European Journal of Clinical Nutrition (2017), 1–7

www.nature.com/ejcn

ORIGINAL ARTICLE

Serum metabolomic profiling highlights pathways associated with liver fat content in a general population sample

M Koch[1,2], S Freitag-Wolf[3], S Schlesinger[1], J Borggrefe[4], JR Hov[5,6,7,8], MK Jensen[2], J Pick[9], MRP Markus[10,11,12], T Höpfner[1], G Jacobs[1,13], S Siegert[14], A Artati[15], G Kastenmüller[16,17], W Römisch-Margl[16], J Adamski[15,17,18], T Illig[19,20], M Nothnagel[14], TH Karlsen[5,6,7,8], S Schreiber[21,22], A Franke[21], M Krawczak[3], U Nöthlings[9,23] and W Lieb[1,13,23]

BACKGROUND/OBJECTIVES: Fatty liver disease (FLD) is an important intermediate trait along the cardiometabolic disease spectrum and strongly associates with type 2 diabetes. Knowledge of biological pathways implicated in FLD is limited. An untargeted metabolomic approach might unravel novel pathways related to FLD.

SUBJECTS/METHODS: In a population-based sample ($n = 555$) from Northern Germany, liver fat content was quantified as liver signal intensity using magnetic resonance imaging. Serum metabolites were determined using a non-targeted approach. Partial least squares regression was applied to derive a metabolomic score, explaining variation in serum metabolites and liver signal intensity. Associations of the metabolomic score with liver signal intensity and FLD were investigated in multivariable-adjusted robust linear and logistic regression models, respectively. Metabolites with a variable importance in the projection >1 were entered in *in silico* overrepresentation and pathway analyses.

RESULTS: In univariate analysis, the metabolomics score explained 23.9% variation in liver signal intensity. A 1-unit increment in the metabolomic score was positively associated with FLD ($n = 219$; odds ratio: 1.36; 95% confidence interval: 1.27–1.45) adjusting for age, sex, education, smoking and physical activity. A simplified score based on the 15 metabolites with highest variable importance in the projection statistic showed similar associations. Overrepresentation and pathway analyses highlighted branched-chain amino acids and derived gamma-glutamyl dipeptides as significant correlates of FLD.

CONCLUSIONS: A serum metabolomic profile was associated with FLD and liver fat content. We identified a simplified metabolomics score, which should be evaluated in prospective studies.

European Journal of Clinical Nutrition advance online publication, 5 April 2017; doi:10.1038/ejcn.2017.43

INTRODUCTION

Progressive fat accumulation in the liver emerges as an important intermediate phenotype along the cardiometabolic disease spectrum. Fatty liver disease (FLD) is defined as the accumulation of liver fat exceeding 5.56%.[1] Prevalences of 35–40% were observed in general population samples[1,2] and even higher prevalence proportions in certain clinical settings.[3] In prospective analyses, FLD is an important predictor of type 2 diabetes, subclinical and clinical cardiovascular diseases,[2] liver cirrhosis[4] and mortality.[5] Abdominal obesity and insulin resistance are strongly associated with FLD.[6] However, the complex pathophysiology of FLD and particularly the metabolic derangements associated with FLD are poorly understood[7,8] and deserve further investigations.

Metabolomics, the quantification of a broad spectrum of small-molecule metabolites, including metabolic intermediates, signaling molecules and secondary metabolites in biological specimens might provide further insights into FLD etiology.[8] So far, only few studies have investigated associations of blood metabolomic profiles with FLD[8] and the studies have been limited to small samples[9] of patients undergoing liver biopsies,[10] or severely obese patients undergoing bariatric surgery.[11] Furthermore, many prior analyses were limited by the assessment of FLD only as a dichotomous trait.[9,12] Thus, the association of metabolomic

[1]Institute of Epidemiology, Kiel University, Kiel, Germany; [2]Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA; [3]Institute of Medical Informatics and Statistics, Kiel University, Kiel, Germany; [4]Department of Radiology, University of Cologne, Cologne, MA, USA; [5]Division of Cancer Medicine, Department of Transplantation Medicine, Surgery and Transplantation, Norwegian PSC Research Center, Oslo University Hospital, Rikshospitalet, Oslo, Norway; [6]K.G. Jebsen Inflammation Research Center, Institute of Clinical Medicine, University of Oslo, Oslo, Norway; [7]Division of Cancer Medicine, Surgery and Transplantation, Research Institute of Internal Medicine, Oslo University Hospital, Oslo, Norway; [8]Division of Cancer Medicine, Section of Gastroenterology, Department of Transplantation Medicine, Surgery and Transplantation, Oslo University Hospital, Rikshospitalet, Oslo, Norway; [9]Nutritional Epidemiology, Department of Nutrition and Food Science, Rheinische Friedrich-Wilhelms-University Bonn, Bonn, Germany; [10]Department of Internal Medicine B, University Medicine Greifswald, Greifswald, Germany; [11]DZHK (German Centre for Cardiovascular Research), Partner Site Greifswald, Greifswald, Germany; [12]Department of Study of Health in Pomerania/Clinical-Epidemiological Research, Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany; [13]PopGen Biobank, University Medical Center Schleswig-Holstein, Kiel, Germany; [14]Cologne Center for Genomics, University of Cologne, Cologne, Germany; [15]Institute of Experimental Genetics, Genome Analysis Center, Helmholtz Zentrum München, Neuherberg, Germany; [16]Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, Neuherberg, Germany; [17]Deutsches Zentrum für Diabetesforschung (DZD), German Centere for Diabetes Research, Neuherberg, Germany; [18]Experimental Genetics, Technical University of Munich, Freising, Germany; [19]Hannover Unified Biobank, Hannover Medical School, Hannover, Germany; [20]Institute for Human Genetics, Hannover Medical School, Hanover, Germany; [21]Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany and [22]Institute of Internal Medicine I, University Medical Center Schleswig-Holstein, Kiel, Germany. Correspondence: Dr M Koch, Department of Nutrition, Harvard T.H. Chan School of Public Health, 655 Huntington Avenue, Boston, MA 02115, USA.
E-mail: mkoch@hsph.harvard.edu
[23]These authors contributed equally to this work.
Received 5 August 2016; revised 22 February 2017; accepted 1 March 2017

profiles and liver fat accumulation in an unselected sample from the community is not well known. Therefore, we investigated the association of serum metabolites with liver signal intensity, a continuous measure of liver fat content determined by magnetic resonance imaging (MRI) in a large sample from the general population. For comparison with previous studies, we further related serum metabolites to FLD investigated as a binary trait.

MATERIALS AND METHODS

Study design and sample

Study participants were drawn from the PopGen control cohort in Northern Germany ($n = 1316$), initially recruited between 2005 and 2007 into the PopGen biobank located in Kiel, Germany. The sample consisted of a representative population sample ($n = 747$) enriched with blood donors ($n = 569$) and served primarily as a reference sample for genetic-epidemiological analyses.[13] At the second examination cycle, conducted from 2010 to 2012 with a response proportion of 71% ($n = 930$), more comprehensive clinical and molecular phenotyping of the study participants was performed, as reported in detail elsewhere.[14] In brief, study participants were invited to the study center for a physical examination, collection of biosamples (blood, stool, urine) and assessment of established cardiometabolic risk factors. In a subsample of 656 participants, MRI measures to determine liver fat content were available. Exclusion of study participants with MRI images not interpretable for liver fat content because of non-compliance to the MRI breathing protocol ($n = 43$), missing information on covariates ($n = 31$), self-reported liver diseases ($n = 26$[14]) or missing blood sample ($n = 1$) resulted in a sample size of 555 individuals. The study was approved by the ethical review board of the Medical Faculty of the University Kiel (Kiel, Germany). Written informed consent was obtained from all study participants.

Exposure and covariate assessment

Non-targeted metabolomics profiling analysis (see Supplementary Methods for details) was performed in the Helmholtz Zentrum München, Germany. Aliquots of a human serum sample (Seralab, London, UK) were measured as a reference in addition to the participants' samples. On the basis of 116 aliquots of the reference, the mean coefficient of variation (CV) for the measured metabolites was 15.8% (Supplementary Table 1). Routine clinical biomarkers were assayed in fresh, unfrozen blood samples in a standard clinical chemistry laboratory at the University hospital Schleswig-Holstein, Campus Kiel, Germany.[14] C-reactive protein (CRP) concentrations were measured by immunoturbidimetry (Hitachi Modular, Roche, Rotkreuz, Switzerland). The detection limit for CRP concentrations was 0.9 mg/l.

Study participants completed detailed self-administered questionnaires, including questions on medical history and medication intake, educational attainment, smoking habits and recreational physical activity. The reported time being physically active was multiplied by the respective metabolic equivalent task (MET) intensity level and summed up.[15] Type 2 diabetes was defined as either self-reported type 2 diabetes diagnosed by a physician, reported intake of anti-diabetic medication, glycated hemoglobin $\geqslant 6.5\%$ or by fasting blood glucose level $\geqslant 126$ mg/dl.[16] Weight and height were measured with participants dressed in light clothing without shoes, subtracting 2 kg from weight measurements to account for clothing. Body mass index (BMI) was calculated as weight in kilograms divided by height in meters squared. Waist circumference was measured at the midpoint between the lower ribs and the iliac crest on the anterior axillary line in resting expiratory position. Alcohol consumption over the past year was assessed with a self-administered food frequency questionnaire available as a web-based version, and optionally as a paper version.[17] This questionnaire was provided by the Department of Epidemiology of the German Institute of Human Nutrition Potsdam-Rehbrücke.[18] In previous analyses, we have assessed the association of dietary intake and liver fat accumulation.[14]

Outcome assessment

Liver fat content was quantified by T1-weighted gradient echo MRI according to previously published methods[14] as the relative liver signal intensity difference of the liver on out-of-phase images compared to in-phase images in arbitrary units.[14] A FLD indicator variable (log liver signal intensity $\geqslant 3.0$) was defined according to a predefined cutoff[1] corresponding to the maximum Youden Index derived using spectroscopic determined FLD (liver fat $\geqslant 5.56\%$[19]) as the reference method.[14] Liver signal intensity was natural logarithmically transformed because of its skewed distribution.

Statistical analysis

For each metabolite, the levels measured semi-quantitatively as raw ion counts were divided by the median value of the samples' run day to account for instrumental day-to-day variations. Of 601 metabolites detected, 69 metabolites with less than 10 distinct non-missing values were excluded from the present analyses. For the remaining 532 metabolites, missing values were imputed with the minimum detected value.[20] Per metabolite, the median number of imputed values was 8.

Trend tests for continuous characteristics across liver signal intensity quartiles were tested for statistical significance using the Wald test in a linear regression model. For this purpose, we used the median liver signal intensity value within quartiles as a continuous variable. Trends of categorical variables across quartiles of liver signal intensity were assessed for statistical significance using a Cochran-Armitage test.

Partial least squares regression. The partial least square (PLS) regression procedure in SAS[21] was applied to examine the joint association of serum metabolites with liver signal intensity. All metabolites with a variable importance projection VIP value >1 ($n = 156$ metabolites) were used in a second PLS model to derive a metabolomics score explaining variation in serum metabolites relevant for variation in liver signal intensity.

Linear and logistic regression. To assess the association between metabolites and liver signal intensity or FLD, we performed 2 sets of analyses. First, the association of a metabolomic score with liver fat intensity and FLD was analyzed using multiple robust linear regression and logistic regression models, respectively. Second, the top 15 metabolites (with the highest VIP values) were associated individually with liver signal intensity and FLD using robust linear and logistic regression models, as appropriate. Model 1 was adjusted for sex and age, model 2 was further adjusted for the following *a priori*-selected potential confounding factors: years of education, smoking status and physical activity. We investigated potential effect modifications by age, sex, type diabetes status, BMI, waist circumference (<88 cm versus $\geqslant 88$ cm in women; <102 cm versus $\geqslant 102$ cm in men) and alcohol consumption (<20 g per day versus $\geqslant 20$ g per day in women; <30 g per day versus $\geqslant 30$ g per day in men) by including respective multiplicative interaction terms in the regression models and by stratified analysis.[22] In sensitivity analyses, we additionally adjusted for BMI, CRP, glycated hemoglobin, contraceptive use, lipid lowering and antihypertensive medication intake. For this analysis, for participants with a CRP value below the detection limit (0.9 mg/l), we imputed a value of 0.45 mg/l.

Overrepresentation and pathway analysis. Additionally, we performed a metabolite set enrichment analysis,[23] and a metabolomic pathway analysis[24] using MetaboAnalyst software version 3.0.[25] Metabolites with VIP values >1 in the PLS regression model[26] of known identity were considered for overrepresentation and pathway analyses. Since glucose, fructose, mannose, galactose, allose and altrose all have the same mass and retention time, these metabolites cannot be detected independently by the Metabolon platform used.[20] Therefore, we obtained a cumulative value for these sugars. For the overrepresentation and pathway analyses, we included glucose as representative of this group in our analyses. The overrepresentation analysis is based on a hypergeometric test, evaluating whether the metabolites of an entered metabolite list are overrepresented in predefined pathway-associated metabolite sets.[27] One-sided, Bonferroni-corrected *P*-values are reported, quantifying the probability to reveal at least a specific number of metabolites in the entered metabolite list belonging to one metabolite set.[27] In pathway topology analysis, the importance of a metabolite as a node within a specific network was quantified based on a pathway impact score, ranging between 0 and 1.[24]

RESULTS

Participants' characteristics across quartiles of liver signal intensity are shown in Table 1. Age, anthropometrical measures, alanine aminotransferase, triglycerides and the metabolomic score were all higher with higher liver signal intensity (reflecting a higher degree of fat accumulation in the liver; *P* for trend <0.001). In

Table 1. Characteristics of the study participants according to quartiles of the liver signal intensity (sample size $n = 555$)

Participant characteristics	*Quartile of liver signal intensity*				P *for trend*[a]
	1 (n = *138)*	*2* (n = *139)*	*3* (n = *139)*	*4* (n = *139)*	
Liver signal intensity (median (Q1, Q3))	2.54 (2.32, 2.63)	2.81 (2.76, 2.87)	3.02 (2.96, 3.08)	3.41 (3.28, 3.68)	—
Fatty liver disease (*n* (%))[b]	0 (0.0)	0 (0.0)	80 (57.6)	139 (100.0)	< 0.001
Age, years (median (Q1, Q3))	59.2 (49.3, 67.6)	60.2 (49.8, 69.1)	64.7 (56.0, 71.8)	64.8 (58.5, 71.5)	< 0.001
Men (*n* (%))	83 (60.1)	74 (53.2)	77 (55.4)	85 (61.2)	0.780
Education (n *(%))*					
⩽ 9 years	36 (26.1)	39 (28.1)	45 (32.4)	50 (36.0)	0.052
10 years	44 (31.9)	49 (35.3)	46 (33.1)	47 (33.8)	0.841
⩾ 11 years	58 (42.0)	51 (36.7)	48 (34.5)	42 (30.2)	0.039
Smoking status (n *(%))*					
Never	73 (52.9)	67 (48.2)	63 (45.3)	59 (42.5)	0.071
Former	55 (39.9)	56 (40.3)	60 (43.2)	68 (48.9)	0.110
Current	10 (7.3)	16 (11.5)	16 (11.5)	12 (8.6)	0.714
Smoking duration, years (median (Q1, Q3))	0.0 (0.0, 15.0)	2.0 (0.0, 18.0)	5.0 (0.0, 20.0)	10.0 (0.0, 25.0)	0.015
Alcohol intake, g per day	8.5 (3.3, 15.8)	7.9 (3.5, 16.6)	7.9 (2.6, 15.5)	12.4 (3.0, 24.9)	< 0.001
Physical activity, MET h per week (median (Q1, Q3))	89.3 (56.0, 136.5)	94.8 (62.0, 135.7)	90.1 (56.6, 130.5)	94.9 (63.9, 136.7)	0.953
BMI, kg/m^2 (median (Q1, Q3))	25.3 (23.1, 27.6)	25.9 (23.3, 28.4)	26.3 (24.2, 28.8)	29.2 (26.8, 31.9)	< 0.001
Waist circumference (Median (Q1, Q3))	90.3 (83.6, 99.7)	93.5 (83.0, 102.0)	95.7 (87.8, 102.0)	103.2 (95.1, 110.5)	< 0.001
Alanine aminotransferase, U/l (median (Q1, Q3))	21 (16, 26)	20 (16, 29)	20 (17, 27)	29 (21, 37)	< 0.001
Triglycerides, mmol/l (median (Q1, Q3))	0.94 (0.71, 1.43)	1.14 (0.88, 1.43)	1.21 (0.90, 1.49)	1.43 (1.06, 2.12)	< 0.001
Type 2 diabetes (*n* (%))[c]	6 (4.4)	7 (5.0)	14 (10.1)	32 (23.0)	< 0.001
Glycated hemoglobin, %	5.6 (5.4, 5.9)	5.5 (5.3, 5.8)	5.6 (5.4, 5.9)	5.8 (5.5, 6.0)	< 0.001
C-reactive protein, mg/l[d]	1.9 (1.2, 3.0)	2.1 (1.4, 3.4)	1.9 (1.3, 3.1)	2.0 (1.5, 3.9)	0.380
Metabolomic score (median (Q1, Q3))	− 2.17 (−4.17, − 0.07)	− 1.66 (−3.59, 0.32)	− 0.03 (−2.57, 2.45)	3.45 (1.10, 6.51)	< 0.001

Abbreviations: Q1, quartile 1; Q3, quartile 3; BMI, body mass index; MET, metabolic equivalent task. [a]*P* for trend values across quartiles of liver signal intensity were based on the Cochran-Armitage test for categorical variables and linear regression analysis for continuous variables with the median liver signal intensity variable within quartiles. [b]Liver signal intensity ⩾ 3.0. [c]Defined using baseline and follow-up information as either self-reported type 2 diabetes diagnosed by a physician, anti-diabetic medication, glycated hemoglobin ⩾6.5% or by fasting blood glucose level ⩾ 126 mg/dl. [d]$n = 328$ participants with CRP values above the detection limit (0.9 mg/l).

higher liver signal intensity quartiles, individuals were more likely to have a lower educational attainment (*P* for trend = 0.039), and to have type 2 diabetes (*P* for trend < 0.001). A total of 16 women reported intake of contraceptives, 76 participants reported intake of lipid lowering medication and 133 participants reported intake of antihypertensive medication.

Association of metabolite score with liver signal intensity and FLD

Out of the 532 metabolites considered in PLS regression, 156 metabolites had a VIP value > 1. A scatter diagram illustrates the relation between liver signal intensity and the metabolomic score (Figure 1). In unadjusted robust linear regression analysis, the metabolomic score based on the 156 metabolites explained 23.9% variation in liver signal intensity. The score included 6 different gamma-glutamyl-amino acids with a VIP value > 1, which were all positively related to liver signal intensity in PLS regression analysis: gamma-glutamylisoleucine, gamma-glutamylleucine, gamma-glutamylmethionine, gamma-glutamylphenylalanine, gamma-glutamyltyrosine and gamma-glutamylvaline.

In the multivariable-adjusted models, a 1-unit increment in the metabolomic score was positively associated with liver signal intensity (regression coefficient (ß): 0.05; 95% CI: 0.05, 0.06) and FLD (OR: 1.36; 95% CI: 1.27, 1.45; Table 2). There was no evidence for effect modification by age, sex, type 2 diabetes and alcohol intake (all *P* for interaction > 0.05). However, the association of the metabolomic score with liver signal intensity, and FLD, respectively, was significantly modified by BMI (*P* for interaction < 0.001). The association between the metabolomic score and liver fat content/FLD was much stronger in individuals with overweight and obesity as compared to individuals with normal weight (see Table 2). Similarly, the association was significantly modified by waist circumference (*P* for interaction < 0.001). The metabolomics score was more strongly associated with liver signal intensity and FLD in women and men with a waist circumference > 88 cm and > 102 cm, respectively, as compared to participants with lower waist circumference.

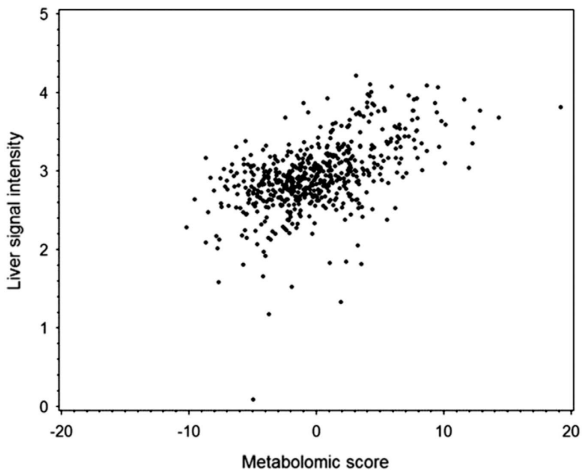

Figure 1. Scatter diagram of liver signal intensity and the metabolomic score.

In sensitivity analyses, we observed that additional adjustment for BMI, CRP, glycated hemoglobin and medication intake did not substantially alter the results (the effect estimates for the association per 1-unit increase of the metabolomic score with liver signal intensity, or FLD were ß: 0.05; 95% CI: 0.04, 0.05 and OR: 1.33; 95% CI: 1.24, 1.43; as compared to ß: 0.05; 95% CI: 0.05, 0.06 and OR: 1.36; 95% CI: 1.27, 1.45 without the additional adjustment).

Table 2. Difference in liver signal intensity and ORs (95% CIs) for fatty liver disease per 1-unit increase in the metabolomic score stratified by age, sex, BMI, alcohol consumption and diabetes status (sample size $n = 555$)

	Total n	*Regression coefficient (95% CI)*		*No. of cases*	*OR (95% CI) for fatty liver disease*	
		Model 1[a]	*Model 2*		*Model 1*	*Model 2*
Overall	555	0.05 (0.05, 0.06)	0.05 (0.05, 0.06)	219	1.35 (1.27, 1.43)	1.36 (1.27, 1.45)
Age						
< 62.3 years	277	0.06 (0.05, 0.07)	0.06 (0.05, 0.07)	134	1.38 (1.26, 1.51)	1.40 (1.27, 1.54)
⩾62.3 years	278	0.05 (0.04, 0.06)	0.05 (0.04, 0.06)	85	1.33 (1.22, 1.45)	1.35 (1.23, 1.47)
P for interaction		0.719	0.728		0.612	0.553
Sex						
M	319	0.06 (0.05, 0.06)	0.06 (0.05, 0.07)	81	1.37 (1.26, 1.49)	1.39 (1.28, 1.51)
F	236	0.05 (0.04, 0,06)	0.05 (0.04, 0.06)	138	1.32 (1.20, 1.46)	1.33 (1.20, 1.47)
P for interaction		0.644	0.660		0.566	0.610
BMI[b]						
Normal weight (18.5–24.9 kg/m^2)	172	0.02 (0.01, 0.03)	0.02 (0.01, 0.03)	41	1.11 (1.00, 1.24)	1.12 (1.00, 1.25)
Overweight (25.0–29.9 kg/m^2)	267	0.06 (0.04, 0.07)	0.06 (0.04, 0.07)	108	1.41 (1.27, 1.56)	1.41 (1.27, 1.57)
Obese (⩾30.0 kg/m^2)	112	0.07 (0.05, 0.08)	0.07 (0.05, 0.09)	68	1.68 (1.37, 2.07)	2.01 (1.50, 2.70)
P for interaction		< 0.001	< 0.001		< 0.001	< 0.001
Waist circumference						
< 88 cm in women, < 102 cm in men	293	0.03 (0.02, 0.04)	0.03 (0.02, 0.04)	80	1.18 (1.09, 1.28)	1.18 (1.09, 1.29)
⩾88 cm in women, ⩽ 102 cm in men	262	0.06 (0.05, 0.07)	0.06 (0.05, 0.08)	139	1.51 (1.35, 1.68)	1.56 (1.39, 1.75)
P for interaction		< 0.001	< 0.001		0.001	0.001
Type 2 diabetes						
No	496	0.05 (0.04, 0.06)	0.05 (0.04, 0.06)	180	1.32 (1.24, 1.41)	1.33 (1.24, 1.42)
Yes	59	0.06 (0.04, 0.08)	0.06 (0.04, 0.09)	39	1.67 (1.24, 2.26)	2.01 (1.35, 3.00)
P for interaction		0.369	0.331		0.123	0.127
Alcohol consumption, g/d[3]						
< 20 g in women, < 30 g in men	476	0.06 (0.05, 0.06)	0.06 (0.05, 0.06)	184	1.36 (1.27, 1.45)	1.36 (1.27, 1.46)
⩾ 20 g in women, ⩾ 30 g in men	79	0.05 (0.03, 0.06)	0.05 (0.03, 0.06)	35	1.33 (1.13, 1.56)	1.34 (1.12, 1.59)
P for interaction		0.331	0.310		0.704	0.638

Abbreviations: BMI, body mass index; F, female, M, male. [a]Model 1 was adjusted for age (years) and sex (male, female). Model 2 was adjusted for age (years), sex (male, female) except sex strata, education (⩽9 years, 10 years, ⩾11 years), smoking status (never, former, current), smoking duration (years) and physical activity (metabolic equivalent task h per week). [b]$n = 4$ individuals with underweight (BMI < 18.5 kg/m^2) were excluded from the analysis.

Individual associations of most significant metabolites with liver signal intensity and FLD

The relative concentrations of the top 15 metabolites according to quartiles of liver signal intensity are displayed in Supplementary Table 2. Specifically, the relative concentrations of cysteine-glutathione disulfide, and of two metabolites of unknown identity were lower with higher liver signal intensity quartile, whereas relative concentrations of glutamate, isoleucine, valine, gamma-glutamylvaline, leucine, tyrosine, gamma-glutymylisoleucine, 3-methyl-2-oxovalerate, propionylcarnitine, gamma-glutamylleucine, urate and cyclo(leucylprolyl) were higher with higher liver signal intensity (all *P* for trend < 0.001). Additional information on these top 15 metabolites, including VIP values, PLS regression coefficients and relevant pathways, is provided in Table 3, Supplementary Table 3). The highest VIP values were obtained for glutamate (VIP = 3.35) and isoleucine (VIP = 3.11).

Overrepresentation and pathway analysis

Of the 156 metabolites with a VIP value greater than one, 111 metabolites were of known identity in the Metabolon Reference Library Database.[28] Of these, 93 metabolites registered in the Human Metabolome Database[28] and in the MetaboAnalyst metabolite set library for metabolomic pathway-associated metabolite sets were entered into overrepresentation analysis using MetaboAnalyst.[25] The only enriched metabolite set was the protein biosynthesis metabolite set, which was six-fold enriched (corrected $P = 2.64E^{-07}$; Supplementary Figure 1). Whereas two hits—metabolites of the entered list matching the metabolites of the set—would have been expected by chance, 13 metabolites matched the protein biosynthesis metabolite set.

Pathways enriched with the 93 metabolites entered were the aminoacyl-tRNA biosynthesis pathway (corrected $P = 7.50E^{-07}$), the nitrogen metabolism pathway (corrected $P = 0.02$), and the valine, leucine and isoleucine biosynthesis pathway (corrected $P = 0.03$) (Supplementary Table 4, Supplementary Figure 2). Of these pathways, the highest pathway impact values were observed for the aminoacyl-tRNA biosynthesis pathway (impact score = 0.17) and the valine, leucine and isoleucine biosynthesis pathway (impact score = 0.15).

DISCUSSION

Principal findings

In this study, we report associations of fasting serum metabolites with continuous liver fat content, quantified by MRI, in a sample from the general population. Our main observations are as follows. First, a metabolomic score was positively associated with liver signal intensity and FLD, even after adjusting for relevant confounders. The observed associations were stronger in obese individuals compared to individuals with overweight or normal BMI and in individuals with higher waist circumference compared

Table 3. Most important metabolites[a] in the PLS regression analysis for liver signal intensity (sample size $n = 555$)

Metabolites	*Subpathway or brief description*	*Regression coefficient for standardized metabolites*	*VIP statistic*
Inverse association with liver signal intensity			
Cysteine-glutathione disulfide	Glutathione metabolism	−0.019	3.05
X—16137	Unknown	−0.018	2.87
X—13543	Unknown	−0.016	2.60
Positive association with liver signal intensity			
Glutamate	Glutamate metabolism	0.021	3.35
Isoleucine	Leucine, isoleucine and valine metabolism	0.019	3.11
Valine	Leucine, isoleucine and valine metabolism	0.019	3.05
Gamma-glutamylvaline	Gamma-glutamyl amino acid	0.018	2.93
Leucine	Leucine, isoleucine and valine metabolism	0.018	2.82
Tyrosine	Phenylalanine and tyrosine metabolism	0.017	2.76
Gamma-glutamylisoleucine	Gamma-glutamyl amino acid	0.016	2.59
3-Methyl-2-oxovalerate	Leucine, isoleucine and valine metabolism	0.016	2.57
Propionylcarnitine	Fatty acid metabolism and branched-chain amino acid metabolism	0.015	2.48
Gamma-glutamylleucine	Gamma-glutamyl amino acid	0.015	2.48
Urate	Purine metabolism, (hypo)xanthine/Inosine containing	0.015	2.47
Cyclo(leucylprolyl)	Dipeptide	0.015	2.41

PLS, partial least squares; VIP, variable importance in the projection; X, number, metabolites of unknown identity. [a]The variable importance in the projection statistic [23] was used to assess the importance of a metabolite in the PLS model. Metabolites with the 15 highest VIP statistics in the PLS model were included in the table.

to individuals with lower waist circumference. Second, the metabolomic score was associated with liver signal intensity also in individuals free of FLD. Third, the 15 metabolites most strongly associated with liver signal intensity comprised cysteine-glutathione disulfide (inversely related to liver signal intensity), and glutamate, isoleucine, valine, gamma-glutamylvaline, leucine, tyrosine, gamma-glutamylisoleucine, 3-methyl-2-oxovalerate, gamma-glutamylleucine, urate and cyclo(leucylprolyl) (all positively related to liver signal intensity), as well as three metabolites of unknown identity. Fourth, strongly associated metabolites were overrepresented in metabolite sets relevant for protein biosynthesis with particular impact on the aminoacyl-tRNA biosynthesis pathway and the valine, leucine and isoleucine biosynthesis pathway.

In the context of current literature

Glutathione metabolism, branched-chain amino acids and liver fat. FLD has been linked to a broad spectrum of metabolic complications, including dyslipidemia and dysglycemia[6] but comprehensive blood metabolomic profiles associated with FLD are largely unknown.[8] Prior studies investigating metabolimic profiles of FLD have been limited by relatively small sets of clinical patients[9,11,29] undergoing liver biopsies or bariatric surgery.[11] and by the assessment of FLD as a dichotomous trait.[9,12] In a relatively small clinical sample ($n = 60$), different metabolites have been investigated in relation to steatohepatitis and FLD.[9] Specifically, the concentration of plasma cysteine-glutathione disulfide was lower and the concentrations of different glutamyl dipeptides and amino acids (including glutamate, tyrosine and isoleucine) were higher in individuals with histological-confirmed hepatic steatosis ($n = 11$) compared to controls (no evidence for FLD on ultrasound; $n = 25$).[9] The authors discuss that lower cysteine-glutathione disulfide levels and higher gamma-glutamyl peptide level—both of which also linked to FLD in the present study—might be attributable to an increased glutathione turnover in response to oxidative stress.[9,12] In concordance with our study, the concentration of branched-chain amino acids were positively associated with the 10-year risk of FLD detected by ultrasound in the Young Finns Study.[30]

We expand these prior reports by demonstrating that these biomarkers are indeed associated with FLD, but also with liver fat on a continuous scale (determined by MRI) in a large sample from the general population ($n = 555$). The association between glutamate and liver fat content is further supported by observational and the experimental data from the Framingham Heart Study and the Malmö Diet and Cancer Study that linked glutamate to metabolic traits (for example, obesity, insulin resistance, high blood pressure and dyslipidemia).[31] However, the association of a non-targeted metabolomics panel with liver fat has not been reported previously in a large community-dwelling sample.

In a prior analysis from our group, limited to a targeted set of metabolites ($n = 138$) and a subsample ($n = 230$), we reported that branched-chain amino acids (including isoleucine, leucine, valine) discriminate well between individuals with and without ultrasonographic evidence of FLD.[32] The present analysis, by contrast, was conducted in a larger sample ($n = 555$), used liver signal intensity on a continuous scale and a much broader set of metabolites ($n = 532$), applying a non-targeted metabolomic approach.

Lipids and liver fat. Lipid accumulation in the liver is one of the hallmarks of FLD.[33] Using a targeted mass-spectrometry-based lipidomic approach, Puri *et al*.[34] identified circulating plasma lipids, associated with nonalcoholic FLD, including saturated fatty acids, monounsaturated fatty acids and specific polyunsaturated fatty acids. In our analyses, however, only one lipid biomarker, namely propionylcarnitine, was among the 15 most important (with the highest VIP values) metabolites associated with liver signal intensity. Propionylcarnitine, however, was not assessed in the former targeted lipidomic analysis.[34]

Effect modification by BMI and waist circumference. We obtained evidence that the association between the metabolomic score and liver signal intensity was modified by BMI and waist circumference. Specifically, the association between the

metabolomic score and liver fat was stronger in obese as compared to normal-weight individuals and in participants with a higher waist circumference compared to participants with a lower waist circumference. Consistent with this observation, a BMI-dependent association between serum metabolites and liver steatosis has been reported in 467 individuals biopsied for laparoscopic cholecystectomy, bariatric surgery or suspected nonalcoholic FLD.[10]

Strengths and limitations

Major strengths of the present study are the assessment of serum metabolites over a broad range of metabolic pathways by applying a non-targeted metabolomic approach in a relatively large population-based setting, and the assessment of liver fat content on a continuous scale using MRI. To minimize measurement error a standardized sample handling protocol was applied in the pre-analytical phase, and samples were stored at − 80 °C until metabolome analyses were conducted. Metabolomic scores can be derived on one hand based on prior physiological knowledge and thus include only metabolites that have been linked to liver fat accumulation in prior publications. However, also metabolites beyond aspects already known may be of relevance in the pathophysiology of FLD. Empirical approaches enable the assessment of metabolomics profiles in a comprehensive and unbiased manner. Furthermore, as opposed to traditional analyses that focus on single metabolites, this approach takes interactions of metabolites into account. Limitations of the study are as follows. The cross-sectional design allows no conclusion with regard to time sequence and causality of the reported associations. The chemical identity of some of the metabolites analyzed is unknown and has to be identified in future studies. As we relied on one single measurement of metabolites, intra-individual variation in serum metabolites concentrations over time could not be assessed within the present study. However, reliability of concentrations of serum metabolite over periods of months[35] and years[36] has been demonstrated . Though the association of carbohydrates with liver fat accumulation might vary by carbohydrate type, no distinction of types of carbohydrates was possible based on this non-targeted metabolomics approach that we applied. Although we believe that insulin resistance might play an important role, no direct measure of insulin resistance such as homeostasis model assessment was available in our sample. Given that a considerable proportion of liver fat variance is not explained by the circulating metabolites investigated, further research on the correlates and determinants of liver fat is warranted.

CONCLUSIONS

In the present study, we identified fasting metabolomic profiles associated with FLD and continuous liver fat content, quantified by MRI, in a general population sample. A metabolomic score, consisting mainly of branched-chain amino acids and derived gamma-glutamyl dipeptides, was related to liver signal intensity and FLD. We observed effect modification by BMI and waist circumference in the sense that the association between these metabolites and liver fat content was stronger in obese and overweight individuals compared to normal-weight individuals and in individuals with a higher waist circumference compared to participants with a lower waist circumference. Furthermore, these associations of these metabolites with liver fat (on a continuous scale) persisted when the analyses were restricted to individuals free of FLD. Overrepresentation and pathway analyses highlighted the importance of branched-chain amino acids and derived gamma-glutamyl dipeptides in FLD. These metabolites are considered to be involved in insulin resistance and oxidative stress. Further studies are warranted to validate our observations in other samples and to specifically assess the correlation of these metabolomic markers with established markers of insulin resistance and oxidative stress. It should be assessed in longitudinal settings whether the identified metabolites are predictive of new onset incident FLD or of change of liver fat content over time. Finally, functional studies are warranted to explore the molecular mechanisms linking the observed metabolites to FLD and liver fat content.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank all participants of the PopGen control cohort study for their invaluable contribution to the study. This work was supported by grants from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) Excellence Cluster 'Inflammation at Interfaces' (EXC306, EXC306/2), sysINFLAME (01ZX1306A) and through grants from the German Federal Ministry of Education and Research (01GR0468). The PopGen 2.0 network is supported by a grant from the German Ministry for Education and Research (01EY1103). Manja Koch is recipient of a postdoctoral research fellowship from the DFG (KO 5187/1-1).

REFERENCES

1 Szczepaniak LS, Nurenberg P, Leonard D, Browning JD, Reingold JS, Grundy S *et al.* Magnetic resonance spectroscopy to measure hepatic triglyceride content: prevalence of hepatic steatosis in the general population. *Am J Physiol Endocrinol Metab* 2005; **288**: e462–e468.
2 Markus MR, Baumeister SE, Stritzke J, Dorr M, Wallaschofski H, Volzke H *et al.* Hepatic steatosis is associated with aortic valve sclerosis in the general population: the Study of Health in Pomerania (SHIP). *Arterioscler Thromb Vasc Biol* 2013; **33**: 1690–1695.
3 Portillo Sanchez P, Bril F, Maximos M, Lomonaco R, Biernacki D, Orsak B *et al.* High prevalence of nonalcoholic fatty liver disease in patients with type 2 diabetes mellitus and normal plasma aminotransferase levels. *J Clin Endocrinol Metab* 2015; **100**: 2231–2238.
4 Anstee QM, Targher G, Day CP. Progression of NAFLD to diabetes mellitus, cardiovascular disease or cirrhosis. *Nat Rev Gastroenterol Hepatol* 2013; **10**: 330–344.
5 Dunn W, Xu R, Wingard DL, Rogers C, Angulo P, Younossi ZM *et al.* Suspected nonalcoholic fatty liver disease and mortality risk in a population-based cohort study. *Am J Gastroenterol* 2008; **103**: 2263–2271.
6 Speliotes EK, Massaro JM, Hoffmann U, Vasan RS, Meigs JB, Sahani DV *et al.* Fatty liver is associated with dyslipidemia and dysglycemia independent of visceral fat: the Framingham Heart Study. *Hepatology* 2010; **51**: 1979–1987.
7 Cohen JC, Horton JD, Hobbs HH. Human fatty liver disease: old questions and new insights. *Science* 2011; **332**: 1519–1523.
8 Dumas ME, Kinross J, Nicholson JK. Metabolic phenotyping and systems biology approaches to understanding metabolic syndrome and fatty liver disease. *Gastroenterology* 2014; **146**: 46–62.
9 Kalhan SC, Guo L, Edmison J, Dasarathy S, McCullough AJ, Hanson RW *et al.* Plasma metabolomic profile in nonalcoholic fatty liver disease. *Metabolism* 2011; **60**: 404–413.
10 Barr J, Caballeria J, Martinez-Arranz I, Dominguez-Diez A, Alonso C, Muntane J *et al.* Obesity-dependent metabolic signatures associated with nonalcoholic fatty liver disease progression. *J Proteome Res* 2012; **11**: 2521–2532.
11 Rodriguez-Gallego E, Guirro M, Riera-Borrull M, Hernandez-Aguilera A, Marine-Casado R, Fernandez-Arroyo S *et al.* Mapping of the circulating metabolome reveals alpha-ketoglutarate as a predictor of morbid obesity-associated non-alcoholic fatty liver disease. *Int J Obes* 2014; **39**: 279–287.
12 Soga T, Sugimoto M, Honma M, Mori M, Igarashi K, Kashikura K *et al.* Serum metabolomics reveals gamma-glutamyl dipeptides as biomarkers for discrimination among different forms of liver disease. *J Hepatol* 2011; **55**: 896–905.
13 Nöthlings U, Krawczak M. [PopGen: A population-based biobank with prospective follow-up of a control group]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 2012; **55**: 831–835.
14 Koch M, Borggrefe J, Barbaresko J, Groth G, Jacobs G, Siegert S *et al.* Dietary patterns associated with magnetic resonance imaging-determined liver fat content in a general population study. *Am J Clin Nutr* 2014; **99**: 369–377.
15 Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett DR Jr, Tudor-Locke C *et al.* 2011. Compendium of physical activities: a second update of codes and MET values. *Med Sci Sports Exerc 2011* **43**: 1575–1581.

16 International Expert Committee. International Expert Committee report on the role of the A1Cassay in the diagnosis of diabetes. *Diabetes Care* 2009; **32**: 1327–1334.
17 Nöthlings U, Hoffmann K, Bergmann MM, Boeing H. Fitting portion sizes in a self-administered food frequency questionnaire. *J Nutr* 2007; **137**: 2781–2786.
18 Dehne LI, Klemm C, Henseler G, Hermann-Kunz E. The german food code and nutrient data base (BLS II.2). *Eur J Epidemiol* 1999; **15**: 355–359.
19 Stefan N, Kantartzis K, Haring HU. Causes and metabolic consequences of fatty liver. *Endocr Rev* 2008; **29**: 939–960.
20 Evans AM, DeHaven CD, Barrett T, Mitchell M, Milgram E. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Anal Chem* 2009; **81**: 6656–6667.
21 Hoffmann K, Schulze MB, Schienkiewitz A, Nothlings U, Boeing H. Application of a new statistical method to derive dietary patterns in nutritional epidemiology. *Am J Epidemiol* 2004; **159**: 935–944.
22 Kotronen A, Yki-Jarvinen H. Fatty liver: a novel component of the metabolic syndrome. *Arterioscler Thromb Vasc Biol* 2008; **28**: 27–38.
23 Xia J, Wishart DS. MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. *Nucleic Acids Res* 2010; **38**: W71–W77.
24 Xia J, Wishart DS. MetPA: a web-based metabolomics tool for pathway analysis and visualization. *Bioinformatics* 2010; **26**: 2342–2344.
25 Xia J, Mandal R, Sinelnikov IV, Broadhurst D, Wishart DS. MetaboAnalyst 2.0--a comprehensive server for metabolomic data analysis. *Nucleic Acids Res* 2012; **40**: W127–W133.
26 Mehmood T, Martens H, Saebo S, Warringer J, Snipen L. A Partial Least Squares based algorithm for parsimonious variable selection. *Algorithms Mol Biol* 2011; **6**: 27.
27 Xia J, Wishart DS. Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. *Nat Protoc* 2011; **6**: 743–760.
28 Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N *et al.* HMDB: the human metabolome database. *Nucleic Acids Res* 2007; **35**: D521–D526.
29 Barr J, Vazquez-Chantada M, Alonso C, Perez-Cormenzana M, Mayo R, Galan A *et al.* Liquid chromatography-mass spectrometry-based parallel metabolic profiling of human and mouse model serum reveals putative biomarkers associated with the progression of nonalcoholic fatty liver disease. *J Proteome Res* 2010; **9**: 4501–4512.
30 Kaikkonen JE, Wurtz P, Suomela E, Lehtovirta M, Kangas AJ, Jula A *et al.* Metabolic profiling of fatty liver in young and middle-aged adults: Cross-sectional and prospective analyses of the Young Finns Study. *Hepatology* 2017; **65**: 491–500.
31 Cheng S, Rhee EP, Larson MG, Lewis GD, McCabe EL, Shen D *et al.* Metabolite profiling identifies pathways associated with metabolic risk in humans. *Circulation* 2012; **125**: 2222–2231.
32 Siegert S, Yu Z, Wang-Sattler R, Illig T, Adamski J, Hampe J *et al.* Diagnosing fatty liver disease: a comparative evaluation of metabolic markers, phenotypes, genotypes and established biomarkers. *PLoS ONE* 2013; **8**: e76813.
33 Rinella ME. Nonalcoholic fatty liver disease: a systematic review. *JAMA* 2015; **313**: 2263–2273.
34 Puri P, Wiest MM, Cheung O, Mirshahi F, Sargeant C, Min HK *et al.* The plasma lipidomic signature of nonalcoholic steatohepatitis. *Hepatology* 2009; **50**: 1827–1838.
35 Floegel A, Drogan D, Wang-Sattler R, Prehn C, Illig T, Adamski J *et al.* Reliability of serum metabolite concentrations over a 4-month period using a targeted metabolomic approach. *PLoS ONE* 2011; **6**: e21103.
36 Yousri NA, Kastenmuller G, Gieger C, Shin SY, Erte I, Menni C *et al.* Long term conservation of human metabolic phenotypes and link to heritability. *Metabolomics* 2014; **10**: 1005–1017.

Supplementary Information accompanies this paper on *European Journal of Clinical Nutrition* website (http://www.nature.com/ejcn)