Can a genotype for lactase persistence be used as a tool to assess the relationship between renal cell carcinoma and milk drinking? Pitfalls in the application of Mendelian randomization

Nic Timpson[1], **Paul Brennan**[2], **Valérie Gaborieau**[2], **Lee Moore**[3], **David Zaridze**[4], **Vsevolod Matveev**[4], **Neonilia Szeszenia-Dabrowska**[5], **Jolanta Lissowska**[6], **Eleonora Fabianova**[7], **Dana Mates**[8], **Vladimir Bencko**[9], **Lenka Foretova**[10], **Vladimir Janout**[11], **Wong-Ho Chow**[3], **Nathaniel Rothman**[3], **Paolo Boffetta**[2], **Roger Harbord**[12], and **George Davey Smith**[1]

[1]The MRC Centre for Causal Analysis in Translational Epidemiology (CAiTE), University of Bristol, Oakfield House, Bristol, BS8 2BN. [2]International Agency for Research on Cancer, 50 Cours Albert Thomas, 69372 Lyon CEDEX 08, France. [3]National Cancer Institute (NCI), Division of Cancer Epidemiology and Genetics, MD 20892 Bethesda, MD, USA. [4]Russian N.N. Blokhin Cancer Research Centre, Department of Epidemiology and Prevention, Moscow, Russia. [5]Institute of Occupational Medicine, Department of Epidemiology, 90950 Lodz, Poland. [6]Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Department of Cancer Epidemiology and Prevention, 02-781 Warsaw, Poland. [7]Regional Authority of Public Health, 97401 Banska Bystrica, Slovakia. [8]Institute of Public Health, 050463 Bucharest, Romania. [9]Charles University, First Faculty of Medicine, Institute of Hygiene and Epidemiology, 12800 Prague 2, Czech Republic. [10]Mazaryk Memorial Cancer Institute, Department of Cancer Epidemiology and Genetics, 65653 Brno, Czech Republic. [11]Palacky University, Faculty of Medicine, Department of Preventive Medicine, CZ-77515 Olomouc, Czech Republic. [12]Department of Social Medicine, University of Bristol, UK.

Abstract

Background—Increased risk of renal cell carcinoma (RCC) with milk consumption has been reported from observational studies including that involved here. Whether this represents causal association or whether it is the result of confounding or bias is unclear. We aimed to assess the potential for genetic variation in lactase persistence to be used as a tool for the interrogation of these relationships.

Methods—Using a large, hospital based case control study, we use a combination of observational genetic and phenotypic data to determine whether the *MCM6* -13910 C/T(rs4988235) variant may be used as an non-confounded and unbiased marker for milk consumption.

Corresponding author: Dr Nicholas Timpson MRC Centre for Causal Analysis in Translational Epidemiology, Bristol University, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, UK. n.j.timpson@bris.ac.uk Tel: +44 117 9287352 Fax: +44 117 928 7236 .
Conflicts: None

Results—Consumption of milk during adulthood was associated with increased risk of RCC (OR=1.35 95% CI 1.03, 1.76 p=0.03). Among controls, consumption of milk was associated with the lactase persistence genotype at rs4988235 (2.39[1.81, 3.15], p=6.9*10^{-10}), however the same genotype was not associated with RCC (OR=1.01 95% CI 0.83, 1.22 p=0.9). In controls, milk consumption was associated with confounding factors including smoking, and educational attainment, while the lactase persistence genotype at rs4988235 genotype showed negligible association with confounding factors.

Conclusions—The absence of an association between the *MCM6* genotype and RCC suggests that observational associations between milk consumption and RCC may be due to confounding or bias. However, if the association between genotype and behavioral exposure is weak, then the power of this test may be low. The nature of intermediate risk factor instrumentation is an important consideration in the undertaking and interpretation of this type of causal analysis experiment.

Keywords

Mendelian randomization; lactase persistence; renal carcinoma

Introduction

Consumption of milk has been reported to be a potential risk factor for renal cell carcinoma (RCC)(1). The causality of this association is difficult to assess in the absence of randomized control trials as milk consumption is likely to be associated with other dietary and lifestyle factors that may themselves be associated with RCC. Other study designs (prospective studies and population based case/control studies(2, 3)) can contribute to the assessment of milk drinking as a risk factor for RCC, however these are subject to the known limitations of observational epidemiology (4, 5) and where done, have not always yielded corroboratory results(6).

A potential solution to this problem of confounding is Mendelian randomization (MR)(7, 8). MR relies on the use of genetic markers associated with modifiable exposures of interest (in this case milk drinking) as non-confounded and unbiased markers of exposure (Figure 1). Assuming the genetic marker is not related to confounding features and is associated with the outcome only through its association with the exposure, then identifying an association between genotype and outcome will test the hypothesis of a true non-confounded association between exposure and outcome(8).

At the level of the population, wide-spread habitual milk drinking is thought largely to reflect the ability to hydrolyze lactose, the principal carbohydrate in milk(9). This ability is lost after weaning in nearly all mammals and for most human populations and this loss is associated with lactose intolerance. While most human populations have high prevalence of lactose intolerance, Northern Europeans tend to have high proportions of lactose tolerance(10). The latter reflects the persistence of the enzyme lactase into adulthood and is thought to be derived from selective pressures brought about by the domestication of livestock, generating strong patterns of advantage for this ability(11, 12).

The population distribution of lactase persistence has been well traced, and a genetic variant associated with lactase persistence has been identified(13). This association is derived from an extended region of linkage disequilibrium (LD) on chromosome 2q21 which contains the associated variant ~14 kb upstream of the lactase coding region (LCT) in the *MCM6* gene. Whilst two variants are recognized as associated with lactase persistence, the extended LD in this region places the most correlated allele(14, 15) on a common haplotypic background which captures nearly all variation in this region and denotes the select emergence of this variant type. There is evidence for the association of the *MCM6* -13910 C/T (henceforth termed rs4988235) variant with lactase persistence and at a population level there is a strong association between prevalence of lactase persistence and consumption of milk. At an individual level, however, work looking at the association between physically assessed lactose tolerance and milk drinking has shown this relationship to be relatively weak(16-27).

We have previously reported an association between milk consumption and RCC in a multicentre case-control study conducted in Russia, Czech Republic, Romania and Poland(28). In this current analysis we investigate the relationship between *MCM6* variation and actual milk consumption in efforts to clarify the potential for this variation to be used as a proxy measure for this risk factor for RCC. We intend to then use this proxy measure as an instrument to assess the causal nature of the association between milk consumption and RCC risk. Given a confirmed relationship between genetically prescribed lactase persistence and milk consumption, we aim to assess the association between RCC risk and the same genetic variation acting as a proxy measure for milk consumption. Assuming that assessment of milk consumption in this way will not suffer the same limitations seen in conventional observational analyses, results from this analysis provide evidence for the presence of causal a relationship between milk consumption and RCC risk. We hope also to comment on the feasibility of using *MCM6* variation as a marker of milk consumption and through this make more general comments as to the importance of the genetic proxy/risk factor relationship in the application of MR.

Materials and methods

The population

Between August 1999 and January 2003, we conducted a hospital-based case-control study of RCC in Russia (Moscow); Romania (Bucharest); Poland (Lodz); and the Czech Republic (Prague, Olomouc, Ceske Budejovice, and Brno). A total of 1,097 newly diagnosed and histologically confirmed RCC cases (ICD-0-2 codes C64) between the ages of 20 and 79 years were recruited. Trained medical staff reviewed medical records to extract relevant diagnostic information, including date and method of diagnosis, histologic type, tumour location, stage and grade.

Eligible controls were patients admitted to the same hospital as cases for conditions unrelated to smoking or genitourinary disorders (except for benign prostatic hyperplasia) who were frequency-matched on age to cases. No single disease made up more than 20% of the control group. Both cases and controls had to be residents of the study areas for at least one year at the time of recruitment. The response rate among eligible subjects who were

requested to participate ranged from 90.0% to 98.6% for cases and from 90.3% to 96.1% for controls.

All study subjects and their physicians provided written informed consent. This study was approved by the institutional review boards of all participating centers.

Standardized lifestyle and food frequency questionnaires were piloted in all centers prior to use and interviews were conducted in-person by trained personnel to elicit information on demographic characteristics, education, exposure to tobacco smoke, alcohol consumption, dietary practices, anthropometry, medical history, family history, and occupational history.

Milk drinking and food consumption

The dietary component of the questionnaire comprised 23 food items, with frequency of consumption (and score) assessed for each item (never [0], <once per month [1], <once per week [2], 1-2 times per week [3], 3-5 times per week [4], and daily [5]). The questionnaire was repeated for two different time periods: 1) the year prior to interview, and 2) prior to political and market changes in 1989 (1991 in Russia). These scores were united into the groupings 0 versus 1+2+3+4+5 in order to yield a dichotomous assessment of adult milk consumption which represented never versus ever consumption patterns. One subject had to be excluded from milk analysis due to missing values. Information on lactose-free milk consumption was not available.

Genotyping

After DNA extraction, genotyping for rs4988235 was performed by the 5′ nuclease assay (TaqMan). DNA from cases and controls were blinded and randomized on PCR plates to avoid any potential bias and duplicate genotyping performed for a random 10% of the total series for genotyping quality control. Genotyping call rates were similar for cases and controls being > 95% for both the cases and controls that remained in our analysis.

Analyses

From these samples, 953 cases and 2396 controls were available with observational data, whilst for genetic analyses, 915 cases and 2346 controls were available with genotypes.

To test for a potential relationship between milk consumption and variation at *MCM6*, we performed logistic regression of the dominant model coded genotypes at rs4988235 (i.e. CC versus CT/TT, non-persistence versus any carriage of lactase persistence alleles) and categorized milk drinking status. Analyses were performed both with and without the covariates sex, alcohol consumption (ever/never), smoking (ever/never), the categorical variable educational attainment (low/medium/high) and the continuous variable age. To test for potential relationships between milk consumption/genotype and RCC including potential confounders, we performed logistic regression of case/control status including the same potentially confounding features.

For analyses across all studies, individual study estimates were combined by meta-analysis. In this case, point estimates and standard errors derived from logistic regression were meta-analyzed using a random effects model using the "metan" user-written command in

Stata(29). With meta-analyzed results, both *p*-values for heterogeneity and an I^2 statistic representing the variance attributable to between study differences were simultaneously calculated.

All statistics were performed using Stata version 10 (StataCorp LP, 2007).

Results

Descriptive statistics and milk drinking profiles for all study participants are included in Table 1 (characteristics for individuals without genetic data, n=~155 overall, did not vary substantively for descriptive characteristics). Minor allele frequencies for rs4988235 *within controls* were observed to be 0.28 in Romania, 0.40 in Poland, 0.35 in Russia and 0.46 in the Czech Republic. The minor allele for all populations was the "T" (persistence) allele at rs4988235, consistent with that in southern and eastern Europe, but opposite to that observed in regions further north and west (the "C" allele was found at a frequency of 0.26 in the UK(17)). No strong evidence for departure of recorded genotype frequencies from Hardy-Weinberg equilibrium was found (p>0.05) except nominally in Romania (p=0.02). Whilst country specific MAF estimates are different, they do reflect intermediate frequencies of the order anticipated within Eastern European populations (Supplementary Table S1).

Differences were observed in the consumption patterns of milk in differing allele groups. In controls, genotype was seen to be associated with milk drinking OR 2.39 [1.81, 3.15], p=6.9*10^{-10}. In all countries, a higher proportion of individuals reported never having consumed milk within those carrying the reported lactase non-persistent CC genotype at rs4988235 (Table 2). Tests of heterogeneity showed there to be no consistent evidence of difference in the association between lactase persistence genotype and milk drinking between countries (Table 2). Romania was the only country not to demonstrate association between genotype and milk drinking tendency.

There was an elevated risk of RCC among those consuming milk as opposed to never consumers (OR=1.35, 95% CI 1.03, 1.76 p=0.03). This was largely driven by the strong, observed, relationship between milk consumption and cancer risk in the Czech Republic (OR=1.68, 95% CI 1.13, 2.49 p=0.01) where the frequency of the lactase persistence driving allele and adherence to it was the greatest (Table 3).

Despite observed differences between the risk of RCC with differing milk consumption patterns and between lactase persistent genotype and milk consumption patterns, no substantial differences were observed between rs4988235 genotype and the risk of RCC either in analyses by country or in the sample as a whole (overall odds of RCC by genotypic group OR=1.01, 95% CI 0.83, 1.22 p=0.9) (Table 4).

Analysis of variables which could potentially have confounded results between milk consumption and the risk of RCC yielded evidence for association between educational attainment (p=0.001) and milk consumption in all countries(Table 5). There was nominal, although not systematic, representation of this relationship and others within results for country specific data. The strongest of these was for the Czech Republic where, milk

consumption was associated with educational attainment and smoking (p=0.02 and 0.007 respectively).

In contrast, analysis between rs4988235 genotype and the same potentially confounding factors did not generally yield evidence of association. However, there was nominal evidence for association of genotype with education (p=0.03, Table 5), which was largely lost after country specific analysis.

Discussion

We aimed to analyze the relationship between milk consumption and RCC by employing a MR framework in order to avoid confounding and bias that may be influencing observational reports of a link between milk consumption and the risk of RCC. In this large, case control study from 4 central and eastern European countries which have intermediate frequencies for rs4988235, we found that whilst there was evidence for an association between milk consumption and RCC, the use of a non-confounded, proxy marker of milk consumption (i.e. a genetic marker associated with milk consumption levels) did not support this finding.

Our study was designed to assess the relationships between milk drinking and RCC and to bring to attention practical issues encountered in the application of Mendelian randomization. Importantly, despite its size, our study had low power to detect or reject a possible causal association between genotype and cancer. This was due to the relatively weak relationship between the genotype and milk consumption (an often ignored characteristic in the examination of lactase persistence genotypes) and the modest observational association between milk consumption and RCC: a study with approximately 37000 cases and 37000 controls would be needed to achieve 80% power under the same framework (see Supplementary material for method of calculation). Part of this impairment of power is likely to be due to a large number of risk exposed control participants (those who carried the lactase non-persistent genotype yet reported drinking milk) and this illustrates the importance of correlation between genotype and risk factor of interest in MR experiments.

A feature of these data was the apparent lack of association between lactase non-persistence associated genotypes and milk avoidance in Romania. Romania was the only country in this work *not* to show a robust relationship between variation at rs4988235 and milk drinking behavior. We have no prior reason to expect different biological properties within this population and this finding may indicate one of two likely scenarios. Firstly and most likely, it may be that the combination of relative small sub-sample size and errors in the reporting of milk drinking that are presenting as a lack of observed association. Alternatively, cultural pressures may be acting to force a departure from the milk drinking behavior one would expect given the presence of this variation. This is a phenomenon that has been used to explain situations elsewhere where populations contain only rare lactase non-persistence(17), however this mechanism could be in operation within populations of intermediate allele frequency for this variant.

A further observation of interest within this analysis was the nominal association between the lactase persistence genotype and educational patterning across Europe. Relationships between genetic variation and confounding features such as this can indicate an impairment in their use as instrumental variables through the re-introduction of environmental confounding(7). In this case it is unlikely that the observed trend has great impact on the overall interpretation of the lack of association between *MCM6* genotype and RCC risk, however it is of interest in light of the use of rs4988235 as a population specific marker. When looking at the descriptive properties of educational attainment (Table 1), there is a suggestion for both difference between countries and the possibility of a gradient across Europe (west/east for this factor as opposed to the accepted east/west for lactase persistence(11)). With the expected gradient in lactase persistence allele frequencies by geography (previously observed) being opposite to that suggested for educational achievement, it becomes less surprising that some level of association is suggested between *MCM6* variation and this factor. However, relating to this study where there is no reliable association between *MCM6* genotypes and RCC risk, this observation may not be of critical importance.

Important aspects raised by this work are sample size and what in this case may be loosely termed the "penetrance" of genetic effect. In this study of over 900 cases of RCC, it is possible to assess direct associations between risk exposure and outcome with reasonable accuracy. However, it is has not been possible to achieve this for genetic proxy markers for exposure (i.e. genetic markers predicting milk consumption) due to poor correlation between genotype and exposure. Although we do observe a lack of association between milk consumption related genotypes and RCC risk, the ability of this to comment directly on the causality of putative observational associations between milk consumption and RCC risk, is limited.

Based on evidence from the associations between genotype and both milk drinking and cancer risk, work presented here may justify caution with respect to the interpretation of associations between milk consumption and cancer risk. However, whilst the translation of *MCM6* variation to lactase persistence may yield true physiological relationships, these appear not to strongly influence actual milk drinking patterns in the populations assessed and this impairs the accuracy of our reassessment of milk as a risk factor for RCC. Importantly, this work provides practical guidance for the employment of MR methods for the dissection of more complex, binary traits. An important lesson from this analysis is that in order to achieve suitable power to allow formal analysis of such an MR framework, clear effects, robust instruments and large sample sizes are required.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. WCRF. Food, nutrition, physical activity and the prevention of cancer: a global perspective. Research AIfC. , editor. Washington DC: 2007.

2. Feskanich D, Willett WC, Stampfer MJ, Colditz GA. Milk, dietary calcium, and bone fractures in women: a 12-year prospective study. Am J Public Health. Jun.1997 87:992–7. [PubMed: 9224182]

3. Riboli E, Kaaks R. The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. International Journal of Epidemiology. 1997; 26(Suppl 1):S6–14. [PubMed: 9126529]

4. Davey Smith G. Reflections on the limitations to epidemiology. Journal of Clinical Epidemiology. 2001; 54:325–31. [PubMed: 11297881]

5. Davey Smith G, Ebrahim S. 'Mendelian Randomisation': can genetic epidemiology contribute to understanding environmental determinants of disease? International Jounal of Epidemiology. 2003; 32:1–22.

6. Lee JE, Hunter DJ, Spiegelman D, Adami HO, Bernstein L, van den Brandt PA, et al. Intakes of coffee, tea, milk, soda and juice and renal cell cancer in a pooled analysis of 13 prospective studies. International Journal of Cancer. 2007; 121:2246–53.

7. Lawlor DA, Harbord RM, Sterne JAC, Timpson NJ, Davey Smith G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. Statistics in Medicine. 2008; 27:1133–63. [PubMed: 17886233]

8. Davey Smith G, Ebrahim S. Mendelian Randomisation: prospects, potentials and limitations. International Jounal of Epidemiology. 2004; 33:30–42.

9. Villako K, Maaroos H. Clinical picture of hypolactasia and lactose intolerance. Scand J Gastroenterol Suppl. 1994; 202:36–54. [PubMed: 8042017]

10. Swallow DM. Genetics of lactase persistence and lactose intolerance. Annual Review of Genetics. 2003; 37:197–219.

11. Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, et al. Genetic signatures of strong recent positive selection at the lactase gene. American Journal of Human Genetics. 2004; 74:1111–20. [PubMed: 15114531]

12. Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, et al. Convergent adaptation of human lactase persistence in Africa and Europe. Nature Genetics. 2007; 39(1):31–40. [PubMed: 17159977]

13. Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, Jarvela I. Identification of a variant associated with adult-type hypolactasia. Nature Genetics. 2002; 30:233–7. [PubMed: 11788828]

14. Poulter M, Hollox E, Harvey CB, Mulcare C, Peuhkuri K, Kajander K, et al. The causal element for the lactase persistence/non-persistence polymorphism is located in a 1Mb region of linkage disequilibrium in Europeans. Annals of Human Genetics. 2003; 67:298–311. [PubMed: 12914565]

15. Stearns, SCE.; Koella, JCE. Evolution in Health and Disease. 2 ed. OUP; Oxford: 2008.

16. Harma M, Alhava E. Is lactose malabsorption a risk factor in fractures of the elderly? Annales Chirurgiae et Gynaecologiae. 1988; 77:180–3. [PubMed: 3267105]

17. Davey Smith G, Lawlor DA, Timpson N, Baban J, Keissling M, Day IN, et al. Lactase persistence related genetic variant: population substructure and health outcomes. European Journal of Human Genetics. 2009; 17:357–67. [PubMed: 18797476]

18. Corazza GR, Benati G, Di Sario A, Tarozzi C, Strocchi A, Passeri M, et al. Lactose intolerance and bone mass in postmenopausal Italian women. British Journal of Nutrition. 1995; 73:479–87. [PubMed: 7766570]

19. Stephenson LS, Latham MC. Lactose intolerance and milk consumption: the relation of tolerance to symptoms. American Journal of Clinical Nutrition. 1974; 27:296–303. [PubMed: 4405996]

20. Obermayer-Pietsch BM, Gugatschka M, Reitter S, Plank W, Strele A, Walter D, et al. Adult-type hypolactasia and calcium availability: decreased calcium intake or impaired calcium absorption? Osteoporos Int. Apr; 2007 18(4):445–51. Apr. [PubMed: 17103297]

21. Mainguet P, Faille I, Destrebecq L, Devogelaer JP, Nagant de Deuxchaisnes C. Lactose intolerance, calcium intake, and osteopenia. Lancet. 1991; 338:1156–7. [PubMed: 1682584]

22. Rasinpera H, Savilahti E, Enattah NS, Kuokkanen M, Totterman N, Lindahl H, et al. A genetic test which can be used to diagnose adult-type hypolactasia in children. Gut. 2004; 53:1571–6. [PubMed: 15479673]

23. Newcomer AD, Hodgson SF, McGill DB, Thomas PJ. Lactase deficiency: prevalence in osteoporosis. Annals of Internal Medicine. 1978; 89:218–20. [PubMed: 677585]

24. Horowitz M, Wishart J, Mundy L, Nordin BE. Lactose and calcium absorption in postmenopausal osteoporosis. Archives of Internal Medicine. 1987; 147:534–6. [PubMed: 3827431]

25. Obermayer-Pietsch BM, Bonelli CM, Walter DE, Kuhn RJ, Fahrleitner-Pammer A, Berghold A, et al. Genetic predisposition for adult lactose intolerance and relation to diet, bone density, and bone fractures. J Bone Miner Res. 2004; 19:42–7. [PubMed: 14753735]

26. Newcomer AD, Thomas PJ, McGill DB, Hofmann AF. Lactase deficiency: a common genetic trait of the American Indian. Gastroenterology. 1977; 72:234–7. [PubMed: 576136]

27. Savaiano DA, Boushey CJ, McCabe GP, Savaiano DA, Boushey CJ, McCabe GP. Lactose intolerance symptoms assessed by meta-analysis: a grain of truth that leads to exaggeration. Journal of Nutrition. 2006; 136:1107–13. [PubMed: 16549489]

28. Hsu CC, Chow W-H, Boffetta P, Moore L, Zaridze D, Moukeria A, et al. Dietary risk factors for kidney cancer in Eastern and Central Europe. American Journal of Epidemiology. 2007; 166:62–70. [PubMed: 17456477]

29. Harris R. metan: fixed- and random-effects meta-analysis. Stata Journal. 2008; 8:3–28.

Modifiable risk
factor

MILK DRINKING

Behavioural
effect

Observational
association

Unconfounded effect free from
reverse causation?

MCM6 **GENOTYPE**

**RENAL CELL
CARCINOMA RISK**

Non-confounded and unbiased
proxy of milk consumption

Outcome
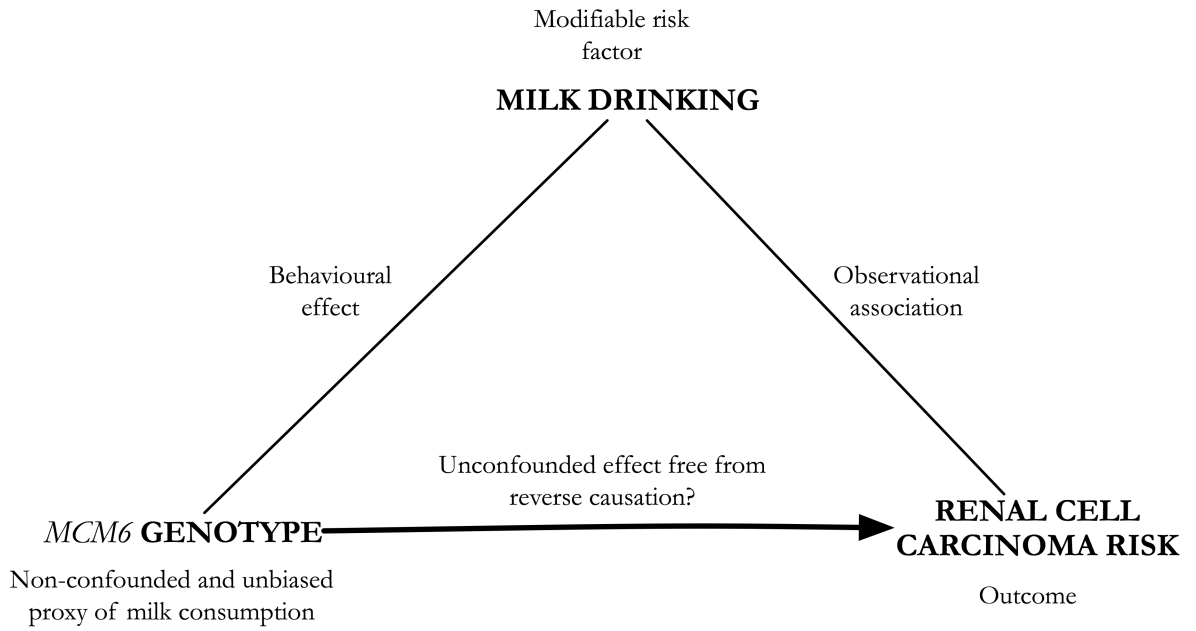
Figure 1. Mendelian randomization framework for the analysis of RCC risk by milk consumption
In this framework, the observational association between milk drinking and RCC is scrutinized by the use of genetic variation that is related to the exposure of interest (milk drinking) and potentially to the outcome of risk (RCC), but not to other possibly confounding factors. As such, genotype may act here as an "instrument" for the reassessment of the originally tentative observational finding.

Table 1

Characteristics of the control participants in each of the 4 countries.

Country	Variable	Cases n	Controls n
all	Case/control status (%)	953 (28.5)	2,396 (71.5)
	Mean age (95%CI)	59.4 (58.8, 60.1)	59.5 (59.1, 59.9)
	Sex (% men)	564 (59.2)	1,715 (71.6)
	Education (% high vs rest)	293 (30.9)	601 (25.2)
	Alcohol drinking (% never)	94 (9.9)	203 (8.5)
	Tobacco smoking (% never vs rest)	447 (47.1)	834 (34.8)
	Milk consumption (% ever)	841 (88.3)	2,030 (84.7)
Romania	Case/control status (%)	90 (33.6)	178 (66.4)
	Mean age (95%CI)	59.5 (57.2, 61.8)	57.5 (55.8, 59.3)
	Sex (% men)	60 (66.7)	115 (64.6)
	Education (% high vs rest)	26 (28.9)	30 (16.9)
	Alcohol drinking (% never)	9 (10.0)	23 (12.9)
	Tobacco smoking (% never vs rest)	34 (37.8)	82 (46.1)
	Milk consumption (% ever)	88 (97.8)	174 (97.8)
Poland	Case/control status (%)	81 (9.1)	805 (90.9)
	Mean age (95%CI)	59.9 (57.8, 62.0)	59.7 (59.1, 60.4)
	Sex (% men)	49 (60.5)	549 (68.2)
	Education (% high vs rest)	22 (27.2)	183 (22.8)
	Alcohol drinking (% never)	7 (8.6)	56 (7.0)
	Tobacco smoking (% never vs rest)	30 (37.0)	228 (28.3)
	Milk consumption (% ever)	69 (85.2)	690 (85.7)
Russia	Case/control status (%)	288 (26.5)	797 (73.5)
	Mean age (95%CI)	58.5 (57.2, 59.7)	59.2 (58.5, 59.9)

Country	Variable	Cases n	Controls n
	Sex (% men)	148 (51.4)	643 (80.7)
	Education (% high vs rest)	135 (46.9)	215 (27.0)
	Alcohol drinking (% never)	27 (9.4)	43 (5.4)
	Tobacco smoking (% never vs rest)	160 (55.6)	263 (33.0)
	Milk consumption (% ever)	239 (83.3)	648 (81.3)
C.Republic	Case/control status (%)	494 (44.5)	616 (55.5)
	Mean age (95%CI)	59.9 (59.0, 60.8)	60.1 (59.3, 60.9)
	Sex (% men)	307 (62.2)	408 (66.2)
	Education (% high vs rest)	110 (22.5)	173 (28.2)
	Alcohol drinking (% never)	51 (10.4)	81 (13.2)
	Tobacco smoking (% never vs rest)	223 (45.4)	261 (42.4)
	Milk consumption (% ever)	445 (90.1)	518 (84.1)

(Milk consumption is defined from the categories: never [0], <once per month [1], <once per week [2], 1-2 times per week [3], 3-5 times per week [4], and daily [5]). Scores are united into 0 versus 1+2+3+4+5.)

Table 2

Observed relationship between milk consumption in controls and variation at rs4988235 in Eastern European populations.

		Non-persistent CC		Persistent CT+TT		P_{unadj}	OR_{adj}	P_{adj}
		N	%	N	%			
Romania	**Ever**	93	97.9	75	97.4	0.8	0.75 (0.10, 5.97)	0.8
	Never	2	2.1	2	2.6			
Poland	**Ever**	75	70.1	113	83.1	0.02	2.58 (1.31, 5.11)	0.006
	Never	32	29.9	23	16.9			
Russia	**Ever**	231	78.0	445	89.9	<0.0001	2.50 (1.60, 3.88)	<0.0001
	Never	65	22.0	50	10.1			
Czech Rep.	**Ever**	244	75.3	396	85.7	0.0002	2.38 (1.55, 3.63)	<0.0001
	Never	80	24.7	66	14.3			
All countries		Meta-analysis (n=1992)					2.39 (1.81, 3.15)	6.9e-10
Heterogeneity				I^2=0%			P_{het}=0.7	

Numbers and proportion of *controls* by genotype and milk drinking category. P_{unadj} represents an unadjusted chi-square test, whilst OR_{adj} and P_{adj} represent a logistic regression analysis for the odds of being in the persistence group by milk drinking status adjusted for age, sex, education, smoking and drinking.

Table 3

Observed relationship between milk drinking and renal cell carcinoma group in Eastern European populations.

	Cases		Controls		OR (95%CI) (ever vs never)	P
	Ever	Never	Ever	Never		
Romania	88 (97.8)	2 (2.2)	174 (97.8)	4 (2.2)	1.57 (0.17, 14.73)	0.7
Poland	69 (85.2)	12 (14.8)	690 (85.7)	115 (14.3)	0.96 (0.50, 1.87)	0.9
Russia	239 (83.3)	48 (16.7)	648 (81.3)	149 (18.7)	1.18 (0.75, 1.84)	0.5
Czech Rep.	445 (90.1)	49 (9.9)	518 (84.1)	98 (15.9)	1.68 (1.13, 2.49)	0.01
All countries	Meta-analysis				1.35 (1.03, 1.76)	0.03
Heterogeneity	I^2=0%				P_{het}=0.5	

Numbers and proportion of individuals by RCC status and milk drinking category. P represents logistic regression adjusted for age, sex, education, smoking and drinking.

Table 4

Observed relationship between renal cell carcinoma risk and rs4988235 genotype group in Eastern European populations.

	Cases		Controls		OR (95%CI)	p
	CC	CT+TT	CC	CT+TT	(CT+TT vs CC)	
Romania	51 (59.3)	35 (40.7)	95 (55.2)	77 (44.8)	0.86 (0.48, 1.53)	0.6
Poland	30 (37.0)	51 (63.0)	296 (37.4)	495 (62.6)	1.08 (0.67, 1.76)	0.8
Russia	121 (42.5)	164 (57.5)	324 (41.2)	462 (58.8)	0.87 (0.63, 1.23)	0.4
Czech Rep.	126 (27.2)	337 (72.8)	176 (29.5)	421 (70.5)	1.15 (0.85, 1.55)	0.4
All countries	Meta-analysis				1.01 (0.83, 1.22)	0.9
Heterogeneity	I^2=0%				P_{het}=0.6	

Numbers and proportion of individuals by RCC status and lactase persistence genotype. P represents logistic regression adjusted for age, sex, education, smoking and drinking.

Table 5

Observed relationship between milk consumption and genotype in Eastern European countries and potentially confounding factors to the relationship between milk consumption and renal cell carcinoma risk.

Country	Variable	OR Milk Consumption (95% CI)	n	p	OR Genotype (95% CI)	n	p
Romania	Education	0.89 (0.15, 5.42)	178	0.9	1.49 (0.86, 2.59)	172	0.2
	Alcohol drinking	0.46 (0.05, 4.60)	170	0.5	0.40 (0.15, 1.08)	164	0.07
	Tobacco smoking	0.87 (0.29, 2.62)	178	0.8	1.012 (0.73, 1.42)	172	0.9
Poland	Education	1.02 (0.64, 1.61)	802	0.9	1.40 (1.00, 1.95)	788	0.05
	Alcohol drinking	0.46 (0.24, 0.88)	682	0.02	0.78 (0.45, 1.36)	670	0.3
	Tobacco smoking	1.00 (0.79, 1.28)	805	0.9	1.18 (0.99, 1.41)	791	0.06
Russia	Education	1.55 (1.12, 2.14)	797	0.008	1.04 (0.81, 1.35)	786	0.7
	Alcohol drinking	0.50 (0.25, 1.00)	650	0.05	0.95 (0.51, 1.77)	640	0.9
	Tobacco smoking	1.01 (0.82, 1.23)	797	0.95	0.95 (0.81, 1.11)	786	0.5
C.Republic	Education	1.57 (1.09, 2.27)	613	0.02	1.17 (0.87, 1.57)	594	0.3
	Alcohol drinking	2.08 (0.97, 4.49)	533	0.06	0.80 (0.48, 1.33)	515	0.4
	Tobacco smoking	0.70 (0.54, 0.91)	615	0.007	0.89 (0.72, 1.09)	596	0.3
All	Education	1.41 (1.14, 1.75)	-	0.001	1.19 (1.01, 1.40)	-	0.03
	Alcohol drinking	0.73 (0.32, 1.66)	-	0.4	0.77 (0.57, 1.05)	-	0.1
	Tobacco smoking	0.90 (0.74, 1.10)	-	0.3	1.01 (0.88, 1.15)	-	0.9

Odds ratios presented from logistic regression of lactase persistence genotype categories (rs4988235 CCvsCT/TT) or binary milk consumption status on confounding factors. Education is represented by the categorical variable low/med/high attainment; Alcohol drinking by the binary variable ever/never consumed and Tobacco smoking by the vinary variable ever/never consumed. Results displayed above are restricted to control subjects and are unadjusted. Results for all countries are derived from meta-analyses of estimates by country.