

# Contribution of a genetic risk score to clinical prediction of hepatic steatosis in obese children and adolescents

Chiara Zusi<sup>1,2</sup>, Alessandro Mantovani<sup>2</sup>, Francesca Olivieri<sup>1</sup>, Anita Morandi<sup>1</sup>, Massimiliano Corradi<sup>1</sup>, Emanuele Miraglia Del Giudice<sup>3</sup>, Marco Dauriz<sup>2</sup>, Luca Valenti<sup>4</sup>, Christopher D. Byrne<sup>5,6</sup>, Giovanni Targher<sup>2</sup>, Claudio Maffeis<sup>1</sup>

<sup>1</sup>Pediatric Diabetes and Metabolic Disorders Unit, Department of Surgery, Dentistry, Pediatrics and Gynaecology, University Hospital of Verona, Verona, Italy

<sup>2</sup>Division of Endocrinology, Diabetes and Metabolism, Department of Medicine, University Hospital of Verona, Verona, Italy

<sup>3</sup>Department of Woman, Child, and General and Specialized Surgery, University of Campania “Luigi Vanvitelli”, Naples, Italy

<sup>4</sup>Department of Pathophysiology and Transplantation, University of Milan, and Translational Medicine and Hepatology – Transfusional Center, Fondazione IRCCS Ca' Granda - Ospedale Maggiore Policlinico Milano, Milan, Italy

<sup>5</sup>Nutrition and Metabolism, Faculty of Medicine, University of Southampton, UK

<sup>6</sup>Southampton National Institute for Health Research Biomedical Research Centre, University Hospital Southampton, Southampton General Hospital, Southampton SO16 6YD, UK

**Conflict of Interest Statement:** The Authors have no potential conflicts of interest to disclose.

**Running Title:** Genetics in paediatric NAFLD

**Word count:** 248 abstract; 3433 text (*excluding* title page, abstract, references, tables and figure legends); n=3 **Tables** and n=2 **Figures**

## Address for correspondence:

Prof. Claudio Maffeis  
Pediatric Diabetes and Metabolic Disorders Unit  
University Hospital  
Piazzale A. Stefani, 1  
37126 Verona, Italy  
Tel: +39 045-8127664  
Fax: +39 045-8127660  
E-mail: [claudio.maffeis@univr.it](mailto:claudio.maffeis@univr.it)

## ABSTRACT

**Background:** Nonalcoholic fatty liver disease (NAFLD) is the commonest liver disease in children and adolescents in Western countries. Complex traits arise from the interplay between environmental and genetic factors in the pathogenesis of NAFLD.

**Aims:** We examined the association between NAFLD and eleven single nucleotide polymorphisms (SNPs) at genetic loci potentially associated with liver damage (*GCKR*, *MBOAT7*, *GPR120*), oxidative stress (*SOD2*), lipid metabolism (*PNPLA3*, *TM6SF2*, *LPIN1*, *ELOVL2*, *FADS2*, *MTTP*) and fibrogenesis (*KLF6*) in a paediatric population. A genetic risk score (GRS) was performed taking into account both these SNPs and clinical risk factors.

**Methods:** We recruited a cohort of 514 obese children and adolescents (mean age [ $\pm$ SD]: 11.2 $\pm$ 2.8 years, z-BMI 3.3 $\pm$ 0.8). NAFLD was identified by ultrasonography. Genotyping was performed by TaqMan-Based RT-PCR system.

**Results:** The overall prevalence of NAFLD was 67.5% (347 patients). Among the eleven genotyped SNPs, the genetic variants in *TM6SF2* rs58542926 (OR=4.13, p=0.002), *GCKR* rs1260326 (OR=1.53, p=0.003), *PNPLA3* rs738409 (OR=1.58, p=0.004) and *ELOVL2* rs2236212 (OR=1.34, p=0.047) were significantly associated with a higher risk of NAFLD. Addition of a 11-polymorphism GRS to established clinical risk factors significantly (albeit modestly) improved the discriminatory capability of the regression model for predicting the risk of NAFLD (with SNPs C-statistic 0.81 [95%CI 0.75-0.88] vs. 0.77 [0.70-0.84] without SNPs; p=0.047).

**Conclusions:** NAFLD was strongly associated with three genetic variants, *TM6SF2* rs58542926, *PNPLA3* rs738409 and *GCKR* rs1260326, and more slightly with *ELOVL2* rs2236212, in obese children and adolescents. Addition of a 11-polymorphism GRS to clinical risk factors improved the predictability of NAFLD.

**Keywords:** obesity; nonalcoholic fatty liver disease; NAFLD; genetics; pediatrics

## INTRODUCTION

Nonalcoholic fatty liver disease (NAFLD) is one of the most common hepatic diseases in adolescents in developed countries [1]. Paediatric NAFLD has become a major public health problem worldwide with important implications for future development of liver dysfunction, type 2 diabetes and other cardiometabolic complications [2]. However, if detected early, the progression of NAFLD might be slowed and its extra-hepatic complications might be also prevented.

Obesity and a sedentary lifestyle, together with metabolic syndrome and ethnicity, contribute to the risk of NAFLD [3][4]. In recent years, a number of studies have suggested that genetic susceptibility also plays an important role in NAFLD development and progression in childhood [5][6]. Multiple genetic *loci* associated with the presence and the severity of NAFLD both in adults and in adolescents have been identified [7][8][9]. Although these *loci* account for only a small fraction of a total heritability of NAFLD, it is possible that they may be combined into a genetic risk score (GRS) for enhanced detection of people at higher risk of developing liver disease [10][11] [12][13][14][15][16][17]. As recently reported, in NAFLD the gene-environmental synergy is more important than either single factor alone [18], and combining genetic and clinical information might be useful for risk stratification in clinical practice [19]. Several genetic models for NAFLD prediction have already been investigated in adults [5][17]. However, there remains a paucity of such models in the paediatric population. Analyzing genetic risk scores in paediatric populations could be useful since children and adolescents are less affected by comorbidities than adults, thereby limiting the effect of potential confounders for NAFLD in these subjects. Therefore, the aim of this study was to test whether addition of genetic variants to established clinical risk factors improved risk prediction for NAFLD in a large sample of children and adolescents with obesity. The genetic variants included in our genetic risk score were 11 single nucleotide polymorphisms (SNPs) chosen on the basis of their involvement in lipid handling, insulin signaling, oxidative stress or hepatic fibrogenesis; all of which are potential contributing pathways affecting the development and progression of NAFLD.

## MATERIALS AND METHODS

## **Subjects**

We consecutively recruited unrelated obese children and adolescents at their first visit at the obesity outpatient clinic of the Pediatric Diabetes and Metabolic Disorders Unit of the University Hospital of Verona (Italy). Inclusion criteria were age between 6–18 years, European ancestry, Italian family origin and body mass index (BMI) greater than the age- and sex-specific BMI cutoff for obesity (using the World Health Organization BMI cutoffs as reference) [20][21]. Exclusion criteria were secondary obesity (e.g., Prader-Willi syndrome, Cohen syndrome, mutation in the *MC4R* gene), known chronic hepatic diseases (e.g., hepatitis by virus, drugs, autoimmunity, hemochromatosis, Wilson's disease), alcohol abuse ( $\geq 140$  g/week), and chronic use of any drugs. Detailed information on these exclusion conditions and current use of medications was collected in all patients by structured interviews during medical examination. Based on these criteria, a total of 514 children and adolescents with obesity were included in the final analysis. Informed consent was obtained from children and their parents. The protocol of the study was approved by the local Ethics Committee of the University Hospital of Verona.

## **Anthropometric and biochemical data**

At recruitment, anthropometric, clinical and biochemical data were collected for all children. BMI was calculated as body weight (in kilograms) divided by body height (in meters) squared. BMI values were standardized (z- BMI) using age- and sex-specific median, standard deviation and power of the box-cox transformation (least mean square method) based on WHO norms [20]. Waist circumference was measured as the minimal circumference measurable on the horizontal plane between the lowest portion of the rib cage and the iliac crest. Waist-to-height ratio (WHtR) was calculated and used as an index of body fat distribution, as previously described [22].

Venous blood samples were collected in the morning after an overnight fast of at least 8 hours. Serum liver enzymes [alanine aminotransferase (ALT) and aspartate aminotransferase (AST)], glucose, lipids and other biochemical blood measurements were determined using standard laboratory procedures at the central Laboratory of our hospital. Low-density lipoprotein (LDL)-cholesterol was calculated using the Friedewald's equation.

Hemoglobin A1c (HbA1c) was measured by a high-performance liquid chromatography analyzer on Tosoh-G7 automated analyzer (Tosoh Bioscience Inc., Japan).

### **Liver ultrasonography**

At recruitment, as part of the routine investigation of children in our obesity outpatient clinic, two experienced radiologists (who were blinded to the clinical details of participants) performed ultrasonography in all children. No information was available on estimates of intra- or inter-rater reliability of ultrasonography. Hepatic steatosis was diagnosed according to ultrasonographic characteristics, including diffuse hyper-echogenicity of the liver relative to the kidneys, ultrasonography beam attenuation, or poor visualization of the intrahepatic vessel borders and diaphragm [23]. Liver ultrasonography has a good sensitivity and specificity for detecting moderate and severe hepatic steatosis, and traditionally its sensitivity is thought to be poor when >20%-30% of hepatocytes are steatotic. In this study, a semi-quantitative ultrasonographic scoring of the degree of hepatic steatosis was not available.

### **Genetic analysis**

Genomic DNA was extracted from peripheral blood leukocytes using salting-out procedures. Genotyping was carried out by predesigned TaqMan probes (Applied Biosystem, USA), according to the manufacturers' protocol. Polymorphism genotyping was performed using 7900 HT Real Time PCR (Applied Biosystem, USA). We genotyped eleven genetic variants that have been previously robustly associated with NAFLD onset or progression: rs738409 (*PNPLA3*), rs58542926 (*TM6SF2*), rs1260326 (*GCKR*), rs2236212 (*ELOVL2*), rs116454156 (*GPR120*), rs1535 (*FADS2*), rs13412852 (*LPIN1*), rs641738 (*MBOAT7*), rs1800591 (*MTTP*), rs3750861 (*KLF6*) and rs4880 (*SOD2*). To estimate the overall contribution of the genetic variants under study on the risk of NAFLD and to test the diagnostic accuracy of this approach, a weighted genetic risk score (wGRS) was also derived from the SNP genotypes.

### **Statistical analysis**

Data are expressed as means $\pm$ SD, medians (inter-quartile ranges, IQR) or proportions. The unpaired Student's *t*-test (for normally distributed variables), the Mann-Whitney U-test (for

non-normally distributed variables; i.e., z-BMI, HbA1c and serum aminotransferase levels) and the chi-squared test (for categorical variables) were used to test differences in clinical and biochemical characteristics between children with, and without NAFLD. Individual SNPs and wGRS associations with categorical variables were tested using logistic regression analyses, whereas associations with continuous variables were tested by generalized linear regression analyses; the aforementioned statistical tests were performed both in unadjusted models and in models adjusted for age, sex and z-BMI; the presence of a significant interaction with z-BMI was also tested in these regression models. The Benjamini–Hochberg (BH) step-up procedure was applied for multiple test adjustment [24]. Covariates included in multivariable regression models were selected as potential confounding factors based on their biological plausibility.

We also created two subsets (i.e. derivation and validation subsets) from the target sample, randomly splitting our cohort. The derivation subset (n=251) was used to build four logistic regression models for predicting NAFLD: model M1, a clinical model based on established clinical predictors of NAFLD and including age, sex, z-BMI, log-transformed serum triglycerides, ALT, fasting glucose and insulin levels; model M2, a clinical + genetic model in which the three major NAFLD-related SNPs (rs58542926, rs738409, rs1260326) were included; model M3 that added also rs2236212 (ELOVL2) to model M2; and, finally, a model M4, a clinical *plus* genetic model in which all 11 genotyped SNPs were added to the clinical model. The resulting predictive equations were applied to the validation subset of children (n=263) and compared to each other in terms of predictive accuracy. To evaluate the model accuracy, receiver operating characteristic (ROC) curves, plotting sensitivity against 1–specificity, were also constructed for all possible cut-offs for the aforementioned logistic regression models. The area under the ROC curve (AUC), or C-statistic, was estimated as a measure of the probability that a randomly selected individual had NAFLD. This analysis was also performed to predict the presence of NAFLD *plus* high serum ALT levels; however, in such case, we did not split our cohort into derivation and validation subsets, because of the low number of children with NAFLD *plus* high serum ALT levels. A *p*-value <0.05 was considered statistically significant. Statistical analysis were performed using Plink[25], SPSS v22.0 software package (SPSS Inc., Chicago, USA) and STATA software 14.2.

## RESULTS

### Subjects characteristics

We studied a cohort of 514 (54.3% male) Italian obese children and adolescents, who consecutively underwent ultrasonography for detecting hepatic steatosis. Their mean ( $\pm$ SD) age was  $11.2\pm 2.8$  years (range: 6-18 years), mean z-BMI  $3.3\pm 0.8$  and mean waist circumference  $92.1\pm 14$  cm, respectively. The overall prevalence of NAFLD was 67.5% (347 children). In the entire cohort, 80% of children had serum aminotransferase levels within the reference values, whereas nearly 20% of them had slightly elevated ( $>40$  IU/L) ALT or AST levels, with only 1.5% ( $n=8$ ) with serum ALT or AST levels  $\geq 100$  IU/L.

The main clinical and biochemical characteristics of participants stratified by presence or absence of NAFLD are shown in **Table 1**. Compared to those without NAFLD, children with NAFLD were more likely to be boys, older and with higher BMIs. They also had significantly higher values of systolic blood pressure (SBP), serum triglyceride, AST and ALT concentrations (although most of these children had fairly normal values), but lower HDL-cholesterol levels. In contrast, diastolic blood pressure (DBP), total cholesterol, LDL-cholesterol, fasting glucose, 2-hour OGTT glucose and hemoglobin A1c levels did not significantly differ between the two groups. When we compared wGRS, both in the derivation and validation subsets, children with NAFLD had a significantly greater wGRS than those without the liver disease.

### Associations of genetic variants with NAFLD

The call rate exceeded 99% for all SNPs and genotypes' distributions were in Hardy-Weinberg Equilibrium (HWE). Only four out of 11 SNPs reached a statistical power of at least 80% (rs1260326, rs58542926, rs738409 and rs2236212). **Table 2** shows HWE p-values and the results of the association between NAFLD and the genotyped SNPs. The frequency distributions of *PNPLA3* rs738409, *GCKR* rs1260326, *ELOVL2* rs2236212, and, especially, *TM6SF2* rs58542926 variants remained significantly different between children with and without NAFLD even after adjustment for age, sex and z-BMI. In all these adjusted regression models, no significant interactions with z-BMI were observed in the associations between individual SNPs and NAFLD ( $p$ -value  $\geq 0.20$  for the interaction term in all regression models). These results did not change even after further adjustment for WHtR, a

measure of body fat distribution, which was available in a subgroup of 412 children (data not shown). As also shown in the table, when an adjustment for multiple tests (by using the Benjamini–Hochberg step-up procedure) was made, only the genetic variants in *TM6SF2* rs58542926, *GCKR* rs1260326 and *PNPLA3* rs738409 remained independently associated with NAFLD.

Interestingly, wGRS, calculated on the validation subset, was significantly associated with presence of NAFLD (unadjusted OR of 1.14, 95% CI 1.03-1.25,  $p=0.009$ ). This association remained statistically significant even after adjustment for age, sex, z-BMI, log-transformed serum triglycerides, ALT, fasting glucose and insulin levels (adjusted OR of 1.18, 95% CI 1.04-1.33,  $p=0.010$ ). Additionally, as reported in **Figure 1**, the prevalence of NAFLD increased progressively across wGRS tertiles.

#### **Diagnostic accuracy of the genetic risk score**

In the validation subset, we undertook a ROC curve analysis applying four different logistic regression models in order to identify subjects at higher risk of NAFLD. The clinical model (M1) including age, sex, z-BMI, log-serum triglycerides, ALT, fasting glucose and insulin levels showed an AUC= 0.77 (95%CI 0.70-0.84). In model M2, addition of the three major NAFLD-related SNPs slightly improved the risk prediction for NAFLD (AUC=0.80; 95%CI 0.73-0.86; C-statistic  $p=0.173$  for comparison to model M1). Similar finding was observed in model M3 that was computed as model M2 *plus* rs2236212 (*ELOVL2*) (AUC=0.80; 95%CI 0.74-0.87; C-statistic  $p=0.118$  for comparison to model M1). Finally, the model M4, a clinical + genetic model in which all 11 genotyped SNPs were added, showed an AUC=0.81 (95% CI 0.75-0.88). Notably, compared to the clinical model alone (M1), this final regression model (M4) significantly (albeit modestly) improved risk prediction for NAFLD (C-statistic  $p=0.047$ ).

When we performed the same ROC curve analyses, on the whole cohort, in order to identify the subgroup of children with both NAFLD and elevated (>40 IU/L) serum ALT levels (i.e., a proxy of suspected steatohepatitis [NASH]), we found that the addition to the clinical M1 (*not* including serum ALT levels) of both the three or four major NAFLD-related SNPs (M2: AUC=0.72; 95%CI 0.67-0.77; C-statistic  $p=0.015$  vs. M1; M3: AUC=0.72; 95%CI 0.68-0.77; C-statistic  $p=0.015$  vs. M1) and all genotyped SNPs significantly improved the



risk prediction of NAFLD *plus* elevated ALT levels (M4: AUC=0.74; 95%CI 0.68-0.78; C-statistic  $p=0.005$  vs. M1).

## DISCUSSION

The main and novel finding of this study was that a genetic risk score, based on the combination of 11 genetic risk variants *plus* established clinical risk factors, improved risk prediction for NAFLD in obese children and adolescents by 5.2%, compared to risk prediction based on clinical factors alone. To our knowledge, this is one of the largest studies (involving a mono-ethnic cohort of Caucasian obese children and adolescents) that has examined a broad panel of genetic polymorphisms associated with susceptibility to NAFLD in paediatric obesity.

Our findings are consistent with similar studies that have been performed previously. Nobili *et al.* showed that addition of a 4-polymorphism risk score (including rs738409, rs4880, rs3750861 and rs13412852 variants) to a clinical model performed better than a clinical risk score alone (including age, diastolic blood pressure and serum AST levels) in predicting the risk of NASH in 152 obese children and adolescents with biopsy-proven NAFLD and elevated serum aminotransferase levels [17]. In our study, using a larger polymorphism genetic score (that also included the four aforementioned genetic variants), we found that the addition of the GRS to the established clinical risk factors significantly improved the risk prediction of NAFLD in our cohort of obese children/adolescents and fairly normal serum aminotransferase levels.

The addition of the GRS also significantly improved the predictability of NAFLD with coexistent elevated serum aminotransferase levels (i.e., a proxy of suspected NASH). Furthermore, we observed a graded relationship between increasing GRS and NAFLD prevalence, suggesting the potential clinical relevance of a combination of risk alleles of individually small effects (Figure 1). Overall, all of these findings partly reflect the fact that these genetic variants may act in an additive manner amplifying the phenotypic effect(s), and these data also support the potential contribution of a genetic score approach to improving identification of individuals at increased NAFLD risk [26]. Although in our study the addition of the eleven genotyped SNPs provided a significant, but modest, improvement in the performance of a prediction model for NAFLD, we believe that the potential future availability of new genetic *loci* of susceptibility to NAFLD might further

improve the role of genetics in risk prediction for NAFLD in paediatric obesity.

Furthermore, it is likely that the progressive reduction in the costs of genotyping associated with the validation of its predictive power, will progressively promote the use of genetic scores in clinical practice.

In our study, only four out of the eleven SNPs included in the score (*TM6SF2* rs58542926, *GCKR* rs1260326, *PNPLA3* rs738409 and *ELOVL2* rs2236212) were significantly associated with an increased risk of NAFLD, independently of age, sex and z-BMI. Since, we did not find any significant association between NAFLD and other genotyped SNPs, it is plausible to assume that this could be due to the relatively low sample size and the low frequency of the risk alleles. However, since it is likely that most of our children and adolescents with ultrasound-defined hepatic steatosis and fairly normal serum aminotransferase levels had NAFL (defined as the presence of  $\geq 5\%$  hepatic steatosis on histology, most likely  $\geq 30\%$  as it was evident by ultrasonography [27][28]), and none (or very few) of them had advanced fibrosis or cirrhosis, we believe that other possible explanations for this finding could also be that the use of a more accurate method for measuring intra-hepatic fat content (by magnetic resonance imaging or spectroscopy) could allow detection of significant associations with some of genetic variants [29]. Moreover, the lack of significant associations of *SOD2* and *KLF6* genes with NAFLD traits might be partly due to the predominant role of these genes in the pathogenesis of advanced forms of NAFLD (NASH and fibrosis) [17]. With regards to *PNPLA3*, *TM6SF2* and *GCKR* genes, our findings corroborate the results of recently published studies showing that these three SNPs are strongly associated with the presence of NAFLD in obese children/adolescents of European ancestry [30][31][32][33][34].

In our study, we also showed for the first time a potential involvement of the *ELOVL2* rs2236212 variant in paediatric NAFLD. *ELOVL2* gene encodes for elongase subtype 2, an enzyme involved in the endogenous synthesis of long-chain polyunsaturated fatty acids (n-3 PUFAs). In a large genome-wide association study, Lemaitre *et al.* reported a significant inverse association between polymorphisms in *ELOVL2* gene and plasma n-3 PUFA levels (mostly lower plasma docosahexaenoic acid levels) in populations of European ancestry, underscoring the possible role of the *ELOVL2* variant in phospholipid metabolism [35].

NAFLD is a common disorder and identifying new determinants, such as new genes, may improve understanding of mechanisms potentially implicated in NAFLD development and

progression [3]. Moreover, by identifying individuals at higher risk of developing NAFLD based on their GRS, clinicians could more accurately assess a patient's unique risk of developing NAFLD and, therefore, recommend personalized interventions to modify that risk. Currently, no recommendations for the use of genetic screening have been made in scientific societies' guidelines [36][37][38][39]. Nevertheless, it has been suggested that genotyping could be useful to stratify the prognosis of NAFLD (risk of progression to cirrhosis and hepatocarcinoma) and the probability of responses to specific therapeutic approaches in selected patients [40][41]. However, further research is needed before the use of genetic testing can be routinely recommended for risk stratification of NAFLD patients in clinical practice [41].

Our study has some limitations that should be mentioned. First, the cross-sectional design of the study limits our ability to establish causality and temporality of the observed associations, although this limitation is mitigated by the fact that the genetic variants are inherited and therefore, reverse causality does not apply. We also could not control for the presence of lifestyle factors associated with NAFLD development, such as sweetened drinks intake and physical activity [42]. Second, the diagnosis of NAFLD was based on ultrasonography and exclusion of secondary causes of hepatic steatosis. Ultrasonography is the recommended first-line imaging method for detecting hepatic steatosis in clinical practice, and it enables a reliable and accurate detection of mild-to-moderate hepatic steatosis. Unfortunately, we did not perform a semi-quantitative ultrasound score of the degree of hepatic steatosis, such as the ultrasonographic fatty liver indicator (US-FLI), which may detect more accurately mild steatosis (i.e. minimum amount 10% on histology [43]). Liver ultrasonography was undertaken by two experienced radiologists, but unfortunately, we are not able to comment on the intra- and inter-observer agreement as this data was not collected. In addition, we did not perform liver biopsy, which remains the current 'reference' method for diagnosing and staging NAFLD [23][27]. Our sample was composed mostly by children and adolescents with fairly normal serum aminotransferase levels and undertaking liver biopsies in these subjects is always far more challenging as sometimes it is necessary for the child to undergo a general anaesthetic for the procedure to be undertaken. Third, we did not have any data on vibration-controlled transient elastography (Fibroscan<sup>®</sup>) for measuring liver stiffness in order to non-invasively stage hepatic fibrosis. Fourth, ethnicity limited to Caucasian, which does not allow to export

results to children and adolescents with other ethnic backgrounds. Finally, the lack of a validation cohort cautions against the generalizability of our findings to other paediatric groups.

Our study also has strengths. First, the study population; e.g. obese children and adolescents have much lower obesity-associated comorbidities than obese adults and, by implication this offers the opportunity to explore genetic associations with NAFLD with reduced influence from other potential confounding factors. Second, the sample size of the study was relatively large to explore associations with NAFLD, despite low allele frequencies and small gene-effect sizes. Third, we have included well-established clinical risk factors in the prediction risk score and adjusted for potential confounders.

In conclusion, in a large cohort of unrelated obese children and adolescents, our study has identified three polymorphisms, i.e. *TM6SF2* rs58542926, *PNPLA3* rs738409 and *GCKR* rs1260326, as major genetic variants that are strongly associated with NAFLD. Our data also suggests the potential involvement of the *ELOVL2* rs2236212 variant as this polymorphism was also independently associated with NAFLD. Furthermore, our data suggest that the addition of a 11-polymorphism risk score to multiple established clinical risk factors improved (albeit modestly) risk prediction of NAFLD. We suggest that further research is now needed to validate our findings in other cohorts of children and adolescents, and to examine whether these genetic variants may predict resolution (or progression) of NAFLD with lifestyle interventions or treatments.

## **ACKNOWLEDGMENTS**

We kindly thank the patients and their families who participated in the study. We also thank the dedicated staff of the Pediatric Diabetes and Metabolic Disorders Unit of the University Hospital in Verona for their support during the clinical study.

**Author contributions:** CZ, AMantovani, FO researched and analyzed data and wrote the manuscript. AMorandi, CM and GT researched data and discussed the manuscript. AMorandi and FO researched data. CM, GT, LV, EMDG and CDB edited the manuscript, and provided substantial contribution to the overall discussion. CM, GT and CZ designed the study, researched data, co-wrote, and edited the manuscript. CM is the guarantor of this

work and, as such, had full access to all the data in the study and takes responsibility for the integrity and the accuracy of the data analysis.

**Conflict of Interest Statement:** The Authors have no potential conflicts of interest to disclose.

**Funding Sources:** GT and CM are supported in part by grants from the University School of Medicine of Verona, Verona, Italy. CDB is support in part by the Southampton National Institute for Health Research Biomedical Research Centre.

## REFERENCES

- [1] Anderson EL, Howe LD, Jones HE, Higgins JPT, Lawlor DA, Fraser A. The Prevalence of Non-Alcoholic Fatty Liver Disease in Children and Adolescents: A Systematic Review and Meta-Analysis. *PLoS One* 2015. doi:10.1371/journal.pone.0140908.
- [2] Tiniakos DG, Vos MB, Brunt EM. Nonalcoholic fatty liver disease: pathology and pathogenesis. *Annu Rev Pathol* 2010. doi:10.1146/annurev-pathol-121808-102132.
- [3] Lonardo A, Bellentani S, Argo CK, Ballestri S, Byrne CD, Caldwell SH, et al. Epidemiological modifiers of non-alcoholic fatty liver disease: Focus on high-risk groups. *Dig Liver Dis* 2015. doi:10.1016/j.dld.2015.08.004.
- [4] Feldstein AE, Charatcharoenwitthaya P, Treeprasertsuk S, Benson JT, Enders FB, Angulo P. The natural history of non-alcoholic fatty liver disease in children: A follow-up study for up to 20 years. *Gut* 2009. doi:10.1136/gut.2008.171280.
- [5] Loomba R, Schork N, Chen CH, Bettencourt R, Bhatt A, Ang B, et al. Heritability of Hepatic Fibrosis and Steatosis Based on a Prospective Twin Study. *Gastroenterology* 2015. doi:10.1053/j.gastro.2015.08.011.
- [6] Schwimmer JB, Celedon MA, Lavine JE, Salem R, Campbell N, Schork NJ, et al. Heritability of Nonalcoholic Fatty Liver Disease. *Gastroenterology* 2009. doi:10.1053/j.gastro.2009.01.050.
- [7] Anstee QM, Seth D, Day CP. Genetic Factors That Affect Risk of Alcoholic and Nonalcoholic Fatty Liver Disease. *Gastroenterology* 2016. doi:10.1053/j.gastro.2016.01.037.
- [8] Mann JP, Valenti L, Scorletti E, Byrne CD, Nobili V. Nonalcoholic Fatty Liver Disease in

Children. *Semin Liver Dis* 2018;38:1–13. doi:10.1055/s-0038-1627456.

- [9] Speliotes EK, Yerges-Armstrong LM, Wu J, Hernaez R, Kim LJ, Palmer CD, et al. Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. *PLoS Genet* 2011. doi:10.1371/journal.pgen.1001324.
- [10] Valenti L, Motta BM, Alisi A, Sartorelli R, Buonaiuto G, Dongiovanni P, et al. LPIN1 rs13412852 polymorphism in pediatric nonalcoholic fatty liver disease. *J Pediatr Gastroenterol Nutr* 2012. doi:10.1097/MPG.0b013e3182442a55.
- [11] Anstee QM, Day CP. The Genetics of Nonalcoholic Fatty Liver Disease: Spotlight on PNPLA3 and TM6SF2. *Semin Liver Dis* 2015. doi:10.1055/s-0035-1562947.
- [12] Barman M, Nilsson S, Naluai ÅT, Sandin A, Wold AE, Sandberg AS. Single nucleotide polymorphisms in the FADS gene cluster but not the ELOVL2 gene are associated with serum polyunsaturated fatty acid composition and development of allergy (in a Swedish birth cohort). *Nutrients* 2015. doi:10.3390/nu7125521.
- [13] Mancina RM, Dongiovanni P, Petta S, Pingitore P, Meroni M, Rametta R, et al. The MBOAT7-TMC4 Variant rs641738 Increases Risk of Nonalcoholic Fatty Liver Disease in Individuals of European Descent. *Gastroenterology* 2016. doi:10.1053/j.gastro.2016.01.032.
- [14] Dongiovanni P, Anstee QM, Valenti L. Genetic predisposition in NAFLD and NASH: impact on severity of liver disease and response to treatment. *Curr Pharm Des* 2013. doi:10.2174/13816128113199990381.
- [15] Eslam M, Valenti L, Romeo S. Genetics and epigenetics of NAFLD and NASH: Clinical impact. *J Hepatol* 2018;68:268–79. doi:10.1016/j.jhep.2017.09.003.
- [16] León-Mimila P, Vega-Badillo J, Gutiérrez-Vidal R, Villamil-Ramírez H, Villareal-Molina T, Larrieta-Carrasco E, et al. A genetic risk score is associated with hepatic triglyceride content and non-alcoholic steatohepatitis in Mexicans with morbid obesity. *Exp Mol Pathol* 2015. doi:10.1016/j.yexmp.2015.01.012.
- [17] Nobili V, Donati B, Panera N, Vongsakulyanon A, Alisi A, Dallapiccola B, et al. A 4-polymorphism risk score predicts steatohepatitis in children with nonalcoholic fatty liver disease. *J Pediatr Gastroenterol Nutr* 2014. doi:10.1097/MPG.0000000000000279.
- [18] Stender S, Kozlitina J, Nordestgaard BG, Tybjaerg-Hansen A, Hobbs HH, Cohen JC.

Adiposity amplifies the genetic risk of fatty liver disease conferred by multiple loci. *Nat Genet* 2017. doi:10.1038/ng.3855.

- [19] Vespasiani-Gentilucci U, Gallo P, Dell'Unto C, Volpentesta M, Antonelli-Incalzi R, Picardi A. Promoting genetics in non-alcoholic fatty liver disease: Combined risk score through polymorphisms and clinical variables. *World J Gastroenterol* 2018. doi:10.3748/wjg.v24.i43.4835.
- [20] De Onis M, Onyango AW, Borghi E, Siyam A, Nishida C, Siekmann J. Development of a WHO growth reference for school-aged children and adolescents. *Bull World Health Organ* 2007. doi:10.2471/BLT.
- [21] Valerio G, Maffei C, Saggese G, Ambruzzi MA, Balsamo A, Bellone S, et al. Diagnosis, treatment and prevention of pediatric obesity: Consensus position statement of the Italian Society for Pediatric Endocrinology and Diabetology and the Italian Society of Pediatrics. *Ital J Pediatr* 2018. doi:10.1186/s13052-018-0525-6.
- [22] Maffei C, Banzato C, Talamini G. Waist-to-Height Ratio, a Useful Index to Identify High Metabolic Risk in Overweight Children. *J Pediatr* 2008. doi:10.1016/j.jpeds.2007.09.021.
- [23] Hernaez R, Lazo M, Bonekamp S, Kamel I, Brancati FL, Guallar E, et al. Diagnostic accuracy and reliability of ultrasonography for the detection of fatty liver: A meta-analysis. *Hepatology* 2011. doi:10.1002/hep.24452.
- [24] Hochberg B. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *J R Stat Soc* 1995. doi:10.2307/2346101.
- [25] Purcell S, Neale B, Todd-brown K, Thomas L, Ferreira MAR, Bender D, et al. REPORT PLINK : A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses 2007;81:559–75. doi:10.1086/519795.
- [26] Sadee W, Hartmann K, Seweryn M, Pietrzak M, Handelman SK, Rempala GA. Missing heritability of common diseases and treatments outside the protein-coding exome. *Hum Genet* 2014. doi:10.1007/s00439-014-1476-7.
- [27] Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, et al. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. *Hepatology* 2018;10. doi:10.1002/hep.29367.
- [28] Bril F, Ortiz-Lopez C, Lomonaco R, Orsak B, Freckleton M, Chintapalli K, et al. Clinical

value of liver ultrasound for the diagnosis of nonalcoholic fatty liver disease in overweight and obese patients. *Liver Int* 2015;35:2139–46. doi:10.1111/liv.12840.

- [29] Umamo GR, Caprio S, Di Sessa A, Chalasani N, Dykas DJ, Pierpont B, et al. The rs626283 variant in the MBOAT7 gene is associated with insulin resistance and fatty liver in Caucasian obese youth. *Am J Gastroenterol* 2018. doi:10.1038/ajg.2018.1.
- [30] Goyal NP, Schwimmer JB. The Genetics of Pediatric Nonalcoholic Fatty Liver Disease. *Clin Liver Dis* 2018. doi:10.1016/j.cld.2017.08.002.
- [31] Marzuillo P, Grandone A, Perrone L, del Giudice EM. Understanding the pathophysiological mechanisms in the pediatric non-alcoholic fatty liver disease: The role of genetics. *World J Hepatol* 2015. doi:10.4254/wjh.v7.i11.1439.
- [32] Lin YC, Chang PF, Chang MH, Ni YH. Genetic variants in GCKR and PNPLA3 confer susceptibility to nonalcoholic fatty liver disease in obese individuals. *Am J Clin Nutr* 2014. doi:10.3945/ajcn.113.079749.
- [33] Tricò D, Caprio S, Umamo GR, Pierpont B, Nouws J, Galderisi A, et al. Metabolic Features of Nonalcoholic Fatty Liver (NAFL) in Obese Adolescents: Findings from a Multi-ethnic Cohort. *Hepatology* 2018. doi:10.1002/hep.30035.
- [34] Di Sessa A, Umamo GR, Cirillo G, Del Prete A, Iacomino R, Marzuillo P, et al. The Membrane-bound O-Acyltransferase7 rs641738 Variant in Pediatric Nonalcoholic Fatty Liver Disease. *J Pediatr Gastroenterol Nutr* 2018. doi:10.1097/MPG.0000000000001979.
- [35] Lemaitre RN, Tanaka T, Tang W, Manichaikul A, Foy M, Kabagambe EK, et al. Genetic loci associated with plasma phospholipid N-3 fatty acids: A Meta-Analysis of Genome-Wide association studies from the charge consortium. *PLoS Genet* 2011. doi:10.1371/journal.pgen.1002193.
- [36] European Association for the Study of the Liver (EASL), European Association for the Study of Diabetes (EASD), European Association for the Study of Obesity (EASO). EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. *J Hepatol* 2016. doi:10.1016/j.jhep.2015.11.004.
- [37] Institute for Health and Care Excellence N. Non-alcoholic fatty liver disease (NAFLD): assessment and management | Guidance and guidelines | NICE. *Non-Alcoholic Fat Liver Dis Assess Manag* 2016.
- [38] Leoni S, Tovoli F, Napoli L, Serio I, Ferri S, Bolondi L. Current guidelines for the



management of non-alcoholic fatty liver disease: A systematic review with comparative analysis. *World J Gastroenterol* 2018. doi:10.3748/wjg.v24.i30.3361.

- [39] Vos MB, Abrams SH, Barlow SE, Caprio S, Daniels SR, Kohli R, et al. NASPGHAN Clinical Practice Guideline for the Diagnosis and Treatment of Nonalcoholic Fatty Liver Disease in Children: Recommendations from the Expert Committee on NAFLD (ECON) and the North American Society of Pediatric Gastroenterology, Hepatology and Nu. *J Pediatr Gastroenterol Nutr* 2017. doi:10.1097/MPG.0000000000001482.
- [40] Petta S, Valenti L, Bugianesi E, Targher G, Bellentani S, Bonino F, et al. A “systems medicine” approach to the study of non-alcoholic fatty liver disease. *Dig Liver Dis* 2016. doi:10.1016/j.dld.2015.10.027.
- [41] Loria P, Adinolfi LE, Bellentani S, Bugianesi E, Grieco A, Fargion S, et al. Practice guidelines for the diagnosis and management of nonalcoholic fatty liver disease. A decalogue from the Italian Association for the Study of the Liver (AISF) Expert Committee. *Dig Liver Dis* 2010. doi:10.1016/j.dld.2010.01.021.
- [42] Nobili V, Liccardo D, Bedogni G, Salvatori G, Gnani D, Bersani I, et al. Influence of dietary pattern, physical activity, and I148M PNPLA3 on steatosis severity in at-risk adolescents. *Genes Nutr* 2014. doi:10.1007/s12263-014-0392-8.
- [43] Ballestri S, Nascimbeni F, Baldelli E, Marrazzo A, Romagnoli D, Targher G, et al. Ultrasonographic fatty liver indicator detects mild steatosis and correlates with metabolic/histological parameters in various liver diseases. *Metabolism* 2017. doi:10.1016/j.metabol.2017.04.003.

**Table 1.** Clinical and biochemical characteristics of children and adolescents stratified by NAFLD status.

	<b>With NAFLD (n=347)</b>	<b>Without NAFLD (n=167)</b>	<b>P-value*</b>
<i>n</i> (M/F)	204/143	75/92	0.003
<b>Age</b> (years)	11.3±2.7	10.8±3.0	0.042
<b>Weight</b> (kg)	73.3 [58-91]	62.1 [49.6-76]	<0.0001
<b>z-BMI</b> (kg/m <sup>2</sup> )	3.42 [3.01-3.73]	2.95 [2.49-3.61]	<0.0001
<b>Waist circumference</b> (cm), n=412	94 [87-104]	85 [78-93]	<0.0001
<b>WHtR</b> , n=412	0.64±0.07	0.58±0.07	<0.0001
<b>Systolic blood pressure</b> (mmHg)	118 [110-125]	114 [107-122]	0.007
<b>Diastolic blood pressure</b> (mmHg)	68 [60-75]	66 [60-73]	0.351
<b>Total cholesterol</b> (mmol/L)	4.03 [3.54-4.56]	3.97 [3.53-4.55]	0.684
<b>LDL cholesterol</b> (mmol/L)	2.41±0.6	2.39±0.6	0.939
<b>HDL cholesterol</b> (mmol/L)	1.10 [0.9-1.28]	1.18 [1.1-1.38]	0.001
<b>Triglycerides</b> (mmol/L)	1.02 [0.7-1.43]	0.86 [0.6-1.28]	0.006
<b>ALT</b> (U/L)	28.5 [20-43]	22.0 [17-29]	<0.0001
<b>AST</b> (U/L)	24.0 [19-29]	21.0 [18-26]	0.003
<b>Fasting glucose</b> (mmol/L)	4.61 [4.3-4.9]	4.60 [4.3-4.9]	0.971
<b>HbA<sub>1c</sub> DCCT</b> (%)	5.5 [5.4-5.7]	5.4 [5.3-5.7]	0.257
<b>HbA<sub>1c</sub> IFCC</b> (mmol/mol)	37 [36-39]	36 [34-39]	0.257
<b>Fasting insulin</b> (mU/L)	23.5 [13.4-37.3]	14.0 [8.7-24.6]	<0.001
<b>wGRS</b> (11 variants on validation subset)	12.02±3.08	10.27±2.40	<0.001
<b>wGRS</b> (11 variants on validation subset)	11.62±2.77	10.59±3.03	0.007

Sample size, *n*=514, unless otherwise indicated. Data are expressed as means±SD, medians and interquartile range [IQR] or proportions. \*Differences between the two groups of individuals were tested by the unpaired Student's t-test for normally distributed variables, the Mann-Whitney U-test for non-normally distributed variables or the chi-squared test for categorical variables, respectively.

**Abbreviations:** ALT, alanine aminotransferase; AST, aspartate aminotransferase; z-BMI, z-score body mass index; HbA<sub>1c</sub> DCCT, Diabetes Control and Complication Trial-Aligned Haemoglobin A1c; HbA<sub>1c</sub> IFCC, International Federation of Clinical Chemistry-Aligned Haemoglobin A1c; HbA1c, hemoglobin A1c; WHtR, waist circumference to height ratio; wGRS, weighted Genetic Risk Score.

**Table 2.** Unadjusted and age-, sex- and z-BMI-adjusted associations between nonalcoholic fatty liver (NAFLD) and each of the eleven genotyped single nucleotide polymorphisms.

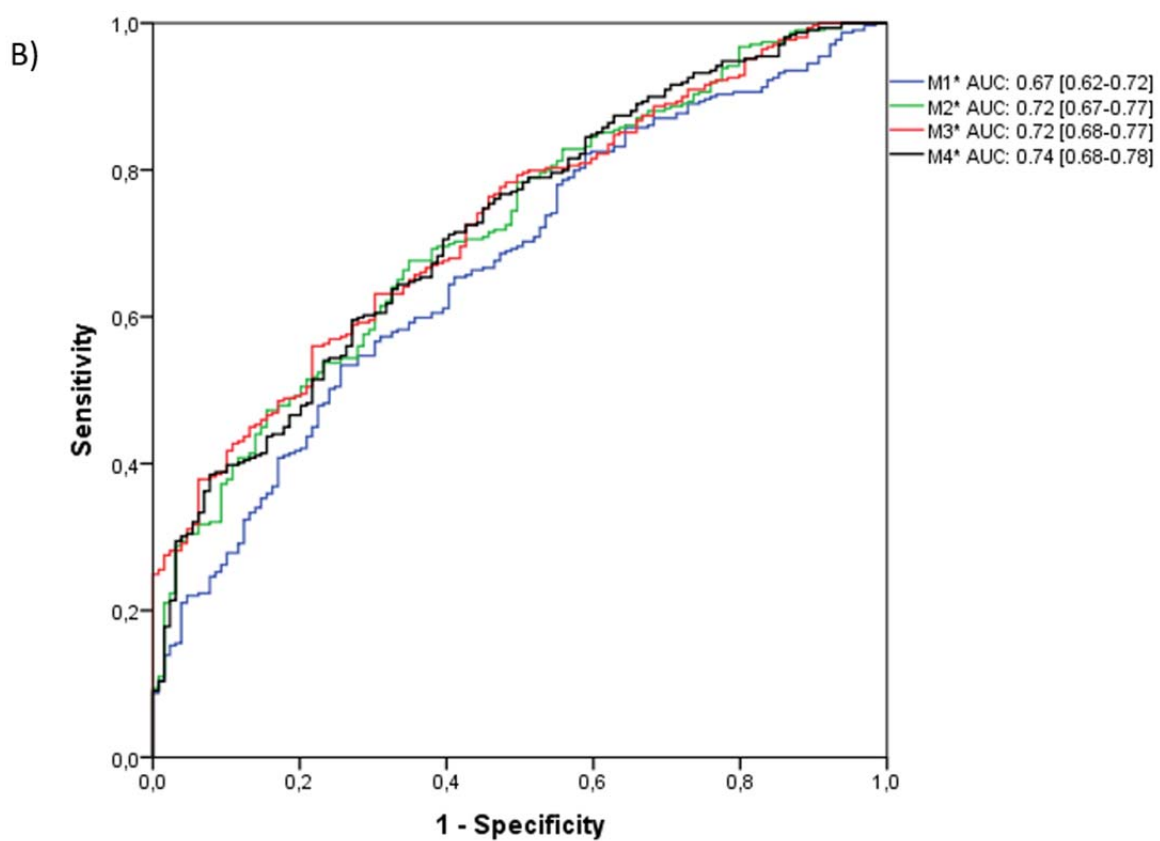
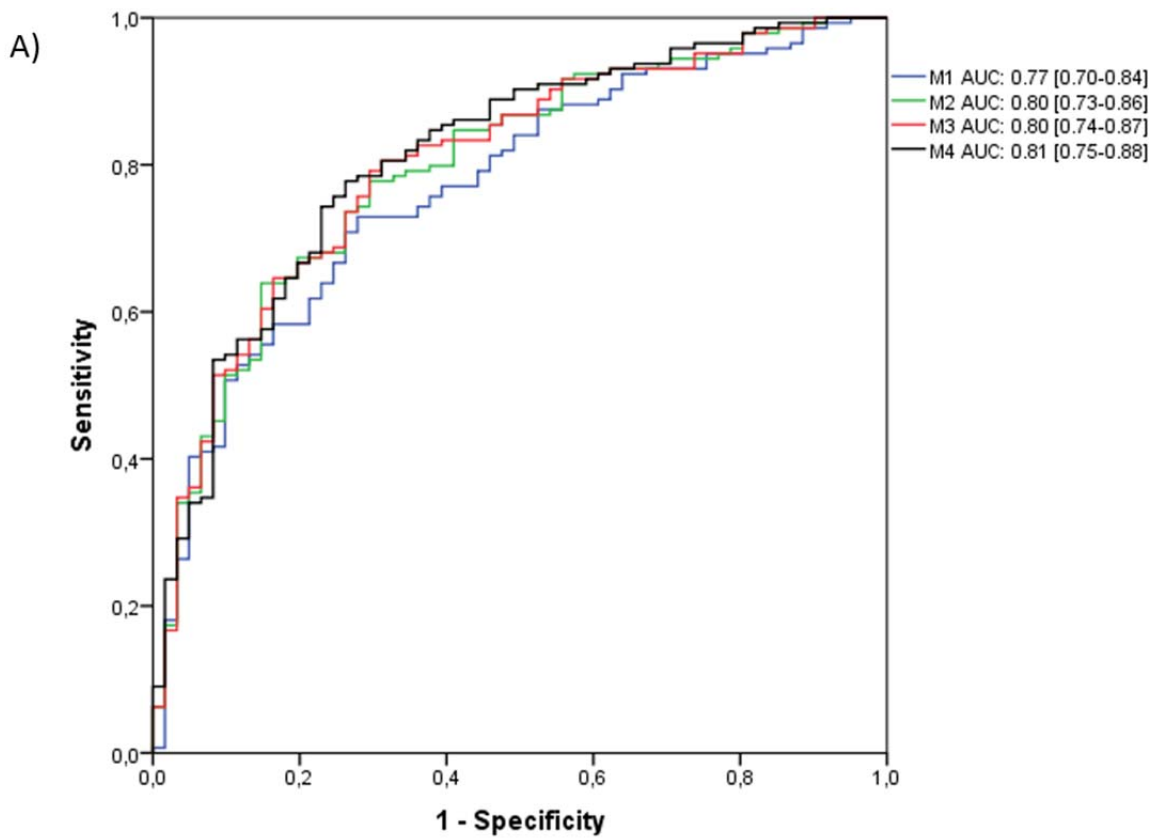
Chr	SNP ID	Gene ID	Minor Allele	MAF	Risk allele	HWE P-value	Unadjusted OR [95% CI]*	P-value*	Adjusted OR [95% CI]**	P-value**
19	rs58542926	<i>TM6SF2</i>	A	0.04	A	0.410	4.15 [1.44-11.9]	0.008 <sup>†</sup>	4.13 [1.41-12.1]	0.002 <sup>†</sup>
2	rs1260326	<i>GCKR</i>	C	0.47	T	0.486	1.50 [1.14-1.96]	0.003 <sup>†</sup>	1.53 [1.15-2.03]	0.003 <sup>†</sup>
22	rs738409	<i>PNPLA3</i>	G	0.31	G	0.734	1.51 [1.12-2.03]	0.007 <sup>†</sup>	1.58 [1.15-2.18]	0.004 <sup>†</sup>
6	rs2236212	<i>ELOVL2</i>	C	0.45	C	0.398	1.35 [1.03-1.77]	0.032	1.34 [1.01-1.80]	0.047
10	rs116454156	<i>GPR120</i>	A	0.02	A	0.618	1.68 [0.61-4.63]	0.317	1.79 [0.58-5.56]	0.289
11	rs1535	<i>FADS2</i>	G	0.29	G	0.471	1.16 [0.87-1.55]	0.309	1.16 [0.85-1.57]	0.342
2	rs13412852	<i>LPIN1</i>	T	0.34	C	0.847	1.13 [0.86-1.48]	0.395	1.13 [0.84-1.50]	0.421
19	rs641738	<i>MBOAT7</i>	T	0.44	T	0.915	1.15 [0.88-1.49]	0.316	1.12 [0.85-1.49]	0.442
4	rs1800591	<i>MTTP</i>	T	0.28	T	0.068	1.02 [0.77-1.34]	0.916	1.08 [0.80-1.45]	0.625
10	rs3750861	<i>KLF6</i>	T	0.08	C	0.109	1.15 [0.73-1.81]	0.556	1.10 [0.68-1.79]	0.697
6	rs4880	<i>SOD2</i>	G	0.49	A	0.748	1.05 [0.80-1.36]	0.743	1.05 [0.80-1.39]	0.721

Sample size,  $n=514$ . SNPs are ordered by P-values (see last column). Presence of NAFLD was the dependent variable in all logistic regression models. \*Data were tested by univariable (unadjusted) logistic regression analysis; \*\*Data were tested by multivariable logistic regression analysis after adjustment for age, sex and z-BMI. <sup>†</sup>Significant P-values after further adjustment for multiplicity (by using the Benjamini–Hochberg step-up procedure).

Abbreviations: HWE: Hardy-Weinberg Equilibrium; z-BMI, z-score body mass index; Chr, chromosome position; SNP, single nucleotide polymorphism; MAF, minor allele frequency; OR; odds ratio; CI, confidence interval.

## FIGURE LEGENDS

**Figure 1.** Prevalence of ultrasound-defined nonalcoholic fatty liver (NAFLD) in obese children and adolescents stratified by tertiles of weighted genetic risk score (wGRS).



**Figure 2. Panel A.** Receiver operating characteristic (ROC) curves with and without genotyped single nucleotide polymorphisms (SNP) in risk prediction models of nonalcoholic fatty liver (NAFLD) in obese children and adolescents. Model M1: clinical regression model including age, sex, z-BMI, serum triglycerides, alanine aminotransferase, glucose and insulin levels (area under curve [AUC]: 0.77, 95% CI 0.70-0.84). Model M2: regression model including the same variables of M1 *plus* the three major SNPs (*GCKR* rs1260326, *TM6SF2* rs58542926 and *PNPLA3* rs738409) associated with NAFLD (AUC: 0.80, 95% CI 0.73-0.86). Model M3: regression model including the same variables of M2 *plus* ELOVL2 rs2236212 (AUC: 0.80, 95% CI 0.74-0.87). Model M4: regression model including the same variables of M1 *plus* all 11 genotyped SNPs (AUC: 0.81, 95% CI 0.75-0.88). Compared to the clinical regression model alone (M1), the model M4 significantly improved risk prediction for NAFLD ( $p=0.047$  by Harrell's C-statistic).

*Panel B.* Receiver operating characteristic (ROC) curves with and without genotyped SNPs in risk prediction models of suspected nonalcoholic steatohepatitis (defined as hepatic steatosis on ultrasound *plus* elevated serum alanine aminotransferase levels) in the whole cohort of obese children and adolescents. \*In Models M1, M2, M3 and M4 were included the same list of variables of the aforementioned logistic regression models (except for exclusion of serum ALT in model M1). NB: In this analysis, we excluded those patients with NAFLD and normal serum ALT levels.