

Genetic variation in the *lactase* gene, dairy product intake and risk for prostate cancer in the European Prospective Investigation into Cancer and Nutrition

Ruth C. Travis¹, Paul N. Appleby¹, Afshan Siddiq², Naomi E. Allen¹, Rudolf Kaaks³, Federico Canzian³, Silke Feller⁴, Anne Tjønneland⁵, Nina Føns Johnsen⁵, Kim Overvad⁶, J. Ramón Quirós⁷, Carlos A. González⁸, Maria-José Sánchez⁹, Nerea Larrañaga¹⁰, Maria-Dolores Chirlaque¹¹, Aurelio Barricarte¹², Kay-Tee Khaw¹³, Nick Wareham¹⁴, Antonia Trichopoulou^{15,16}, Elisavet Valanou¹⁶, Erifili Oustoglou¹⁶, Domenico Palli¹⁷, Sabina Sieri¹⁸, Rosario Tumino¹⁹, Carlotta Sacerdote²⁰, H. B(as). Bueno-de-Mesquita^{21,22}, Pär Stattin²³, Pietro Ferrari²⁴, Mattias Johansson^{25,26}, Teresa Norat², Elio Riboli² and Timothy J Key¹

¹Cancer Epidemiology Unit, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, United Kingdom

²Department of Epidemiology and Public Health, Imperial College, London, United Kingdom

³Division of Cancer Epidemiology, German Cancer Research Centre, Heidelberg, Germany

⁴Department of Epidemiology, German Institute of Human Nutrition, Potsdam-Rehbrücke, Nuthetal, Germany

⁵Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen, Denmark

⁶Department of Cardiology, Center for Cardiovascular Research, Aalborg Hospital, Aarhus University Hospital and Department of Epidemiology, School of Public Health, Aarhus University, Aarhus, Denmark

⁷Public Health Directorate, Asturias, Spain

⁸Unit of Nutrition, Environment and Cancer, Cancer Epidemiology Research Programme, Catalan Institute of Oncology, Barcelona, Spain

⁹Andalusian School of Public Health, Granada (Spain) and CIBER de Epidemiología y Salud Pública (CIBERESP), Spain

¹⁰Public Health Division of Gipuzkoa, Basque Regional Health Department and Ciberesp-Biodonostia, San Sebastian, Spain

¹¹Epidemiology Department, Murcia Health Council, CIBER en Epidemiología y Salud, Pública (CIBERESP), Murcia, Spain

¹²Navarre Public Health Institute, Pamplona, Spain and Consortium for Biomedical Research in Epidemiology and Public Health (CIBER Epidemiología y Salud Pública-CIBERESP), Spain

¹³University of Cambridge, United Kingdom

¹⁴MRC Epidemiology Unit, Cambridge, United Kingdom

¹⁵WHO Collaborating Center for Food and Nutrition Policies, Department of Hygiene, Epidemiology and Medical Statistics, University of Athens Medical School, Athens, Greece

¹⁶Hellenic Health Foundation, Athens, Greece

¹⁷Molecular and Nutritional Epidemiology Unit, Cancer Research and Prevention Institute, ISPO, Florence Italy

¹⁸Nutritional Epidemiology Unit, Fondazione IRCCS Istituto Nazionale Tumori, Milan, Italy

¹⁹Cancer Registry and Histopathology Unit, "Civile - M.P.Arezzo" Hospital, ASP 7 Ragusa, Italy

²⁰Center for Cancer Prevention (CPO-Piemonte), and Human Genetic Foundation (HuGeF), Torino, Italy

²¹National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands

²²Department of Gastroenterology and Hepatology, University Medical Centre, Utrecht, The Netherlands

²³Department of Public Health and Clinical Medicine, Nutritional Research, Umeå University, Umeå, Sweden/ Department of Surgical and Perioperative Sciences, Urology and Andrology, Umeå University, Umeå, Sweden

²⁴Nutrition and Metabolism Section, International Agency for Research on Cancer, The World Health Organization, Lyon, France

²⁵International Agency for Research on Cancer (IARC), Lyon, France

²⁶Department of Surgical and Perioperative Sciences, Urology and Andrology, Umeå University, Umeå, Sweden

Key words: prostate cancer, polymorphism, genetic, lactase, dairy

Grant sponsors: Cancer Research UK, European Commission: Public Health and Consumer Protection Directorate 1993-2004; Research Directorate-General 2005; German Cancer Aid; German Cancer Research Center; German Federal Ministry of Education and Research; Danish Cancer Society; Health Research Fund (FIS) of the Spanish Ministry of Health; ISCIII RETICC RD 06/0020, Spain; Spanish Regional Governments of Andalusia, Asturias, Basque country, Murcia (no. 6236) and Navarra, and Catalan Institute of Oncology; Medical Research Council, UK; the Stroke Association, UK; British Heart Foundation; Department of Health, UK; Food Standards Agency, UK; the Wellcome Trust, UK; the Hellenic Health Foundation; the Stavros Niarchos Foundation; the Hellenic Ministry of Health and Social Solidarity; Italian Association for Research on Cancer (AIRC); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), Statistics Netherlands (The Netherlands).

DOI: 10.1002/ijc.27836

History: Received 6 Feb 2012; Accepted 4 Jul 2012; Online 11 Sep 2012

Correspondence to: Ruth C. Travis, Cancer Epidemiology Unit, Nuffield Department of Clinical Medicine, University of Oxford, Oxford OX3 7LF, United Kingdom, Tel.: +44-0-1865-289600, Fax: +44-0-1865-289610, E-mail: ruth.travis@ceu.ox.ac.uk

High dairy protein intake has been found to be associated with increased prostate cancer risk in the European Prospective Investigation into Cancer and Nutrition (EPIC). To further examine this possible relationship, we investigated the hypothesis that a genetic polymorphism in the *lactase* (*LCT*) gene might be associated with elevated dairy product intake and increased prostate cancer risk in a case-control study nested in EPIC. The C/T-13910 *lactase* variant (rs4988235) was genotyped in 630 men with prostate cancer and 873 matched control participants. Dairy product consumption was assessed by diet questionnaire. Odds ratios (ORs) for prostate cancer in relation to *lactase* genotype were estimated by conditional logistic regression. *Lactase* genotype frequency varied significantly between countries, with frequencies of the T (lactase persistence) allele ranging from 7% in Greece to 79% in Denmark. Intake of milk and total dairy products varied significantly by *lactase* genotype after adjustment for recruitment center; adjusted mean intakes of milk were 44.4, 69.8 and 82.3 g/day among men with CC, CT and TT genotypes, respectively. The *lactase* variant was not significantly associated with prostate cancer risk, both in our data (adjusted OR for TT vs. CC homozygotes: 1.10, 95% CI: 0.76–1.59) and in a meta-analysis of all the published data (combined OR for T allele carriers vs. CC homozygotes: 1.12, 0.96–1.32). These findings show that while variation in the *lactase* gene is associated with milk intake in men, the *lactase* polymorphism does not have a large effect on prostate cancer risk.

What's new?

High dairy protein intake has previously been found to be associated with increased prostate cancer risk in the European Prospective Investigation into Cancer and Nutrition (EPIC). The current study was nested in EPIC, and results from this first Europe-wide study suggest that while the C/T13910 *lactase* polymorphism is associated with milk intake, the variant has no large effect on prostate cancer risk. The data illustrate the challenges of applying mendelian randomisation to explore the relationship between dairy product consumption and cancer risk. Very large studies with both genetic and dietary data are thus needed for investigations using genetic proxies of nutritional exposures.

A high intake of dairy protein has been found to be associated with increased risk for prostate cancer in the European Prospective Investigation into Cancer and Nutrition,¹ although results from other prospective studies of dairy products have been inconsistent.^{2–9} The inconsistencies in the published literature may be in part due to measurement error, with estimates of dairy product intake based on self-reported information obtained by dietary questionnaire. It is also possible that the associations found with dairy products and prostate cancer risks are due to confounding by unknown dietary or lifestyle factors.

The main sugar in dairy products is lactose, and the digestion and absorption of lactose is dependent on the activity of the lactase enzyme (lactase-phlorizin hydrolase) in the wall of the intestine. A C/T polymorphism (rs4988235), which resides 13,910 base pairs upstream of the lactase coding sequence in the *lactase* gene, is the major variant associated with activity of the lactase enzyme in European populations.¹⁰ The ancestral C allele is associated with the developmental downregulation of the lactase enzyme after the first few years of life (also referred to as lactase nonpersistence) and with reduced tolerance of lactose-rich foods in adulthood.¹¹ The T variant is associated with continued lactase production in adulthood (lactase-persistence) and is thought to have arisen around 9000 years ago and become common in northern Europe, probably due to the selective advantage conferred by dairy consumption in the context of the spread of dairy farming.¹² To date, however, the pub-

lished epidemiological data on dairy product consumption in adulthood by *lactase* genotype is limited.^{13–15}

The existence of the *lactase* C/T polymorphism and the limited evidence suggesting a link with dairy product consumption provide an opportunity to explore the Mendelian randomization approach, using *lactase* genotype as a proxy for dairy intake and exploiting the random allocation of alleles from parents to offspring, to overcome potential confounding of the association between dairy product consumption and cancer risk.^{16–18} The proposed hypothesis is that individuals carrying the T allele (lactase-persistence variant) will have a higher intake of lactose-rich dairy products and an increased risk for prostate cancer.

We report here findings from a Europe-wide study of the C/T-13910 *lactase* genotypes in relation to intake of dairy products and risk for prostate cancer, among 630 men with incident prostate cancer and 873 matched control participants participating in the European Prospective Investigation into Cancer and Nutrition (EPIC).

Material and Methods

Participants and data

Between 1992 and 2000, approximately 500,000 individuals (150,000 men) were recruited into the European Prospective Investigation into Cancer and Nutrition (EPIC) from 23 centers in 10 European countries. The methods of recruitment and study design have been described in detail elsewhere.¹⁹ Participants completed an extensive questionnaire on dietary

and nondietary factors at recruitment, and about 400,000 individuals (of whom 137,000 were men) also provided a blood sample. All participants gave written consent, and approval for the study was obtained from the Internal Review Board of the International Agency for Research on Cancer (Lyon, France) and from the local ethics committees in participating countries.

Dietary intake during the year before enrolment was measured by country-specific validated food frequency questionnaires or diet histories, as previously described.¹⁹ For this analysis, dairy products included milk and milk beverages, yoghurt, fromage blanc, cheese, dairy puddings/desserts, dairy creams and butter. Intakes were calculated as g/day. Estimated daily nutrient intakes were calculated by multiplying the nutrient content of each food of a specific portion size by the frequency of consumption as stated on the dietary questionnaire using national food tables from each country as compiled in the EPIC Nutrient DataBase' (ENDB).^{20,21} Measurements of circulating concentrations of certain nutritional and hormonal analytes have been made in previous nested case control studies of prostate cancer in men from EPIC and were available for the current analyses.^{22–26}

This study includes prostate cancer cases occurring after blood collection and individually matched male control participants from the eight participating countries which recruited men: Denmark, Germany, Greece, Italy, the Netherlands, Spain, Sweden and the United Kingdom (UK).

Follow-up for diagnosis of prostate cancer is provided through record linkage with population-based cancer registries in six of the participating countries: Denmark, Italy, the Netherlands, Spain, Sweden and the UK. In Germany and Greece, follow-up is active and is achieved through checks of insurance records and cancer and pathology registries as well as *via* self-reported questionnaires; self-reported incident cancers are verified through medical records. Data on vital status in most EPIC study centers were collected from mortality registries at the regional or national level, in combination with data collected by active follow-up (Greece). The 10th Revision of the International Statistical Classification of Diseases, Injuries and Causes of Death (ICD) was used, and cancer of the prostate was defined as code C61. For each EPIC center closure dates of the study period were defined as the latest dates of complete follow-up for both cancer incidence and vital status (dates varied between centers, from June 1999 to January 2003).

Case patients were men who developed prostate cancer after the date of blood collection and before the end of the study period, defined for each study center by the latest date of follow-up. The cases with no available blood sample and those participants who had missing information on the date of the blood donation or who had a history of another cancer (except nonmelanoma skin cancer) at the time of the blood collection were excluded. After these exclusions, at the time of genotyping for the current study, DNA was available for approximately 600 men with prostate cancer who had had

their DNA extracted for the EPIC component of the Breast and Prostate Cancer Cohort Consortium.

Data on the stage and grade of disease at diagnosis were collected from each center, where possible. A total of 480 cases (76.7%) had information on tumor stage; of these 328 (52.6%) were classified as localized (tumor [T], node [N], metastasis [M] categories T0 or T1 or T2 and N0 or NX and M0, or stage coded in recruitment center as localized), and 152 (24.1%) were classified as advanced (T3 or T4, N1+, M1, or some combination of these, or stage coded in recruitment center as metastatic). Information on histological grade was available for 423 cases (65.2%); of these, 301 (47.8%) were classified as low-grade (Gleason sum <7 or equivalent, *i.e.* coded as moderately or as well differentiated) and 122 (17.4%) were classified as high-grade (Gleason sum \geq 7 or equivalent, *i.e.* coded as poorly differentiated or as undifferentiated).

Each case patient was matched to one control participant, with the exception of cases from Umeå in Sweden, an EPIC-associated cohort in which case patients were matched to two control participants, selected at random among appropriate risk sets consisting of all male cohort members alive and free of cancer (except nonmelanoma skin cancer) at the time of diagnosis of the index case. An incidence density sampling protocol for control selection was used, such that controls could include participants who became a case later in time, while each control participant could also be sampled more than once. Matching criteria included: recruitment center, age at enrolment (\pm 6 months), time of day of blood collection (\pm 1 hr) and time between blood draw and last consumption of food or drinks (<3, 3–6, >6 hr). While these matching criteria are not all necessary for analyses of genotype, they are applied to allow the same sets to be used for analyses of multiple markers of risk, such as nested case control analyses of circulating concentrations of hormonal and nutritional biomarkers.

Genotyping

Genotyping of the C/T-13910 *lactase* variant (rs4988235) was conducted using the TaqMan assay (Applied Biosystems) at Imperial College, London. The internal quality of genotype data was assessed by 2% blinded samples in duplicate and an intralaboratory concordance rate of greater than 97.2% was observed. Empty wells (containing water only) were also included on each plate. Hardy-Weinberg Equilibrium (HWE) checks have been performed among the controls, stratified by country. No deviation in HWE was observed ($p > 0.05$). The SNP was successfully genotyped in 97.8% of samples. In total, genotype data for the current analysis were available for 630 cases: 79 cases in Denmark, 179 in Germany, 9 in Greece, 50 in Italy, 22 in the Netherlands, 69 in Spain, 84 in Sweden and 138 in the UK.

Biomarker assays

Measurements of circulating concentrations of several nutritional and hormonal biomarkers made for previous nested

case control studies of prostate cancer in EPIC were available for a subset of participants in the current study. These analytes were phytanic acid, 25-hydroxyvitamin D, insulin-like growth factor 1 (IGF1), insulin-like growth factor binding protein 3 (IGFBP3), androstenedione, androstenediol glucuronide, testosterone, sex hormone binding globulin (SHBG), calculated free testosterone, enterodiol and enterolactone and details of the laboratory assays have been described elsewhere.^{22–26}

Statistical analysis

The baseline characteristics of participants by *LCT* genotype were compared using analyses of variance and multinomial logistic regression models for continuous and categorical variables, respectively, with adjustment for center of recruitment.

To investigate dairy product intake by the *lactase* C/T polymorphism, analyses of variance were conducted for *LCT* genotypes, with adjustment for recruitment center, in relation to consumption of milk, yoghurt, cheese, butter, dairy based deserts (including ice-cream), total dairy products, dairy protein, total fat, fat from dairy products, calcium intake and circulating concentrations of nutritional and hormonal analytes. Nonconsumers of dairy milk are defined as persons consuming <5 g/day milk and milk beverages.

Conditional logistic regression models were applied to calculate the relative risks (odds ratios) for prostate cancer in relation to the *lactase* C/T genotype. Likelihood ratio chi-square tests were used to test the main effect of genotype on risk (comparing models with genotype terms with a model with no genotype terms), and tests for trend (trend tests of genetic effect) were obtained by scoring *LCT* genotypes as 0, 1 and 2, respectively, according to the number of variant T alleles, and with homozygotes for the C allele as the reference group. The effects of potential confounders, other than matching criteria, which are controlled for by design, were examined by including additional regression terms in the logistic regression models. Potential confounders were chosen *a priori* as factors that may be associated with the risk of prostate cancer developing or being detected and were smoking (never, past, present), body mass index (BMI, kg/m²; in fourths), physical activity (inactive, moderately inactive, moderately active combined with active) (16), alcohol intake (<8, 8–15, 16–39, ≥40 g/day), marital status (married/cohabiting or not married/cohabiting) and education level (primary or none, secondary, degree level). For each of these variables a small proportion of values were unknown; these values were included in the analyses as a separate category.

For tests of heterogeneity in associations of the *lactase* C/T polymorphism with risk of prostate cancer across subgroups and by case characteristics, we used likelihood ratio tests to compare models with a cross-product term between the genotype trend test variable and subgroup membership with models without the cross-product term. Participants were divided into subgroups according to prostate tumor stage (localized or advanced); histologic grade (low grade or

high grade); time to diagnosis (less than 4 years after blood collection, 4 or more years after blood collection); age at blood collection (<60 or ≥60 years); and country of recruitment (8 countries). For tests of heterogeneity of risk by case characteristic (tumor stage, histological grade and time to diagnosis), the controls in each matched set were assigned the characteristics of their case.

To put the results of this study into the context of previous research, we conducted a meta-analysis of our results together with the results of previously published studies. Studies were identified by searching PubMed and the reference lists of relevant articles by using the search terms lactase and prostate cancer. Summary relative risks were estimated as the weighted average of the study-specific RRs for men with the TC or TT *lactase* genotypes compared with those with the CC genotype, with weights determined by the inverse of the variance of each relative risk.

Statistical analyses were performed with the Stata 10 statistical software package.²⁷ All tests of statistical significance were two-sided, and *p* values below 0.05 were considered significant.

Results

630 men diagnosed with prostate cancer from recruitment until the end of follow-up and 873 matched participants without prostate cancer were included in the analyses. Their median age at blood collection was 60 years (range: 43–76 years). Prostate cancer diagnosis followed blood collection by a median of 3.5 years (range: <1 to 9.5 years) and the median age at diagnosis was 64 years (range: 47–82 years).

Table 1 shows the distribution of the *lactase* genotypes in the eight countries participating in the study. Genotype frequencies followed the Hardy–Weinberg equilibrium (HWE) within each country (*p* > 0.05). Genotype frequencies varied markedly across Europe. In Greece none of participants and in Italy 5.6% of participants were homozygous for the T allele and 15% and 29% of participants, respectively, had at least one copy of the T allele. In contrast, in Sweden and Denmark, 55.6% and 63.3%, respectively, were homozygous for the T allele, and 95% and 94% of participants, respectively, had at least one copy of the T allele. Similarly, there was significant variation in allele frequencies between the countries, with frequencies of the T allele ranging from 7% and 17% in Greece and Italy, respectively, to 75% in Sweden and 79% in Denmark (Fig. 1). These genotype and allele frequencies were similar to published frequencies for corresponding regions of Europe.^{10,13–15}

The baseline characteristics of participants by *lactase* genotype are shown in Table 1. After adjustment for recruitment center, the characteristics of participants did not vary significantly by genotype, with the exception of educational attainment; CC homozygotes were less likely to have been educated to secondary level or higher (54.2% educated to secondary level or higher) than were heterozygotes and TT homozygotes (64.0% and 69.6%, respectively).

Table 1. Baseline characteristics by *lactase* genotype among all participants

Characteristic	CC (<i>n</i> = 262, 17.4%)	CT (<i>n</i> = 635, 42.3%)	TT (<i>n</i> = 606, 40.3%)	<i>p</i> Value
Country, n (%)				
Greece	17 (85.0)	3 (15.0)	0 (0.0)	
Italy	77 (71.3)	25 (23.2)	6 (5.6)	
Spain	30 (19.7)	81 (53.3)	41 (27.0)	
Germany	79 (18.2)	221 (50.9)	134 (30.9)	
Holland	6 (13.3)	19 (42.2)	20 (44.4)	
UK	33 (9.32)	145 (41.0)	176 (49.7)	
Sweden	11 (4.7)	92 (39.7)	129 (55.6)	
Denmark	9 (5.7)	49 (31.0)	100 (63.3)	
Mean age at blood collection (years)	60.8 (60.1–61.4)	60.9 (60.5–61.3)	60.9 (60.5–61.3)	0.95
Weight (kg)	79.3 (77.7–80.8)	81.0 (80.1–81.9)	80.1 (79.1–81.0)	0.10
Height (cm)	172.7 (171.8–173.6)	173.3 (172.8–173.8)	173.3 (172.8–173.8)	0.45
BMI (kg/m ²)	26.6 (26.1–27.0)	27.0 (26.7–27.2)	26.6 (26.4–26.9)	0.12
Alcohol consumption (g/d)	16.0 (13.1–18.9)	19.0 (17.4–20.7)	19.6 (17.8–21.3)	0.14
Current smoker (%) ^a	23.0 (17.6–29.4)	20.1 (16.9–23.6)	24.4 (20.7–28.4)	0.21
Physically active (%) ^a	45.6 (38.1–53.3)	47.5 (42.7–52.3)	48.2 (42.8–53.5)	0.88
Married or cohabiting (%) ^a	90.0 (83.4–94.1)	89.8 (85.7–92.8)	93.2 (89.4–95.7)	0.25
Educated to secondary level or higher (%) ^a	54.2 (46.1–62.2)	64.0 (59.0–68.8)	69.6 (64.2–74.6)	0.03

Values are means or percentages and their 95% confidence intervals after adjustment for regional centre.

^aDenotes unknown for some men.

The intake of dairy products and selected nutrients by *lactase* genotype after adjustment for recruitment center is shown in Table 2. Mean intake of milk, total dairy products and calcium increased significantly and the proportion of nonconsumers of milk decreased significantly with increasing number of copies of the *LCT* T allele. Adjusted mean intake of milk was 44.4 g/day among CC homozygotes, 69.8 g/day among heterozygotes and 82.3 g/day among TT homozygotes. There was no statistically significant heterogeneity in the relationship between milk intake and *LCT* genotype by country or recruitment center. Furthermore, even within the largest centers that had recruited participants from a small geographical area there was some evidence of a stepwise relationship; for example, in EPIC-Norfolk, intake of milk was 214.3 g/day among CC homozygotes, 297.8 g/day among heterozygotes and 309.1 g/day among TT homozygotes. Mean intake of protein from dairy products was also higher among *LCT* heterozygotes and TT homozygotes, respectively, although this association only reached borderline significance. No significant associations were observed between *LCT* genotype and intakes of yoghurt, cheese, butter, dairy desserts, dietary fat, or fat from dairy products. Mean intake of ice cream (a component of dairy desserts) was low in all *LCT* genotype groups, and while there was a statistically significant association between *LCT* genotype and ice cream intake (*p* value for difference = 0.02), with greater intake among heterozygotes, there was no significant trend in intake by genotype (*p* trend = 0.4).

Table 3 shows the mean concentrations of selected blood analytes by *lactase* genotype. After adjustment for regional center, there was significant variation by genotype in the circulating concentration of phytanic acid and borderline significant variation in circulating concentrations of 25-hydroxyvitamin D and enterodiol, respectively, with higher levels of phytanic acid and 25-hydroxyvitamin D being observed among men with more copies of the *LCT* T allele and lower levels of enterodiol among carriers of the T allele. No significant association was observed between circulating concentrations of IGF1, IGFBP3, sex hormones, SHBG or enterolactone and *LCT* genotype.

In participants with prostate cancer, the frequency of the *lactase* T allele was 40% and in control participants it was 40.5%. Table 4 shows relative risk of prostate cancer by *lactase* genotype. We observed no statistically significant association between risk for prostate cancer and *lactase* genotype, with or without adjustment for potential confounders in the recessive, codominant (additive) or dominant models. The adjusted odds ratios among CT heterozygotes and TT homozygotes compared with CC homozygotes were 1.03 (95% CI: 0.72–1.46) and 1.10 (0.76–1.59), respectively. There was no statistical evidence of heterogeneity in the trends in risk for prostate cancer by genotype according to the age of the men at blood collection, country of recruitment, prostate tumor stage, histological grade, or time to diagnosis (data not shown). The T allele frequencies were particularly low in Italy and Greece and therefore, in a sensitivity analysis, we also investigated the associations between the *LCT* variant

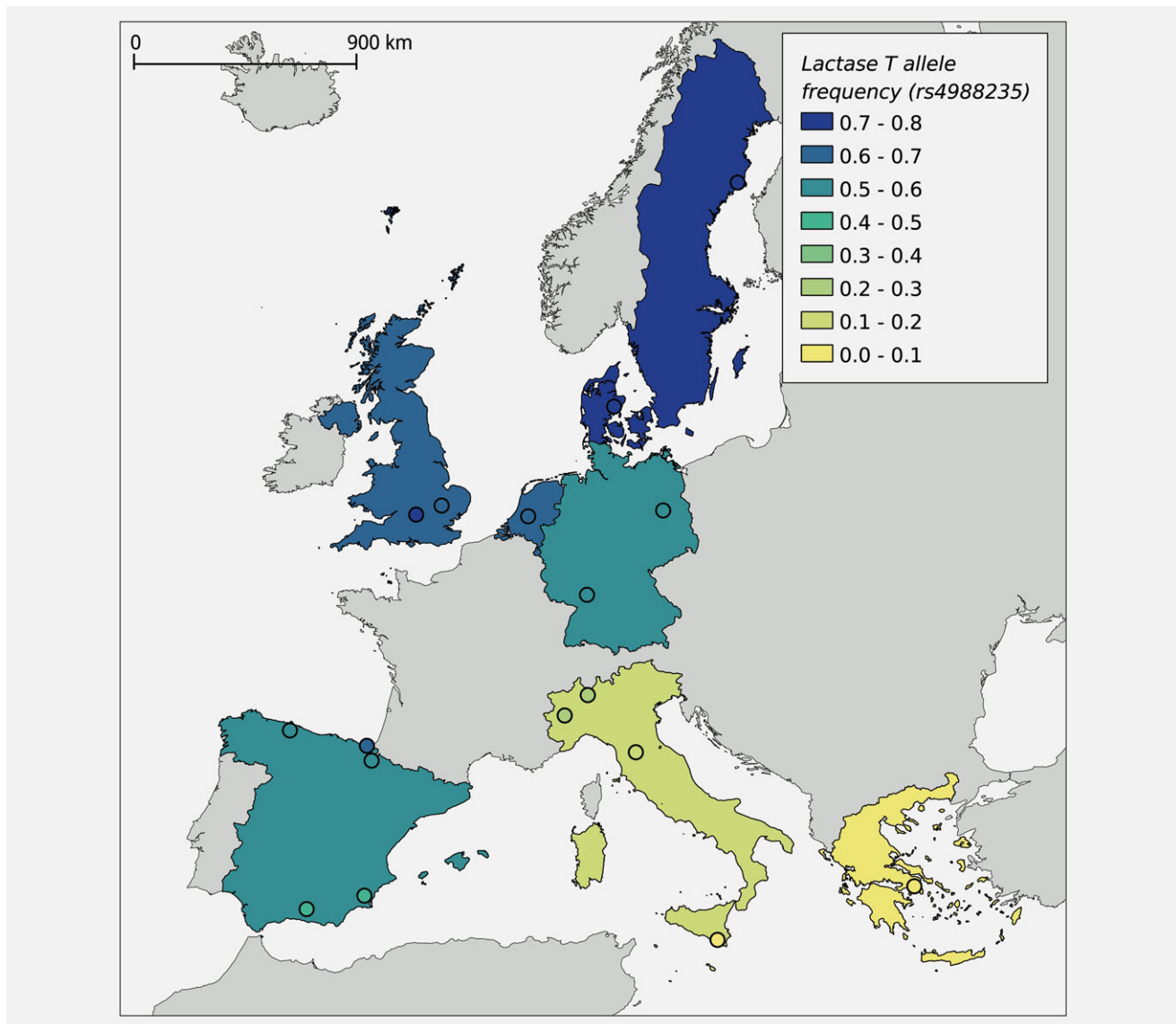


Figure 1. Frequency of the rs4988235 T allele by country in men from the European Prospective Investigation into Cancer and Nutrition. Color gradient indicates the average *lactase* T allele frequency in men from EPIC by country and by specific recruitment centers (indicated by filled circles).

and prostate cancer risk after excluding Greece and Italy; findings were not materially altered (data not shown).

Results from the one previous study of the *LCT* variant and prostate cancer risk,¹⁵ together with the corresponding result from this study and a summary relative risk are shown in Table 5. In Swedish and Finnish populations, respectively, the *lactase* variant was not associated with a significantly increased risk of prostate cancer.¹⁵ Taking our results with these previous findings, the combined odds ratio for prostate cancer in men with at least one copy of the *lactase* T allele compared to men with no copies was 1.12 (95 % CI: 0.96–1.32).

Discussion

We investigated common genetic variation in the *lactase* gene in relation to intake of dairy products and risk for pros-

tate cancer in a large, European study (EPIC). Our results provide evidence for a significant association of the C/T-13910 *lactase* variant with intake of milk but no evidence for a large association between the polymorphism and risk for prostate cancer.

The marked gradient in *LCT* allele frequency across Europe, both between and within countries, which we and others have observed^{10,12–15,28} coincides with substantial variation in diet and lifestyle practices between regions, including dairy product intake, and thus the overall relationship between *LCT* genotype and dairy product intake across Europe may be due to this concomitant cultural variation. Even in the absence of a true diet-genotype association, confounding due to this population stratification (*i.e.* systematic differences in allele frequency by region) might give rise to

Table 2. Intake of selected nutrients and dairy products by *lactase* genotype

Nutrient or dairy product	CC	CT	TT	<i>p</i> for difference ^a	<i>p</i> for trend ^a
Nonconsumers of dairy milk, %	18.8 (14.7–23.0)	15.5 (13.0–18.1)	12.1 (9.2–14.9)	0.034	0.010
Milk and milk beverages, g/day	44.4 (33.4–58.9)	69.8 (59.2–82.2)	82.3 (69.2–97.8)	0.002	0.001
Ditto excluding nonconsumers	135 (116–157)	170 (157–185)	175 (161–190)	0.013	0.018
Yoghurt, g/day	9.5 (6.9–13.2)	13.9 (11.5–16.8)	12.4 (10.1–15.0)	0.14	0.48
Cheese, g/day	17.6 (15.0–20.5)	18.9 (17.3–20.6)	18.3 (16.6–20.1)	0.70	0.88
Butter, g/day	0.75 (0.56–1.02)	0.60 (0.51–0.72)	0.70 (0.58–0.84)	0.34	0.94
Dairy desserts ^b g/day	5.5 (4.5–6.8)	6.2 (5.5–7.0)	5.7 (5.0–6.5)	0.56	0.90
Ice cream, g/day	3.3 (2.7–4.2)	2.6 (2.3–3.0)	3.3 (2.9–3.8)	0.02	0.45
All dairy products ^b g/day	206 (182–233)	253 (236–272)	258 (239–278)	0.010	0.015
Protein from dairy products, g/day	12.7 (11.4–14.2)	14.9 (13.9–15.9)	14.9 (13.9–16.0)	0.05	0.07
Fat, g/day	80.8 (77.0–84.8)	77.6 (75.4–79.8)	80.6 (78.2–83.2)	0.13	0.57
Fat from dairy products, g/day	11.7 (10.4–13.2)	12.7 (11.8–13.6)	11.9 (11.0–12.8)	0.32	0.81
Calcium, mg/day	855 (811–901)	906 (878–935)	922 (891–954)	0.08	0.04

Values are percentages or geometric means and their 95% confidence intervals after adjustment for regional centre.

^a*p* values for difference/trend in mean dietary intake or percentage of nonconsumers by genotype from analysis of variance with 2 d.f. and 1 d.f., respectively. ^bDairy desserts includes cream-based desserts and ice-cream; all dairy products includes milk and milk beverages, yoghurt, fromage blanc, cheeses, dairy puddings desserts, dairy creams. Nonconsumers of dairy milk are defined as persons consuming <5 g/day milk and milk beverages.

Table 3. Geometric mean concentration of selected blood analytes by *lactase* genotype

Analyte	<i>N</i>	CC	CT	TT	<i>p</i> for difference ^a	<i>p</i> for trend ^a
Phytanic acid (ng/mL)	1,062	493 (452–538)	503 (476–531)	551 (518–585)	0.05	0.03
Serum 25–OH vitamin D (nmol/L)	1,062	40.6 (36.0–45.7)	44.2 (40.9–47.8)	47.4 (43.1–52.1)	0.16	0.06
IGF-I (nmol/L)	1,013	21.8 (20.6–23.2)	21.5 (20.7–22.3)	22.0 (21.1–22.9)	0.70	0.71
IGFBP-3 (nmol/L)	1,013	131 (127–135)	129 (126–131)	130 (127–133)	0.69	0.87
Androstenedione (nmol/L)	1,011	4.51 (4.29–4.75)	4.65 (4.50–4.80)	4.66 (4.49–4.82)	0.60	0.43
Androstanediol glucuronide (nmol/L)	1,017	13.5 (12.3–14.7)	12.4 (11.8–13.1)	13.3 (12.5–14.2)	0.15	0.74
Testosterone (nmol/L)	924	15.4 (14.4–16.5)	15.4 (14.8–16.1)	16.2 (15.5–17.0)	0.24	0.14
SHBG (nmol/L)	962	40.8 (38.2–43.5)	41.0 (39.4–42.7)	42.6 (40.7–44.5)	0.40	0.22
Calculated free testosterone (pmol/L)	887	276 (258–294)	270 (260–282)	285 (273–298)	0.22	0.24
Enterolactone (ng/mL)	1062	2.69 (2.21–3.27)	2.91 (2.58–3.29)	2.64 (2.30–3.03)	0.53	0.67
Enterodiol (ng/mL)	1062	0.36 (0.30–0.43)	0.28 (0.25–0.31)	0.28 (0.24–0.32)	0.07	0.08

Values are geometric means and their 95% confidence intervals after adjustment for regional centre.

^a*p* values for difference/trend in geometric mean concentration by genotype from analysis of variance with 2 d.f. and 1 d.f., respectively.

an apparent association between diet or biomarker concentrations and *lactase* genotype in EPIC overall. However, even after adjusting for the regional recruitment center we found that milk consumption, total dairy product intake and calcium intake were related to *LCT* genotype, with evidence for a dose–response relationship with increasing milk and calcium intake with a higher number of T alleles. Our observations of a slightly weaker trend in the intake of dairy protein by *LCT* genotype is consistent with the hypothesized mechanism of lactose intolerance given that dairy product intake predominantly comprises milk by weight, whereas dairy pro-

tein as a percentage of energy intake comprises both protein from lactose-rich milk and protein from the protein-dense but low lactose cheese.

Previously published epidemiological data on the *lactase* variant and milk intake are relatively limited and inconsistent. Our results are in accord with data from 1673 men in the Swedish CAPS study, which reported lower milk intake in CC homozygotes (221 g/day) compared to men carrying the T allele (348 g/day).¹⁵ However, they differ from the findings of a study on 643 Italian men and women from the EPIC-Italy cohort, a small number of whom may be included

Table 4. Relative risk^a (95% confidence interval) of prostate cancer by *lactase* (*LCT*) genotype in EPIC

<i>LCT</i> genotype	Number (%)		OR	(95% CI)	Adjusted OR ^{b c}	(95% CI)
	Controls	Cases				
Co-dominant model						
CC	151 (17.3)	111 (17.6)	1.00	(Reference)	1.00	(Reference)
CT	368 (42.2)	267 (42.4)	1.07	(0.78–1.48)	1.03	(0.72–1.46)
TT	354 (40.5)	252 (40.0)	1.05	(0.75–1.48)	1.10	(0.76–1.59)
P value			0.917 ^c /0.848 ^d		0.847 ^c /0.576 ^d	
Dominant model						
CC	151 (17.3)	111 (17.6)	1.00	(Reference)	1.00	(Reference)
TC/TT	722 (82.7)	519 (82.4)	1.06	(0.78–1.45)	1.06	(0.75–1.48)
P value			0.695 ^c		0.750 ^c	
Recessive model						
CC/TC	519 (59.5)	378 (60.0)	1.00	(Reference)		1.00 (Reference)
TT	354 (40.5)	252 (40.0)	1.00	(0.80–1.25)	1.07	(0.84–1.37)
P value			0.985 ^c		0.581 ^c	

^aCase patients and control participants were matched on recruitment centre, age at enrolment (± 6 months), time of day of blood collection (± 1 hr), follow-up time (as close as possible), time between blood draw and last consumption of food or drinks (<3 , 3–6, >6 hr). ^bAdjustment was made for smoking (never, past, present), physical activity (inactive, moderately inactive, moderately active combined with active), alcohol intake (<8 , 8–15, 16–39, ≥ 40 g/day), marital status (married or cohabiting, not married or cohabiting), education (primary or none, secondary, degree level), and BMI (fourths). ^c*p* value from likelihood ratio chi-square test of the main effect of genotype on risk, comparing models with genotype terms with a model with no genotype terms. ^d*p* value from trend test of genetic effect from likelihood ratio tests scoring *LCT* genotypes as 0, 1, and 2, respectively, according to the number of variant T alleles, and with homozygotes for the C allele as the reference group.

Table 5. Relative risks (and 95% confidence intervals) from studies of the C/T-13910 *lactase* variant (rs4988235) in relation to prostate cancer risk, comparing carriers of the T allele (CT and TT genotypes) with CC homozygotes^a

Author, year	Population	Cases	Controls	Relative risk (95% confidence interval)
Torniainen <i>et al.</i> , 2010	Finnish	1,229	473	1.11 (0.83–1.47)
	Swedish	2,924	1,842	1.16 (0.91–1.46)
Travis <i>et al.</i> , 2012	European	630	873	1.06 (0.75–1.48)
Summary		4,783	3,188	1.12 (0.96–1.32)

^aData available on the comparison of risk in CT and TT versus CC genotypes in Torniainen *et al.*, 2010.

in the current analyses; while results from this Italian study showed a significantly higher intake of ice-cream among TT homozygotes, there was less evidence for an association between *LCT* genotype and milk intake.¹³ The results from two other studies, which had information on milk consumption (never vs. usually drink milk) but not on amount of milk consumed, were also inconsistent.^{14,18} In a study of 3,344 women from the United Kingdom, the likelihood of never drinking milk was not found to be associated with *lactase* genotype.¹⁴ In contrast, among 2396 men and women from Russia, Romania, Poland and the Czech Republic, a higher proportion of CC homozygous individuals reported never consuming milk than did carriers of the T allele.¹⁸ It remains possible that the overall association we observed between milk consumption and *LCT* genotype is due to residual confounding within center, particularly for large centers recruiting over wide geographical areas within which there may be variation in the frequency of the *lactase* C/T al-

lele. However, our findings of no heterogeneity in the relationship between genotype and milk intake by recruitment center, and of associations among men from regional centers which recruited from geographically limited areas, such as EPIC-Norfolk, provide evidence in support of a true relationship between *LCT* genotype and milk consumption.

While the traditional view has been that lactase-persistence is an autosomal dominant trait, our results suggest a stepwise relationship, with a higher milk intake being observed with increasing number of T alleles. Our findings are consistent with those from a clinical study of 52 adults, in which lactase expression in intestinal biopsy samples was found to follow a trimodal distribution.²⁹ The biopsy study found very low levels of lactase activity in individuals who were homozygous for the C allele.²⁹ Our data also suggest that, as has been reported in other studies, individuals with the CC genotype can tolerate some milk consumption.

We also assessed whether the *lactase* variant was associated with circulating concentrations of a number of nutritional and hormonal biomarkers that were available for men in the current study, including phytanic acid, 25-hydroxyvitamin D, IGF1, enterodiol and enterolactone.^{22–26} Plasma concentrations of phytanic acid varied significantly by *LCT* genotype; higher concentrations were observed in TT homozygotes than in heterozygotes and CC homozygotes. These findings are consistent with the fact that phytanic acid is predominantly obtained from foods high in ruminant fat (*i.e.* meat and dairy products from cows, sheep and goats) and provide further indirect support for an association between *lactase* genotype and dairy food intake. We also observed borderline significant associations between *lactase* genotype and circulating concentrations of 25-hydroxyvitamin D and enterodiol; a higher concentration of 25-hydroxyvitamin D was observed with increasing number of copies of the *LCT* T allele and a lower concentration of enterodiol was observed among carriers of the T allele compared to CC homozygotes. Given the multiple statistical tests these may be chance findings and require confirmation in other large studies. In the one previous report on circulating vitamin D levels, no association was found with *lactase* genotype in two elderly Dutch populations (comprising 6,367 and 844 individuals, respectively).³⁰ It is plausible, however, that there is an increased conversion of 25-hydroxy-vitamin-D into 1,25-hydroxy-vitamin D (the hormonal biologically active form of vitamin D) in individuals who have a lower calcium intake because they carry the *LCT* C allele. There might also be some degree of general malabsorption associated with the C allele, which means that there is lower absorption of dietary vitamin D into the body. An association between the *lactase* variant and enterodiol is also possible, given that differences in the digestion of lactose due to the *LCT* variant may alter the activity or composition of the gut microflora, which is responsible for the production of enterodiol from the plant lignan secoisolariciresinol.³¹ Intake of dairy products and calcium has been positively associated with circulating concentrations of insulin-like growth factors;³² however, we found no evidence for an association between *lactase* genotype and serum concentrations of IGF1 and IGFBP3.

Our findings of no significant association between *LCT* genotype and prostate cancer risk are consistent with those from the only previously published study of the *LCT* variant in relation to risk for prostate cancer, which reported no significant associations in nested case control studies among Finnish men (1,229 men with prostate cancer and 473 controls) and Swedish men (2,924 men with prostate cancer and 1,842 controls).¹⁵ Meta-analysis of our results together with these previous findings suggest that the *lactase* variant is not likely to be strongly related to prostate cancer risk, although it is not possible to exclude a small association (summary relative risk 1.12, 95% confidence interval 0.96–1.32). These data provide little support for a strong relationship between dairy product consumption and prostate cancer risk, consist-

ent with the null findings in some but not all of the published studies on the association between dairy product consumption and prostate cancer risk.^{2–7,33–38} It is possible, however, that the relatively small average difference in milk intake between genotype groups (37 g/day between CC and TT homozygotes in EPIC overall, and 81 g/day in the UK where the difference was most marked) is of insufficient magnitude to have a detectable influence on risk for prostate cancer should such an association truly exist. Furthermore, it may be that other components of dairy foods are associated with prostate cancer. The *lactase* genotype is a marker of intake of lactose-rich dairy products (for example, milk, yoghurt, ice cream, low fat cream) rather than being a marker of dairy product intake *per se*, and is not correlated with dietary factors such as intake of cheese or butter.

Dairy intake and the *lactase* gene-variant in relation to prostate cancer has been one of the specific hypotheses discussed with respect to the application of Mendelian Randomization (MR) to assess relationships between environmental exposures and risk for disease.¹⁴ The current analyses, however, demonstrate some of the challenges of applying Mendelian Randomization to explore the relationship between dietary factors and cancer risk.³⁹ Because the expected association between dairy products and prostate cancer is relatively small in magnitude and the difference in total dairy intake by *lactase* genotype is modest, even in this large prospective study and with a relatively common allele of interest (the frequency of the T allele was 0.61 overall), there is relatively limited power to detect an association between the *lactase* genotype and the disease. Based on previous findings from a prospective cohort analysis in EPIC¹ for example, in which a 35 g per day increment in dairy protein intake was associated with an approximately 32% increase in risk for prostate cancer, the 2.2 g per day difference in dairy protein intake between CC and TT homozygotes would be estimated to result in only a 2% elevated risk among TT homozygotes compared with CC homozygotes. With the *LCT* genotype frequencies observed in EPIC, a study of approximately 30,000 cases and 30,000 controls would be needed to detect a relative risk for prostate cancer of this magnitude with 80% power.

The results from this large prospective study in European men suggest that variation in the *lactase* gene is associated with intake of milk but provide no evidence for a large association with risk for prostate cancer. The data presented illustrate the limitations of applying Mendelian Randomization as a research strategy to explore the relationship between dairy product consumption and cancer risk and the need for very large prospective studies with both genetic and dietary data in such studies of genetic proxies of nutritional exposures.

Acknowledgements

We thank the participants in the EPIC study, Carine Biessy and Bertrand Hémon at IARC and Laure Dossus at DKFZ for their expertise in data handling, and Andrew Chadwick at the Cancer Epidemiology Unit for the preparation of the figure.

References

- Allen NE, Key TJ, Appleby PN, et al. Animal foods, protein, calcium and prostate cancer risk: the European Prospective Investigation into Cancer and Nutrition. *Br J Cancer* 2008;98:1574–81.
- Ahn J, Albanes D, Peters U, et al. Dairy products, calcium intake, and risk of prostate cancer in the prostate, lung, colorectal, and ovarian cancer screening trial. *Cancer Epidemiol Biomarkers Prev* 2007;16:2623–30.
- Gao X, LaValley MP, Tucker KL. Prospective studies of dairy product and calcium intakes and prostate cancer risk: a meta-analysis. *J Natl Cancer Inst* 2005;97:1768–77.
- Mitrou PN, Albanes D, Weinstein SJ, et al. A prospective study of dietary calcium, dairy products and prostate cancer risk (Finland). *Int J Cancer* 2007;120:2466–73.
- Park SY, Murphy SP, Wilkens LR, et al. Calcium, vitamin D, and dairy product intake and prostate cancer risk: the Multiethnic Cohort Study. *Am J Epidemiol* 2007;166:1259–69.
- Park Y, Mitrou PN, Kipnis V, et al. Calcium, dairy foods, and risk of incident and fatal prostate cancer: the NIH-AARP Diet and Health Study. *Am J Epidemiol* 2007;166:1270–9.
- Rohrmann S, Platz EA, Kavanaugh CJ, et al. Meat and dairy consumption and subsequent risk of prostate cancer in a US cohort study. *Cancer Causes Control* 2007;18:41–50.
- Qin LQ, Xu JY, Wang PY, et al. Milk consumption is a risk factor for prostate cancer in Western countries: evidence from cohort studies. *Asia Pac J Clin Nutr* 2007;16:467–76.
- Severi G, English DR, Hopper JL, et al. Re: prospective studies of dairy product and calcium intakes and prostate cancer risk: a meta-analysis. *J Natl Cancer Inst* 2006;98:794–5; author reply 5.
- Enattah NS, Sahi T, Savilahti E, et al. Identification of a variant associated with adult-type hypolactasia. *Nat Genet* 2002;30:233–7.
- Weiss KM. The unkindest cup. *Lancet* 2004;363:1489–90.
- Beja-Pereira A, Luikart G, England PR, et al. Gene-culture coevolution between cattle milk protein genes and human lactase genes. *Nat Genet* 2003;35:311–13.
- Sacerdote C, Guarrera S, Smith GD, et al. Lactase persistence and bitter taste response: instrumental variables and mendelian randomization in epidemiologic studies of dietary factors and cancer risk. *Am J Epidemiol* 2007;166:576–81.
- Smith GD, Lawlor DA, Timpson NJ, et al. Lactase persistence-related genetic variant: population substructure and health outcomes. *Eur J Hum Genet* 2009;17:357–67.
- 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;16:956–61.
- Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;32:1–22.
- Davey Smith G, Ebrahim S. What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? *BMJ* 2005;330:1076–9.
- Timpson NJ, Brennan P, Gaborieau V, et al. Can lactase persistence genotype be used to reassess the relationship between renal cell carcinoma and milk drinking? Potentials and problems in the application of Mendelian randomization. *Cancer Epidemiol Biomarkers Prev* 2010;19:1341–8.
- 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;5:1113–24.
- Deharveng G, Charrondiere UR, Slimani N, et al. Comparison of nutrients in the food composition tables available in the nine European countries participating in EPIC. European Prospective Investigation into Cancer and Nutrition. *Eur J Clin Nutr* 1999;53:60–79.
- Slimani N, Deharveng G, Unwin I, et al. The EPIC nutrient database project (ENDB): a first attempt to standardize nutrient databases across the 10 European countries participating in the EPIC study. *Eur J Clin Nutr* 2007;61:1037–56.
- Travis RC, Spencer EA, Allen NE, et al. Plasma phyto-oestrogens and prostate cancer in the European Prospective Investigation into Cancer and Nutrition. *Br J Cancer* 2009;100:1817–23.
- Linseisen J, Rohrmann S, Miller AB, et al. Fruit and vegetable consumption and lung cancer risk: updated information from the European Prospective Investigation into Cancer and Nutrition (EPIC). *Int J Cancer* 2007;121:1103–14.
- Price AJ, Allen NE, Appleby PN, et al. Plasma phytanic acid concentration and risk of prostate cancer: results from the European Prospective Investigation into Cancer and Nutrition. *Am J Clin Nutr* 2010;91:1769–76.
- Travis RC, Crowe FL, Allen NE, et al. Serum vitamin D and risk of prostate cancer in a case-control analysis nested within the European Prospective Investigation into Cancer and Nutrition (EPIC). *Am J Epidemiol* 2009;169:1223–32.
- Travis RC, Key TJ, Allen NE, et al. Serum androgens and prostate cancer among 643 cases and 643 controls in the European Prospective Investigation into Cancer and Nutrition. *Int J Cancer* 2007;121:1331–8.
- StataCorp. Stata Statistical Software: Release 10 College Station, TX: StataCorp LP, 2007.
- Bersaglieri T, Sabeti PC, Patterson N, et al. Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet* 2004;74:1111–20.
- Kuokkanen M, Enattah NS, Oksanen A, et al. Transcriptional regulation of the lactase-phlorizin hydrolase gene by polymorphisms associated with adult-type hypolactasia. *Gut* 2003;52:647–52.
- Koek WN, van Meurs JB, van der Eerden BC, et al. The T-13910C polymorphism in the lactase phlorizin hydrolase gene is associated with differences in serum calcium levels and calcium intake. *J Bone Mineral Res: The Official J Am Soc Bone Mineral Res* 2010;25:1980–7.
- Campbell AK, Matthews SB, Vassel N, et al. Bacterial metabolic 'toxins': a new mechanism for lactose and food intolerance, and irritable bowel syndrome. *Toxicology* 2010;278:268–76.
- Crowe FL, Key TJ, Allen NE, et al. The association between diet and serum concentrations of IGF-I, IGFBP-1, IGFBP-2, and IGFBP-3 in the European Prospective Investigation into Cancer and Nutrition. *Cancer Epidemiol Biomarkers Prev* 2009;18:1333–40.
- Huncharek M, Muscat J, Kupelnick B. Dairy products, dietary calcium and vitamin D intake as risk factors for prostate cancer: a meta-analysis of 26,769 cases from 45 observational studies. *Nutr Cancer* 2008;60:421–41.
- Kesse E, Bertrais S, Astorg P, et al. Dairy products, calcium and phosphorus intake, and the risk of prostate cancer: results of the French prospective SU.VI.MAX (Supplementation en Vitamines et Minéraux Antioxydants) study. *Br J Nutr* 2006;95:539–45.
- Koh KA, Sesso HD, Paffenbarger RS, et al. Dairy products, calcium and prostate cancer risk. *Br J Cancer* 2006;95:1582–5.
- Kurahashi N, Inoue M, Iwasaki M, et al. Dairy product, saturated fatty acid, and calcium intake and prostate cancer in a prospective cohort of Japanese men. *Cancer Epidemiol Biomarkers Prev* 2008;17:930–7.
- Tseng M, Breslow RA, Graubard BI, et al. Dairy, calcium, and vitamin D intakes and prostate cancer risk in the National Health and Nutrition Examination Epidemiologic Follow-up Study cohort. *Am J Clin Nutr* 2005;81:1147–54.
- Tseng M, Giri V, Watkins-Bruner D, et al. Dairy intake and 1,25-dihydroxyvitamin D levels in men at high risk for prostate cancer. *Cancer Causes Control* 2009;20:1947–54.
- von Hinke Kessler Scholder S, Smith GD, Lawlor DA, et al. Mendelian randomization: the use of genes in instrumental variable analyses. *Health Econ* 2011;20:893–6.