



# Metabolites of milk intake: a metabolomic approach in UK twins with findings replicated in two European cohorts

Tess Pallister<sup>1</sup> · Toomas Haller<sup>2</sup> · Barbara Thorand<sup>3</sup> · Elisabeth Altmaier<sup>4,8</sup> · Aedin Cassidy<sup>5</sup> · Tiphaine Martin<sup>1</sup> · Amy Jennings<sup>5</sup> · Robert P. Mohny<sup>6</sup> · Christian Gieger<sup>4,8</sup> · Alexander MacGregor<sup>5</sup> · Gabi Kastenmüller<sup>7</sup> · Andres Metspalu<sup>2</sup> · Tim D. Spector<sup>1</sup> · Cristina Menni<sup>1</sup>

Received: 8 September 2015 / Accepted: 20 July 2016

© The Author(s) 2016. This article is published with open access at [Springerlink.com](http://Springerlink.com)

## Abstract

**Purpose** Milk provides a significant source of calcium, protein, vitamins and other minerals to Western populations throughout life. Due to its widespread use, the metabolic and health impact of milk consumption warrants further investigation and biomarkers would aid epidemiological studies.

**Methods** Milk intake assessed by a validated food frequency questionnaire was analyzed against fasting blood metabolomic profiles from two metabolomic platforms in females from the TwinsUK cohort ( $n = 3559$ ). The top metabolites were then replicated in two independent populations (EGCUT,  $n = 1109$  and KORA,  $n = 1593$ ), and the results from all cohorts were meta-analyzed.

**Results** Four metabolites were significantly associated with milk intake in the TwinsUK cohort after adjustment for multiple testing ( $P < 8.08 \times 10^{-5}$ ) and covariates (BMI, age, batch effects, family relatedness and dietary covariates) and replicated in the independent cohorts. Among the metabolites identified, the carnitine metabolite trimethyl-N-aminovalerate ( $\beta = 0.012$ , SE = 0.002,  $P = 2.98 \times 10^{-12}$ ) and the nucleotide uridine ( $\beta = 0.004$ , SE = 0.001,  $P = 9.86 \times 10^{-6}$ ) were the strongest novel predictive biomarkers from the non-targeted platform. Notably, the association between trimethyl-N-aminovalerate and milk intake was significant in a group of MZ twins discordant for milk intake ( $\beta = 0.050$ , SE = 0.015,  $P = 7.53 \times 10^{-4}$ ) and validated in the urine of 236 UK twins ( $\beta = 0.091$ , SE = 0.032,  $P = 0.004$ ). Two metabolites from the targeted platform, hydroxysphingomyelin C14:1 ( $\beta = 0.034$ , SE = 0.005,  $P = 9.75 \times 10^{-14}$ ) and diacylphosphatidylcholine C28:1 ( $\beta = 0.034$ , SE = 0.004,  $P = 4.53 \times 10^{-16}$ ), were also replicated.

**Electronic supplementary material** The online version of this article (doi:[10.1007/s00394-016-1278-x](https://doi.org/10.1007/s00394-016-1278-x)) contains supplementary material, which is available to authorized users.

✉ Tess Pallister  
[tess.pallister@kcl.ac.uk](mailto:tess.pallister@kcl.ac.uk)

✉ Cristina Menni  
[cristina.menni@kcl.ac.uk](mailto:cristina.menni@kcl.ac.uk)

<sup>1</sup> Department of Twin Research and Genetic Epidemiology, St Thomas Hospital, King's College London, London SE1 7EH, UK

<sup>2</sup> Estonian Genome Center, University of Tartu, Tartu, Estonia

<sup>3</sup> Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany

<sup>4</sup> Institute of Genetic Epidemiology, Helmholtz Zentrum München – German Research Center for Environmental Health, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany

<sup>5</sup> Department of Nutrition, Norwich Medical School, University of East Anglia, Norwich, UK

<sup>6</sup> Metabolon, Inc., Durham, NC, USA

<sup>7</sup> Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany

<sup>8</sup> Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, 85764 Neuherberg, Germany

**Conclusions** We identified and replicated in independent populations four novel biomarkers of milk intake: trimethyl-N-aminovalerate, uridine, hydroxysphingomyelin C14:1 and diacylphosphatidylcholine C28:1. Together, these metabolites have potential to objectively examine and refine milk-disease associations.

**Keywords** Nutrition · Metabolomics · Twins · Biomarkers · Milk

### Abbreviations

DZ	Dizygotic
EGCUT	Estonian Genome Center of the University of Tartu
KORA	Cooperative Health Research in the Region of Augsburg
PC aa C28:1	Diacylphosphatidylcholine C28:1
MZ	Monozygotic
SM(OH) C14:1	Hydroxysphingomyelin C14:1

### Introduction

Milk and dairy products have been consumed throughout life by lactase-producing cultures for thousands of years, aided by the lactase persistence/dairying co-evolution estimated to have occurred 7500 years ago [1]. They provide excellent sources of calcium and other micronutrients for many populations around the world [2]. Studies of the long-term health and metabolic implications of dairy product consumption based on recall have, however, yielded inconsistencies [3]. Epidemiological studies have mainly explored relationships with combined dairy product intake, though different products may promote different physiological effects, which may be due to varying macronutrient contents (reduced lactose and whey fraction), microbiological species (and their metabolic products) or altered nutrient absorption as a result of processing [3]. Of all dairy products, fluid milk is consumed the most frequently in many Western countries though consumption has been declining in some, such as the USA [4] and UK [5] likely due to health concerns about fat. Milk consumption is generally encouraged throughout life, for the growth of strong bones and teeth during childhood through to aging women for the prevention of osteoporosis and fractures [6]. The evidence for the long-term consumption of milk on health outcomes from observational studies has been somewhat inconsistent, with a meta-analysis of 17 prospective studies suggestive of a protective effect on CVD and no relationship with mortality [7], while two large Swedish cohorts identified associations with mortality [8]. Usage of different questionnaires to measure milk intake as well as recall bias complicates large population comparisons; thus,

incorporation of objective biomarkers into epidemiological methods is essential.

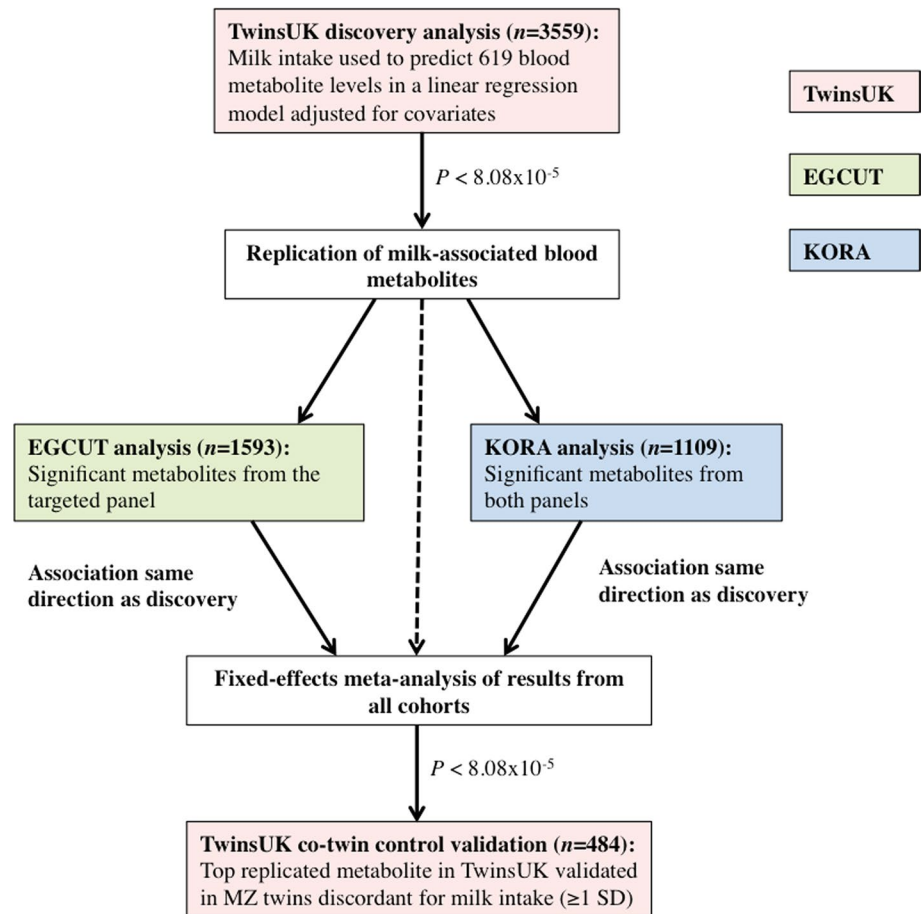
Metabolomics provides a snapshot of the metabolic status of an individual by high-throughput measurement of small metabolites and has the potential to refine diet-disease associations, primarily through food biomarker identification and providing insight into novel nutrient and disease pathways [9]. A recent example involved using metabolites of alcohol to predict indices of cardiovascular health [10]. A limiting factor in many metabolomic studies is the amount of inter-individual variation in metabolite levels, up to 81 % of which is attributable to genetics [11]. Monozygotic (MZ) twins, matched for baseline genetics, age, sex and early life experiences, and discordant for dietary phenotypes, provide a natural case-control experiment to evaluate diet-metabolite associations while controlling for these factors [12]. In this study, we present a novel multi-platform targeted and non-targeted metabolomics analysis of milk consumption to identify potential biomarkers of long-term milk intake, with our main findings replicated in two European populations.

### Subjects and material

A pipeline of the study design is presented in Fig. 1.

#### Discovery sample

Subjects included in the analysis were twins enrolled in the TwinsUK registry, a national register of UK adult female twins [13]. Healthy twins were recruited nation-wide primarily through media campaigns without selecting for particular diseases or traits. In this study, we included 3559 female twins, who completed at least one 131-item FFQ between 1995 and 2007, and had metabolomics and BMI data available within 5 years of FFQ completion. The 131-item Food Frequency Questionnaire (FFQ) was developed and validated against pre-established nutrient biomarkers for the European Prospective Investigation into Diet and Cancer (EPIC) Norfolk [14]. Quality control, subject exclusion criteria and methods for nutrient determination from FFQ data has been previously described [15]. Briefly, submitted FFQs were excluded if greater than 10 food items were left unanswered, or if the total energy intake estimate derived from FFQ as a ratio of the subject's estimated basal metabolic rate (determined by the Harris-Benedict equation [16]) was more than two standard deviations outside the mean of this ratio ( $<0.52$  or  $>2.58$ ). Milk intake was represented in servings per week with one serving equal to 1 pint (585 g). To determine intakes, participants were asked which type of milk they most often used and how much milk they consumed per day: none,  $\frac{1}{4}$  pint (146.25 g),  $\frac{1}{2}$

**Fig. 1** Pipeline of study design

pint (292.5 g),  $\frac{3}{4}$  pint (438.75), 1 pint (585 g) or more than one pint (this was coded as  $1\frac{1}{2}$  pints and equal to 877.5 g). Intake frequencies of all milk types from dairy sources (prepared dried, whole, semi-skimmed, and skimmed cow's milk) were summed, and the variable adjusted for total reported energy intake using the residual method [17]. For other dairy variables, intakes of average serving sizes were assessed on a 9-point scale (ranging from never or less than once per month to 6+ times per day) and frequencies summed into the following broader categories: yoghurt (full fat and low fat), cheese [cottage cheese and dairy cheese (for example, cheddar, brie, edam)], dairy desserts (for example, chocolate mousse, cream caramels), ice cream, creams [single (sour cream) and double (clotted cream)], and butter. Other phenotypic data relevant to the present study include BMI and zygosity (determined by methods outlined previously [13]). The study was approved by the St. Thomas' Hospital Research Ethics committee and all subjects provided informed written consent.

#### Replication cohorts

**EGCUT** The Estonian Biobank is the population-based biobank of the Estonian Genome Center of the University

of Tartu (EGCUT) [18]. All subjects are unrelated volunteers and were recruited randomly by general practitioners (GP) and physicians in hospitals. For this analysis, subjects with milk intake and fasted serum metabolomics measurements, as determined by the targeted mass spectrometry-based Biocrates platform (Biocrates AbsoluteIDQ p150 Kit; [www.biocrates.com](http://www.biocrates.com)), were used ( $n = 1109$ ). All subjects responded to the following question as part of the general questionnaire: "Last week on how many days did you consume milk products?" The response was recorded on a four point scale: (1) never, (2) once or twice, (3) 3–5 times, (4) 6–7 times. The questionnaire was administered on the same day (during the same visit) as the blood sample was drawn.

**KORA** The Cooperative Health Research in the Region of Augsburg (KORA) study includes unrelated individuals from the general population in the region of Augsburg, Germany [19]. As part of the fourth survey (S4), a total of 4261 persons aged 25–74 years were examined between 1999 and 2001. For the present analysis, 1593 individuals with serum levels of the significant metabolites associated with milk intake identified in TwinsUK by the non-targeted Metabolon and targeted Biocrates platforms and information on milk intake were used. Dietary intake was assessed

using a validated FFQ as part of a standardized interview at the same day blood samples for measurement of metabolites was taken [20]. The FFQ consisted of 24 food items and asked subjects to recall their “average intake” without specifying a time frame in the following six frequency categories: almost daily, several times per week, about once a week, several times per month, once a month or less, never. The food item “milk intake including buttermilk” was used in the present analysis. Prior to analysis, the milk intake variable was recoded for ease of interpretation so that the least frequency (never) was coded as 1 and the highest frequency (almost daily) was coded as 6, and the rest of the categories changed accordingly (once a month or less, 2; several times per month, 3; about once a week, 4; and several times per week, 5).

### Metabolomic profiling

Following six or more hours of fasting, blood samples were taken. The samples were frozen at  $-45^{\circ}\text{C}$  until metabolomics profiling, further details on sample protocols can be found in Supplementary text 1.

#### Metabolon

Non-targeted mass spectroscopy-based metabolomic profiling was conducted by the metabolomics provider Metabolon, Inc. (Durham, NC) on 3559 fasting blood samples and 236 spot urine samples taken at the same time as plasma from TwinsUK, as previously described [21, 22]. The Metabolon platform includes 280 structurally named biochemicals (known metabolites) categorized into the following broad categories: amino acids, carbohydrates, vitamins, lipids, nucleotides, peptides, and xenobiotics. An asterisk (\*) at the end of the metabolite name indicates the metabolite identity has not been confirmed by accurate mass data. Further details on the handling of instrumental variability can be found in Supplementary text 1.

#### Biocrates

A targeted metabolomic assay was also performed in a subset of 858 participants, with samples overlapping with Metabolon profiling, in the TwinsUK study using the Biocrates Absolute IDQ™-kit p150 (BIOCRATES Life Sciences, AG, Innsbruck, Austria) as previously described [23, 24]. Briefly, the flow injection analysis (FIA) tandem mass spectrometry (MS/MS) method is used to quantify 163 known small molecule metabolites simultaneously by multiple reaction monitoring. Quantification of the metabolites is then achieved by reference to appropriate internal standards.

The Biocrates dataset contains acylcarnitines ( $\text{C}_x\text{:y}$ ), hydroxylacylcarnitines [ $\text{C}(\text{OH})_x\text{:y}$ ] and dicarboxylacylcarnitines ( $\text{C}_x\text{:y}\text{-DC}$ ); amino acids; sphingomyelins ( $\text{SM}_x\text{:y}$ ) and sphingomyelin-derivatives [ $\text{SM}(\text{OH})_x\text{:y}$ ]; and glycerophospholipids (PC).

The Biocrates platform measures absolute metabolite value (mM).

### Quality control of the metabolomic dataset

**Metabolon** Quality control on the metabolomics dataset was performed as previously described [21, 22]. Briefly, raw data were median-normalized by dividing metabolite concentrations by the day median of that metabolite and then inverse-normalized. Thus, in the case where instrument instability was not caught, the impact would be limited to a given batch of samples within the study. Metabolites with more than 20 % of values missing were excluded to avoid false-positive associations. Minimum run day measures were imputed to the missing values.

**Biocrates** The metabolite serum concentrations were log transformed as these were right-skewed.

### Statistical analysis

Statistical analysis was carried out using Stata version 12.

For each metabolite, random intercept linear regression analysis was first undertaken in the TwinsUK sample adjusting for age, metabolite batch, BMI, family relatedness and potential dietary confounders:

$$Y_i = \beta_0 + \beta_i X_{ij} + \gamma_i \text{age}_{ij} + \delta_i \text{BMI}_{ij} + \zeta_j + \varepsilon_{ij}$$

where  $Y_i$  is the metabolite and  $X_{ij}$  the dietary phenotype of twin  $j$  from pair  $i$ ,  $\zeta_j$ , is the family-specific error component that captures the unobserved heterogeneity or family characteristics. Dietary confounders were identified by using a backward stepwise regression model including a set of dietary phenotypes (food group, alcohol and unsaturated fat [to account for oil consumption]) intakes that were significantly associated with milk intake using a cut-off threshold  $P < 0.001$ . We adjusted for multiple testing using Bonferroni correction thus giving a significant threshold of  $8.08 \times 10^{-5}$  ( $0.05/(619 \text{ detected metabolites})$ ).

We replicated each significant milk-metabolite association in the two independent cohorts, KORA and EGCUT, using linear regression adjusting for age, sex, BMI and fasted status (only in KORA). We then combined the results using inverse variance fixed-effects meta-analysis. Associations that passed the Bonferroni cut-off for multiple testing ( $P < 8.08 \times 10^{-5}$ ) were considered replicated.

For the top replicated metabolite from TwinsUK, the same linear regression analysis (though not adjusted for

dietary covariates) was undertaken in a sample of MZ twin pairs for milk intake (twin pairs at least one SD apart in milk intake). The MZ discordant twin pair design provides a powerful means to examine associations while controlling for age, sex and baseline genetic sequence. Associations between the replicated milk-associated blood metabolite from TwinsUK with intakes of other dairy products in the whole TwinsUK population were determined using the same linear regression model described above, though not adjusted for dietary confounders.

The beta coefficients presented in the results of each linear regression analysis represent the amount of milk consumed that corresponds to a 1 SD increase in the metabolite level.

### Binary classification tests

The replicated metabolites (adjusted for covariates) were fitted into a logistic regression models each with the lower quintile of milk intake assigned a negative outcome (0) and the top (high intake) quintile of milk intake was considered a positive outcome (1). A binary classification test was then conducted to evaluate the utility of the metabolites. The ability of the metabolites to correctly classify subjects consuming a high serving of milk per day (sensitivity; true positive rate) and correctly classify subjects consuming the lower quintile of servings of milk per day (specificity; false-positive rate) of the model was predicted, and the receiver operating characteristic curve (ROC) was generated by plotting the true positive rate against the false-positive rate at multiple threshold settings. We further compared the models for replicated metabolites against dairy fat biomarkers pentadecanoic (C15:0) and heptadecanoic acid (C17:0) (recently confirmed by a dietary intervention study [25]) by testing the equality of the ROC area for each.

A second binary classification test was performed to assess the utility of the replicated metabolites to predict genotypic lactase persistence (positive outcome, 1) versus non (negative outcome, 0). The replicated metabolites (adjusted for covariates), reported milk intake, lactose intake and blood levels of dairy fat biomarkers (adjusted for covariates), were each fitted into a logistic regression model to classify lactase persistent individuals versus non-persistent according to genotype (SNP rs4988235 on the *MCM6* gene: CC, lactase non-persistent; TT or CT lactase persistent). The ability of the replicated metabolites, milk fat biomarkers, and reported lactose and milk intakes to correctly classify lactase persistent subjects (sensitivity; true positive rate) and correctly classify lactase non-persistent subjects (specificity; false-positive rate) of the model were predicted and the receiver operating characteristic curves (ROC) produced. The equality of the ROC area for

**Table 1** Characteristics of the participants by study group

	TwinsUK ( <i>n</i> = 3559)	KORA ( <i>n</i> = 1593)	EGCUT ( <i>n</i> = 1109)
Age (years)	55.3 (13.4)	64.1 (5.5)	37.9 (15.7)
BMI (kg/m <sup>2</sup> )	26.0 (4.9)	28.5 (4.3)	25.2 (4.6)
M:F	0:3559	824:769	555:554
Milk intake <sup>a</sup>	3.6 (2.2)	3.5 (2.0)	3.5 (0.7)

Characteristics are expressed in mean (SD) for age, BMI and milk intake and the male to female ratio (M:F) for sex

<sup>a</sup> Milk intake is expressed in servings per week for TwinsUK, the average frequency category of milk intake for KORA (never, 1; once a month or less, 2; several times per month, 3; about once a week, 4; several times per week, 5; almost daily, 6), and the number of days per week on which milk products were consumed for EGCUT

each model was tested against the ROC area for the replicated metabolites.

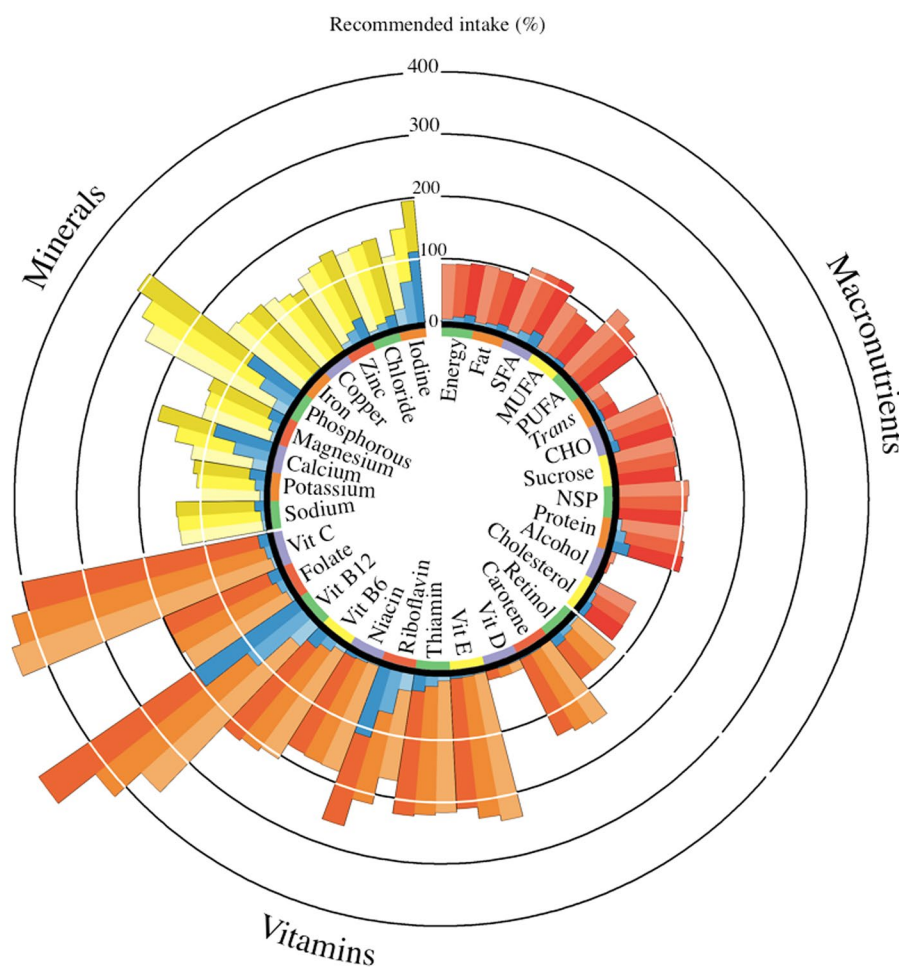
## Results

The demographic characteristics of the study populations are presented in Table 1. The TwinsUK sample was an all-female population on the upper end of middle age, the KORA sample was older and EGCUT younger and both had an approximately even sex ratio. Average BMI for all groups was in the overweight classification (>25 kg/m<sup>2</sup>). Total nutrient intakes for TwinsUK are represented as a percentage of UK recommended intakes by tertile of milk intake (Fig. 2). For subjects consuming the top tertile of milk (mean weekly servings: 6.07, SD = 1.70), milk intake contributed to, or nearly, 100 % of the recommended intakes of vitamin B12 (191 %), riboflavin (110 %), iodine (110 %), calcium (91 %) and phosphorous (90 %). The corresponding increases in nutrient profiles were seen for increasing tertiles of milk intake ( $P < 0.001$ ). In the same group ( $P < 0.001$ ), decreases in non-milk specific nutrients, vitamin C, carotene, vitamin E, MUFA, PUFA, non-starch polysaccharide (NSP) and alcohol were observed; therefore, intakes of relevant foods and nutrients were accounted for in the subsequent metabolomic analysis.

### Metabolites associated with milk consumption

A total of nine metabolites passed the Bonferroni cutoff for association ( $P < 8.08 \times 10^{-5}$ ) from the targeted (Biocrates) and non-targeted (Metabolon) panels in the TwinsUK population (Table 2). To ensure genetic relatedness of the twin pairs did not influence our results, we ran the same linear regression analysis including zygosity as a covariate and our results were unchanged.





**Fig. 2** Total nutrient intakes for TwinsUK represented as a percentage of UK recommended intakes by tertile of milk intake. Total nutrient intakes for TwinsUK represented as a percentage of UK recommended intakes by tertile of milk intake. Average nutrient intakes by increasing tertile of milk intake from clockwise (lightest to darkest) were assessed for percentage of the recommended intakes for 55-year-old women (according to the UK Dietary Reference Values [47]). The *blue color* represents the amount milk consumption contributes to the intake of each nutrient. Using milk intake by tertile as the predictor of the residual energy adjusted nutrient intakes in

a linear regression statistically significant trends ( $P < 0.001$ ) were observed for all nutrients, except cholesterol, sodium, retinol, vitamin D, niacin and folate. Carotene and retinol are represented as percentage of the recommended intake for total retinol equivalents. There is no UK DRV for vitamin D; therefore, 10 ug/d was used. Mean weekly residual-adjusted milk servings by tertile (grams per day in parentheses): tertile 1, mean[SD] = 1.27[1.04] (111.5[80.7]); tertile 2, 3.48[0.50] (295.2[35.2]); tertile 3, 6.07[1.70] (507.0[131.9]). *Trans*, *trans* fatty acids; *NSP* non-starch polysaccharides

### Targeted metabolomics

The glycerophospholipid diacylphosphatidylcholine C28:1 (PC aa C28:1;  $\beta = 0.024$ , SE = 0.005,  $P = 7.24 \times 10^{-7}$ ) and the sphingolipid hydroxysphingomyelin C14:1 (SM(OH) C14:1;  $\beta = 0.024$ , SE = 0.005,  $P = 1.42 \times 10^{-7}$ ) from the Biocrates platform were significantly associated with milk intake in TwinsUK and were replicated both in EGCUT and KORA (meta-analysis results: PC aa C28:1 ( $\beta = 0.034$ , SE = 0.004,  $P = 4.53 \times 10^{-16}$ ) and SM(OH) C14:1 ( $\beta = 0.034$ , SE = 0.005,  $P = 9.75 \times 10^{-14}$ )). See Table 2.

### Non-targeted metabolomics

Seven known non-targeted metabolites were significantly associated with milk intake in TwinsUK and two were also independently replicated in KORA (meta-analysis results: trimethyl-N-aminovalerate ( $\beta = 0.012$ , SE = 0.002,  $P = 2.98 \times 10^{-12}$ ) and uridine ( $\beta = 0.004$ , SE = 0.001,  $P = 9.86 \times 10^{-6}$ )). We find a consistent and statistically significant association of trimethyl-N-aminovalerate and milk intake in a subpopulation of 484 MZ twins discordant for milk intake ( $\beta = 0.050$ , SE = 0.015,  $P = 7.53 \times 10^{-4}$ ) (Fig. 3).

**Table 2** List of metabolites and their associations with milk intake in the TwinUK cohort ( $n = 3559$ ) and replication cohorts KORA ( $n = 1593$ ) and EGCUT ( $n = 1109$ ), and fixed-effects meta-analysis

Metabolite	Pathway	Super pathway	TwinUK <sup>a</sup>		KORA <sup>b</sup>		EGCUT <sup>c</sup>		Fixed-effects meta-analysis <sup>d</sup>	
			Beta (SE)	P	Beta (SE)	P	Beta (SE)	P	Beta	P
Non-targeted										
Trimethyl-N-aminovale- rate (5-trimethylaminovale- rate)	I	Carnitine metabolism	0.089 (0.008)	$6.03 \times 10^{-30}$	0.008 (0.002)	$6.88 \times 10^{-6}$	NA		0.012 (0.002)	$2.98 \times 10^{-12}$ *
Uridine	n	Pyrimidine metabo- lism, uracil contain- ing	0.035 (0.008)	$2.07 \times 10^{-5}$	0.004 (0.001)	$7.74 \times 10^{-5}$	NA		0.004 (0.001)	$9.86 \times 10^{-6}$ *
Phenylalanine	a-a	Phenylalanine and tyrosine metabolism	0.044 (0.008)	$8.29 \times 10^{-8}$	0.001 (0.001)	$9.50 \times 10^{-2}$	NA		0.002 (0.001)	$2.70 \times 10^{-2}$
Tyrosine	a-a	Phenylalanine and tyrosine metabolism	0.035 (0.008)	$1.16 \times 10^{-5}$	0.002 (0.001)	$1.31 \times 10^{-1}$	NA		0.002 (0.001)	$3.52 \times 10^{-2}$
Valine	a-a	Valine, leucine and isoleucine metabo- lism	0.032 (0.008)	$6.89 \times 10^{-5}$	0.000 (0.001)	$7.67 \times 10^{-1}$	NA		0.001 (0.001)	$4.76 \times 10^{-1}$
1,5-Anhydroglucitol	ch	Glycolysis, gluco- neogenesis, pyruvate metabolism	-0.041 (0.008)	$4.07 \times 10^{-7}$	-0.002 (0.002)	$4.07 \times 10^{-1}$	NA		-0.005 (0.002)	$3.13 \times 10^{-2}$
Erythronate	ch	Aminosugars metabo- lism	-0.032 (0.008)	$6.16 \times 10^{-5}$	0.001 (0.001)	$4.75 \times 10^{-1}$	NA			
Targeted										
Diacylphosphatidylcholine C28:1	I	Glycerol-phospholipid	0.024 (0.005)	$7.24 \times 10^{-7}$	0.060 (0.010)	$3.43 \times 10^{-9}$	0.061 (0.013)	$2.00 \times 10^{-6}$	0.034 (0.004)	$4.53 \times 10^{-16}$ *
Hydroxy-sphingomyelin C14:1	I	Sphingolipid	0.024 (0.005)	$1.42 \times 10^{-7}$	0.223 (0.029)	$3.17 \times 10^{-14}$	0.066 (0.013)	$1.28 \times 10^{-6}$	0.034 (0.005)	$9.75 \times 10^{-14}$ *

Beta coefficients presented for the results of each linear regression analysis represent the milk intake frequency that corresponds to a 1 SD increase in the metabolite level / lipids, *n* nucleotide, *a-a* amino acid, *ch*, carbohydrate, *KORA* Cooperative Health Research in the Region of Augsburg, *EGCUT* Estonian Genome Center of the University of Tartu

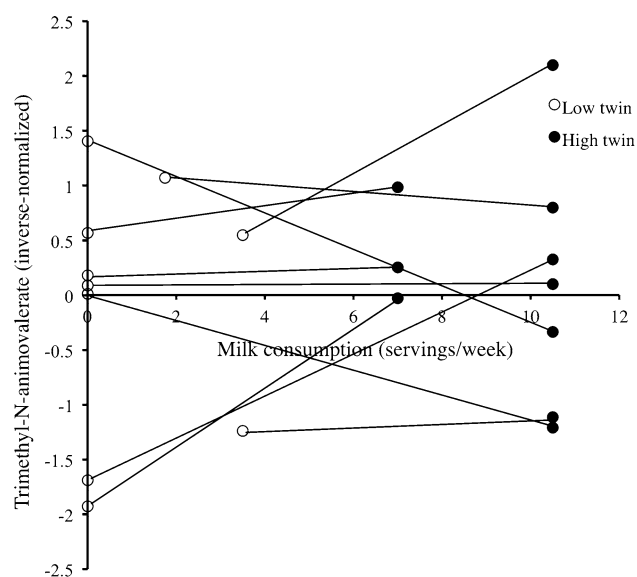
\* Passes the significance threshold for multiple testing ( $P = 8.08 \times 10^{-5}$ )

<sup>a</sup> Milk intakes (servings per week) derived from Food Frequency Questionnaires were completed within  $\pm 5$  years of blood sample collection and fitted as the predictor of metabolite levels in a linear regression. Model adjusted for age, BMI, batch effects, family relatedness and dietary covariates (intake of other dairy products, alcohol, fruit and fruit juices, vegetables, cereals, tea and coffee, and total unsaturated fat). Significance threshold:  $P = 8.08 \times 10^{-5}$

<sup>b</sup> Milk intakes derived from questionnaire completed at the same time of blood sample collection were used as the predictor of metabolite levels in a linear regression. Model adjusted for age, BMI, sex and fasting status. Associations are significant if they are in the same direction as the TwinsUK sample

<sup>c</sup> Milk intakes derived from questionnaire completed at the same time of blood sample collection were used as the predictor of metabolite levels in a linear regression. Model adjusted for age, BMI and sex. Associations are significant if they are in the same direction as the TwinsUK sample

<sup>d</sup> Fixed-effects meta-analysis conducted on milk intake and metabolite associations passing the Bonferroni cut-off ( $0.05/(619 \text{ detected metabolites} \times 1 \text{ diet phenotype}) = 8.08 \times 10^{-5}$ ) from the TwinsUK population and in the same direction in the replication populations



**Fig. 3** Trimethyl-N-aminovalerate versus milk intake for the 10 most discordant twin pairs for reported milk intake. Trimethyl-N-aminovalerate versus milk intake for the 10 most discordant twin pairs for reported milk intake. Milk intake (*x*-axis) versus inverse-normalized levels of blood trimethyl-N-aminovalerate (*y*-axis; unadjusted for age or BMI) are shown for the top 10 twin pairs ( $n = 20$ ) most discordant for milk intake. Twin pairs are connected by a *line*. For 7 out of 10 twin pairs, a higher intake of milk for Twin 2 corresponds to a higher level of trimethyl-N-aminovalerate in the blood (or vice versa)

### Associations between replicated metabolites and dairy product intakes

Replicated metabolites were further examined for their specificity to measure milk intake in TwinsUK. To separate milk or general dairy product intake, we repeated the analysis against weekly servings of other dairy products in the whole TwinUK sample (i.e., yoghurt, cheese, dairy desserts, ice cream, creams and butter; Supplementary Table S1). The association with total dairy product intake and trimethyl-N-aminovalerate passed the significance threshold for multiple testing ( $\beta = 0.012$ ,  $SE = 0.002$ ,  $P = 1.52 \times 10^{-8}$ ), though the large reduction in the strength of association compared to total milk intake suggests trimethyl-N-aminovalerate is not as good a marker of total dairy product consumption as for milk intake alone. No associations between uridine and other or total dairy product intakes passed the significance threshold for multiple testing. The two lipid metabolites, SM(OH) C14:1 and PC aa C28:1, were both significantly associated with butter intake (SM(OH) C14:1:  $\beta = 0.013$ ,  $SE = 0.002$ ,  $P = 2.67 \times 10^{-11}$ ; and PC aa C28:1:  $\beta = 0.011$ ,  $SE = 0.002$ ,  $P = 8.28 \times 10^{-9}$ ), and SM(OH) C14:1 was also significantly associated with cream intake ( $\beta = 0.031$ ,  $SE = 0.007$ ,  $P = 1.90 \times 10^{-6}$ ).

The associations with individual dairy products suggest the lipid metabolites may perform well as markers of dairy fat intake. When total milk intake is separated into high and low fat (Supplementary Table S2), trimethyl-N-aminovalerate and uridine are significantly associated only with low-fat milk intake (trimethyl-N-aminovalerate:  $\beta = 0.078$ ,  $SE = 0.007$ ,  $P = 1.99 \times 10^{-30}$ ; and uridine:  $\beta = 0.032$ ,  $SE = 0.007$ ,  $P = 5.53 \times 10^{-6}$ ), while SM(OH) C14:1 and PC aa C28:1 are not significantly associated with either. When total dairy product intake is separated into high- and low-fat dairy products (Supplementary Table S2), trimethyl-N-aminovalerate is significantly associated only with low-fat dairy intake ( $\beta = 0.028$ ,  $SE = 0.003$ ,  $P = 4.61 \times 10^{-16}$ ), while SM(OH) C14:1 and PC aa C28:1 are significantly associated with only high-fat dairy product intake (SM(OH) C14:1:  $\beta = 0.008$ ,  $SE = 0.001$ ,  $P = 2.23 \times 10^{-9}$ ; and PC aa C28:1:  $\beta = 0.008$ ,  $SE = 0.001$ ,  $P = 5.77 \times 10^{-9}$ ).

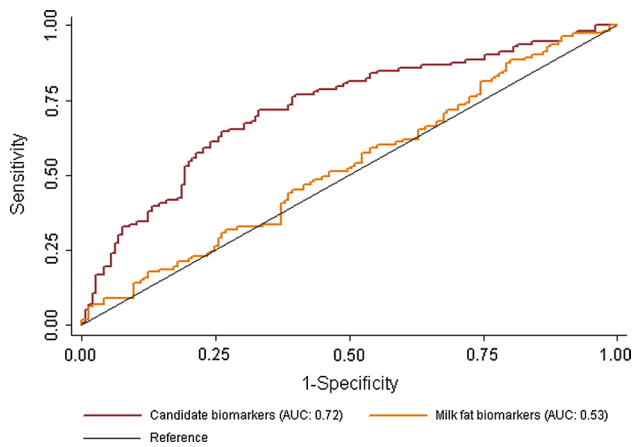
### Urinary excretion

We examined urinary metabolites of 236 twins whose samples were collected at the same time as blood and confirmed a significant association with trimethyl-N-aminovalerate excretion ( $\beta = 0.091$ ,  $SE = 0.032$ ,  $P = 0.004$ ) and milk intake thus suggesting that urine could also be used to evaluate trimethyl-N-aminovalerate levels.

### Potential utility of milk-associated metabolites to act as biomarkers

The utility of SM(OH) C14:1, PC aa C28:1, uridine and trimethyl-N-aminovalerate (candidate biomarkers), to act as markers of milk intake was then assessed by conducting a binary classification test. Twins consuming the bottom quintile of milk (low intake, mean: 0.7 servings/wk) were classed as a negative outcome and top (high intake, 6.8 servings/wk) quintiles classed as positive outcomes. Figure 4 shows the ROC curves for both models as well as models for recently confirmed milk fat biomarkers pentadecanoic and heptadecanoic acid [25] for comparison. The ability of the candidate biomarkers to correctly classify twins reporting a high intake of milk per day was 58 % (sensitivity), whereas 78 % of twins consuming a low milk intake were correctly classified (specificity), overall 69 % of subjects were correctly classified in this model. The AUC for this model was 0.73 (95 % CI 0.66, 0.79). The AUC for the candidate milk biomarkers was significantly greater than milk fat biomarkers ( $\chi^2: 21.08$ ,  $P < 0.0001$ ), the AUC for this model was 0.53 (0.46, 0.60), Supplementary Table S3 contains the results of the full analysis.





**Fig. 4** ROC curves for candidate biomarkers and milk fat biomarkers predictive ability to correctly classify twins reporting low and high intakes of milk per week. Receiver operating characteristic curves for candidate biomarkers and milk fat biomarkers predictive ability to correctly classify twins reporting low and high intakes of milk per week. Candidate biomarkers include blood levels of trimethyl-N-aminovalerate, uridine, hydroxysphingomyelin C14:1 and diacylphosphatidylcholine C28:1. Milk fat biomarkers include blood levels of pentadecanoic (C15:0) and heptadecanoic (C17:0) acids. *AUC* area under the receiver operating characteristic curve

Because FFQ data are imprecise due to a number of known biases, we attempted to compare the utility of each model using genotype to avoid this confounding. Individuals who have genotypic lactase non-persistence (lactose intolerance) consume on average less milk products than those who carry the lactase persistent alleles as they produce less lactase resulting in GI upset when lactose is consumed in higher amounts. The results of the binary classification test using lactase genes for each model (candidate biomarkers, milk fat biomarkers, reported lactose and milk intake) are presented in Table 3, and the ROC curves are presented in Fig. 5. The AUC for the candidate biomarkers

**Table 3** Results of the binary classification test for the candidate biomarkers, milk fat biomarkers, reported lactose and milk intakes to correctly classify lactase persistent and non-persistent twins

	Sensitivity (%)	Specificity (%)	Correctly classified (%)	AUC (95 % CI)	Versus candidate biomarkers	
					$\chi^2$	<i>P</i>
Milk intake	76	36	56	0.53 (0.45, 0.60)	5.12	0.024
Lactose intake	58	50	54	0.53 (0.46, 0.60)	4.53	0.033
Milk fat biomarkers	54	54	54	0.55 (0.48, 0.63)	1.47	0.226
Candidate biomarkers	62	66	64	0.62 (0.55, 0.69)		

The candidate biomarkers (adjusted for covariates), reported milk intake, lactose intake and blood levels of dairy fat biomarkers (adjusted for covariates), were each fitted into a logistic regression model to classify lactase persistent individuals (1, positive outcome) versus non (0, negative outcomes) according to genotype (SNP rs4988235 on the *MCM6* gene: CC, *n* = 63 lactase non-persistent; TT or CT, *n* = 577 lactase persistent). The equality of the receiver operating characteristic area (AUC) for each model was tested against the ROC area for the candidate biomarkers. Candidate biomarkers include blood levels of trimethyl-N-aminovalerate, uridine, hydroxysphingomyelin C14:1 and diacylphosphatidylcholine C28:1. Milk fat biomarkers include blood levels of pentadecanoic (C15:0) and heptadecanoic (C17:0) acids

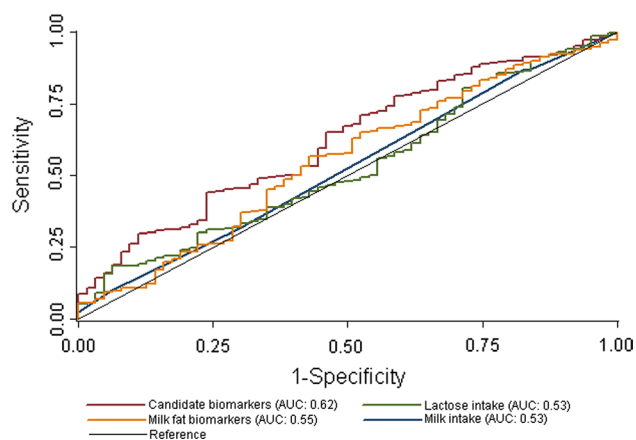
*AUC* area under the receiver operating characteristic curve

was the largest (0.62[0.55, 0.69]) and significantly different than reported milk ( $\chi^2$ : 5.12, *P* = 0.024) and lactose ( $\chi^2$ : 4.53, *P* = 0.033) intakes though not for milk fat biomarkers, suggesting the candidate biomarkers capture milk intake more accurately than reported intakes. The AUC for milk fat biomarkers (0.55[0.48, 0.63]) was not significantly different than reported intakes, suggesting it is not an improvement over FFQ to measure milk intake. The AUC for the candidate biomarkers compared to the milk fat biomarkers was not statistically different.

## Discussion

Using targeted and non-targeted metabolomics approaches, we identified four novel metabolites associated with milk intake, which we replicated in independent populations: SM(OH) C14:1 and PC aa C28:1, uridine and trimethyl-N-aminovalerate. We have reported a strong and novel positive association between circulating levels of trimethyl-N-aminovalerate with milk consumption, and we subsequently used the co-twin control method to confirm the strength of this association independent of genetics. We found these metabolites performed better than FFQ at identifying genotypic lactase persistent individuals.

The association between reported milk consumption and trimethyl-N-aminovalerate is novel and difficult to explain due to the lack of literature on this metabolite. Trimethyl-N-aminovalerate (also known as 5-trimethylaminovalerate, N $\delta$ -Trimethyl-5-aminopentanoate, 5-N-trimethylaminopentanoate, or  $\gamma$ -butyrobetaine[GBB]-5) is very likely a methylated product of 5-aminovalerate (a lysine or proline degradation product produced by gut microflora [26]). Both molecules exhibit structural similarities to carnitine; moreover,  $\gamma$ -butyrobetaine hydroxylase (BBOX) catalyzes the conversion of gamma butyrobetaine to carnitine [27],



**Fig. 5** ROC curves for candidate and milk fat biomarkers, reported lactose and milk intake predictive ability to correctly classify twins with genotypic lactase persistence and non-persistence. Receiver operating characteristic curves for candidate biomarkers, milk fat biomarkers, reported lactose and milk intake predictive ability to correctly classify twins with genotypic lactase persistence and non-persistence. Candidate biomarkers include blood levels of trimethyl-N-aminovalerate, uridine, hydroxy sphingomyelin C14:1 and diacylphosphatidylcholine C28:1. Milk fat biomarkers include blood levels of pentadecanoic (C15:0) and heptadecanoic (C17:0) acids. AUC area under the receiver operating characteristic curve

which is involved in the generation of metabolic energy from long-chain fatty acids. We did not, however, observe a positive association between blood carnitine and its metabolites and reported milk intake. Thus, these molecules seem to be connected through microbial metabolism (that may take place in the dairy cow rumen or by milk spoilage bacteria) involving BBOX enzymes. Why this association is strong with milk intake and not with other dairy products (cheese, yoghurt, butter or cream) is unclear, though may be related to the different properties of milk, such as the higher protein, lactose and water content, non-fermentation by beneficial microorganisms and increased susceptibility to spoilage [27]. Investigations into the metabolome of milk have yet to identify trimethyl-N-aminovalerate, although this may be related to different metabolomics platforms used [28, 29].

We previously reported that circulating trimethyl-N-aminovalerate (previously misidentified as 3-dehydrocarnitine) levels are strongly associated with gene variants on the *MARCH8* gene on chromosome 10 (top hit SNP = rs2291429,  $\beta = 0.02$ ,  $P = 8.7 \times 10^{-11}$ ) [30], which was associated with higher milk intake in our population ( $\beta = 0.165$ , SE = 0.076,  $P = 0.031$ ). The *MARCH8* or cellular modulator of immune recognition (c-Mir) gene has roles in immune regulation [31, 32]. Components of cow's milk protein, caseins,  $\alpha$ -lactalbumin and  $\beta$ -lactoglobulin, are known allergens [33]. These data provide further supportive evidence that milk intake has potential

immune-mediated effects, though the relationship between this gene, trimethyl-N-aminovalerate and milk consumption warrants further investigation.

Despite the limited discourse on the origin of trimethyl-N-aminovalerate in humans, this metabolite (misidentified as 3-dehydrocarnitine) has been detected and associated with health indices in several recent studies using the Metabolon platform [34–36]. One small ( $n = 73$ ) study of older overweight adults with limited mobility, sampled from a randomized, double-blind, controlled trial evaluating the efficacy of whey protein on resistance exercise-induced muscle mass gains [37], identified a positive association between serum trimethyl-N-aminovalerate and abdominal adiposity and intra-muscular adipose tissue in the whole group [34]. In the KORA cohort ( $n = 1762$ ), higher serum trimethyl-N-aminovalerate levels were observed in subjects using fibrates ( $n = 5$ ) versus not [35]. Use of this metabolite as a milk intake surrogate could produce other interesting epidemiological findings.

Uridine is a nucleotide base involved in pyrimidine metabolism. Uridine is directly involved in the conversion of galactose to glucose, though it is not clear that this process could raise human blood uridine levels. Bovine milk is abundant in uridine metabolites, with uridine 5'-monophosphate (5'UMP) accounting for between 86 and 90 % of all nucleotides [38]. Though studies are limited for the bioavailability of food uridine metabolites in humans, oral feeding of 5'UMP has shown to rapidly raise serum uridine in gerbils [39] and is also the predominant uridine metabolite in human breast milk and infant formula. Interestingly, higher blood uridine levels were associated with lower carotid–femoral pulse-wave velocity, a measure of arterial stiffness, in 1797 twins from the TwinsUK cohort [40]. Moreover, expression levels in fat of the purinergic receptor P2Y gene were associated with circulating uridine in TwinsUK, which modulates endothelial nitric oxide synthase phosphorylation [41].

Total circulating phospholipid myristic acid (14:0) has recently been used as a biomarker of dairy fat intake on CVD risk and incident coronary heart disease in a multi-ethnic US adult cohort ( $n = 2837$ ), although no associations were found [42]. We identified positive associations between milk consumption and SM(OH) C14:1 and PC aa C28:1, and these phospholipids are typically composed of myristoleic and myristic (only for PC aa C28:1) acid moieties that are found in characteristically high concentrations in milk fat as structures in the milk fat globule membrane [43]. These reflect findings from the EPIC-Potsdam subcohort ( $n = 2380$ ) where a pattern with high intake of butter and low intake of margarine accounted for 8 % of the variation in circulating levels of SM(OH) C14:1 [44]. These markers may be more useful for measuring the intake of higher fat milk products and would not adequately

to measure skimmed milk intake for instance; however, the inclusion of these metabolites did improve the predictive performance of our models.

This study being cross-sectional has some limitations. Questionnaires were used to assess milk consumption for all groups, which have been plagued by issues of participant reporting [45], though this allowed direct comparisons across the populations used. Validation of the EPIC-Norfolk FFQ used for TwinsUK against 7-day diet diaries demonstrated a tendency to over-report milk intake (350 g per day for FFQ, 200 g per day according to diary) [46], making precise quantification of milk intake unlikely. However, intakes derived from FFQ were found to rank participants in extremes well [46]. We attempted to overcome this issue by applying a novel approach using genotype to evaluate our models predictive performance against confirmed biomarkers of milk fat intake and reported lactose and milk intakes. Future studies should employ interventions to confirm our findings, and the discovery of other metabolites could further improve the predictive ability. Moreover, the origin of the association between milk intake and trimethyl-N-aminovalerate requires further investigation. In particular, metabolomic profiling of milk should be undertaken in future studies. Our study has multiple strengths. Our discovery population had a large sample size with the main results replicated in two independent populations from different countries and cultures of both sexes, suggesting generalizability. We were able to use discordant MZ twins, controlled for confounders of age, sex, the primary DNA sequence, and a shared early environment, to confirm the strength of our top association.

## Conclusions

With the use of high-throughput multi-platform metabolomics, we have identified and successfully replicated in independent populations four novel metabolites associated with milk consumption. Moreover, our evaluation of the ability of these metabolites to predict genotypic lactase persistence versus non-persistence suggests they perform better than milk fat biomarkers or reported lactose and milk intake as markers of milk intake. This study demonstrates the advantages of metabolomics in nutritional epidemiology, and these novel metabolites have the potential to act as objective biomarkers to strengthen epidemiological studies determining the impact of milk intake on health.

**Acknowledgments** The authors wish to express their appreciation to all study participants on the TwinsUK, KORA and EGCUT studies for donating their blood samples and time.

**Funding** TwinsUK was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007-2013).

The study also receives support from the National Institute for Health Research (NIHR) Clinical Research Facility at Guy's & St Thomas' NHS Foundation Trust and NIHR Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. Metabolomic analysis was funded by Pfizer. Aedín Cassidy is a Royal Society Wolfson Research Merit Award Holder. The KORA research platform (KORA, Cooperative Health Research in the Region of Augsburg) was initiated and financed by the Helmholtz Zentrum München-German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. This research has partly received funding from the European Union's Seventh Framework Programme (FP7-Health-F5-2012) under Grant agreement No 305280 (MIMomics). The EGCUT studies were financed by University of Tartu (Grant "Center of Translational Genomics"), by Estonian Government (Grants #TP1GV14060L, #ETF9353) and by European Commission through the European Regional Development Fund in the frame of grant "Centre of Excellence in Genomics", by an EFSU New Horizons grant and Estonian Research Infrastructure's Roadmap and through FP7 Grant #313010.

**Authors' contributions** TP, TDS and CM designed research; RPM conducted research; BT, EA, CG, GK, AMacG, AC and AJ provided essential materials; TP, TH and CM analyzed data or performed statistical analysis; TP, TDS and CM wrote the paper; TP, TDS and CM had primary responsibility for final content; AC, AJ, TM, TDS, AMacG and AM reviewed and edited the manuscript.

## Compliance with ethical standards

**Conflict of interest** Robert P. Mohny is an employee of Metabolon, Inc. All other authors had no conflicting interests.

**Ethical approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

**Informed consent** Informed consent was obtained from all individual participants included in the study.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

1. Itan Y, Powell A, Beaumont MA, Burger J, Thomas MG (2009) The origins of lactase persistence in Europe. *PLoS Comput Biol* 5:e1000491. doi:10.1371/journal.pcbi.1000491
2. Prentice AM (2014) Dairy products in global public health. *Am J Clin Nutr* 99:1212s–1216s. doi:10.3945/ajcn.113.073437
3. Astrup A (2014) Yogurt and dairy product consumption to prevent cardiometabolic diseases: epidemiologic and experimental studies. *Am J Clin Nutr* 99:1235s–1242s. doi:10.3945/ajcn.113.073015

4. Stewart H, Dong D, Carlson A (2013) Why are Americans consuming less fluid milk? A look at generational differences in intake frequency. In: Agriculture USDo (ed), p 35
5. Food statistics team (2013) Family food: 2013. In: UK department for Environment FaRA (ed), London
6. Rizzoli R (2014) Dairy products, yogurts, and bone health. *Am J Clin Nutr* 99:1256s–1262s. doi:[10.3945/ajcn.113.073056](https://doi.org/10.3945/ajcn.113.073056)
7. Soedamah-Muthu SS, Ding EL, Al-Delaimy WK, Hu FB, Engberink MF, Willett WC, Geleijnse JM (2011) Milk and dairy consumption and incidence of cardiovascular diseases and all-cause mortality: dose-response meta-analysis of prospective cohort studies. *Am J Clin Nutr* 93:158–171. doi:[10.3945/ajcn.2010.29866](https://doi.org/10.3945/ajcn.2010.29866)
8. Michaelsson K, Wolk A, Langenskiöld S, Basu S, Warensjö Lemming E, Melhus H, Byberg L (2014) Milk intake and risk of mortality and fractures in women and men: cohort studies. *BMJ (Clin Res ed.)* 349:g6015. doi:[10.1136/bmj.g6015](https://doi.org/10.1136/bmj.g6015)
9. Guertin KA, Moore SC, Sampson JN, Huang WY, Xiao Q, Stolzenberg-Solomon RZ, Sinha R, Cross AJ (2014) Metabolomics in nutritional epidemiology: identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. *Am J Clin Nutr* 100:208–217. doi:[10.3945/ajcn.113.078758](https://doi.org/10.3945/ajcn.113.078758)
10. Zheng Y, Yu B, Alexander D, Steffen LM, Nettleton JA, Boerwinkle E (2014) Metabolomic patterns and alcohol consumption in African Americans in the Atherosclerosis Risk in Communities Study. *Am J Clin Nutr* 99:1470–1478. doi:[10.3945/ajcn.113.074070](https://doi.org/10.3945/ajcn.113.074070)
11. Shin SY, Fauman EB, Petersen AK, Krumsiek J, Santos R, Huang J, Arnold M, Erte I, Forgetta V, Yang TP, Walter K, Menni C, Chen L, Vasquez L, Valdes AM, Hyde CL, Wang V, Ziemek D, Roberts P, Xi L, Grundberg E, Waldenberger M, Richards JB, Mohny RP, Milburn MV, John SL, Trimmer J, Theis FJ (2014) An atlas of genetic influences on human blood metabolites. *Nat Genet* 46:543–550. doi:[10.1038/ng.2982](https://doi.org/10.1038/ng.2982)
12. Menni C, Zhai G, Macgregor A, Prehn C, Romisch-Margl W, Suhre K, Adamski J, Cassidy A, Illig T, Spector TD, Valdes AM (2013) Targeted metabolomics profiles are strongly correlated with nutritional patterns in women. *Metabolomics* 9:506–514. doi:[10.1007/s11306-012-0469-6](https://doi.org/10.1007/s11306-012-0469-6)
13. Moayyeri A, Hammond CJ, Hart DJ, Spector TD (2013) The UK Adult Twin Registry (TwinsUK Resource). *Twin Res Hum Genet* 16:144–149. doi:[10.1017/thg.2012.89](https://doi.org/10.1017/thg.2012.89)
14. Bingham SA, Welch AA, McTaggart A, Mulligan AA, Runswick SA, Luben R, Oakes S, Khaw KT, Wareham N, Day NE (2001) Nutritional methods in the European Prospective Investigation of Cancer in Norfolk. *Public Health Nutr* 4:847–858
15. Teucher B, Skinner J, Skidmore PM, Cassidy A, Fairweather-Tait SJ, Hooper L, Roe MA, Foxall R, Oyston SL, Cherkas LF, Perks UC, Spector TD, MacGregor AJ (2007) Dietary patterns and heritability of food choice in a UK female twin cohort. *Twin Res Hum Genet* 10:734–748. doi:[10.1375/twin.10.5.734](https://doi.org/10.1375/twin.10.5.734)
16. Fraenkel G (1951) Effect and distribution of vitamin B<sub>12</sub>. *Arch Biochem Biophys* 34:457–467
17. Willett W, Stampfer MJ (1986) Total energy intake: implications for epidemiologic analyses. *Am J Epidemiol* 124:17–27
18. Leitsalu L, Haller T, Esko T, Tammesoo ML, Alavere H, Snieder H, Perola M, Ng PC, Magi R, Milani L, Fischer K, Metspalu A (2014) Cohort profile: Estonian biobank of the Estonian Genome Center, University of Tartu. *Int J Epidemiol*. doi:[10.1093/ije/dyt268](https://doi.org/10.1093/ije/dyt268)
19. Holle R, Happich M, Lowel H, Wichmann HE (2005) KORA—a research platform for population based health research. *Gesundheitswesen (Bundesverband der Ärzte des Öffentlichen Gesundheitsdienstes (Germany))* 67(Suppl 1):S19–S25. doi:[10.1055/s-2005-858235](https://doi.org/10.1055/s-2005-858235)
20. Winkler G, Doring A (1998) Validation of a short qualitative food frequency list used in several German large scale surveys. *Z Ernahrungswiss* 37:234–241
21. Menni C, Kastenmuller G, Petersen AK, Bell JT, Psatha M, Tsai PC, Gieger C, Schulz H, Erte I, John S, Brosnan MJ, Wilson SG, Tsaprouni L, Lim EM, Stuckey B, Deloukas P, Mohny R, Suhre K, Spector TD, Valdes AM (2013) Metabolomic markers reveal novel pathways of ageing and early development in human populations. *Int J Epidemiol* 42:1111–1119. doi:[10.1093/ije/dyt094](https://doi.org/10.1093/ije/dyt094)
22. Menni C, Fauman E, Erte I, Perry JR, Kastenmuller G, Shin SY, Petersen AK, Hyde C, Psatha M, Ward KJ, Yuan W, Milburn M, Palmer CN, Frayling TM, Trimmer J, Bell JT, Gieger C, Mohny RP, Brosnan MJ, Suhre K, Soranzo N, Spector TD (2013) Biomarkers for type 2 diabetes and impaired fasting glucose using a nontargeted metabolomics approach. *Diabetes* 62:4270–4276. doi:[10.2337/db13-0570](https://doi.org/10.2337/db13-0570)
23. Illig T, Gieger C, Zhai G, Romisch-Margl W, Wang-Sattler R, Prehn C, Altmaier E, Kastenmuller G, Kato BS, Mewes HW, Meitinger T, de Angelis MH, Kronenberg F, Soranzo N, Wichmann HE, Spector TD, Adamski J, Suhre K (2010) A genome-wide perspective of genetic variation in human metabolism. *Nat Genet* 42:137–141. doi:[10.1038/ng.507](https://doi.org/10.1038/ng.507)
24. Römisch-Margl W, Prehn C, Bogumil R, Röhring C, Suhre K, Adamski J (2012) Procedure for tissue sample preparation and metabolite extraction for high-throughput targeted metabolomics. *Metabolomics* 8:133–142. doi:[10.1007/s11306-011-0293-4](https://doi.org/10.1007/s11306-011-0293-4)
25. Abdullah MM, Cyr A, Lepine MC, Labonte ME, Couture P, Jones PJ, Lamarche B (2015) Recommended dairy product intake modulates circulating fatty acid profile in healthy adults: a multi-centre cross-over study. *Br J Nutr* 113:435–444. doi:[10.1017/s0007114514003894](https://doi.org/10.1017/s0007114514003894)
26. Martin FP, Sprenger N, Montoliu I, Rezzi S, Kochhar S, Nicholson JK (2010) Dietary modulation of gut functional ecology studied by fecal metabolomics. *J Proteome Res* 9:5284–5295. doi:[10.1021/pr100554m](https://doi.org/10.1021/pr100554m)
27. Rydzik AM, Leung IK, Kochan GT, Loik ND, Henry L, McDonough MA, Claridge TD, Schofield CJ (2014) Comparison of the substrate selectivity and biochemical properties of human and bacterial gamma-butyrobetaine hydroxylase. *Org Biomol Chem* 12:6354–6358. doi:[10.1039/c4ob01167h](https://doi.org/10.1039/c4ob01167h)
28. Klein MS, Almstetter MF, Schlamberger G, Nurnberger N, Dettmer K, Oefner PJ, Meyer HH, Wiedemann S, Gronwald W (2010) Nuclear magnetic resonance and mass spectrometry-based milk metabolomics in dairy cows during early and late lactation. *J Dairy Sci* 93:1539–1550. doi:[10.3168/jds.2009-2563](https://doi.org/10.3168/jds.2009-2563)
29. Lu J, Antunes Fernandes E, Paez Cano AE, Vinitwatanakun J, Boeren S, van Hooijdonk T, van Kneegsel A, Vervoort J, Hettinga KA (2013) Changes in milk proteome and metabolome associated with dry period length, energy balance, and lactation stage in postparturient dairy cows. *J Proteome Res* 12:3288–3296. doi:[10.1021/pr4001306](https://doi.org/10.1021/pr4001306)
30. Shin SY, Fauman EB, Petersen AK, Krumsiek J, Santos R, Huang J, Arnold M, Erte I, Forgetta V, Yang TP, Walter K, Menni C, Chen L, Vasquez L, Valdes AM, Hyde CL, Wang V, Ziemek D, Roberts P, Xi L, Grundberg E, Waldenberger M, Richards JB, Mohny RP, Milburn MV, John SL, Trimmer J, Theis FJ (2014) An atlas of genetic influences on human blood metabolites. *Nature* 46:543–550. doi:[10.1038/ng.2982](https://doi.org/10.1038/ng.2982)
31. Goto E, Ishido S, Sato Y, Ohgimoto S, Ohgimoto K, Nagano-Fujii M, Hotta H (2003) c-MIR, a human E3 ubiquitin ligase, is a functional homolog of herpesvirus proteins MIR1 and MIR2 and has similar activity. *J Biol Chem* 278:14657–14668. doi:[10.1074/jbc.M211285200](https://doi.org/10.1074/jbc.M211285200)
32. Ohmura-Hoshino M, Matsuki Y, Aoki M, Goto E, Mito M, Uematsu M, Kakiuchi T, Hotta H, Ishido S (2006) Inhibition



- of MHC class II expression and immune responses by c-MIR. *J Immunol* 177:341–354
33. Tsabouri S, Douros K, Priftis KN (2014) Cow's milk allergenicity. *Endocr metab immun disord drug targ* 14:16–26
  34. Lustgarten MS, Price LL, Phillips EM, Fielding RA (2013) Serum glycine is associated with regional body fat and insulin resistance in functionally-limited older adults. *PLoS ONE* 8:e84034. doi:[10.1371/journal.pone.0084034](https://doi.org/10.1371/journal.pone.0084034)
  35. Altmaier E, Fobo G, Heier M, Thorand B, Meisinger C, Romisch-Margl W, Waldenberger M, Gieger C, Illig T, Adamski J, Suhre K, Kastenmuller G (2014) Metabolomics approach reveals effects of antihypertensives and lipid-lowering drugs on the human metabolism. *Eur J Epidemiol* 29:325–336. doi:[10.1007/s10654-014-9910-7](https://doi.org/10.1007/s10654-014-9910-7)
  36. Goedert JJ, Sampson JN, Moore SC, Xiao Q, Xiong X, Hayes RB, Ahn J, Shi J, Sinha R (2014) Fecal metabolomics: assay performance and association with colorectal cancer. *Carcinogenesis* 35:2089–2096. doi:[10.1093/carcin/bgu131](https://doi.org/10.1093/carcin/bgu131)
  37. Chale A, Cloutier GJ, Hau C, Phillips EM, Dallal GE, Fielding RA (2013) Efficacy of whey protein supplementation on resistance exercise-induced changes in lean mass, muscle strength, and physical function in mobility-limited older adults. *J Gerontol Ser A Biol Sci Med Sci* 68:682–690. doi:[10.1093/gerona/gls221](https://doi.org/10.1093/gerona/gls221)
  38. Mateo CD, Peters DN, Stein HH (2004) Nucleotides in sow colostrum and milk at different stages of lactation. *J Anim Sci* 82:1339–1342
  39. Cansev M, Watkins CJ, van der Beek EM, Wurtman RJ (2005) Oral uridine-5'-monophosphate (UMP) increases brain CDP-choline levels in gerbils. *Brain Res* 1058:101–108. doi:[10.1016/j.brainres.2005.07.054](https://doi.org/10.1016/j.brainres.2005.07.054)
  40. Menni C, Mangino M, Cecelja M, Psatha M, Brosnan MJ, Trimmer J, Mohnhey RP, Chowienczyk P, Padmanabhan S, Spector TD, Valdes AM (2014) Metabolomic study of carotid-femoral pulse-wave velocity in women. *J Hypertens*. doi:[10.1097/hjh.0000000000000467](https://doi.org/10.1097/hjh.0000000000000467)
  41. da Silva CG, Specht A, Wegiel B, Ferran C, Kaczmarek E (2009) Mechanism of purinergic activation of endothelial nitric oxide synthase in endothelial cells. *Circulation* 119:871–879. doi:[10.1161/circulationaha.108.764571](https://doi.org/10.1161/circulationaha.108.764571)
  42. de Oliveira Otto MC, Nettleton JA, Lemaitre RN, Steffen LM, Kromhout D, Rich SS, Tsai MY, Jacobs DR, Mozaffarian D (2013) Biomarkers of dairy fatty acids and risk of cardiovascular disease in the multi-ethnic study of atherosclerosis. *J Am Heart Assoc* 2:e000092. doi:[10.1161/jaha.113.000092](https://doi.org/10.1161/jaha.113.000092)
  43. Mansson HL (2008) Fatty acids in bovine milk fat. *Food Nutr Res*. doi:[10.3402/fnr.v52i0.1821](https://doi.org/10.3402/fnr.v52i0.1821)
  44. Floegel A, von Ruesten A, Drogan D, Schulze MB, Prehn C, Adamski J, Pischon T, Boeing H (2013) Variation of serum metabolites related to habitual diet: a targeted metabolomic approach in EPIC-Potsdam. *Eur J Clin Nutr* 67:1100–1108. doi:[10.1038/ejcn.2013.147](https://doi.org/10.1038/ejcn.2013.147)
  45. Westerterp KR, Goris AH (2002) Validity of the assessment of dietary intake: problems of misreporting. *Curr Opin Clin Nutr Metab Care* 5:489–493
  46. Bingham SA, Gill C, Welch A, Cassidy A, Runswick SA, Oakes S, Lubin R, Thurnham DI, Key TJ, Roe L, Khaw KT, Day NE (1997) Validation of dietary assessment methods in the UK arm of EPIC using weighed records, and 24-hour urinary nitrogen and potassium and serum vitamin C and carotenoids as biomarkers. *Int J Epidemiol* 26(Suppl 1):S137–S151
  47. Health Do (1991) Dietary reference values for food energy and nutrients for the United Kingdom. In: Report of the panel on dietary reference values of the committee on medical aspects of food policy. Report on Health and Social Subjects 41. HMSO, London