



Dairy Product Intake and Risk of Type 2 Diabetes in EPIC-InterAct: A Mendelian Randomization Study

<https://doi.org/10.2337/dc18-2034>

OBJECTIVE

To estimate the causal association between intake of dairy products and incident type 2 diabetes.

RESEARCH DESIGN AND METHODS

The analysis included 21,820 European individuals (9,686 diabetes cases) of the EPIC-InterAct case-cohort study. Participants were genotyped, and rs4988235 (LCT-12910C>T), a SNP for lactase persistence (LP) which enables digestion of dairy sugar, i.e., lactose, was imputed. Baseline dietary intakes were assessed with diet questionnaires. We investigated the associations between imputed SNP dosage for rs4988235 and intake of dairy products and other foods through linear regression. Mendelian randomization (MR) estimates for the milk-diabetes relationship were obtained through a two-stage least squares regression.

RESULTS

Each additional LP allele was associated with a higher intake of milk (β 17.1 g/day, 95% CI 10.6–23.6) and milk beverages (β 2.8 g/day, 95% CI 1.0–4.5) but not with intake of other dairy products. Other dietary intakes associated with rs4988235 included fruits (β –7.0 g/day, 95% CI –12.4 to –1.7 per additional LP allele), nonalcoholic beverages (β –18.0 g/day, 95% CI –34.4 to –1.6), and wine (β –4.8 g/day, 95% CI –9.1 to –0.6). In instrumental variable analysis, LP-associated milk intake was not associated with diabetes (hazard ratio 0.99_{per 15 g/day}, 95% CI 0.93–1.05).

CONCLUSIONS

rs4988235 was associated with milk intake but not with intake of other dairy products. This MR study does not suggest that milk intake is associated with diabetes, which is consistent with previous observational and genetic associations. LP may be associated with intake of other foods as well, but owing to the modest associations we consider it unlikely that this has caused the observed null result.

Eating healthily on a daily basis is a major step to prevent development of type 2 diabetes (1). Higher intake of dairy products has been associated with a lower risk of diabetes in a meta-analysis of observational studies (2). Yogurt and cheese intake particularly were associated with lower diabetes risk, whereas milk intake was not, with substantial heterogeneity for most dairy products. In European populations, a direct association between milk intake and risk of diabetes was suggested, but this was

Linda E.T. Vissers,¹ Iivonne Sluijs,¹
Yvonne T. van der Schouw,¹
Nita G. Forouhi,² Fumiaki Imamura,²
Stephen Burgess,³ Aurelio Barricarte,⁴
Heiner Boeing,⁵ Catalina Bonet,⁶
Maria-Dolores Chirlaque,^{7,8}
Guy Fagherazzi,⁹ Paul W. Franks,¹⁰
Heinz Freisling,¹¹ Marc J. Gunter,¹¹
J. Ramón Quirós,¹² Daniel B. Ibsen,¹³
Rudolf Kaaks,¹⁴ Timothy Key,¹⁵
Kay T. Khaw,³ Tilman Kühn,¹⁴
Olatz Mokoroa,^{8,16} Peter M. Nilsson,¹⁰
Kim Overvad,¹⁷ Valeria Pala,¹⁸
Domenico Palli,¹⁹ Salvatore Panico,²⁰
Carlotta Sacerdote,²¹
Annemieke M.W. Spijkerman,²²
Anne Tjonneland,²³ Rosario Tumino,²⁴
Miguel Rodríguez-Barranco,^{8,25}
Olov Rolandsson,²⁶ Elio Riboli,²⁷
Stephen J. Sharp,² Claudia Langenberg,²
and Nicholas J. Wareham²

¹Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands

²MRC Epidemiology Unit, University of Cambridge, Cambridge, U.K.

³Department of Public Health and Primary Care, School of Clinical Medicine, University of Cambridge, Cambridge, U.K.

⁴Navarre Public Health Institute (ISPN), Pamplona, Spain

⁵German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany

⁶Unit of Nutrition and Cancer, Cancer Epidemiology Research Program, Institut Català d'Oncologia, Barcelona, Spain

⁷Department of Epidemiology, Regional Health Council, IMIB-Arixaca, University of Murcia, Murcia, Spain

⁸CIBER in Epidemiology and Public Health (CIBERESP), Madrid, Spain

⁹CESP UMR1018, INSERM, Institut Gustave Roussy, Paris South University–Paris-Saclay University, Villejuif, France

¹⁰Lund University, Malmö, Sweden

¹¹Section of Nutrition and Metabolism, International Agency for Research on Cancer, Lyon, France

¹²Public Health Directorate, Asturias, Spain

¹³Section for Epidemiology, Department of Public Health, Aarhus University, Aarhus, Denmark

¹⁴German Cancer Research Center (DKFZ), Heidelberg, Germany

¹⁵University of Oxford, Oxford, U.K.

¹⁶Public Health Division of Gipuzkoa, Biodonostia Research Institute, San Sebastian, Spain

¹⁷Department of Cardiology, Aalborg University Hospital, Aalborg, Denmark

not statistically significant (2). Protective components of dairy products may be whey proteins, odd-chain fatty acids, and the high nutrient density of dairy (3). Also, interactions within the dairy food matrix may modify the metabolic effects of dairy consumption (4).

However, potential confounding and reverse causation cannot be excluded (5). Owing to these limitations, the causal role of dairy products in diabetes prevention remains debatable.

The relationship between dairy products and risk of diabetes could be investigated by applying a Mendelian randomization (MR) approach (5), using genetic variability in the *MCM6* gene associated with lactase persistence (LP) in adults as an instrumental variable (IV). Lactase is necessary to break down the sugars that are found in dairy products, i.e., lactose. SNPs in the *MCM6* region have been associated with LP (6). rs4988235 (LCT-12910C>T) has been associated with LP in European populations (6,7) and has been associated with a higher intake of milk in European cohorts (8–12), albeit not in all (13).

Previous MR studies reported no association between LP-associated milk intake and diabetes (8,14). However, variation in the *MCM6* gene is likely to lead to population stratification (6,10), which would introduce bias to an MR analysis, and previous MR studies (8,14) did not sufficiently adjust for population substructure. Also, previous studies did not investigate whether rs4988235 was specifically associated with dairy product intake after adjusting for population substructure.

We therefore investigated whether rs4988235 associated with intake of dairy products and other foods in a pan-European study in eight countries with

different dietary habits (15). We adjusted for genetic principal components (PCs) and study center to adjust for population substructure (16). Next, we used rs4988235 in an IV analysis to investigate whether there is a causal relationship between the LP-associated exposure and risk of diabetes.

RESEARCH DESIGN AND METHODS

Study Design and Population

European Prospective Investigation into Cancer (EPIC)-InterAct is a prospective case-cohort study nested within eight European countries of the EPIC study (17). From 340,234 adults of the EPIC study for whom baseline blood samples were available, EPIC-InterAct randomly selected a subcohort of 16,154 participants and identified 12,403 incident cases of type 2 diabetes between 1991 and 2007, including 778 cases from the subcohort by design.

We excluded 5,287 participants who were not successfully genotyped (14% of samples had no DNA, 58% had insufficient DNA, 8% failed initial PCR, and 20% failed array-based genotyping). Of the 5,287 participants who failed genotyping, 57% developed diabetes during follow-up vs. 55% of 22,492 participants who were successfully genotyped.

We additionally excluded participants with missing information on dairy product intake ($n = 592$) and one participant per set of relatives (identity by descent >0.1875 , $n = 80$), leading to a total population of 21,820 (Supplementary Fig. 1).

Genotyping and GWAS Search

DNA was extracted from buffy coat from a citrated blood sample on an automated Autopure LS DNA extraction system (Qiagen, Hilden, Germany) with Puregene chemistry (Qiagen).

Participant samples were genotyped with the Illumina HumanCoreExome-24 chip array, version 1, and Illumina HumanCoreExome-12 chip array, version 1 ($n = 12,792$), or with the Illumina660W-Quad BeadChip ($n = 8,955$). Sample exclusion criteria were low call rate (threshold $<95.4\%$ in Illumina660W-Quad BeadChip and $<98\%$ in HumanCoreExome arrays), discordance between self-reported sex and sex based on X chromosome heterozygosity, outliers for heterozygosity, lack of concordance with previous genotyping results, or non-European genetic ancestry.

Before imputation, SNPs were filtered to remove those with minor allele count <2 , call rate $<95\%$, or Hardy Weinberg P value $<1e-6$. Imputation to the Haplotype Reference Consortium, version 1.0, panel using IMPUTE, version 2.3.2, was performed at the Wellcome Trust Centre for Human Genetics.

For the present analyses, we used the imputed LP SNP rs4988235 (LCT-12910C>T) (7). Imputation “info,” a measure that reflects certainty of imputation (18), was 0.89 on the Illumina660W-Quad BeadChip and 0.45 on the Illumina HumanCoreExome chips for rs4988235.

PhenoScanner (19) was used to search for phenotypes that have been associated with rs4988235 or its proxies ($R^2 > 0.8$) in previous GWAS. These phenotypes could be mediators of an association between milk intake and diabetes but could also indicate pleiotropy. Pleiotropy occurs when genetic variation affects two or more phenotypic traits that are seemingly unrelated and can lead to violation of assumptions made in MR studies (5). The effect allele (T) of rs4988235 has been associated (at $P < 5 * 10^{-8}$) with a greater hip circumference (20), greater height (21), and lower

¹⁸Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy

¹⁹Cancer Risk Factors and LifeStyle Epidemiology Unit, Institute for Cancer Research, Prevention and Clinical Network (ISPRO), Florence, Italy

²⁰Dipartimento di Medicina Clinica e Chirurgia, Federico II University, Naples, Italy

²¹Unit of Cancer Epidemiology, University Hospital “Città della Salute e della Scienza,” University of Turin, and Center for Cancer Prevention (CPO), Torino, Italy

²²National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands

²³Danish Cancer Society Research Center, Copenhagen, Denmark

²⁴ASP Ragusa, Ragusa, Italy

²⁵Escuela Andaluza de Salud Pública, Instituto de Investigación Biosanitaria IBS.GRANADA, Hospitales Universitarios de Granada, Universidad de Granada, Granada, Spain

²⁶Umeå University, Umeå, Sweden

²⁷School of Public Health, Imperial College London, London, U.K.

Corresponding author: Ivonne Sluijs, i.sluijs-2@umcutrecht.nl

Received 26 September 2018 and accepted 9 January 2019

This article contains Supplementary Data online at <http://care.diabetesjournals.org/lookup/suppl/doi:10.2337/dc18-2034/-/DC1>.

© 2019 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. More information is available at <http://www.diabetesjournals.org/content/license>.

LDL and total cholesterol (22) (Supplementary Table 1). We did not find evidence of an association between rs4988235 and phenotypes that are unrelated to dairy product intake.

Dietary Measurement

Dietary intake over the previous 12 months before study inclusion was assessed at baseline through diet questionnaires, which varied per country or study center. These questionnaires were developed to reflect local eating habits and validated locally (23–25).

Information on intake of milk ([semi]-skimmed or full fat, regardless of fermentation), yogurt and thick fermented milk (e.g., sour milk), and cheese was available for the full subcohort ($n = 12,722$). Availability of consumption data for other dairy products differed by country and/or center, depending on the cohort-specific questionnaire. Information on intake of dairy creams (e.g., whipped cream), curd (e.g., quark, cottage cheese), milk-based puddings (e.g., custard), milk beverages (e.g., chocolate milk), and milk for coffee and creamers was available for 11,536, 10,372, 10,372, 6,867, and 2,959 subcohort participants, respectively. Information on consumption of nonmilk dairy was calculated by summing intake of all dairy products other than milk.

Measurement of fatty acids has previously been described in detail (26). In short, fatty acids C15:0 and C17:0 were measured in phospholipids in plasma samples that were stored at baseline at -196°C or -150°C , using gas chromatography (Agilent Technologies, Santa Clara, CA) equipped with flame ionization detection.

Covariates

Baseline information on education, lifestyle, and medical history was obtained from self-administered questionnaires. Weight, height, and hip and waist circumference were measured during a visit to a study center. BMI (weight in kilograms divided by the square of height in meters) and waist-to-hip ratio (WHR) (waist circumference/hip circumference) were calculated, adjusted for clothing by subtracting 1.5 kg for weight and 2.0 cm for circumferences for people who were normally dressed without shoes. Physical activity was classified as inactive, moderately inactive, moderately active, and

active, according to the Cambridge Physical Activity Index (27).

The lipids HDL cholesterol, triglycerides, and lipoprotein (a) were measured in serum using a Cobas enzymatic assay (Roche Diagnostics, Mannheim, Germany) on a Roche Hitachi Modular P analyzer. LDL cholesterol was calculated by Friedewald equation.

Erythrocyte HbA_{1c} was measured using Tosoh (HLC-723G8) ion-exchange high-performance liquid chromatography on a Tosoh G8.

Diabetes

Ascertainment of incident type 2 diabetes has previously been described (17) and involved a review of the existing EPIC data sets at each center using multiple sources of evidence, including self-report, linkage to primary and secondary care registers, drug registers, hospital admissions, and mortality data. To increase specificity of case definition, we sought confirmation of type 2 diabetes diagnosis where there was only one source of evidence, including individual medical record review in some centers. Follow-up was censored at date of diagnosis, 31 December 2007, or the date of death—whichever occurred first.

Data Analysis

Baseline characteristics and dietary intakes of the subcohort and by most probable rs4988235 genotype were presented as percentages, mean \pm SD, or median (25th percentile–75th percentile [p25–p75]). We additionally reported dietary intakes per country. Hardy-Weinberg equilibrium (HWE) based on most probable LP genotype was examined in the subcohort and after stratification to country.

We first checked two IV assumptions, namely, whether the IV is reliably associated with exposure and whether the IV is independent of confounders of the exposure-outcome relationship. The third IV assumption is that the IV solely influences the outcome via a causal pathway that includes the exposure of interest and cannot be checked (5). IV assumptions were checked only in a subcohort participants because the subcohort is population based. We then proceeded to the IV analysis in the full case-cohort population to answer the main research question: Is there a causal

relationship between dairy product intake and risk of diabetes?

rs4988235 and Intake of Dairy Products

rs4988235 was modeled continuously with SNP dosage, assuming an additive effect of LP alleles. Linear regression was used to examine the association between LP and dairy products. If rs4988235 was associated with multiple dairy products, the association between the LP SNP and a composite dairy product exposure (e.g., milk and milk beverages) was examined. The model with the highest F statistic was used for subsequent IV analyses. The F statistic is often used to assess the risk of weak instrument bias in MR studies, but simulation studies have shown that bias when using one IV is very low, even for expected F statistics around 5 (28).

To support results based on diet questionnaire data, we also performed a linear regression between rs4988235 and plasma levels of saturated fats C15:0 and C17:0. These fatty acids are mainly found in dairy products, and plasma measurements of C15:0 and C17:0 have previously been associated with dairy product intake, although there seems to be some endogenous production of these fatty acids as well (29).

We adjusted for the first three genetic PCs, study center, genotyping platform, sex, and age in the linear regression models.

rs4988235 and Diabetes Risk Factors

To examine potential pleiotropy of rs4988235 in EPIC-InterAct, we investigated the relationship between rs4988235 and baseline lifestyle factors, anthropometric measurements, and blood lipids. These phenotypes could either be intermediates in the causal chain of an association between milk intake and diabetes or pleiotropic effects of the LP SNP.

Linear regression was used for continuous variables and logistic regression for dichotomous variables. Triglycerides and lipoprotein a were log transformed before analysis. Analyses were adjusted for genotyping platform, sex, age, the first three genetic PCs, and study center.

The avoidance or consumption of dairy products may have consequences for the whole dietary pattern. More specifically, if a person avoids dairy products, this could lead to a reduction in total energy

intake, replacement of dairy by other beverages or foods, or a combination of the two scenarios. It is also possible that dairy products are consumed together with other beverages or foods.

We therefore investigated the relationship between rs4988235 and dietary intake of whole foods other than dairy, using linear regression in manner identical to that for dairy products. We did not correct for multiple testing in the analysis of these closely related dietary exposures.

rs4988235 and Risk of Diabetes and MR Analysis

First, we investigated whether rs4988235 was related to HbA_{1c} levels in the subcohort through linear regression, adjusting for genotyping platform, sex, age, first three genetic PCs, and study center. The relationship between rs4988235 and risk of incident diabetes was examined per country through a Prentice-weighted Cox regression (30), using age as underlying time scale, with adjustment for genotyping platform, sex, age, the first three genetic PCs, and study center. Country-specific results were pooled with inverse variance weights in a random effects meta-analysis using restricted maximum likelihood estimation.

We then performed the MR analysis, a two-stage least squares IV analysis. LP-associated dairy product intake for each participant was calculated by using rs4988235, genotyping platform, sex, age, the first three genetic PCs, and study center as predictors in a linear regression model. Next, we investigated the association between predicted dairy product intake and diabetes in a Prentice-weighted Cox regression per country and subsequently pooled results. We investigated associations per 15 g genetically predicted dairy intake to obtain interpretable 95% CIs.

Because a regular two-stage least squares IV analysis does not take variance of the gene-exposure association into account and this could influence our findings (31), we additionally performed the MR analysis among 10,000 bootstrap samples and obtained a 95% CI through the percentile method. Also, we repeated MR analysis under the assumption of dominance of LP, after exclusion of noncase subjects with an HbA_{1c} >6.5% (48 mmol/mol) at baseline,

in participants from countries where rs4988235 was in HWE and in participants genotyped on the Illumina660W-Quad BeadChip. All analyses were performed in R, version 3.4.1 (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

The subcohort consisted of 12,722 participants with a mean (SD) age of 52 (9) years; 61.6% was female. In EPIC-InterAct, median milk intake was 162 g/day (p25–p75 27–300), and nonconsumption of milk was rare (7.1%) (Table 1). Milk intake differed between countries, ranging from 15 g/day (0–150) in France to 200 g/day (107–307) in the U.K. Nonconsumption of milk ranged from 47.6% in France to 0% in Denmark (Supplementary Tables 2–5).

Prevalence of homozygous LP (rs4988235 T/T genotype) differed within Europe, with a range from 7.4% in Italy to 53.9% in Sweden (Supplementary Table 6). LP SNP genotypes were in HWE in the total subcohort, but deviation from HWE at a $P < 0.05$ significance level was observed in Italy, Spain, the U.K., Germany, and Denmark (Supplementary Table 6). Baseline characteristics of the subcohort by rs4988235 genotype showed that with increasing number of LP (T) alleles, participants were older, had a higher systolic blood pressure, and were more often physically active and highly educated (Supplementary Table 7). Also, consumption of dairy products, potatoes, margarine, sugar, nonalcoholic beverages, and coffee was higher, whereas intake of pasta/rice, cereal products, fruit, vegetables,

Table 1—Baseline characteristics and dairy intake of the subcohort

	Subcohort	Missing
<i>N</i> (persons)	12,722	
Baseline characteristics		
Age (years)	52 ± 9	
Male sex	38.4	
Current smoker	26.2	
Former smoker	26.7	
Hypertension	18.4	2.4
Systolic blood pressure (mmHg)	133 ± 20	21.2
Diastolic blood pressure (mmHg)	82 ± 11	21.2
Total cholesterol (mmol/L)	5.9 ± 1.1	4.3
HDL cholesterol (mmol/L)	1.5 ± 0.4	4.3
LDL cholesterol (mmol/L)	3.8 ± 1.0	5.5
Triglycerides (mmol/L)	1.1 (0.8–1.6)	4.3
Lipoprotein a (mg/L)	384 (200–684)	5.6
Physically inactive	22.6	1.2
BMI (kg/m ²)	26.0 ± 4.2	0.5
WHR	0.85 ± 0.09	7.5
HbA _{1c} >6.5% (>48 mmol/mol)	1.4	1.6
Low level of education†	40.6	1.6
Premenopausal status‡	41.6	
History of myocardial infarction	1.4	1.6
History of stroke	0.9	8.6
Dairy product intake		
Energy (kcal)	2,056 (1,675–2,517)	No intake
Milk (g/day)	162 (37–300)	7.1
Milk beverages (g/day)	1 (0–7)	72.2
Milk for coffee and creamers (g/day)	0 (0–14)	89.5
Dairy creams (g/day)	1 (0–4)	36.2
Milk-based puddings (g/day)	3 (0–15)	54.0
Curd (g/day)	0 (0–7)	61.3
Yogurt, thick fermented milk (g/day)	28 (0–100)	25.9
Ice cream (g/day)	3 (0–9)	23.9
Cheese (g/day)	28 (15–50)	5.1
Fatty acid measurements		
C15:0 (mol%)	0.21 ± 0.07	0.8
C17:0 (mol%)	0.40 ± 0.09	0.8

Data are expressed as mean ± SD, median (p25–p75), or percentage of participants with available data for variable. †No education or only primary school education. ‡Percentage among women.

legumes, and vegetable oils was lower (Supplementary Table 8).

rs4988235 and Intake of Dairy Products

After adjustment for sex, age, PCs, study center, and genotyping platform, one additional LP allele was associated with a higher intake of milk (β 17.1 g/day, 95% CI 10.6–23.6) and milk beverages (β 2.8 g/day, 95% CI 1.0–4.5) but not with intake of other dairy products (Table 2). *F* statistic of the model predicting milk intake was 74.0, and predicting a composite end point of milk and milk beverage intake decreased the *F* statistic to 64.5.

rs4988235 was associated with higher plasma C15:0 (β 0.003% of total fats, 95% CI 0.0005–0.005). A similar association was suggested for C17:0 levels (β 0.002% of total fats, 95% CI –0.0007 to 0.005), but this was not statistically significant.

rs4988235 and Diabetes Risk Factors

rs4988235 was associated with a modestly lower HDL cholesterol (β –0.017, 95% CI –0.029 to –0.004) (Table 3). After multivariable adjustment, dietary intakes associated with rs4988235 were potatoes (β 3.0, 95% CI 0.9–5.1), cereal (β –3.5, 95% CI –6.7 to –0.3), fruits (β –7.0, 95% CI –12.4 to –1.7), poultry

(β –0.8, 95% CI –1.4 to –0.2), nonalcoholic beverages (β –18.0, 95% CI –34.4 to –1.6) and wine (β –4.8, 95% CI –9.1 to –0.6) (Table 2). rs4988235 was not associated with energy intake.

rs4988235 and Risk of Diabetes and MR Analysis

rs4988235 was not associated with baseline HbA_{1c} levels (β –0.07%, 95% CI –0.22 to 0.07; *n* = 12,519) or with risk of developing diabetes (hazard ratio [HR]_{per additional LP allele} 0.99, 95% CI 0.94–1.04). MR analysis suggested no association between milk intake and diabetes (HR 0.99_{per 15 g/day}, 95% CI 0.93–1.05; *I*² = 44%) (Table 4 and Supplementary Fig. 2).

Results did not differ when repeating the IV analysis among 10,000 bootstrap samples (HR_{per 15 g/day} 0.99, 95% CI 0.94–1.04), and assuming a dominant effect of LP did not alter conclusions (HR_{per 15 g/day} 0.99, 95% CI 0.92–1.07). Restriction of analysis to participants genotyped on the Illumina660W-Quad BeadChip did not change conclusions; exclusion of 73 participants with an HbA_{1c} >6.5% (48 mmol/mol) or limiting analysis to countries in HWE did not change conclusions either (Supplementary Table 9).

CONCLUSIONS

In this European case-cohort study including 9,686 incident type 2 diabetes cases, we did not find evidence for a causal relationship between milk intake and diabetes.

The LP SNP rs4988235 was associated with milk intake but not with intake of other dairy products after multivariable adjustment. rs4988235 was associated with a slightly lower HDL cholesterol, a higher intake of potatoes, and a lower intake of cereal, fruit, poultry, wine, and nonalcoholic beverages.

The null association that we observed between milk intake and diabetes is in line with observational studies (2), including from EPIC-InterAct (32), and studies investigating the association between variation in the *MCM6* gene and diabetes (8,12,14,33,34).

Our results are consistent with findings from an MR study in a Danish cohort (odds ratio [OR]_{per 250 g/week} 0.99, 95% CI 0.93–1.06 [assuming LP dominant effect] [1,355 cases]) (8), and a two-sample MR study using end point data from DIAGRAM (DIAbetes Genetics Replication

Table 2—Association after multivariable adjustment between rs4988235 and dietary intake among 12,722 subcohort participants

	β^*	95% CI	<i>P</i>	<i>n</i>
Energy (kcal/day)	5.2	–12.6 to 23.0	0.57	12,722
Milk (g/day)	17.1	10.6–23.6	2 * 10 ^{–7}	12,722
Nonmilk dairy (g/day)	3.8	0.5–7.1	0.02	12,722
Milk beverages (g/day)	2.8	1.0–4.5	2 * 10 ^{–3}	6,867
Milk for coffee (g/day)	0.10	–0.72 to 0.91	0.82	2,959
Dairy creams (g/day)	0.06	–0.13 to 0.25	0.53	11,536
Cream desserts (g/day)	0.12	–0.70 to 0.94	0.77	10,372
Curd (g/day)	–0.15	–0.80 to 0.51	0.66	10,372
Yogurt (g/day)	2.2	–0.5 to 4.9	0.11	12,722
Ice cream (g/day)	–0.04	–0.41 to 0.33	0.82	12,722
Cheese (g/day)	–0.09	–1.05 to 0.87	0.85	12,722
Potatoes (g/day)	3.0	0.9–5.1	5 * 10 ^{–3}	12,722
Pasta/rice (g/day)	–0.2	–1.8 to 1.3	0.78	12,722
Bread (g/day)	–1.4	–3.7 to 0.9	0.22	12,722
Cereal (g/day)	–3.5	–6.7 to –0.3	0.03	12,722
Vegetables (g/day)	–0.3	–3.6 to 3.0	0.86	12,722
Legumes (g/day)	–0.14	–0.74 to 0.46	0.65	12,722
Fruits (g/day)	–7.0	–12.4 to –1.7	0.01	12,722
Nuts (g/day)	–0.05	–0.29 to 0.20	0.72	12,722
Red meat (g/day)	–0.01	–0.96 to 0.95	0.99	12,722
Poultry (g/day)	–0.8	–1.4 to –0.2	6 * 10 ^{–3}	12,722
Processed meat (g/day)	0.01	–0.92 to 0.93	0.99	12,722
Vegetable oils (g/day)	0.07	–0.17 to 0.31	0.57	12,722
Margarine (g/day)	0.3	–0.1 to 0.8	0.16	12,722
Butter (g/day)	0.13	–0.12 to 0.38	0.32	12,722
Sugar (g/day)	0.9	–0.7 to 2.6	0.25	12,722
Cake/cookies (g/day)	–1.3	–2.8 to 0.1	0.06	12,722
Beverages (g/day) [†]	–18.0	–34.3 to –1.6	0.03	12,722
Soft drinks (g/day)	–0.3	–5.3 to 4.6	0.90	12,722
Juice (g/day)	–0.2	–3.6 to 3.3	0.93	12,722
Coffee (g/day)	4.4	–5.7 to 14.5	0.39	12,722
Tea (g/day)	–6.5	–14.2 to 1.2	0.10	12,722
Alcohol (g/day)	–0.3	–0.8 to 0.3	0.34	12,722
Wine (g/day)	–4.8	–9.1 to –0.6	0.03	12,722

* β derived from linear regression model, adjusted for first three genetic PCs, study center, genotyping platform, sex, and age. [†]Sum of all nonalcoholic beverages, excluding water and milk.

And Meta-analysis) ($OR_{per\ 66\ g/day}$ 0.92, 95% CI 0.83, 1.02 [assuming LP additive effect] [26,488 cases]) (14).

These findings combined support the notion that the inverse association between milk intake and diabetes as seen in some studies (2) is likely caused by residual confounding.

Strengths of this study include the large population-based case cohort from different regions within Europe with a broad range of LP prevalence. We confirmed the gene-exposure (rs4988235-milk) relationship with odd-chain fatty acids measurements and investigated the association between the LP SNP and a wide range of dietary intakes. We showed robustness of our findings in sensitivity analyses.

Despite its strengths, there are study limitations worth noting. First, 19% of EPIC-InterAct participants were excluded from our analyses owing to missing genetic information. However, baseline characteristics (17) and the proportion of diabetes cases did not differ substantially between participants with and participants without genotype information, so it is unlikely that this has led to selection bias.

A second limitation is that intake of dairy products was determined with a diet questionnaire, which would mainly lead to nondifferential measurement error for this analysis. However, local validation studies have shown a reasonable to high relative validity of the diet questionnaire compared with 24-h recalls for milk and milk products and a fair to reasonable validity for cheese (35,36).

Also, we relied on imputed data to ascertain a SNP loading for rs4988235 with suboptimal imputation info scores (18) for participants genotyped to the Illumina HumanCoreExome chips, which might lead to lower IV strength due to misclassification. However, restriction of analysis to participants genotyped to the Illumina660W-Quad BeadChip did not change results. We also observed deviation from HWE ($P < 0.05$) in most countries, while there was no deviation from HWE in the total cohort. Testing for deviation from HWE at $P < 0.05$ has been proposed to assess whether ascertainment bias may be present (37), which would occur if LP-associated characteristics determine study inclusion. This form of selection bias is unlikely, since

EPIC-InterAct is comprised of population-based cohorts (17) and we did not find evidence of an association between rs4988235 and early death in our GWAS search (Supplementary Table 1). Also, deviation from HWE may be explained by genetic mixture of subpopulations with different allele frequencies (37), as is the case for rs4988235 in EPIC-InterAct. If this is the case, appropriate correction for population substructure prevents bias. In EPIC-InterAct, participant characteristics and dietary intake by rs4988235 genotype differed substantially (Supplementary Tables 7 and 8), but many of these associations were not found after adjustments for study center and genetic PCs (Tables 2 and 3). Also, a sensitivity analysis in countries without deviation from HWE did not alter conclusions.

Owing to aforementioned limitations of this study, we cannot exclude the possibility that a small effect of milk intake on diabetes risk is present, despite our null finding.

In EPIC-InterAct, LP due to rs4988235 was associated with a higher milk intake and lower intake of other nonalcoholic beverages. We also observed modest associations between the LP SNP and higher intake of potatoes and a lower intake of cereal, fruit, poultry, and wine. We found no association between rs4988235 and total energy intake. Our results suggest that LP may be associated with a composite of

nutritional factors, rather than only with milk intake, although the association of the LP SNP with milk is much stronger than with other dietary products. These associations should be interpreted cautiously, as they may be due to chance, since we did not correct for multiple testing and we cannot exclude presence of residual population stratification. Repeating these analyses in other cohorts is required before drawing firm conclusions about a potential LP-associated dietary pattern.

We also observed that all Danish participants consumed milk, including those who were genetically unable to produce lactase, whereas ~40% of the French LP population did not consume milk. An explanation for this could be that there are cultural habits that lead to milk consumption in lactase non-persistent people or milk avoidance in LP people, as hypothesized previously (8,38).

Given the known heterogeneity of gene-exposure association for LP (8–10,12), we propose, in line with others (38), to demonstrate the association between an LP SNP and milk intake before using this IV in an MR study. To attribute the association of an LP SNP with disease to the correct exposure, one should know the association between the LP SNP and possible replacements for milk, total energy intake, and other (dairy) foods in the population from which gene-outcome data were derived. Otherwise, causal inference

Table 3—Association between rs4988235 and diabetes risk factors among 12,722 subcohort participants

	Estimate*	95% CI	P	n
BMI (kg/m ²)	0.03	−0.09 to 0.16	0.61	12,652
WHR	0.0017	−0.0003 to 0.0038	0.10	11,771
Systolic blood pressure (mmHg)	0.41	−0.23 to 1.06	0.21	10,027
Diastolic blood pressure (mmHg)	0.20	−0.18 to 0.58	0.30	10,026
Presence of hypertension	0.02	−0.06 to 0.10	0.63	12,677
Total cholesterol (mmol/L)	−0.014	−0.048 to 0.021	0.44	12,171
HDL cholesterol (mmol/L)	−0.017	−0.029 to −0.004	0.01	12,173
LDL cholesterol (mmol/L)	0.001	−0.031 to 0.032	0.97	12,028
Triglycerides (mmol/L)†	0.009	−0.007 to 0.025	0.28	12,172
Lipoprotein a (mg/L)†	−0.01	−0.05 to 0.03	0.65	12,014
Current or former smoker	1.06	0.99–1.13	0.09	12,722
Physical inactivity	0.99	0.91–1.07	0.81	12,722
Low level of education	1.03	0.95–1.11	0.47	12,722
Peri- or postmenopausal status	1.00	0.85–1.17	0.97	7,841

*Estimate is OR derived from logistic regression model (for smoking, physical inactivity, low level of education, and post- or perimenopausal status) or β derived from linear regression (all other variables), adjusted for first three genetic PCs, study center, genotyping platform, sex, and age.
†Log transformed.

Table 4—Association between genetic LP and diabetes (gene-outcome) and association between genetically predicted milk intake and diabetes among 21,820 EPIC-InterAct participants

	HR*	95% CI	n
Gene-outcome	0.99	0.94–1.04	21,820
Genetically predicted milk intake (per 15 g)	0.99	0.93–1.05	21,820

*HR for diabetes in gene-outcome is expressed per additional LP allele. HR for diabetes in IV analyses is expressed per 15 g of genetically predicted milk intake. Analyses are performed per country, using age as underlying time scale, and are adjusted for sex, genetic variability (first three PCs), study center, and genotyping platform. The reported overall estimates were obtained by pooling country-specific results in random effects meta-analysis.

on the exposure level may lead to an incorrect conclusion (39).

Providing additional information on the LP-associated exposure also facilitates comparison between studies. In our GWAS search, we found that each additional effect (T) allele of rs4988235 was associated with lower LDL and total cholesterol (22), which we did not observe in EPIC-InterAct. We cannot exclude that this is due to lower power in our analysis.

In conclusion, rs4988235 was associated with intake of milk and milk beverages, but not with intake of other dairy products in EPIC-InterAct. The MR analysis provided no evidence for a causal relationship between milk intake and type 2 diabetes, which is in line with previous genetic and observational studies. We consider it unlikely that modest associations between LP and other dietary intakes have caused this null result. No conclusion can be drawn regarding causality of the relationship of dairy products other than milk with diabetes.

Acknowledgments. The authors thank all EPIC participants and staff for their contribution to the study. The authors thank Nicola Kerrison (MRC Epidemiology Unit) for managing the data for the InterAct project. The authors thank staff from the Technical, Field Epidemiology, and Data Functional Group Teams of the MRC Epidemiology Unit in for carrying out sample preparation, DNA provision and quality control, genotyping, and data-handling work. The authors specifically thank S. Dawson for coordinating the sample provision for biomarker measurements; A. Britten for coordinating DNA sample provision and genotyping of candidate markers; N. Kerrison, C. Gillson, and A. Britten for data provision and genotyping quality control; and M. Sims for writing the technical laboratory specification for the intermediate pathway biomarker measurements and for overseeing the laboratory work.

Funding. Funding for the InterAct project was provided by the European Union Sixth

Framework Programme (grant number LSHM_CT_2006_037197). In addition, InterAct investigators acknowledge funding from the following agencies: for N.G.F., F.I., C.L., and N.J.W., MRC Epidemiology Unit core support (MC_UU_12015/5 and MC_UU_12015/1), and for N.G.F. and N.J.W., National Institute for Health Research Cambridge Biomedical Research Centre (IS-BRC-1215-20014). Verification of diabetes cases was additionally funded by NL Agency grant IGE05012 and an Incentive Grant from the Board of University Medical Center Utrecht (the Netherlands) (to I.S. and Y.T.v.d.S.). P.W.F. acknowledges funding from the Swedish Research Council, the Swedish Heart Lung Foundation, and the Swedish Diabetes Association. J.R.Q. acknowledges funding from the Asturias Regional Government. R.K. acknowledges funding from Deutsche Krebshilfe. T.Ke. acknowledges funding from the U.K. Medical Research Council (MR/M012190/1) and the Wellcome Trust Our Planet, Our Health (Livestock, Environment and People [LEAP 205212/Z/16/Z]). S.P. acknowledges funding from Associazione Italiana per la Ricerca sul Cancro. A.M.W.S. acknowledges funding from the Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund, and Statistics Netherlands (the Netherlands). K.O. and A.T. acknowledge funding from the Danish Cancer Society. R.T. acknowledges funding from AIRE-ONLUS Ragusa, AVIS-Ragusa, and the Sicilian Regional Government.

Duality of Interest. P.W.F. acknowledges funding from Novo Nordisk. No other potential conflicts of interest relevant to this article were reported.

Author Contributions. L.E.T.V. analyzed data and drafted the manuscript. L.E.T.V. and I.S. take responsibility for the manuscript contents. L.E.T.V., I.S., and Y.T.v.d.S. had access to all data for this study. L.E.T.V., I.S., Y.T.v.d.S., N.G.F., F.I., S.B., A.B., H.B., C.B., M.-D.C., G.F., P.W.F., H.F., M.J.G., J.R.Q., D.B.I., R.K., T.Ke., K.T.K., T.Kü., O.M., P.M.N., K.O., V.P., D.P., S.P., C.S., A.M.W.S., A.T., R.T., M.R.-B., O.R., E.R., S.J.S., C.L., and N.J.W. contributed to study conception and design and interpretation of data, revising the article critically for important intellectual content and final approval of the version to be published. L.E.T.V. and I.S. are the guarantors of this work and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Prior Presentation. Parts of this study were presented in abstract form at the Mendelian Randomization Conference, Bristol, U.K., 11–13 July 2017, and the European Society of Cardiology Congress, Barcelona, Spain, 26–30 August 2017.

References

1. Franz MJ, Bantle JP, Beebe CA, et al. Evidence-based nutrition principles and recommendations for the treatment and prevention of diabetes and related complications. *Diabetes Care* 2002;25:148–198
2. Gijsbers L, Ding EL, Malik VS, de Goede J, Geleijnse JM, Soedamah-Muthu SS. Consumption of dairy foods and diabetes incidence: a dose-response meta-analysis of observational studies. *Am J Clin Nutr* 2016;103:1111–1124
3. Rice BH, Cifelli CJ, Pikosky MA, Miller GD. Dairy components and risk factors for cardiometabolic syndrome: recent evidence and opportunities for future research. *Adv Nutr* 2011;2:396–407
4. Thorning TK, Bertram HC, Bonjour JP, et al. Whole dairy matrix or single nutrients in assessment of health effects: current evidence and knowledge gaps. *Am J Clin Nutr* 2017;105:1033–1045
5. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014;23:R89–R98
6. Itan Y, Jones BL, Ingram CJ, Swallow DM, Thomas MG. A worldwide correlation of lactase persistence phenotype and genotypes. *BMC Evol Biol* 2010;10:36
7. Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, Järvelä I. Identification of a variant associated with adult-type hypolactasia. *Nat Genet* 2002;30:233–237
8. Bergholdt HK, Nordestgaard BG, Ellervik C. Milk intake is not associated with low risk of diabetes or overweight-obesity: a Mendelian randomization study in 97,811 Danish individuals. *Am J Clin Nutr* 2015;102:487–496
9. Tornaiainen S, Hedelin M, Autio V, et al. Lactase persistence, dietary intake of milk, and the risk for prostate cancer in Sweden and Finland. *Cancer Epidemiol Biomarkers Prev* 2007;16:956–961
10. Smith GD, Lawlor DA, Timpson NJ, et al. Lactase persistence-related genetic variant: population substructure and health outcomes. *Eur J Hum Genet* 2009;17:357–367
11. Travis RC, Appleby PN, Siddiq A, et al. Genetic variation in the lactase gene, dairy product intake and risk for prostate cancer in the European prospective investigation into cancer and nutrition. *Int J Cancer* 2013;132:1901–1910
12. Lamri A, Poli A, Emery N, et al. The lactase persistence genotype is associated with body mass index and dairy consumption in the D.E.S.I. R. study. *Metabolism* 2013;62:1323–1329
13. Corella D, Arregui M, Coltell O, et al. Association of the LCT-13910C>T polymorphism with obesity and its modulation by dairy products in a Mediterranean population. *Obesity (Silver Spring)* 2011;19:1707–1714
14. Yang Q, Lin SL, Au Yeung SL, et al. Genetically predicted milk consumption and bone health, ischemic heart disease and type 2 diabetes: a Mendelian randomization study. *Eur J Clin Nutr* 2017;71:1008–1012

15. Hjärtåker A, Lagiou A, Slimani N, et al. Consumption of dairy products in the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort: data from 35 955 24-hour dietary recalls in 10 European countries. *Public Health Nutr* 2002;5:1259–1271
16. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904–909
17. Langenberg C, Sharp SJ, Franks PW, et al. Gene-lifestyle interaction and type 2 diabetes: the EPIC interact case-cohort study. *PLoS Med* 2014;11:e1001647
18. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet* 2010;11:499–511
19. Staley JR, Blackshaw J, Kamat MA, et al. PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics* 2016;32:3207–3209
20. Shungin D, Winkler TW, Croteau-Chonka DC, et al.; ADIPOGen Consortium; CARDIOGRAM-plusC4D Consortium; CKDGen Consortium; GE-FOS Consortium; GENIE Consortium; GLGC; ICBP; International Endogene Consortium; LifeLines Cohort Study; MAGIC Investigators; MuTHER Consortium; PAGE Consortium; ReproGen Consortium. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* 2015;518:187–196
21. Wood AR, Esko T, Yang J, et al.; Electronic Medical Records and Genomics (eMEMERGE) Consortium; MIGen Consortium; PAGE Consortium; LifeLines Cohort Study. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* 2014;46:1173–1186
22. Willer CJ, Schmidt EM, Sengupta S, et al.; Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. *Nat Genet* 2013;45:1274–1283
23. Riboli E, Hunt KJ, Slimani N, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr* 2002;5:1113–1124
24. Kroke A, Klipstein-Grobusch K, Voss S, et al. Validation of a self-administered food-frequency questionnaire administered in the European Prospective Investigation into Cancer and Nutrition (EPIC) Study: comparison of energy, protein, and macronutrient intakes estimated with the doubly labeled water, urinary nitrogen, and repeated 24-h dietary recall methods. *Am J Clin Nutr* 1999;70:439–447
25. Margetts BM, Pietinen P. European Prospective Investigation into Cancer and Nutrition: validity studies on dietary assessment methods. *Int J Epidemiol* 1997;26(Suppl. 1):S1–S5
26. Forouhi NG, Koulman A, Sharp SJ, et al. Differences in the prospective association between individual plasma phospholipid saturated fatty acids and incident type 2 diabetes: the EPIC-InterAct case-cohort study. *Lancet Diabetes Endocrinol* 2014;2:810–818
27. Wareham NJ, Jakes RW, Rennie KL, et al. Validity and repeatability of a simple index derived from the short physical activity questionnaire used in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *Public Health Nutr* 2003;6:407–413
28. Burgess S, Thompson SG; CRP CHD Genetics Collaboration. Avoiding bias from weak instruments in Mendelian randomization studies. *Int J Epidemiol* 2011;40:755–764
29. Risérus U, Marklund M. Milk fat biomarkers and cardiometabolic disease. *Curr Opin Lipidol* 2017;28:46–51
30. Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* 1986;73:1–11
31. Boef AG, Dekkers OM, le Cessie S. Mendelian randomization studies: a review of the approaches used and the quality of reporting. *Int J Epidemiol* 2015;44:496–511
32. Sluijs I, Forouhi NG, Beulens JW, et al.; InterAct Consortium. The amount and type of dairy product intake and incident type 2 diabetes: results from the EPIC-InterAct Study. *Am J Clin Nutr* 2012;96:382–390
33. Mahajan A, Go MJ, Zhang W, et al.; DIAbetes Genetics Replication And Meta-Analysis (DIAGRAM) Consortium; Asian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium; South Asian Type 2 Diabetes (SAT2D) Consortium; Mexican American Type 2 Diabetes (MAT2D) Consortium; Type 2 Diabetes Genetic Exploration by Nex-Generation Sequencing in Muylti-Ethnic Samples (T2D-GENES) Consortium. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet* 2014;46:234–244
34. Enattah NS, Forsblom C, Rasinperä H, Tuomi T, Groop PH, Järvelä I; FinnDiane Study Group. The genetic variant of lactase persistence C (-13910) T as a risk factor for type I and II diabetes in the Finnish population. *Eur J Clin Nutr* 2004;58:1319–1322
35. Ocké MC, Bueno-de-Mesquita HB, Goddijn HE, et al. The Dutch EPIC food frequency questionnaire. I. Description of the questionnaire, and relative validity and reproducibility for food groups. *Int J Epidemiol* 1997;26(Suppl. 1):S37–S48
36. Pisani P, Faggiano F, Krogh V, Palli D, Vineis P, Berrino F. Relative validity and reproducibility of a food frequency dietary questionnaire for use in the Italian EPIC centres. *Int J Epidemiol* 1997;26(Suppl. 1):S152–S160
37. Rodriguez S, Gaunt TR, Day IN. Hardy-Weinberg equilibrium testing of biological ascertainment for Mendelian randomization studies. *Am J Epidemiol* 2009;169:505–514
38. Smith GD, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. *Int J Epidemiol* 2004;33:30–42
39. Holmes MV, Ala-Korpela M, Smith GD. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. *Nat Rev Cardiol* 2017;14:577–590