a range of practice settings/specialties including hepatology, gastroenterology, and internal medicine .

The key elements of a biomarker signature include its biological plausibility, robustness of assay performance across different clinical conditions, diagnostic performance for its intended use in the target population, and the potential implications of misclassification. Each individual component of NIS4 has a strong biological plausibility for NASH and fibrosis. Among NIS4 variables, miR-34a-5p was the most discriminating for identification of at risk NASH. As previously published, (Cermelli S, et al. *PLoS One*. 2011;6(8):e23937; Liu XL, et al. *World J Gastroenterol*. 2016;22(44):9844-9852;;) this report showed an association between circulating levels of miR-34a-5p and NASH histology (i.e., NAS, fibrosis stage). While it cannot be excluded that excess adipose tissue may contribute to circulating miR-34a-5p, (Pan Y, et al. *J Clin Invest*. 2019;129(2):834-849; Ortega FJ, et al. PLoS One. 2010 Feb 2;5(2):e9022) overexpression of miR-34a-5p in the liver of NASH patients has been documented, (Cheung O, et al. *Hepatology*. 2008;48(6):1810-20) supporting the hepatic origin of this biomarker. In line with this observation, the discriminating potency of NIS4 to detect at risk NASH was independent of age and comorbidities, including obesity. Several reports suggest a role of miR-34a-5p in the pathogenesis of NASH, notably of steatosis and lipotoxicity, hepatocyte apoptosis, and fibrosis. (Ding J, et al. *Sci Rep*. 2015;5:13729; Castro RE, et al. *J Hepatol*. 2013 Jan;58(1):119-25; Min HK et al. Cell Metab 2012 15(5):665-674) A2M and YKL-40 are both established biomarkers for fibrosis. (Naveau S, et al. *Dig Dis Sci*. 1994;39(11):2426-32; Kumagai E, et al. *Sci Rep*. 2016;6:35282) HbA1c is a classic marker of glycaemic control and insulin resistance, two clinical parameters shown to be fibrosis drivers in NASH. (Sherwani SI, et al. *Biomark Insights*. 2016;11:95-104) While the exact role of YKL-40 in fatty liver disease is not well understood, the cellular source of YKL-40 appears to be activated macrophages, as confirmed by immunofluorescence staining in vivo. (Kumagai E, et al. *Sci Rep*. 2016;6:35282) YKL-40 serum levels were associated with hepatic fibrosis in patients with NAFLD. (Kumagai E, et al. *Sci Rep*. 2016;6:35282)

As the intended use of NIS4 is that of a diagnostic test for detection of at risk NASH in populations with metabolic risk factors, both ruling out individuals at low risk and ruling in individuals at high risk of disease progression are important, albeit in different clinical contexts. The optimal NIS4 test configuration and associated decision threshold (i.e., cutoff) will likely be dependent on multiple factors including the clinical setting (e.g., primary care, specialty care, tertiary care) and resulting pre-test prevalence of the target condition, clinical comorbidities (e.g., percentage of patients with type 2 diabetes), and ethnic composition (e.g., white, Asian, Hispanic), which may influence the prevalence of at risk NASH compared with what was observed in the META-ANALYSIS cohort as simulated in Supplemental Table S5. This cohort comprised two independent cohorts derived from two different multicentre international interventional trials, which captured a large variety of patients suspected to have NASH; however, a bias towards preferential screening of patients perceived as more severe cannot be excluded. Such bias was minimised in the GOLDEN-505 cohort, since in this trial all non-cirrhotic patients with NASH were eligible irrespective of disease severity and fibrosis stage. In the RESOLVE-IT-DIAG cohort, all screened patients were considered, including those who did not have at risk NASH. Lastly, although a relatively low proportion of patients with cirrhosis were included in the META-ANALYSIS cohort, the overall patient distribution across disease activity (NAS 0-8) and fibrosis (F 0-4) can be assumed to be close to the patient population referred to specialists due to suspected disease (i.e., NIS4 intended use population). In line with this assumption, patient distribution across the disease activity and fibrosis spectrum was comparable to that reported in observational patient registries in the US and Europe. (Kleiner DE, at al. *JAMA Netw Open*. 2019; 2;2(10):e19125652019; McPherson S, et al. *J Hepatol.* 2015;62(5):1148-55)

In the META-ANALYSIS cohort, a low NIS4 cutoff (<0.36) ruled out patients who did not have at risk NASH with high sensitivity (85%) and an associated high NPV (80%). NIS4 as a rule out test (i.e., NIS4 values below the low cutoff) is optimal, and would theoretically elicit an even higher NPV in clinical settings where the prevalence of at risk NASH is lower than that observed in the META-ANALYSIS cohort. On the other hand, a high NIS4 cutoff (>0.62) ruled in at risk NASH with a high specificity (85%) and an associated high PPV (81%). Among patients ruled in, >80% would potentially be eligible for therapeutic interventions based on the histological presence of at risk

NASH, while false positives could potentially lead to misdiagnosis and possible mistreatment. As a rule in test, NIS4 (i.e., NIS4 values above the high cutoff) is optimal, and would theoretically elicit an even higher PPV in clinical settings where the prevalence of at risk NASH is higher than in the META-ANALYSIS cohort.

Another important consideration in diagnostic test development is the clinical decision-making that it may inform. While NIS4 >0.62 is strongly suggestive of at risk NASH and sufficiently accurate to drive a clinical decision in most cases, the physician will ultimately decide whether additional assessments are necessary. Notably, converging evidence could be generated by different non-invasive methods such as elastography; however, liver biopsy could still be necessary from a clinical standpoint in certain circumstances (e.g., complex cases). Nevertheless, 75% of patients were correctly classified using NIS4, suggesting that this new biomarker panel may substantially reduce the need for liver biopsies in clinical practice.

We anticipate that patients with NIS4 <0.62 could potentially be retested, as the disease state may change in these individuals. The frequency of retesting would ultimately be influenced by score value, individual patient needs, risk factors for disease progression, physician decision, and supportive evidence of repeat NIS4 testing, but a 1- to 2-year interval can be proposed by analogy with recommendations for non-invasive fibrosis markers.

While the inclusion of HbA1c in NIS4 raises the possibility that diabetes treatment may affect test accuracy, comparison of AUROC performance of individual variables show that, compared with miR-34a-5p, YKL-40, or A2M, HbA1c has a meaningful but limited impact on the overall test performance. However, further research is required to better understand the potential impact of glucose-lowering agents on the potential use of NIS4 to monitor disease stability, progression, or regression in patients with type 2 diabetes.

NIS4 was specifically designed and optimized to identify at risk NASH because, from a clinical practice perspective, these patients are the ones that would be in need for pharmacotherapy once therapies currently in Phase 3 of clinical development become available. Current non-invasive blood-based tests, as well as transient elastography, were initially developed to identify patients with advanced fibrosis (i.e., NASH CRN stages 3 and 4). Unlike NIS4, these tests will therefore miss

patients at the F2 stage, while also not informing on presence/absence of NASH activity. This will result in missing a sizeable fraction of patients that could be candidates for pharmacotherapy. Another important aspect is that NIS4 maintains robust clinical performance irrespective of age, gender, BMI, presence or absence of abnormal liver enzymes, and comorbidities as compared with other non-invasive tests. This is expected to reduce the number of false positives or false negatives and, therefore, the proportion of misclassified individuals.

One limitation of this study is the restricted FibroScan™ data available across GOLDEN-505 and RESOLVE-IT-DIAG cohorts, which precluded direct diagnostic performance comparisons with transient elastography and other imaging modalities; however, as a blood-based test, NIS4 has the potential added benefit of wide availability and not requiring test-specific operator training. An additional limitation is the low proportion of patients with cirrhosis across cohorts.

The components of NIS4 can be measured by widely deployed methodologies, some of which are already in clinical routine use, such as A2M or HbA1c. In this study, YKL-40 was measured with a simple available immunoassay, while miR-34a-5p was measured with a well-known and widely implemented RT-qPCR method.

In summary, we have developed and extensively validated a new blood-based biomarker panel specifically designed to detect at risk NASH (which comprises both NAS ≥4 and F ≥2) in patients with metabolic risk factors. There is a high unmet need to non-invasively identify the subset of individuals who are at higher risk of disease progression and should be considered for future therapeutic intervention. Such non-invasive test would considerably increase the number of NASH patients that could benefit from future NASH therapies, as an invasive biopsy is a major barrier both to large scale diagnosis and access to treatment. In this context, NIS4 has the potential to be deployed within the framework of clinical laboratories, to achieve straightforward integration into future clinical care, and to be more cost-effective and patient accessible than biopsy. In doing so, NIS4 can also help improve the diagnostic rate of NASH, uncover the true prevalence of at risk NASH in patients with suspected disease, and help healthcare providers identify those most in need of future therapeutic interventions.

Acknowledgments

Medical writing assistance was provided by Florencia Schapiro for Syneos Health, and supported by Genfit.

References

Authors: tables and figures start in next page

Tables and Figures

Table 1. Patient characteristics of independent cohorts

	GOLDEN-505 Cohort	RESOLVE-IT-DIAG Cohort	META-ANALYSIS Cohort	ANGERS Cohort
Type or recruiting centres	Hepatology, gastroenterology	Hepatology, gastroenterology, internal medicine, cardiology, infectious disease	Hepatology, gastroenterology, internal medicine, cardiology, infectious disease	Hepatology, gastroenterology
N	239	475	714	228
Sex, male, %	54	45	48	62
Age (years), mean±SD	52 ± 12	55 ± 12	54 ± 12	54 ± 13
BMI (kg/m^2), mean±SD	31.31 ± 4.62	34.07 ± 6.06	33.15 ± 5.77	32.89 ± 6.34
Ethnicity, % (N)				
White	88.70 (212)	70.95 (337)	76.89 (549)	
Hispanic	2.09 (5)	18.32 (87)	12.89 (92)	N/A
Black	3.77 (9)	2.53 (12)	2.94 (21)	
Other	5.44 (13)	8.21 (39)	7.28 (52)	
Type 2 diabetes, %	35	41	39	41
Dyslipidaemia, %	53	50	51	69
Arterial hypertension, %	53	60	58	75
ALT (IU/L), mean±SD	63.59 ± 41.2	64.0 ± 43.9	63.88 ± 42.99	63.54 ± 40.1
AST (IU/L), mean±SD	42.79 ± 26.63	45.97 ± 29.8	44.91 ± 28.81	46.08 ± 40.27
GGT (IU/L), mean±SD	77.14 ± 84.19	73.81 ± 84.06	74.92 ± 84.06	111.1 ± 165.8
ALP (IU/L), mean±SD	76.7 ± 22.24	82.84 ± 31.36	80.74 ± 28.76	79.43 ± 29.45
Glucose (mmol/L), mean±SD	5.94 ± 1.70	6.10 ± 1.79	6.05 ± 1.77	6.92 ± 2.8
HbA1c (%), mean±SD	6.04 ± 0.91	6.18 ± 0.95	6.13 ± 0.94	6.47 ± 1.29
TG (mmol/L), mean±SD	1.89 ± 1.07	1.97 ± 1.15	1.94 ± 1.12	1.99 ± 1.47
TC (mmol/L), mean±SD	4.95 ± 1.11	4.91 ± 1.25	4.92 ± 1.21	5.11 ± 1.28
HDL-C (mmol/L), mean±SD	1.26 ± 0.33	1.24 ± 0.37	1.25 ± 0.35	1.11 ± 0.3
LDL-C (mmol/L), mean±SD	2.84 ± 0.94	2.76 ± 1.02	2.79 ± 0.99	3.11 ± 1.09
Patients with at risk NASH, %	43.5	54.7	51.0	37.7
Fibrosis stage, %				
Stage=0	15	12	13	13
Stage=1	36	29	31	29
Stage=2	27	28	28	27
Stage=3	22	29	27	24
Stage=4	0	2	1	7
Fibrosis stage, mean	1.56	1.80	1.72	1.85
NAS category, %				
0–1	0	8	5	14
2–3	13	10	11	33
4–5	52	48	49	42
≥6	35	34	35	11
NAS, mean	5.0	4.7	4.8	3.5

BMI=body mass index; SD=standard deviation; ALT=alanine aminotransferase; AST=aspartate aminotransferase; GGT=gamma-glutamyl transferase; ALP=alkaline phosphatase; IU=international unit; HbA1c=haemoglobin A1c; TG=triglycerides; TC=total cholesterol; HDL-C=high density lipoprotein cholesterol; LDL-C=low density lipoprotein cholesterol; NAS=nonalcoholic fatty liver disease activity; N/A=not available.

Table 2. Head-to-head comparison of miR-34a-5p diagnostic performance with other single biomarkers to detect patients with at risk NASH (NAS ≥4 and F ≥2) in the META-ANALYSIS cohort (N=714). *P* value for comparison of each individual marker performance with miR-34a-5p according to DeLong's test. ROC curves are shown in Figure 1

Comparator Individual Markers	Number of Patients with Both miR-34a-5p and Comparator Marker Data	Number of Patients with at Risk NASH	AUROC of Comparator Marker (95% CI)	AUROC of miR-34a-5p (95% CI)	*P* Value for Comparison
miR-34a-5p	714	364	-	0.77 (0.73, 0.80)	-
A2M	714	364	0.70 (0.66, 0.74)	0.77 (0.73, 0.80)	0.0093
HbA1c	714	364	0.65 (0.61, 0.69)	0.77 (0.73, 0.80)	<0.0001
YKL-40	714	364	0.70 (0.66, 0.74)	0.77 (0.73, 0.80)	0.0082
CK18-M65	626	358	0.70 (0.66, 0.74)	0.73 (0.69, 0.77)	0.2079
CK18-M30	683	343	0.73 (0.69, 0.76)	0.76 (0.73, 0.80)	0.0563
ALT	714	364	0.66 (0.62, 0.70)	0.77 (0.73, 0.80)	<0.0001
AST	714	364	0.74 (0.70, 0.78)	0.77 (0.73, 0.80)	0.0997
P3NP	711	362	0.74 (0.70, 0.77)	0.77 (0.73, 0.80)	0.1688
TIMP1	710	361	0.69 (0.65, 0.72)	0.76 (0.73, 0.80)	0.0009
Hyaluronic acid	710	362	0.70 (0.66, 0.74)	0.77 (0.73, 0.80)	0.0124
miR-122-5p	462	230	0.63 (0.58, 0.68)	0.76 (0.72, 0.81)	<0.0001

ALT=alanine aminotransferase; AST=aspartate aminotransferase; HbA1c=haemoglobin A1c; CK18=cytokeratin-18; A2M=alpha-2 macroglobulin; P3NP=procollagen type III N-terminal peptide; TIMP1=tissue inhibitor matrix metalloproteinase 1; miR=microRNA; ROC=receiver operating characteristic; NASH=nonalcoholic steatohepatitis; NAS=nonalcoholic fatty liver disease activity score; F=fibrosis stage

Table 3. NIS4 performance metrics to discriminate patients with (NAS ≥4 and F ≥2) or without (NAS <4 and/or F <2) at risk NASH

Cohort	GOLDEN-505 Training	RESOLVE-IT-DIAG Validation	ANGERS Validation	META-ANALYSIS	META-ANALYSIS			ANGERS Validation		
Algorithm	Original	Original	Original	Original	Refined			Refined		
N	239	475	227	714	714			227		
At risk NASH, % (N)	43.5% (104)	54.7% (260)	37.9% (86)	51% (364)	51% (364)			37.9% (86)		
AUROC (95% CI)	0.80 (0.74, 0.85)	0.83 (0.80, 0.86)	0.77 (0.70, 0.82)	0.82 (0.79, 0.85)	0.83 (0.80, 0.86)			0.77 (0.71, 0.83)		
Cutoff	**Balanced**				**Low**	**Balanced**	**High**	**Low**	**Balanced**	**High**
	0.43	0.43	0.43	0.43	0.36	0.50	0.62	0.36	0.50	0.62
Total accuracy, % (95% CI)	74.90 (69.40, 80.39)	74.11 (70.17, 78.04)	70.04 (64.09, 76.00)	74.37 (71.17, 77.57)	72.9 (69.71, 76.23)	74.65 (71.46, 77.84)	72.55 (69.28, 75.82)	71.37 (65.48, 77.25)	72.69 (66.89, 78.48)	71.81 (65.95, 77.66)
Sensitivity, % (95% CI)	74.04 (65.61, 82.46)	76.54 (71.39, 81.69)	69.77 (60.06, 79.47)	75.82 (71.43, 80.22)	85.16 (81.51, 88.82)	73.90 (69.39, 78.41)	60.44 (55.42, 65.46)	77.91 (69.14, 86.68)	70.93 (61.33, 80.53)	58.14 (47.71, 68.57)
Specificity, % (95% CI)	75.56 (68.31, 82.81)	71.16 (65.11, 77.22)	70.21 (62.66, 77.76)	72.86 (68.20, 77.52)	60.29 (55.16, 65.41)	75.43 (70.92, 79.94)	85.14 (81.42, 88.87)	67.38 (59.64, 75.11)	73.76 (66.50, 81.02)	80.14 (73.56, 86.73)
PPV (95% CI)	70.00 (61.44, 78.56)	76.25 (71.08, 81.41)	58.82 (49.27, 68.37)	74.39 (69.95, 78.83)	69.04 (64.77, 73.32)	75.77 (71.32, 80.23)	80.88 (76.2, 85.56)	59.29 (50.23, 68.35)	62.24 (52.65, 71.84)	64.10 (53.46, 74.75)
NPV (95% CI)	79.07 (72.05, 86.09)	71.50 (65.45, 77.54)	79.20 (72.08, 86.32)	74.34 (69.72, 78.97)	79.62 (74.77, 84.47)	73.54 (68.97, 78.10)	67.42 (63.05, 71.79)	83.33 (76.49, 90.17)	80.62 (73.80, 87.44)	75.84 (68.97, 82.72)
LR+ (95% CI)	3.03 (2.23, 4.19)	2.65 (2.15, 3.33)	2.34 (1.76, 3.14)	2.79 (2.34, 3.36)	2.14 (1.88, 2.47)	3.01 (2.49, 3.66)	4.07 (3.1, 5.32)	2.39 (1.85, 3.13)	2.70 (2.00, 3.70)	2.93 (2.02, 4.28)
LR- (95% CI)	0.34 (0.24, 0.47)	0.33 (0.26, 0.41)	0.43 (0.30, 0.59)	0.33 (0.27, 0.40)	0.25 (0.19, 0.32)	0.35 (0.29, 0.41)	0.46 (0.40, 0.53)	0.33 (0.21, 0.48)	0.39 (0.27, 0.54)	0.52 (0.40, 0.67)

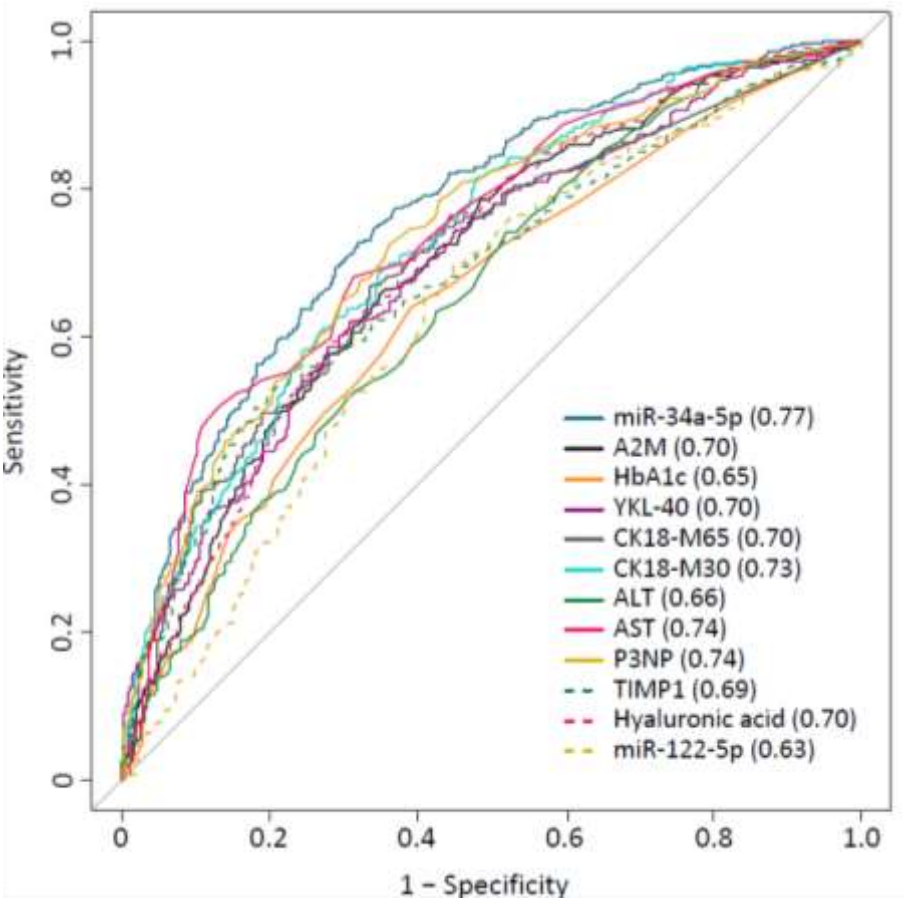
miR=microRNA; A2M=alpha-2 macroglobulin; HbA1c=haemoglobin A1c; NAS=NAFLD activity score; NAFLD=nonalcoholic fatty liver disease; F=fibrosis stage; CI=confidence interval; AUROC=area under receiver operating curve; PPV=positive predictive value; NPV= negative predictive value; LR+=positive likelihood ratio; LR-=negative likelihood ratio; NASH=nonalcoholic steatohepatitis; ROC=receiver operating characteristic.

Table 4. Head-to-head comparison of NIS4 performance with non-invasive NASH/fibrosis blood-based tests to identify patients with at risk NASH (NAS ≥4 and F ≥2) or (NAS ≥4 and F ≥3) within the META-ANALYSIS cohort (N=714); NIS4 and NASH/fibrosis tests were compared in the same patient samples. Samples with missing values for the reference test were not considered. Paired DeLong's test was used for comparison between AUROC values

Comparator Biomarker Panel	Number of Patients with Both Comparator and NIS4 Data	Number of Patients with Condition	Score AUROC (95% CI)	NIS4 AUROC (95% CI)	*P* value
To Identify at Risk NASH (NAS ≥4 and F ≥2)					
NIS4	714	364	-	0.83 (0.80, 0.85)	-
FIB-4	710	364	0.75 (0.71, 0.78)	0.83 (0.80, 0.86)	<0.0001
NFS	702	357	0.69 (0.65, 0.73)	0.83 (0.80, 0.86)	<0.0001
ELF	708	361	0.75 (0.71, 0.78)	0.83 (0.80, 0.86)	<0.0001
BARD	714	364	0.59 (0.55, 0.63)	0.83 (0.80, 0.85)	<0.0001
APRI	710	364	0.75 (0.71, 0.78)	0.83 (0.79, 0.86)	<0.0001
FibroTest™	708	361	0.69 (0.65, 0.72)	0.83 (0.79, 0.86)	<0.0001
FibroMeter™	544	345	0.74 (0.70, 0.78)	0.79 (0.75, 0.83)	0.0144
To Identify at Risk NASH (NAS ≥4 and F ≥3)					
NIS4	714	182	-	0.80 (0.77, 0.84)	-
FIB-4	710	182	0.75 (0.71, 0.79)	0.80 (0.77, 0.84)	0.0199
NFS	702	179	0.71 (0.67, 0.76)	0.80 (0.77, 0.84)	0.0005
ELF	708	182	0.76 (0.72, 0.80)	0.80 (0.76, 0.84)	0.0415
BARD	714	182	0.64 (0.60, 0.69)	0.80 (0.77, 0.84)	<0.0001
APRI	710	182	0.74 (0.70, 0.78)	0.80 (0.77, 0.84)	0.0019
FibroTest™	708	180	0.69 (0.65, 0.73)	0.80 (0.76, 0.84)	<0.0001
FibroMeter™	544	171	0.70 (0.65 , 0.74)	0.76 (0.71, 0.80)	0.0140

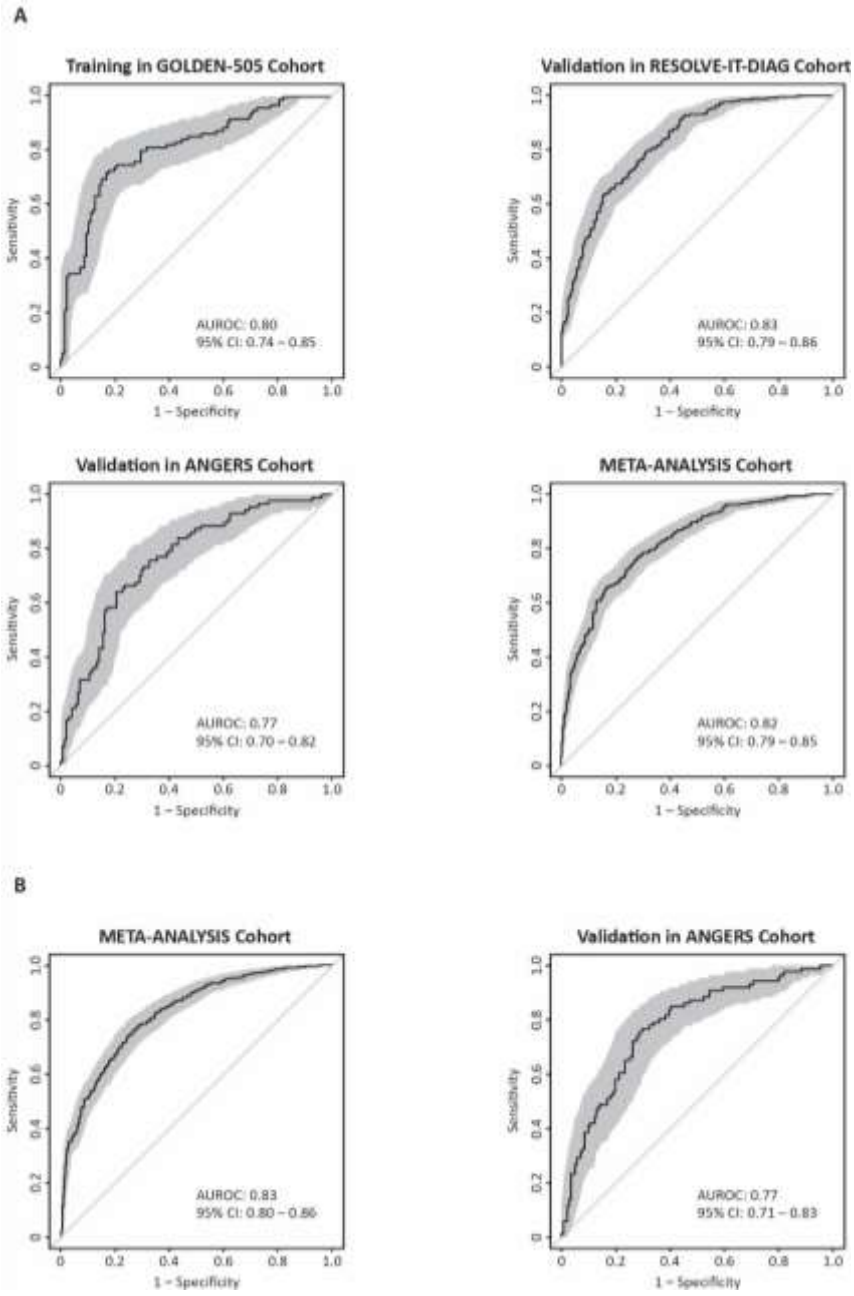
NASH=nonalcoholic steatohepatitis; NAS=NAFLD activity score; NAFLD=nonalcoholic fatty liver disease; F=fibrosis stage; CI=confidence interval; AUROC=area under receiver operating curve; FIB-4=Fibrosis-4; NFS=NAFLD Fibrosis Score; ELF=Enhanced Liver Fibrosis; APRI=AST to Platelet Ratio Index; AST=aspartate aminotransferase; ROC=receiver operating characteristic.

Figure 1. Comparison of miR-34a-5p with serum biomarkers to discriminate patients with at risk NASH (NAS ≥4 and F ≥2) in the META-ANALYSIS cohort (N=714). AUROC values are provided in parentheses



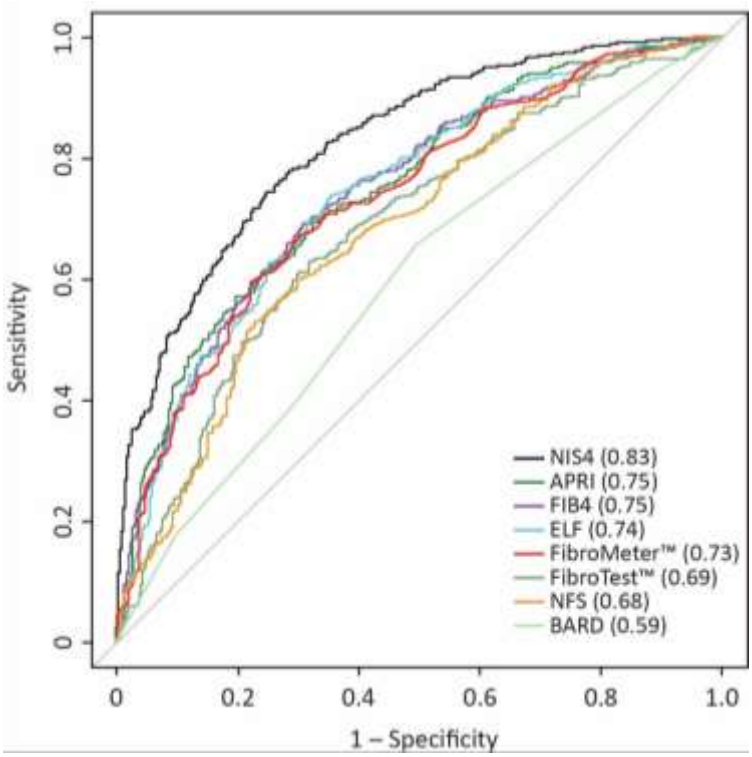
Footnote: NASH=nonalcoholic steatohepatitis; NAS=NAFLD activity score; NAFLD=nonalcoholic fatty liver disease; F=fibrosis stage; AUROC=area under receiver operating curve; A2M=alpha-2 macroglobulin; HbA1c=haemoglobin A1c; CK18=cytokeratin 18; ALT=alanine aminotransferase; AST=aspartate transaminase; P3NP=procollagen type III N-terminal peptide; TIMP1= tissue inhibitor matrix metalloproteinase 1.

Figure 2. Training and validation of NIS4 in the identification of at risk NASH (NAS ≥4 and F ≥2). **A.** GOLDEN-505 cohort-derived NIS4: ROC curves of NIS4 algorithm in the GOLDEN-505 training cohort (N=239), RESOLVE-IT-DIAG validation cohort (N=475), ANGERS validation cohort (N=227), and META-ANALYSIS cohort (N=714); **B.** Refined NIS4: ROC curves of refined NIS4 algorithm in the META-ANALYSIS cohort (N=714), and ANGERS validation cohort (N=227). Shaded areas represent 95% CI obtained after analyses of 1000 bootstrap samples



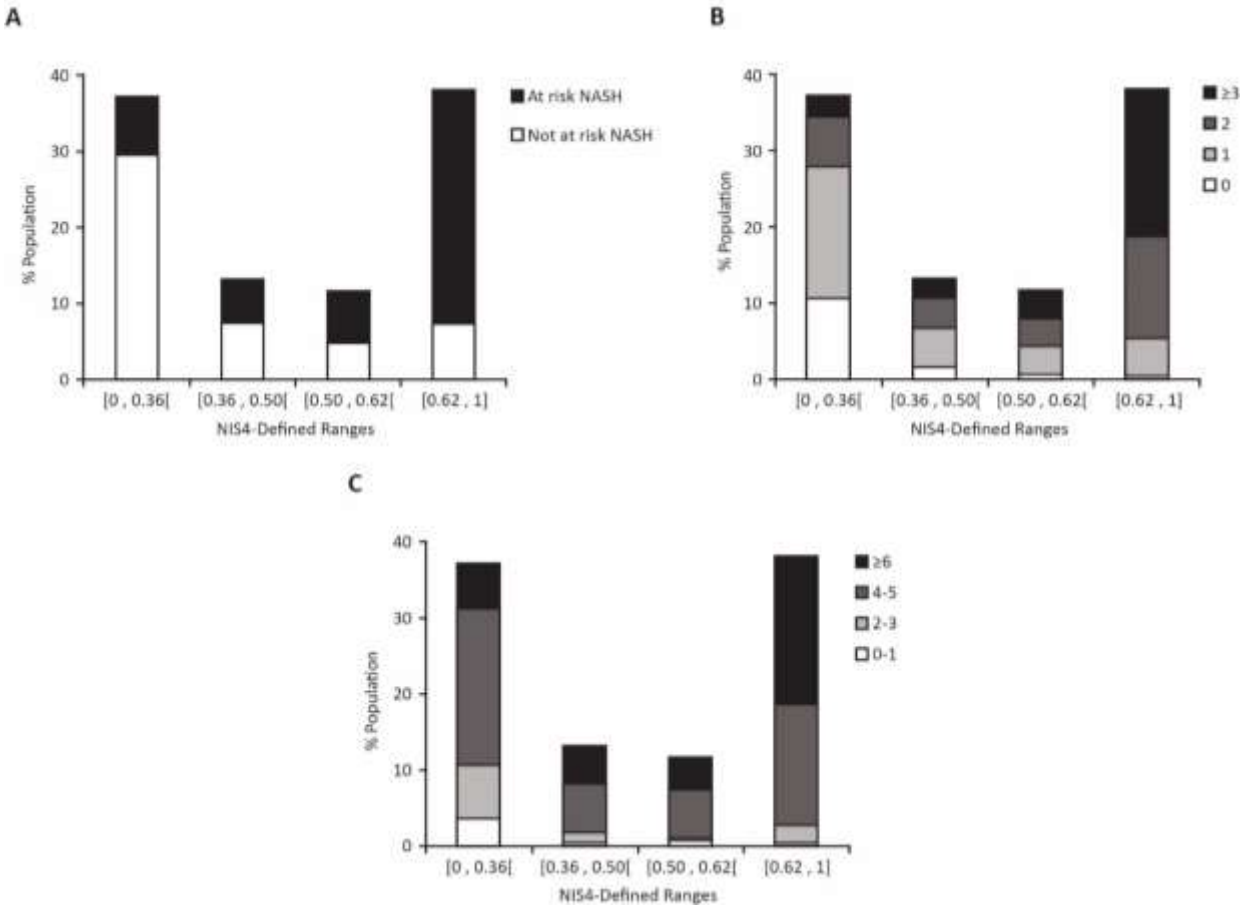
Footnote: NASH=nonalcoholic steatohepatitis; NAS=NAFLD activity score; NAFLD=nonalcoholic fatty liver disease; F=fibrosis stage; ROC=receiver operating characteristic; AUROC=area under receiver operating curve; CI=confidence interval.

Figure 3. Comparison of ROCs and AUROCs obtained in the META-ANALYSIS cohort (N=714) for NIS4 and other blood-based diagnostic scores for identification of patients with at risk NASH (NAS ≥4 and F ≥2). AUROC values are provided in parentheses



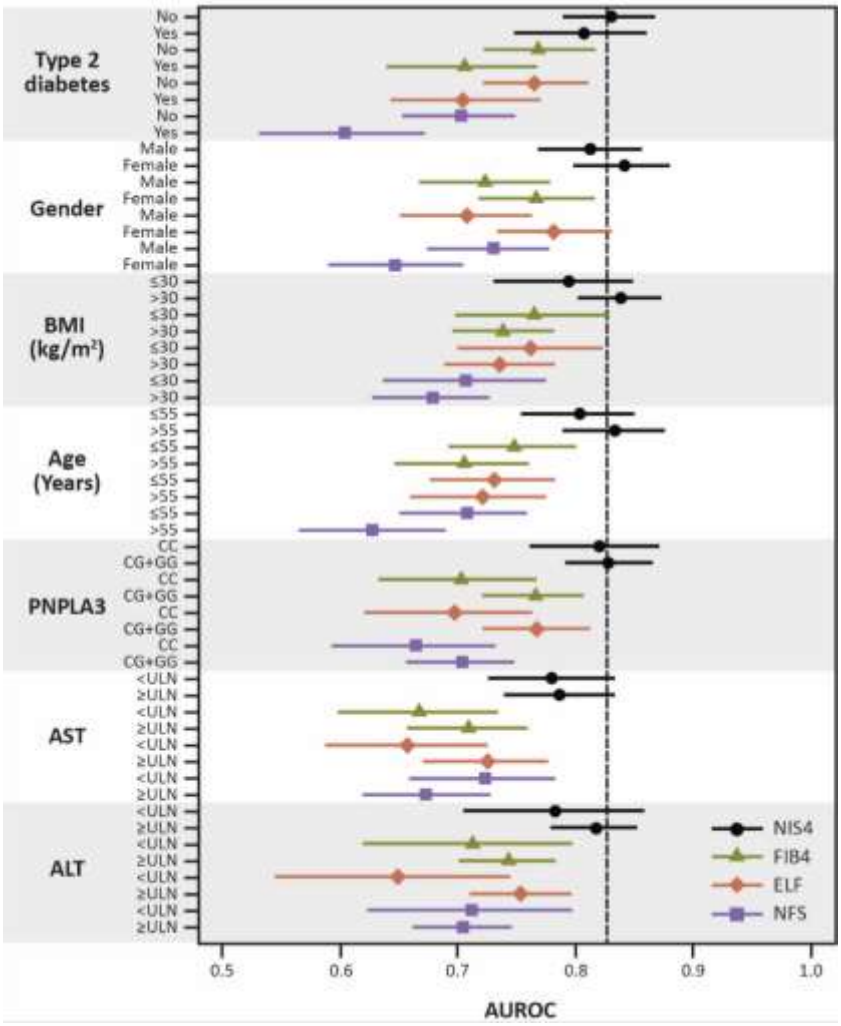
Footnote: NASH=nonalcoholic steatohepatitis; NAS=NAFLD activity score; NAFLD=nonalcoholic fatty liver disease; F=fibrosis stage; ROCs=receiver operating characteristics; AUROCs=area under receiver operating curves; APRI=AST to Platelet Ratio Index; AST=aspartate aminotransferase; FIB4=Fibrosis-4; NFS=NAFLD Fibrosis Score; NAFLD=nonalcoholic fatty liver disease; ELF=Enhanced Liver Fibrosis.

Figure 4. Distribution of histological features in NIS4-defined categories within the META-ANALYSIS cohort (N=714): NIS4=0.36 (low cutoff), NIS4=0.50 (balanced cutoff), and NIS4=0.62 (high cutoff). Results are expressed as percentage of the total population. **A.** Distribution of patients with at risk NASH (NAS ≥4 and F ≥2) across NIS4-defined ranges; **B.** Fibrosis stage distribution across NIS4-defined ranges; **C.** NAS severity distribution across NIS4-defined ranges



Footnote: NASH=nonalcoholic steatohepatitis; NAS=NAFLD activity score; NAFLD=nonalcoholic fatty liver disease; F=fibrosis stage.

Figure 5. Subpopulation AUROC analysis: AUROC of NIS4, FIB4, ELF, and NFS for identification of patients with at risk NASH (NAS ≥4 and F ≥2) in subpopulations of the META-ANALYSIS cohort (N=714). The dashed vertical line represents the AUROC in the total population (0.83). Horizontal lines represent 95% CI



Footnote: AUROC=area under receiver operating curve; FIB4=Fibrosis 4; ELF=Enhanced Liver Fibrosis; NFS=NAFLD Fibrosis Score; NAFLD=nonalcoholic fatty liver disease; NASH=nonalcoholic steatohepatitis; NAS=nonalcoholic fatty liver disease activity score; F=fibrosis stage; CI=confidence interval; BMI=body mass index; PNPLA3=patatin-like phospholipase domain-containing protein 3; AST=aspartate aminotransferase; ALT=alanine aminotransferase.