DR. PABLO JOSÉ GIRAUDI (Orcid ID : 0000-0003-2852-6648)

Article type : Original Articles

Editor : Helena Cortez-Pinto

# A simple *in silico* strategy identifies candidate biomarkers for the diagnosis of liver fibrosis in morbidly obese subjects

Pablo J. Giraudi[1], Sabrina E. Gambaro[1], Sofia Ornelas Arroyo[1], Carla M.Chackelevicius[1], Michela Giuricin[2], Marta Silvestri[2,6], Daniele Macor[3], Lory S. Crocé[3], Deborah Bonazza[4], Giorgio Soardo[5], Nicolò de Manzini[2,6], Fabrizio Zanconati[4], Claudio Tiribelli[1], Silvia Palmisano[2,6] and Natalia Rosso[1]

[1] Fondazione Italiana Fegato, Centro Studi Fegato, Area Science Park Basovizza Bldg.Q SS14 Km 163.5, Trieste, Italy.
[2] Università degli Studi di Trieste, Ospedale di Cattinara, Chirurgia Generale
[3] Università degli Studi di Trieste, Clinica Patologie del fegato, Dip. Medicina Ospedale Cattinara
[4] School of Anatomic Pathology, University of Udine and Trieste; Department of Medical, Surgical and Health Sciences, University of Trieste
[5] Dipartimento di Scienze Mediche Sperimentali e Cliniche, Azienda Ospedaliero Universitaria Santa Maria della Misericordia di Udine
[6] Università degli Studi di Trieste, Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute.

Corresponding author:
Pablo J. Giraudi
Area Scienze Park, Basovizza, Ed Q, SS14 Km 163.5, 34012, Trieste, Italy
Phone: +39 040 375 7923
Email: pablo.giraudi@fegato.it

**Electronic** Number of figures and tables: 4 figures, 3 tables

Disclosures: No authors report conflicts of interest/financial-disclosures

Financial Support:

Finanziamento Ricerca di Ateneo - U05SPFRA14 - FRA 2014 (CdA dd. 19.12.2014) and Fondazione Italiana Fegato. PJG by Fondazione Umberto Veronesi (Grants 2015 and 2016), SEG by Project 297 nutrizione 297 CTGAS (CUPB91C12000000001) and CMC by the Italian Ministry of Foreign Affairs.

List of Abbreviations:

NAFLD, non-alcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis; PPI, protein-protein interactions; MO, morbidly obese; BMI, body mass index; AST, aspartate aminotransferase; ALT, alanine aminotransferase; GGT, γ-glutamyltransferase; TAG, triglyceride; HDL, high density lipoprotein; LDL, low density lipoprotein; HbA1c, glycated hemoglobin; FIB-4, fibrosis-4 score; APRI, AST to platelet-ratio-index; CD44, CD44-antigen; SPARC, secreted protein acidic and rich in cysteine; EGFR, epidermal growth factor receptor; IGF2, insulin-like growth factor 2.

**Abstract**

Background and aims

Non-alcoholic fatty liver disease (NAFLD) is a chronic liver disorder, tightly associated with obesity.

The histological spectrum of the disease ranges from simple steatosis to steatohepatitis, with different stages of fibrosis, and fibrosis stage is the most significant predictor of mortality in NAFLD. Liver biopsy continues to be the gold standard for its diagnosis and reliable non-invasive diagnostic tools are unavailable. We investigated the accuracy of candidate proteins, identified by an *in silico* approach, as biomarkers for diagnosis of fibrosis.

Methods

71 morbidly obese (MO) subjects with biopsy-proven NAFLD were enrolled, and the cohort was subdivided according to minimal (F0/F1) or moderate (F2/F3) fibrosis. The plasmatic level of CD44 antigen (CD44), secreted protein acidic and rich in cysteine (SPARC), epidermal growth factor receptor (EGFR) and insulin-like growth factor 2 (IGF2) were determined by ELISA. Significant associations between plasmatic levels and histological fibrosis were determined by correlation analysis and the diagnostic accuracy by the area under receiver operating characteristic curves (AUROC).

Results

82% of the subjects had F0/F1 and 18% with F2/F3 fibrosis. Plasmatic levels of IGF2, EGFR and their ratio (EGFR/IGF2) were associated with liver fibrosis, correlating inversely for IGF2 ($p<0.006$) and directly ($p<0.018$; $p<0.0001$) for EGFR and EGFR/IGF2, respectively. The IGF2 marker had the best diagnostic accuracy for moderate fibrosis (AUROC 0.83), followed by EGFR/IGF2 ratio (AUROC 0.79) and EGFR (AUROC 0.71).

Conclusions: Our study support the potential utility of IGF2 and EGFR as non-invasive diagnostic biomarkers for liver fibrosis in morbidly obese subjects.

**Abstract Electronic** Keywords: liver fibrosis, biomarkers*, in silico* strategy, morbid obese

**Key points**

1) Fibrosis stage is the most significant predictor of mortality in non-alcoholic fatty liver disease. The diagnostic gold standard is liver biopsy. Reliable non-invasive diagnostic tools are unavailable.
2) Our *in silico* strategy identifies novel biomarkers to diagnose early stages of fibrosis.
3) Plasmatic levels of candidates allowed discriminating moderate fibrosis in morbidly obese subjects.

4) Our data would be included in an algorithm with others serum markers to reduce the need for liver biopsy.

## Introduction

According to European Health Interview Survey (2016), almost 1 adult in 6 in the EU is considered obese [1]. Morbidly obese (MO) subjects are at particular risk for the development of non-alcoholic fatty liver disease (NAFLD) [2 3]. Several epidemiological studies have linked NAFLD to unhealthy diet and sedentary behaviours [4 5].

NAFLD includes different stages, ranging from simple steatosis to non-alcoholic steatohepatitis (NASH). The latter is characterised by steatosis plus necroinflammation and can have different stages of fibrosis ranging from absent to cirrhosis [4]. Unfortunately, despite the increase in awareness of this disease, there are still no reliable non-invasive diagnostic tests and liver biopsy remains the gold standard. However, it is invasive, complications may occur and require hospitalisation. For these reasons, there is an urgent clinical need to develop non-invasive assays for the staging of liver fibrosis in NAFLD/NASH.

The discovery of new serum biomarkers to be used either separately/combined in a panel of markers could contribute not only to the diagnosis but also to follow-up the progression/remission of the disease. Nevertheless, the identification of novel biomarkers for liver fibrosis can be a daunting work due to the multiple factors involved in disease progression. Currently, the study of the interactome at gene/protein level is possible through the use of high-throughput and bioinformatic tools, such as Cytoscape software [6]. This software allows analysing the biological information about protein-protein interactions (PPI) stored in different molecular databases, such as IntAct, MINT, UniProt, etc. Thus, *in silico*

analysis of biological networks represents an alternative option to elucidate novel biomarkers, as previously described by Page [7] and AbdulHameed [8].

Considering the aforementioned issues, we applied an *in silico* strategy to identify new effective biomarkers for the diagnosis of moderate/advanced liver fibrosis stages. We then assessed their accuracy in a cohort of MO subjects with different stages of fibrosis.

## Materials and Methods

### Patient cohort and study protocol

71 MO subjects undergoing bariatric surgery were prospectively and consecutively enrolled by a multidisciplinary team (surgeons, dieticians, hepatologists and psychiatrists). All consenting patients were included in accordance with the international guidelines: age 18 to 65 years, a body mass index (BMI) of 40 kg/m$^2$ or between 35 and 40 kg/m$^2$ with obesity-related co-morbidities, well-informed and motivated patients with acceptable operative risks, failure of nonsurgical treatments, declared compliance to follow lifelong medical surveillance [9]. Liver biopsy was performed in all subjects at the time of the surgical procedure. The exclusion criteria were: previous diagnosis of others forms of chronic liver disease, including suspected/confirmed hepatocellular carcinoma; alcoholic liver disease (> 25 g/day alcohol consumption) or known HBV, HCV and HIV positivity. MO subjects gave their written informed consent before participating in this study, approved by protocol N. 22979 Local Ethical Committee (Comitato Etico Regionale Unico, FVG, SSN).

In addition, blood samples from informed consenting healthy lean subjects and from patients with F3 NASH and advanced metabolic-related cirrhosis were included in the study, and considered as negative and positive controls, respectively.

**Clinical-Biochemical Assessment**

Anthropometric parameters, such as age, sex (M/F) and BMI ($kg/m^2$) were scored during the baseline visit. Blood samples were collected after overnight fasting for the further assessment of liver biochemistry, glucose, glycated haemoglobin (HbA1c) and lipids. The homeostatic model assessment for insulin resistance (HOMA-IR) was calculated as described by Matthews [10]. Diabetes was diagnosed according to the ESC-EASD guidelines [11]. Surrogate markers' scores of liver fibrosis were calculated as described by Sumida for FIB-4 [12], Calès for APRI and FibroMeter [13] and Harrison for BARD index [14].

**Liver biopsy and histopathology**

Liver wedge biopsies were performed on the left lobe and two pathologists interpreted them. Steatosis was graded according to the amount of fat present in the hepatocytes on haematoxylin/eosin staining. Biopsies showing no or minimal (<5%) steatosis and absent injury or fibrosis were considered as normal. The samples that showed more than 5% steatosis were labelled as NAFLD. The histological diagnosis of NASH and fibrosis was made in accordance with Kleiner-Brunt criteria [15 16].

***In silico* biological network analysis**

To obtain biological networks for each gene/protein of interest involved in fibrogenesis, we used Cytoscape [6]. UniProtKB identifiers and protein-protein interaction (PPI) data from curated databases were used in network creation [17]. Each biological network is constituted by a central node (for example cytokeratin-18 (CK-18)) that interacts with protein partners. Central nodes correspond to the proteins involved in the fibrotic process, and their relevant information is reported in STable 1 (Supporting information). In summary, the selected proteins, used to obtain each biological network, can be categorized into three

different groups based on their role in the NAFL disorder: a) those reported as putative markers for the progression of the disease, such as CK-18, adipocyte fatty acid binding protein (AFABP), fibroblast growth factor 21 (FGF21), insulin-like growth factor-binding protein 3 (IBP-3) and lymphocyte cytosolic protein 1 (LCP-1) [18 19 20]; b) those reported to be involved in phenotype modulation (activation/reversion) of hepatic stellate cells (HSC), such as galectin-1 (GALS1), ubiquitin conjugation factor E4B (UBE4B), vitronectin (VTN) and alpha smooth muscle actin (α-SMA), laminin subunit beta 1 (LAMB1) [21]; and c) those involved in extracellular matrix remodeling such as osteopontin (SPP1), collagen alpha-1 (III) chain (COL1A3), matrix-metalloproteinase-2 (MMP2) and tissue inhibitor of metalloproteinase-2 (TIMP2) [22 23].

Once defined the network for each protein the total network was generated by the integration of each singular one, the layout is presented in Fig. 1.

**Plasma CD44, SPARC, IGF-2 and EGFR**

Plasmatic levels of candidate biomarkers were measured by ELISA commercial kits (further details in Supporting information).

**Statistical analysis**

Continuous variables were expressed as mean ± (standard deviation) or median (interquartile range), and categorical as numbers or percentages. Categorical variables were analyzed using chi-square tests with correction, when appropriate. Independent *t*-test and ANOVA were used for normally distributed continuous variables. Non-parametric tests (Mann–Whitney and Kruskal–Wallis with post-hoc analysis) were applied for continuous variables that failed to pass D'Agostino&Pearson omnibus normality test.

Correlation analysis was performed using Pearson or Spearman's correlation coefficients
to estimate the association of plasmatic candidates' levels and several factors of interest. Statistical analysis was performed by using GraphPad Prism 5.01. Multivariable analysis using multiple linear regression models were performed to determine the independent factors associated with candidates plasmatic levels, using GraphPad Instat3.

The diagnostic performance of candidates was assessed by receiver operating characteristic (ROC) curves. The area under the ROC (AUROC) was used to compare the accuracy between different fibrosis diagnostic tests. The sensitivity, specificity, positive predictive values (PPVs) and negative predictive values (NPVs) for relevant cut-offs were also calculated. ROC analysis was performed using MedCalc Statistical Software 16.4.3.

**Results:**

***Subjects' demographic***

The main demographic, clinical and biochemical features of the cohort are reported in Table 1. The MO cohort presented alterations mainly in glucose homoeostasis with abnormal values for fasting glucose, glycated haemoglobin and HOMA-IR (113 ± 25 mg/dL, 6.3 ± 1 μU/mL and 5.3 ± 4 respectively) respect to CTRLs. Approximately, 20% of the subjects had type 2 diabetes. No differences in cholesterol and triglycerides were observed among groups. As expected, cirrhosis group had altered levels of GGT, and altered glucose homoeostasis. The hepatic histological features of MO cohort are shown in Table 1. Briefly, 58 (82%) MO subjects had no significant/minimal (F0/F1) and 13 (18%) significant/moderate (F2/F3) fibrosis.

### *In silico* identification of candidate biomarkers

From the in silico analysis, four candidates were selected: insulin growth factor 2 (IGF2), secreted protein acidic and rich in cysteine (SPARC), CD44-antigen (CD44) and epidermal growth factor receptor (EGFR). Those were chosen since they are soluble proteins, supposed to be released in the plasma from the liver, and because they link several central nodes/proteins involved in fibrogenesis. Specifically, IGF2 interacts with VTN (involved in the modulation of HSC phenotype) and IBP-3 (reported as a potential biomarker for NAFLD progression). EGFR links directly three proteins: IBP3 (mentioned above), LGALS1 and UBE4B, the last two involved in the activation of HSC. Thus, following IGF2/EGFR candidates, we are also able to consider eventual variations of the others connected proteins (like VTN, IBP-3, etc). These other proteins, in turn, regulate several biological processes associated with the progression of the fibrotic process. Thus, the selection of these candidates improved our probability of obtaining reliable markers to follow liver fibrosis (evidenced in red, Fig. 1 and STable1). To further validate the association of our candidates with fibrosis, gene expression analysis were performed in liver biopsies (STable 2, SFig. 1A and 1B).

### Biochemical data and correlation to fibrosis

Serum levels of AST, ALT and GGT, as well as the AST/ALT ratio, are reported in Fig. 2. AST levels were slightly increased in subjects with cirrhosis respect to the F0/F1 in MO cohort (41.3 ± 24.4 *vs.* 23.3 ± 12.7, $p<0.01$; Fig. 2A). Additionally, GGT serum levels were significantly higher in subjects with cirrhosis than in those with F0/F1 and CTRLs (124 ± 77 *vs.* 32.7 ± 23.8 and 24.9 ± 13.2, respectively, $p<0.01$) (Fig. 2B), but no difference was observed between the different stages of fibrosis. ALT levels and the AST/ALT ratio were similar among the groups (Fig. 2C and 2D). When different surrogated indexes of fibrosis (APRI, FIB-4, FibroMeter and BARD) were calculated, the results were not able to

distinguish between fibrosis' stages (SFig. 2–Supporting information), confirming that those indexes do not reflect the stage of hepatic fibrosis in the MO cohort.

**Plasmatic levels of candidate biomarkers and correlation with fibrosis**

The plasmatic concentration of our candidate was determined by ELISA (Fig. 3). IGF2 levels were significantly decreased as fibrosis progress (Fig. 3A). The median IGF2 level in CTRLs and cirrhosis was 6.80 (interquartile range, 4.82 – 9.40) ng/mL and 0.92 (0.79-1.20) ng/mL, respectively ($p<0.001$). Interestingly, IGF2 was able to distinguish F0/F1 from F2/F3 in the MO cohort (2.20 (1.70-2.80) and 1.45 (0.45-1.82), respectively; $p<0.05$).

EGFR levels were significantly increased with the progression of hepatic fibrosis compared to CTRLs (Fig. 3B). The median EGFR levels in subjects with F0/F1, F2/F3 and CTRLs were 110 ng/mL (81.5-125.5), 115 ng/mL (107.4-143.0) and 48.5 ng/mL (44.7-63.0), respectively (F0/F1 and F2/F3 *vs* CTRLs, $p<0.001$). Surprisingly, cirrhosis group showed similar EGFR levels to MO subjects (116.2 ng/mL (90.9 - 127), been only significantly different from lean controls ($p<0.01$).

Regarding the plasmatic levels of CD44, a trend of increase (not statistically significant) was observed with the progression of liver damage. SPARC plasmatic levels did not show any significant variation in the MO cohort (Fig. 3D and 3E).

To enhance the sensitivity of the two informative candidate biomarkers (IGF2 and EGFR), we calculated their ratio (Fig. 3C). The EGFR/IGF2 ratio showed a positive association with the stage of fibrosis and allowed to differentiate subjects with fibrosis (F0/F1, F2/F3) and cirrhosis from CTRLs. The median values were 47.0 (30.5-69.2) for F0/F1, 82.0 (54.0-332.0) for F2/F3, 122.2 (91.6 -137.6) for cirrhosis and 5.9 (3.8-9.5) for CTRLs (F0/F1 *vs.* CTRLs, $p<0.01$; F2/F3 and Cirrhosis *vs.* CTRLs, $p<0.001$). Moreover,

EGFR/IGF2 values were significantly different between subjects with cirrhosis or significant fibrosis (F2/F3) from those with minimal fibrosis (F0/F1) (Cirrhosis *vs.* F0/F1, $p<0.05$; F2/F1 *vs.* F0/F1, $p<0.05$). As expected, candidate biomarkers in plasma did not show any correlation with steatosis in the MO cohort (SFig. 3–Supporting information).

**Diagnosis of liver fibrosis using plasmatic biomarkers in MO subjects**

Pearson's or Spearman's correlation analysis was performed to evaluate the association between the plasmatic levels of our candidates with BMI, lipid profile, HOMA-IR, FLI and liver histological scores for steatosis, inflammation and fibrosis. IGF2 plasmatic level had a significant negative correlation with both lobular inflammation ($p=0.024$) and fibrosis ($p=0.006$) (Table 2 and Supporting information-SFig. 4A and B). On the other hand, EGFR correlates negatively with total cholesterol ($p=0.032$) and positively with lobular inflammation ($p=0.021$) and fibrosis ($p=0.018$) (Table 2 and Supporting information-SFig. 4C, 4D and 4E). EGFR/IGF2 ratio showed a positive correlation with lobular inflammation ($p=0.027$) and fibrosis ($p<0.0001$) (Table 2 and Supporting information-SFig. 4F and 4G). CD44 and SPARC plasmatic levels showed no correlation with the parameters under analysis (STable 3-Supporting information).

When multivariable analysis using a multiple linear regression model was applied, fibrosis was the main contributor associated with IGF2 plasmatic levels and with the EGFR/IGF2 ratio, while, all factors equally contributed in the case of EGFR (STable 4-Supporting information).

The diagnostic accuracy for liver fibrosis of our candidates was compared with those of scoring systems such as FIB-4 and Fibrometer by using AUROC analysis (Fig. 4). IGF2 showed the best diagnostic accuracy for significant/moderate fibrosis (AUROC 0.83),

followed by the EGFR/IGF2 ratio (AUROC 0.79), EGFR (AUROC 0.71), FibroMeter score (AUROC 0.64) and FIB-4 score (0.63). The sensitivity, specificity, PPVs and NPVs of each test for optimal cut-off values are reported in Table 3. Overall, IGF2 had the highest accuracy in detecting significant fibrosis while FIB-4 score had the lowest. At the optimal threshold of 1.9 ng/mL, IGF2 had a 86% sensitivity and 74% specificity in our MO cohort. Thus, IGF2 and EGFR or their ratio had a higher specificity and sensitivity than the currently used surrogate indexes based on routine laboratory tests.

**Discussion**

Since the available non-invasive tools for the diagnosis of NAFLD/NASH are still inconclusive, we prospectively explored the reliability of putative biomarkers for the detection of liver fibrosis in a cohort of severely obese individuals. In line with the range reported by other studies in bariatric subjects [24 25], we observed a high prevalence of NASH (62%), 82% with minimal fibrosis and 18% with moderate fibrosis.

Fibrosis stage was recently established to be the most important prognostic factor for liver-related outcomes and mortality [26]. Even though several simple, non-invasive clinical indexes have been proposed to diagnose fibrosis in subjects with NAFLD (extensively reviewed by Kaswala) [27], they are neither accurate nor reliable enough to substitute the diagnostic gold standard (liver biopsy). Several studies suggested elastography techniques (transient ultrasound elastography, acoustic radiation force impulse imaging or supersonic shear wave elastography) as the most effective, safe, quick and cheapest imaging tests. Unfortunately, their use is actually limited by the characteristic of the patient and in severely obese subjects, these techniques are not applicable even using the XL probe [28]. Magnetic resonance elastography (MRI), provide a highly accurate measurement of fibrosis, inflammation and steatosis (recently reviewed by Han), however, its application in clinical

practice is limited by the scarce availability (academic centres) and its high cost [29]. Thus, our study aimed to contribute to providing accurate serum biomarkers which in combination with imaging techniques, would be accurate, safe and reliable in the diagnosis and monitor fibrosis.

The most relevant finding of this study is that the plasmatic level of two of our candidates (IGF2 and EGFR) are closely associated with the stage of liver fibrosis. IGF2 is inversely correlated with the degree of lobular inflammation and fibrosis and previous studies showed a reduction of IGF2 plasmatic level in subjects with cirrhosis, inversely correlated with the hepatic damage [30]. Moreover, evidence about the association of lower IGF2 levels with the stage of fibrosis was provided in a paediatric NAFLD cohort [31] and, more recently, in a larger adults NAFLD population [32]. Our study confirms these findings and extends them to a cohort of severely obese adults. Regarding fatty liver, and contrarily to the study of Ajmera [32], we did not show any association between plasmatic levels of IGF2 and the degree of liver steatosis. Currently, the role of IGF2 in the fibrotic process is not fully unraveling mainly due to the lack of data regarding its autocrine/paracrine actions in the liver.

We found that an increased level of EGFR is associated with fibrosis, but to a lesser extent than IGF2. Correlation studies show a direct and indirect relationship with liver inflammation and total plasma cholesterol, respectively. The EGFR signalling axis has received much attention, because of the high density of EGFR in hepatocytes. Several studies proposed a key role of EGFR during liver regeneration, cirrhosis and HCC, highlighting its role in the development of liver damage [33]. Less known is the role of circulating EGFRs in liver disease, and only one study described its diagnostic use in HCC [34]. Further studies will be necessary to clarify the source(s) and function(s) of circulating EGFRs in inflammation and fibrosis. Although several studies reported a positive correlation between EGFR

activation or the presence of EGF ligands and the total cholesterol levels in humans [35] [36], no information was available about the associations between plasmatic soluble EGFRs and serum cholesterol before this study.

The candidate biomarkers individuated in this pilot study do not distinguish between a fatty liver and steatohepatitis, and it should not be used to diagnose NASH. Our analysis was focused on the potential use of these biomarkers to detect significant fibrosis (stage>F2) in severe obese subjects, and their performance for diagnosis was established through receiver operating characteristic (ROC) curves. The AUROC of IGF2 was 0.83 for the diagnosis of significant fibrosis and was superior to the others candidates and the non-invasive surrogate markers reported till now. Additionally, we observed a high NPVs (≥92%) for IGF2 and EGFR with modest corresponding PPVs (ranging from 44% - 58%) for the diagnosis of fibrosis. Beyond the modest PPVs, the accuracy of our candidates as tests for moderate fibrosis could be improved if prevalence is higher than 20% (reported range in MO subjects - from 8% to 60%-) [25]. Information on diagnostic accuracy of surrogate markers for liver fibrosis in severely obese subjects is scarce. Cleva reported AUROC data of 0.52, 0.88 and 0.99 for AST/ALT, Age-platelet and APRI, respectively, when used for the diagnosis of advanced fibrosis (≥F3) [37]. ALT and HbA1c were combined in a ROC statistical model and used to predict the presence of fibrosis (≥F1) with AUROC of 0.90 [38]. By using FIB-4, NFS and Fibrotest; a diagnostic performance for advanced fibrosis with an AUROC of 0.77, 0.75 and 0.72 respectively, was reported [24] [39].

The main strength of this study is the well characterised morbidly obese cohort with the biopsy-proven liver disease. Among the limitations, it should be mentioned: the relatively small sample size, the low prevalence of advanced/severe fibrosis or cirrhosis in the MO cohort and the lack of histological data (for ethical issues) in lean controls. In conclusion, this

study proposes IGF2 and EGFR as accurate biomarkers for the diagnosis of significant to moderate fibrosis in MO subjects. The introduction of these biomarkers in clinical practice, either alone or combined with others serum markers in a score, may reduce the need for liver biopsy. Larger prospective studies are needed to confirm this conclusion.

References:

1. News releases - Eurostat. 2016. http://ec.europa.eu/eurostat/documents/2995521/7700898/3-20102016-BP-EN.pdf/c26b037b-d5f3-4c05-89c1-00bf0b98d646. Accessed May 11, 2017

2. Papadia FS, Marinari GM, Camerini G, et al. Liver damage in severely obese patients: a clinical-biochemical-morphologic study on 1,000 liver biopsies. *Obes Surg*. 2004;14:952–958.

3. Morita S, Neto DDS, Morita FHA, Morita NK, Lobo SMA. Prevalence of Non-alcoholic Fatty Liver Disease and Steatohepatitis Risk Factors in Patients Undergoing Bariatric Surgery. *Obes Surg*. 2015;25:2335–2343.

4. Adams LA, Lymp JF, St Sauver J, et al. The natural history of nonalcoholic fatty liver disease: a population-based cohort study. *Gastroenterology*. 2005;129:113–121.

5. Zelber-Sagi S, Nitzan-Kaluski D, Goldsmith R, et al. Long term nutritional intake and the risk for non-alcoholic fatty liver disease (NAFLD): a population based study. *J Hepatol*. 2007;47:711–717.

6. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13:2498–2504.

7. Page S, Birerdinc A, Estep M, et al. Knowledge-based identification of soluble biomarkers: hepatic fibrosis in NAFLD as an example. *PloS One*. 2013;8:e56009.

8. AbdulHameed MDM, Tawa GJ, Kumar K, et al. Systems level analysis and identification of pathways and networks associated with liver fibrosis. *PloS One*. 2014;9:e112193.

9. Gastrointestinal surgery for severe obesity. *Consens Statement Natl Inst Health Consens Dev Conf*. 1991;9:1–20.

10. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia*. 1985;28:412–419.

11. Authors/Task Force Members, Rydén L, Grant PJ, et al. ESC Guidelines on diabetes, pre-diabetes, and cardiovascular diseases developed in collaboration with the EASD: the Task Force on diabetes, pre-diabetes, and cardiovascular diseases of the European Society of

Cardiology (ESC) and developed in collaboration with the European Association for the Study of Diabetes (EASD). *Eur Heart J*. 2013;34:3035–3087.

12. Sumida Y, Yoneda M, Hyogo H, et al. Validation of the FIB4 index in a Japanese nonalcoholic fatty liver disease population. *BMC Gastroenterol*. 2012;12:2.

13. Calès P, Lainé F, Boursier J, et al. Comparison of blood tests for liver fibrosis specific or not to NAFLD. *J Hepatol*. 2009;50:165–173.

14. Harrison SA, Oliver D, Arnold HL, Gogia S, Neuschwander-Tetri BA. Development and validation of a simple NAFLD clinical scoring system for identifying patients without advanced disease. *Gut*. 2008;57:1441–1447.

15. Kleiner DE, Brunt EM. Nonalcoholic fatty liver disease: pathologic patterns and biopsy evaluation in clinical research. *Semin Liver Dis*. 2012;32:3–13.

16. Kleiner DE, Brunt EM, Van Natta M, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatol Baltim Md*. 2005;41:1313–1321.

17. Orchard S, Kerrien S, Abbani S, et al. Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. *Nat Methods*. 2012;9:345–350.

18. Shen J, Chan HL-Y, Wong GL-H, et al. Non-invasive diagnosis of non-alcoholic steatohepatitis by combined serum biomarkers. *J Hepatol*. 2012;56:1363–1370.

19. Vuppalanchi R, Jain AK, Deppe R, et al. Relationship between changes in serum levels of keratin 18 and changes in liver histology in children and adults with nonalcoholic fatty liver disease. *Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc*. 2014;12:2121-2130.e1–2.

20. Miller MH, Walsh SV, Atrih A, Huang JT-J, Ferguson MAJ, Dillon JF. Serum proteome of nonalcoholic fatty liver disease: a multimodal approach to discovery of biomarkers of nonalcoholic steatohepatitis. *J Gastroenterol Hepatol*. 2014;29:1839–1847.

21. Zhang H, Wu P, Chen F, et al. SILAC-based quantitative proteomic analysis of secretome between activated and reverted hepatic stellate cells. *Proteomics*. 2014;14:1977–1986.

22. Tanwar S, Trembling PM, Guha IN, et al. Validation of terminal peptide of procollagen III for the detection and assessment of nonalcoholic steatohepatitis in patients with nonalcoholic fatty liver disease. *Hepatol Baltim Md*. 2013;57:103–111.

23. Wang X, Lopategi A, Ge X, et al. Osteopontin induces ductular reaction contributing to liver fibrosis. *Gut*. 2014;63:1805–1818.

24. Luger M, Kruschitz R, Kienbacher C, et al. Prevalence of Liver Fibrosis and its Association with Non-invasive Fibrosis and Metabolic Markers in Morbidly Obese Patients with Vitamin D Deficiency. *Obes Surg*. 2016. doi:10.1007/s11695-016-2123-2

25. Machado M, Marques-Vidal P, Cortez-Pinto H. Hepatic histology in obese patients undergoing bariatric surgery. *J Hepatol*. 2006;45:600–606.

26. Ekstedt M, Hagström H, Nasr P, et al. Fibrosis stage is the strongest predictor for disease-specific mortality in NAFLD after up to 33 years of follow-up. *Hepatol Baltim Md*. 2015;61:1547–1554.

27. Kaswala DH, Lai M, Afdhal NH. Fibrosis Assessment in Nonalcoholic Fatty Liver Disease (NAFLD) in 2016. *Dig Dis Sci*. 2016;61:1356–1364.

28. Wong VW-S, Vergniol J, Wong GL-H, et al. Liver stiffness measurement using XL probe in patients with nonalcoholic fatty liver disease. *Am J Gastroenterol*. 2012;107:1862–1871.

29. Han MAT, Saouaf R, Ayoub W, Todo T, Mena E, Noureddin M. Magnetic resonance imaging and transient elastography in the management of Nonalcoholic Fatty Liver Disease (NAFLD). *Expert Rev Clin Pharmacol*. 2017;10:379–390.

30. Rehem RNAMA, El-Shikh WMHM. Serum IGF-1, IGF-2 and IGFBP-3 as parameters in the assessment of liver dysfunction in patients with hepatic cirrhosis and in the diagnosis of hepatocellular carcinoma. *Hepatogastroenterology*. 2011;58:949–954.

31. Cianfarani S, Inzaghi E, Alisi A, Germani D, Puglianiello A, Nobili V. Insulin-like growth factor-I and -II levels are associated with the progression of nonalcoholic fatty liver disease in obese children. *J Pediatr*. 2014;165:92–98.

32. Ajmera V, Perito ER, Bass NM, et al. Novel Plasma Biomarkers Associated with Liver Disease Severity in Adults with Nonalcoholic Fatty Liver Disease. *Hepatol Baltim Md*. 2016. doi:10.1002/hep.28776

33. Komposch K, Sibilia M. EGFR Signaling in Liver Diseases. *Int J Mol Sci*. 2015;17.

34. Divella R, Daniele A, Gadaleta C, et al. Circulating transforming growth factor-β and epidermal growth factor receptor as related to virus infection in liver carcinogenesis. *Anticancer Res*. 2012;32:141–145.

35. Sánchez-Vizcaíno E, Vehí C, Campreciós G, Morcillo C, Soley M, Ramírez I. Heparin-binding EGF-like growth factor in human serum. Association with high blood cholesterol and heart hypertrophy. *Growth Factors Chur Switz*. 2010;28:98–103.

36. Berrahmoune H, Lamont JV, Herbeth B, et al. Association between EGF and lipid concentrations: a benefit role in the atherosclerotic process? *Clin Chim Acta Int J Clin Chem*. 2009;402:196–198.

37. de Cleva R, Duarte LF, Crenitte MRF, de Oliveira CPM, Pajecki D, Santo MA. Use of noninvasive markers to predict advanced fibrosis/cirrhosis in severe obesity. *Surg Obes Relat Dis Off J Am Soc Bariatr Surg*. 2016;12:862–867.

38. Gholam PM, Flancbaum L, Machan JT, Charney DA, Kotler DP. Nonalcoholic fatty liver disease in severely obese subjects. *Am J Gastroenterol*. 2007;102:399–408.

39. Poynard T, Lassailly G, Diaz E, et al. Performance of biomarkers FibroTest, ActiTest, SteatoTest, and NashTest in patients with severe obesity: meta analysis of individual patient data. *PloS One*. 2012;7:e30325.

**Figure legends:**

**Fig. 1. Protein-Protein Interaction biological network of the mediators involved in liver fibrosis visualised using Cytoscape open software.** Individuated candidate biomarkers are highlighted in red.

**Fig. 2. Boxplot levels of serum surrogate markers *versus* the stage of fibrosis.** A) Aspartate aminotransferase (AST); B) Gamma-glutamyl transpeptidase (GGT); C) Alanine aminotransferase (ALT); and D) Ratio AST/ALT. F0/F1 (non significan/minimal fibrosis) and F2/F3 (significant/moderate fibrosis) in the MO cohort. **Significant at $p<0.01$.

**Fig. 3. Boxplot of candidate biomarkers plasmatic levels *versus* the stage of fibrosis.** A) IGF2; B) EGFR; C) EGFR/IGF2 ratio, D) CD44 and E) SPARC. Data were expressed as Median (interquartile range (IQR)) and statistical analysis using ANOVA test. ***Significant at $p<0.001$; **significant at $p<0.01$ and *significant at $p<0.05$.

**Fig. 4. Receiver operating characteristics (ROC) curves for the non-invasive markers for the diagnosis of significant fibrosis (Kleiner-Brunt fibrosis stage 2-3).**

**Table 1. Demographic and laboratory characteristics of all subjects**

| | MO | CTRLs (Lean) | Cirrhosis |
|---|---|---|---|
| | (n=71) | (n=11) | (n=14) |
| Age (years) | 43 ± 12* | 33 ± 4 | 69 ± 8*** |
| Gender (female) | 47 (66 %) | 6 (54 %) | 7 (50 %) |
| BMI (Kg/m$^2$) | 44 ± 7*** | 23 ± 2 | 29 ± 5 |
| ALT (IU/L) | 33 ± 31 | 23 ± 12 | 38 ± 21 |
| AST (IU/L) | 24 ± 14 | 25 ± 11 | 41 ± 24 |
| GGT (IU/L) | 33 ± 24 | 25 ± 14 | 124 ± 78** |
| Fasting Glucose (mg/dL) | 113 ± 25* | 93 ± 11 | 124 ± 41* |
| Fasting Insulin (µU/mL) | 20 ± 14 | 8 ± 5 | 13 ± 6 |
| HbA1c (%) | 6.3 ± 1* | N/A | 7 ± 2* |
| HOMA-IR | 5.3 ± 4** | 1.6 ± 0.4 | 3.7 ± 2 |
| Diabetes | 15 (21%) | 0 | 8 (57%) |
| Cholesterol (mg/dL) | 200 ± 40 | 182 ± 34 | 190 ± 54 |
| Triglycerides (mg/dL) | 141 ± 91 | 93 ± 35 | 136 ± 64 |
| Steatosis 0/1/2/3 | 7/29/16/19 | N/A | N/A |
| Lobular Inflammation 0/1/2 | 14/45/12 | N/A | N/A |
| Balloning 0/1/2 | 16/18/37 | N/A | N/A |
| Fibrosis 0/1/2/3/4 | 8/50/12/1/0 | N/A | 0/0/0/6/8 |
| NAFL/NASH | 27 (38%)/44 (62%) | N/A | 60%/40% |

BMI, body mass index, ALT, alanine aminotransferase; AST, aspartate aminotransferase, GGT, gamma-glutamyl transferase, HbA1c, glycated hemoglobin, HOMA-IR, homeostatic model assessment of insulin resistance, N/A, not available. ***p<0.001; **p<0.01 and *p<0.05 were considered statistically significant vs CTRLs

**Table 2. Correlations with IGF2, EGFR and EGFR/IGF2 ratio in MO cohort**

| | IGF2 | | EGFR | | EGFR/IGF2 ratio | |
|---|---|---|---|---|---|---|
| **Parameter** | rho | *p* value* | rho | *p* value* | rho | *p* value* |
| **BMI** | -0.15 | 0.249 | 0.01 | 0.914 | 0.15 | 0.240 |
| **Triglycerides** | 0.06 | 0.637 | -0.05 | 0.698 | -0.08 | 0.550 |
| **Total cholesterol** | -0.04 | 0.760 | **-0.36** | **0.004**** | 0.02 | 0.865 |
| **HOMA-IR** | -0.16 | 0.249 | 0.19 | 0.149 | 0.21 | 0.135 |
| **FLI** | -0.16 | 0.246 | -0.0006 | 0.996 | 0.17 | 0.210 |
| **Steatosis** | -0.12 | 0.387 | -0.06 | 0.619 | 0.14 | 0.275 |
| **NAS** | -0.18 | 0.173 | 0.08 | 0.540 | 0.18 | 0.178 |
| **Lobular Inflammation** | **-0.30** | **0.024*** | **0.28** | **0.021*** | **0.30** | **0.027*** |
| **Ballooning** | -0.06 | 0.660 | -0.01 | 0.921 | -0.20 | 0.150 |
| **AST/ALT ratio** | 0.02 | 0.875 | -0.14 | 0.238 | -0.03 | 0.840 |
| **GGT** | -0.03 | 0.844 | 0.03 | 0.809 | 0.04 | 0.764 |
| **FIB-4** | 0.05 | 0.734 | -0.13 | 0.297 | 0.05 | 0.717 |
| **Fibrometer** | -0.03 | 0.819 | 0.004 | 0.974 | 0.10 | 0.488 |
| **Fibrosis** | **-0.36** | **0.006**** | **0.29** | **0.018*** | **0.63** | **< 0.0001** |

Steatosis and fibrosis scores were according to Kleiner-Brunt histological classification. *p* value corresponds to *Ho: rho = 0* (the two variables do not vary together at all). ***p<0.001, **p<0.01 and *p<0.05 were considered statistically significant.

**Table 3: Comparison of the performance of each test for the diagnosis of significant fibrosis in the MO cohort**

| Biomarker/Test | AUROC (95% CI) | Cut-off | Sens (%) | Spec (%) | PPV | NPV |
|---|---|---|---|---|---|---|
| **IGF2** | 0.83 (0.70-0.92) | 1.9 | 85.7 | 73.7 | 58.3 | 92.3 |
| **EGFR** | 0.71 (0.59-0.82) | 102.5 | 94.4 | 52.9 | 46.2 | 95.7 |
| **EGFR/IGF2** | 0.79 (0.67-0.88) | 58 | 73.3 | 73.6 | 54.3 | 86.6 |
| **Fibrometer** | 0.64 (0.51-0.75) | 51 | 76.9 | 64.0 | 47.8 | 86.6 |
| **FIB-4** | 0.63 (0.50-0.74) | 0.78 | 66.7 | 64.1 | 44.4 | 81.8 |

AUROC, Sens, sensitivity; Spec, specificity; PPV, positive predictive value; NPV, negative predictive value.