Research project proposal:
Metabolomics Profile in NAFLD patients: the weight of Steatosis, Steatohepatitis, Cirrhosis and Hepatocellular Carcinoma
Acronym: **NAME (NAfld MEtabolomics)**

Authors:

Marcello Persico[1], Mario Masarone[1], Jacopo Troisi[1], Antonio Grieco[2]
[1]Department of Medicine and Surgery, University of Salerno, Italy; [2] Department of Internal Medicine, Università Cattolica del Sacro Cuore, Roma, Italy

Introduction

The term Non Alcoholic Fatty Liver Disease (NAFLD) has been increasingly evoked in Hepatology ambulatories and clinics in the last decade. It defines the presence of any significant (>5% of hepatocytes) amount of fat accumulation in the liver in the absence of an "unsafe" quantity of alcohol consumption and any other cause of liver diseases [1]. This term includes at least two different clinical entities: a form that represents only the accumulation of fats in the liver (also named steatosis or Non Alcoholic Fatty Liver -NAFL), and the Non Alcoholic Steato-Hepatitis (NASH) which is characterized by steatosis along with necroinflammation and fibrosis. This latter entity is a "progressive" form that has histological features that make it hardly distinguishable from alcoholic liver disease. In fact, it is characterized by the presence of hepatocytes ballooning, lobular inflammation, perisinusoidal and perivenular fibrosis at liver histology [2]. Even if these two entities always go under the same definition of NAFLD, they don't share the same natural history: the first being a "benign" presentation with no (or very rare) progression, the second being responsible of liver cirrhosis, Hepatocellular Carcinoma (HCC) and liver-related deaths [3]. Nevertheless, these two conditions seem to share the same risk factors. NAFLD is a unique "challenge" for the hepatologists and, due to the to changes in dietary habits and increased sedentary lifestyle, has seen a worldwide increment in the last years, making it one of the most frequent liver diseases in the world [4]. It is generally considered a "benign disease" with low rates of progression to fibrosis, cirrhosis and hepatocellular carcinoma [5]. Nevertheless, due to the high number of affected patients, the prevalence of related cirrhosis increased overtime, and actually it represents the third cause of liver transplantation in the USA [6]. Moreover, even if the incidence of Hepatocellular Carcinoma (HCC) in NAFLD patients is lower than that in HCV/HBV cirrhotic patients, the absolute burden of NASH-related HCC is higher, due to the higher number of patients with NAFLD in respect to HCV infected ones [7]. It is very likely that the importance of this disease will continue to increase in the future, when the new therapies and prevention programs for hepatitis C and B will further reduce the size of viral infections of the liver. For these reasons, it is very important to recognize the mechanisms underlying its onset and progression in the liver. Even if many insights on this topic were made in the last years, many aspects of the pathophysiological mechanisms underlying this disease remain to be explored.
Moreover, nowadays to differentiate the presence of a simple, non evolutive, liver steatosis (NAFL) from a potentially worsening steatohepatitis still represents a diagnostic issue. In fact, the definitive differential diagnosis between these two entities still relay on a such invasive technique as liver biopsy, which, in the clinical practice, is difficult to propose to a large number of patients, of whom only a minority is potentially affected by the evolutionary form of the disease (NASH). [4] To address this issue, various clinical scores have been proposed for the use in NAFLD patients, such as "Fatty Liver Index", "NAFLD Fibrosis Score", and analogues [8,9]. Even if an approach of this type has the advantage of being a non invasive method for discriminating between potentially benign and evolutionary diseases, it has various limitations, based primarily on the lack of

validation on large cohorts and different populations, and on the scarce power to evaluate "intermediate presentations". These methods have, in fact, a large "grey zone" in which they fail to address the real risk of the patients. Other non invasive methods relay on imaging techniques (such as Fibroscan, Continuous Attenuation Parameter – CAP, Ultrasonography and Magnetic Resonance) that can, at their best, quantify the liver fat content rather than differentiate between simple steatosis and steatohepatitis [10,11, 12].

In this setting it would be very useful to individuate one, or more, specific biomarkers, to differentiate NAFLD patients from general population and, between NAFLD presentations, simple steatosis (NAFL) from steatohepatitis (NASH) and, eventually, hepatocellular carcinoma (NASH-HCC).

In the recent years, a powerful tool to search for specific biomarkers has been validated: the untargeted metabolomics. Metabolomics examines the small molecules (with molecular weight less than 1,5 KDa) that represent the complex of metabolites coming from the whole organic metabolism, in order to find metabolic markers that could eventually identify a specific disease presentation.[13]

This technique has the advantage to evaluate more accurately the "phenotype" of a disease, in respect to genes, transcripts and proteins, which very likely undergo to epigenetic, transcriptional and pre- and post- translational modifications [14].

In the field of NAFLD very few studies, on a small number of patients, have recently sketched a partial metabolomics profile of such disease [15, 16, 17]. For this reason, it may be of high interest to further evaluate the plasma metabolomics profile of a cohort of biopsy proven NAFLD patients in order to find one, or more, specific biomarkers capable to differentiate them from healthy controls and, between NAFLD patients, to discriminate progressive forms (NASH), and eventually NAFLD-cirrhosis or hepatocellular carcinoma (NASH-HCC), from non progressive liver steatosis (NAFL).

Study design
Multicenter, no-profit, non interventional cross-sectional observational study.

Endpoint
Primary endpoint: to evaluate plasma metabolomics profile of a cohort of NAFLD patients

Secondary endpoint: To differentiate plasma metabolomics profiles of NAFL, NASH, NAFLD-cirrhotic and NASH-HCC patients

Patients and methods
A cohort of biopsy proven NAFLD patients will be enrolled in all the centers involved in the study. In the same way, a cohort of age- and sex-matched healthy indviduals without any known liver disease (viral and/or metabolic) and without metabolic syndrome and/or diabetes will be also enrolled in the centers involved in the study. Patients, and controls, will be stratified for age, sex and BMI. NAFLD patients will be divided on the basis of the histological presence of steatosis or steatohepatitis diagnosed by the Kleiner and Brunt criteria [18,19], and for the eventual presence, assessed by at least two concordant imaging techniques, of hepatocellular carcinoma. In this way, the plasma metabolomics profile will be evaluated in five groups:

1) Non alcoholic Simple steatosis (NAFL)
2) Non alcoholic Steatohepatitis (NASH)

3) Nonalcoholic Steatosis based Cirrhosis
4) Non alcoholic steatohepatitis associated Hepatocellular carcinoma (NASH-HCC)
5) Controls

Of every patient (and controls) will be recorded: clinical history with alcohol consumption and smoking habits registration, physical examination, arterial pressure, waist circumference, Body mass index, blood glucose, total and fractioned cholesterol, triglycerides, AST, ALT, GGT, ALP, blood count, metabolic syndrome evaluation by NCEP-ATPIII criteria, abdomen ultrasonography with the evaluation of the bright liver echo pattern presence [10]. Liver tissue samples will be collected by performing a hepatic percutaneous biopsy with Surecut 17G needles, via the intercostal route using an echo-guided or echo-assisted method. Liver specimens will be used for histological examination if they will be at least 1.5-cm long and contain >5 portal spaces. Biopsies will be evaluated with the Kleiner score [18] for necroinflammation grading and fibrosis staging and by the Brunt score [19] for the presence and extent of steatosis by skilled pathologists. Each patient will be included in the study after giving an informed consent. The study will be submitted for approval by the local ethical committees.

Statistical analysis

Statistical analyses will be performed using the Statistical Program for Social Sciences (SPSS®) ver.16.0 for Macintosh® (SPSS Inc., Chicago, Illinois, USA). Student t-test and Mann-Whitney U test will be performed to compare continuous variables, chi-square with Yates correction or Fisher-exact test to compare categorical variables. Univariate and multivariate analyses will be also performed to test independent variables affecting the endothelial dysfunction, by performing ANOVA, linear regressions and binary logistic regressions where applicable. Statistical significance will be defined when "*p<0,05*" in a "two-tailed" test with a 95% Confidence Interval.

Statistic multivariate data analysis (PLS-DA) will be performed on normalized and correct (on Internal Standard peak area) chromatogram using R (Foundation for Statistical Computing, Vienna, Austria). Mean centering and unit variance scaling will be applied for all analyses.
PLS is a supervised method that uses multivariate regression techniques to extract via linear combination of original variables (X) the information that can predict class membership (Y). PLS regression will be performed using *plsr* function provided by R *pls* package [20]. Classification and cross-validation will be performed using the corresponding *wrapper* function offered by the *caret* package [21]. To assess the significance of class discrimination, a permutation test will be performed. In each permutation, a PLS-DA model will be built between the data (X) and the permuted class labels (Y) using the optimal number of components determined by cross validation for the model based on the original class assignment. Two types of test statistics will done to measure class discrimination. The first one is based on prediction accuracy during training. The second one is separation distance based on the ratio between group sum of the squares and the within group sum of squares (B/W-ratio). If the observed test statistics is part of the distribution based on the permuted class assignments, class discrimination cannot be considered significant

from a statistical point of view [22]. There are two variable importance measures in PLS-DA. The first, Variable Importance in Projection (VIP) is a weighted sum of squares of the PLS loadings, taking into account the amount of explained Y-variation in each dimension. VIP scores will be calculated for each component. The other importance measure is based on the weighted sum of PLS-regression. The weights are a function of the reduction of the sums of squares across the number of PLS components. The average of the feature coefficients will be used to indicate the overall coefficient-based importance.

The metabolic pathway will be made using MetScap application [23] of the software Cytoscape [24].

Sample size calculation

Due to the fact that, in common with other 'omics' techniques, in metabolomic phenotyping, there is no accepted approach to sample size determination, in large part due to the unknown nature of the expected effect. In such hypothesis free science, neither the number or class of important analytes nor the effect size are known a priori, so, almost all study were based on "available cases".

We try to estimate the sample size to get the 80% statistical power, with a multivariate simulation, which deals effectively with the highly correlated structure and high-dimensionality of metabolic phenotyping data as suggest by Blaise et al. [25].

First, we performed a pilot study based on 13 cases (3 with NAFLD, 4 with NASH, 3 with NAFLD derived Cirrhosis and 3 with Cryptogenic cirrhosis). These cases were used to test the preliminary hypothesis that the metabolic signature can consent a class separation. For sample size calculation, a large data set is simulated based on the characteristics of the pilot study, than, an effect of a given size, corresponding to a discrete classification was added. Different sample sizes was modeled by randomly selecting data sets of various sizes from the simulated data. These considerations allow us to estimate a sample size of 60 patients and 60 controls to classify respect the NAFL patients with a false discovery rate (FDR) of 0.1

Methods

METABOLITE EXTRACTION AND DERIVATIZATION

The metabolome extraction, purification and derivatization will be carried by the MetaboPrep GC kit (Theoreo, Montecorvino Pugliano, Italy) according to manufacturer instructions. Instrumental analyses will be carried with a GC-MS system (GC-2010 Plus gas chromatograph coupled to a QP2010 Plus single quadrupole mass spectrometer; Shimadzu Corp., Kyoto, Japan).

GCMS ANALYSIS

Samples of 2 μL from the derivatized solution will be injected into the GC-MS system (GC-2010 Plus gas chromatograph coupled to a 2010 Plus single quadrupole mass spectrometer; Shimadzu Corp., Kyoto, Japan). Chromatographic separation will be achieved with a 30 m 0.25 mm CP-Sil 8 CB fused silica capillary GC column with 1.00 μm film thickness from Agilent (Agilent, J&W), with helium as carrier gas. The initial oven temperature of 100 °C will be held for 1 min and then raised

at 4 °C/min to 320 °C with further 4 minutes of hold time. The gas flow will be set to achieve a constant linear velocity of 39 cm/s and the split flow was set to 1:5. The mass spectrometer will be operated in electron impact (70 eV) in full scan mode in the interval of 35-600 m/z with a scan velocity of 3333 amu/sec and a solvent cut time of 4.5 minute. The complete GC program duration will be 60 minutes.

References

1. Angulo P: Nonalcoholic fatty liver disease. N Engl J Med 2002;346:1221–1231.

2. Brunt EM. Pathology of nonalcoholic fatty liver disease. Nat Rev Gastroenterol Hepatol 2010; 7: 195–203.

3. Matteoni CA, Younossi ZM, Gramlich T, Boparai N, Liu YC, McCullough AJ. Nonalcoholic fatty liver disease: a spectrum of clinical and pathological severity. Gastroenterology 1999; 116: 1413-19

4. Chalasani N, Younossi Z, Lavine JE, Diehl AM, Brunt EM, Cusi K, et al. The diagnosis and management of non-alcoholic fatty liver disease: practice Guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association. Hepatology. 2012;55: 2005-2023

5. Matteoni CA, Younossi ZM, Gramlich T, Boparai N, Liu YC, McCullough AJ. Nonalcoholic fatty liver disease: a spectrum of clinical and pathological severity. Gastroenterology. 1999;116: 1413-1419.

6. White DL, Kanwal F, El-Serag HB. Association between nonalcoholic fatty liver disease and risk for hepatocellular cancer, based on systematic review. Clin Gastroenterol Hepatol. 2012;10: 1342-1359.

7. Masarone M, Federico A, Abenavoli L, Loguercio C, Persico M. Non alcoholic fatty liver: epidemiology and natural history. Rev Recent Clin Trials. 2014;9(3):126-33.

8. Bedogni G, Bellentani S, Miglioli L, et al. The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population. BMC Gastroenterol. 2006;6:33.

9. Angulo P, Hui JM, Marchesini G, Bugianesi E, George J, Farrell GC, Enders F, Saksena S, Burt AD, Bida JP, Lindor K, Sanderson SO, Lenzi M, Adams LA, Kench J, Therneau TM, Day CP The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. Hepatology. 2007 Apr;45(4):846-54.

10. Palmentieri B et al. DLD 2006

11. Sasso M, Tengher-Barna I, Ziol M et al. J Viral Hepat 2012 Apr; 19(4): 244-253

12. Huwart L et al, Gastroenterology, 2008:32-40

13. Griffith Wj et al. Targeted Metabolomics. Angewandte Chemie Int Ed Engl 2010; 49: 5426-5445

14. Patti GJ et al Nat Rev Mol Cell Biol 13; 263-269

15. Cano A, Alonso C. Deciphering non-alcoholic fatty liver disease through metabolomics. Biochem Soc Trans. 2014 Oct;42(5):1447-52. doi: 10.1042/BST20140138

16. Kalhan SC et al. Metabolism Clinicl and Experimental 60 (2011): 404-413

17. Voiculescu M, Nanau RM, Neuman MG. Non-invasive biomarkers in non-alcoholic steatohepatitis-induced hepatocellular carcinoma. J Gastrointestin Liver Dis. 2014 Dec;23(4):425-9. doi: 10.15403/jgld.2014.1121.234.bna.

18. D.E. Kleiner, E.M. Brunt, M. Van Natta, C. Behling, M.J. Contos, O.W. Cummings, L.D. Ferrell, Y.C. Liu, M.S. Torbenson, A. Unalp-Arida, M. Yeh, A.J. McCullough, A.J. Sanyal, N.S.C.R. Network. Design and validation of a histological scoring system for nonalcoholic fatty liver disease, Hepatology 41 (2005) 1313-1321

19. Brunt EM, Janney CG, Di Bisceglie AM, Neuschwander-Tetri BA, Bacon BR. Nonalcoholic Steatohepatitis: a proposal for grading and staging the histological lesions. Am J Gast-roenterol 1999; 30:847–850.

20. Ron Wehrens and Bjorn-Helge Mevik.pls: Partial Least Squares Regression (PLSR) and principal Component Regres- sion (PCR), 2007, R package version 2.1-0

21. Max Kuhn. Contributions from Jed Wing and Steve Weston and Andre Williams.caret: Classification and Regression Training, 2008, R package version 3.45

22. Bijlsma et al.Large-Scale Human Metabolomics Studies: A Strategy for Data (Pre-) Processing and Validation, Anal Chem. 2006, 78 567 – 574

23. Karnovsky A, Weymouth T, Hull T, Tarcea VG, Scardoni G, Laudanna C, Sartor MA, Stringer KA, Jagadish HV, Burant C, Athey B, Omenn GS. Metscape 2 bioinformatics tool for the analysis and visualization of metabolomics and gene expression data. Bioinformatics, 28:3 (373-80). 2012 Feb 1.

24. Nishida K, Ono K, Kanaya S, Takahashi K. KEGGscape: a Cytoscape app for pathway data integration. F1000Res, 3: (144). 2014.

25. Benjamin J. Blaise, Gonçalo Correia, Adrienne Tin, J. Hunter Young, Anne-Claire Vergnaud, Matthew Lewis, Jake T. M. Pearce, Paul Elliott, Jeremy K. Nicholson, Elaine Holmes, and Timothy M. D. Ebbels. Power Analysis and Sample Size Determination in Metabolic Phenotyping. Anal. Chem. 2016, 88, 5179–5188