American Journal of Epidemiology
Vol. 166, No. 5
DOI: 10.1093/aje/kwm113
Advance Access publication June 27, 2007

Original Contribution

Lactase Persistence and Bitter Taste Response: Instrumental Variables and Mendelian Randomization in Epidemiologic Studies of Dietary Factors and Cancer Risk

Carlotta Sacerdote[1], Simonetta Guarrera[2], George Davey Smith[3], Sara Grioni[4], Vittorio Krogh[4], Giovanna Masala[5], Amalia Mattiello[6], Domenico Palli[5], Salvatore Panico[6], Rosario Tumino[7], Fabrizio Veglia[2], Giuseppe Matullo[2], and Paolo Vineis[2,8]

[1] Centre for Cancer Epidemiology and Prevention (CPO Piemonte), Turin, Italy.
[2] Institute for Scientific Interchange (ISI) Foundation, Turin, Italy.
[3] MRC Centre for Causal Analyses in Translational Epidemiology, Department of Social Medicine, University of Bristol, Bristol, United Kingdom.
[4] Epidemiology Unit, National Cancer Institute, Milan, Italy.
[5] Molecular and Nutritional Epidemiology Unit, Centre for the Study of Cancer Prevention (CSPO), Scientific Institute of Tuscany, Florence, Italy.
[6] Department of Clinical and Experimental Medicine, Federico II University, Naples, Italy.
[7] Ragusa Cancer Registry, Ragusa, Italy.
[8] Department of Epidemiology and Public Health, Imperial College London, London, United Kingdom.

Received for publication October 16, 2006; accepted for publication March 7, 2007.

Consumption of dairy products seems to increase the risk of cancer at several sites, while intake of cruciferous vegetables could have protective effects. However, these dietary intakes are subject to measurement error, and associations with cancer could be due to confounders. Mendelian randomization has been suggested as a way to overcome confounding by exploiting the random allocation of alleles from parents to offspring. In mid-2006, the authors conducted a study of allele frequencies for the lactase (*LCT*) and taste receptor, type 2, member 38 (*TAS2R38*) genes, including 634 volunteers recruited (1992–1998) from the Italian branch of the European Prospective Investigation into Cancer and Nutrition. The authors hypothesized that there would be a lower milk intake among carriers of the *LCT* CC genotype and a different intake of cruciferous vegetables among carriers of the *TAS2R38* variant. Overall, the frequency of the *LCT* T allele was higher in northern Italy than in southern Italy. Food intake was associated with gene variants. An association was evident for ice cream and *LCT* variants ($p = 0.004$); less so for milk intake. In addition, the *TAS2R38* variant showed a geographic gradient and an association with cruciferous vegetable intake. These results suggest that the *LCT* and *TAS2R38* variants are good candidates for Mendelian randomization studies of cancer and other health outcomes.

diet; lactase; neoplasms; polymorphism, genetic; taste; T2R taste receptors

Abbreviations: EPIC, European Prospective Investigation into Cancer and Nutrition; *LCT*, lactase [gene]; *TAS2R38*, taste receptor, type 2, member 38 [gene].

It has been suggested that intake of milk could increase the risk of prostate and testicular cancers (1). The hypothesized mechanism for the development of prostate cancer is the inhibitory effect of calcium on the conversion of

Correspondence to Dr. Paolo Vineis, Department of Epidemiology and Public Health, Imperial College London, Norfolk Place, London W2 1PG, United Kingdom (e-mail: p.vineis@imperial.ac.uk).

TABLE 1. Distribution of lactase (*LCT*) genotypes and taste receptor, type 2, member 38 (*TAS2R38*) haplotype variants by study center in the Italian branch of the European Prospective Investigation into Cancer and Nutrition, 2006

Study center	*LCT* genotype						*TAS2R38* haplotype			
	CC (nondigesters)		CT (poor digesters)		TT (digesters)		PA/PA, AV/PA, or AA/PA (tasters)		AV/AV (nontasters)	
	No.	%	No.	%	No.	%	No.	%	No.	%
Torino	67	63.8	27	25.7	11	10.5	66	65.4	35	34.7
Varese	62	59.1	32	30.5	11	10.7	69	67.7	33	32.4
Firenze	157	77.0	43	21.1	4	2.0	146	74.9	49	25.1
Napoli	44	88.0	5	10.0	1	2.0	39	79.6	10	20.4
Ragusa	122	86.5	16	11.4	3	2.1	99	68.8	45	31.3
p value*			<0.0001					0.22		

* *p* value from χ^2 test.

25-hydroxyvitamin D to 1,25-dihydroxyvitamin D. The latter has an antiproliferative effect in human prostate cells (2). The presence of female hormones in milk could be a determinant of the relation between dairy product intake and testicular cancer (3). Intake of cruciferous vegetables, possibly through the induction of phase II metabolizing enzymes, could have a protective effect against cancer at several sites, including the prostate, lung, and colorectum (4).

However, both milk and cruciferous vegetable intakes are subject to considerable measurement error, and the association with cancer could be due to confounders such as exposures/behaviors associated with social class. Mendelian randomization has been proposed (5) as a way to overcome confounding by exploiting the random allocation of alleles from parents to offspring. The association between a gene variant and a disease is not subject to the confounding by behavioral or socioeconomic factors that has clearly led to misleading findings in conventional observational epidemiologic studies, nor does reverse causation or other biases inherent in observational research apply to studies based on Mendelian randomization (6). If one can show that a genetic variant (e.g., one related to lactase persistence) that affects milk intake is associated with cancer, this will be indirect but unconfounded evidence for a role of milk in carcinogenesis. The same applies to a variant of a gene that influences intake of cruciferous vegetables (e.g., the taste receptor, type 2, member 38 (*TAS2R38*) gene).

Lactase persistence is a dominant trait controlled by the lactase (*LCT*) gene. People who are homozygous for the recessive C allele of the intron C > T (position −13910) polymorphism (rs4988235) have almost undetectable levels of intestinal lactase production compared with people with *LCT* TC and *LCT* TT (7–11). For evolutionary reasons that are not totally clear, humans generally undergo a physiologic shutdown of lactase activity after weaning (8). However, in some populations, a mutation occurred—apparently around 5,000–10,000 years ago—that led to lactase persistence. People with lactase persistence are much more common in Northern Europe; for example, the frequency of the T allele is 80 percent in the United Kingdom but only 10–20 percent

TABLE 2. Distribution of lactase (*LCT*) genotypes and taste receptor, type 2, member 38 (*TAS2R38*) haplotype variants by area of birth and area of residence and related odds ratios in the Italian branch of the European Prospective Investigation into Cancer and Nutrition, 2006

	Area of birth						Area of residence					
	North		Center-South		OR*	95% CI*	North		Center-South		OR	95% CI
	No.	%	No.	%			No.	%	No.	%		
LCT genotype												
TT (digesters)	21	12.3	8	1.9	1.00†		22	10.5	8	2.0	1.00†	
CT (poor digesters) or CC (nondigesters)	150	87.7	422	98.1	7.38	3.32, 18.07	188	89.5	387	98.0	5.66	2.57, 13.77
TAS2R38 haplotype												
AV/AV (nontasters)	58	35.2	112	26.5	1.00†		68	33.5	104	26.8	1.00†	
PA/PA, AV/PA, or AA/PA (tasters)	107	64.9	311	73.5	0.66	0.45, 0.98	135	66.5	284	73.2	0.73	0.50, 1.05

* OR, odds ratio; CI, confidence interval.
† Reference category.

TABLE 3. Mean and median intakes of dairy products by lactase (*LCT*) genotype in the Italian branch of the European Prospective Investigation into Cancer and Nutrition, 2006

LCT genotype	Dairy product intake (g/day)				
	Milk	Yogurt	Cheese	Butter	Ice cream
CC (nondigesters)					
Mean (SD*)	110.4 (146.4)	34.2 (85.5)	62.3 (45.2)	1.8 (3.1)	19.6 (20.9)
Median	90.9	2.1	53.1	0.6	12.9
CT (poor digesters)					
Mean (SD)	126.0 (145.8)	31.8 (44.2)	57.3 (43.2)	1.2 (2.5)	17.7 (15.5)
Median	108.0	8.3	50.6	0.6	12.9
TT (digesters)					
Mean (SD)	130.3 (144.4)	44.8 (69.8)	62.1 (46.3)	1.4 (2.3)	36.9 (26.3)
Median	120.0	12.5	53.0	0.4	22.7
p value†	0.50	0.72	0.55	0.07	0.004

* SD, standard deviation.
† *p* value from analysis of variance.

in Southern Europe (9). Grand et al. (10) have found a haplotype covering approximately 1 million nucleotide bases that includes the *LCT* gene. In spite of the unusual block length (not being broken down by meiotic recombination), the presence of a shared haplotype and the presence of the same DNA variants in nonpersistence alleles in different, distantly related populations together suggest that the persistence variant is relatively old (11). Enattah et al. (11) have hypothesized that this haplotype underwent strong selective pressure around 5,000–10,000 years ago, when dairy farming arose. A separate origin for lactase persistence in African populations has recently been demonstrated (12).

Phenylthiocarbamide is a bitter chemical whose bitter taste is not appreciated by approximately 25 percent of the population. The *TAS2R38* gene is a member of the bitter taste receptor family. Two single nucleotide polymorphisms in the *TAS2R38* gene define the bitter taste response: Pro49Ala C > G (rs713598) and Ala262Val C > T (rs1726866). These polymorphisms allow construction of two haplotypes, which refer to taster and nontaster status, respectively (13). The molecular basis of the differential response to phenylthiocarbamide determined by these haplotypes has been recently demonstrated (14). Bitter-tasting compounds in cruciferous vegetables resemble phenylthiocarbamide; therefore, the analyzed *TAS2R38* gene variants may be related to cruciferous vegetable intake (15).

Before directly testing the association between *LCT* and *TAS2R38* variants and cancer, we have conducted a study on the geographic distribution of allele frequencies at different locations in Italy and the association between variant alleles and the intake of milk derivatives or cruciferous vegetables. We hypothesized 1) a north-to-south decreasing gradient for the T allele frequency for *LCT*, based on previous observations of such a gradient in Europe; 2) a lower intake of milk (but not yogurt or cheese, in which lactose is hydrolyzed) among carriers of the CC genotype for *LCT* (lactase nonpersistence); and 3) an association between cruciferous vegetable intake and variant *TAS2R38* haplotypes.

MATERIALS AND METHODS

The European Prospective Investigation into Cancer and Nutrition (EPIC) is a multicenter European study in which approximately 520,000 healthy volunteers were recruited, between 1992 and 1998, in 10 European countries (Sweden, Denmark, the Netherlands, Norway, the United Kingdom, France, Germany, Spain, Italy, and Greece). EPIC investigators have collected very detailed dietary and lifestyle histories through self-administered questionnaires, plus a 24-hour dietary recall administered through person-to-person interview, anthropologic measurements, and a 30- to 40-ml blood sample. Details on EPIC study procedures are given elsewhere (16).

In mid-2006, we analyzed polymorphisms in the *LCT* and *TAS2R38* genes in 634 healthy subjects randomly sampled from the Italian EPIC cohort. In these 634 persons, we also measured bulky DNA adducts by means of P32-postlabeling.

TABLE 4. Mean and median intakes of cruciferous vegetables by taste receptor, type 2, member 38 (*TAS2R38*) haplotype in the Italian branch of the European Prospective Investigation into Cancer and Nutrition, 2006

TAS2R38 haplotype	Cruciferous vegetable intake (g/day)	
	Mean (SD*)	Median
PA/PA, AV/PA, or AA/PA (tasters)	10.2 (15.0)	5.1
AV/AV (nontasters)	13.6 (21.2)	8.2
p value†	0.02	

* SD, standard deviation.
† *p* value from Wilcoxon's two-sample test.

TABLE 5. Mean intake of dairy products by study center and lactase (*LCT*) genotype in the Italian branch of the European Prospective Investigation into Cancer and Nutrition, 2006

Study center and *LCT* genotype	Dairy product intake (g/day)				
	Milk	Yogurt	Cheese	Butter	Ice cream
Varese					
CC (nondigesters)	112.3 (139.1)*	38.4 (67.7)	70.5 (59.0)	2.8 (4.8)	18.3 (16.6)
CT (poor digesters)	186.2 (178.1)	30.3 (42.7)	64.6 (42.2)	1.6 (2.0)	19.0 (22.7)
TT (digesters)	145.7 (173.6)	64.7 (98.3)	57.1 (40.0)	0.3 (0.5)	46.7 (109.6)
Torino					
CC (nondigesters)	104.5 (101.1)	33.4 (42.7)	65.4 (46.3)	1.5 (3.1)	15.8 (3.1)
CT (poor digesters)	109.1 (116.0)	42.2 (48.2)	45.6 (32.7)	1.1 (1.1)	8.1 (6.9)
TT (digesters)	126.5 (125.2)	40.1 (50.1)	68.8 (55.5)	1.8 (2.5)	12.5 (8.5)
Firenze					
CC (nondigesters)	144.4 (186.7)	51.9 (122.1)	55.6 (33.9)	1.9 (3.0)	18.6 (22.8)
CT (poor digesters)	116.3 (142.7)	36.3 (47.5)	58.3 (34.3)	1.1 (1.3)	15.7 (14.4)
TT (digesters)	159.2 (174.8)	34.4 (39.5)	58.6 (24.5)	3.6 (3.8)	35.8 (28.0)
Ragusa					
CC (nondigesters)	68.8 (91.7)	9.6 (31.1)	65.2 (48.5)	1.2 (1.7)	20.0 (21.7)
CT (poor digesters)	59.1 (80.7)	6.2 (14.5)	59.2 (48.2)	0.9 (1.6)	10.8 (9.8)
TT (digesters)	48.9 (63.0)	2.8 (4.8)	60.3 (72.8)	0.7 (2.3)	34.3 (48.3)

* Numbers in parentheses, standard deviation.

A 5′-nuclease assay (TaqMan) was used to genotype all of the polymorphisms, using fluorogenic MGB (minor groove binder) probes purchased by Applied Biosystems (Foster City, California). (Information on probes and polymerase chain reaction conditions for genotyped single nucleotide polymorphisms can be obtained from the authors upon request.) We analyzed the *TAS2R38* Pro49Ala (C > G) and Ala262Val (T > C) missense polymorphisms and the *LCT* T/C intron polymorphism.

Haplotypes were reconstructed and their frequencies were estimated from genotype data using a Bayesian statistical method based on the coalescence theory, implemented in the PHASE package, version 2.0 (17); individual phases were assigned to each participant through the use of default settings. *TAS2R38* AV/AA haplotypes with lower frequency and undefined taster status have been excluded.

We computed mean and median intakes of milk, yogurt, cheese, butter, and ice cream according to *LCT* variants and estimated the statistical strength of the differences by analysis of variance. We repeated the same analyses for cruciferous vegetables and *TAS2R38* haplotypes, estimating the statistical strength of the differences using Wilcoxon's two-sample test. Hardy-Weinberg equilibrium was evaluated by chi-squared test.

RESULTS

Both the *LCT* and the *TAS2R38* genotypes were in Hardy-Weinberg equilibrium. Table 1 shows the distribution of *LCT* and *TAS2R38* haplotypes by EPIC study center. Overall, the frequency of the *LCT* gene T allele was approximately 21 percent in northern Italy and 9 percent in southern Italy, that is, much lower than in Northern Europe. In particular, the TT genotype was very rare in southern Italy (around 2 percent). The regional differences were less pronounced for the *TAS2R38* gene. As would be expected, the area of birth and area of residence of the EPIC participants were strong predictors of *LCT* genotypes and *TAS2R38* haplotype variants (table 2). Interestingly, migrants from southern Italy who were living near the northern

TABLE 6. Mean intake of cruciferous vegetables by study center and taste receptor, type 2, member 38 (*TAS2R38*) haplotype in the Italian branch of the European Prospective Investigation into Cancer and Nutrition, 2006

Study center and *TAS2R38* haplotype	Cruciferous vegetable intake (g/day)
Varese	
PA/PA, AV/PA, or AA/PA (tasters)	7.1 (6.5)*
AV/AV (nontasters)	9.0 (11.0)
Torino	
PA/PA, AV/PA, or AA/PA (tasters)	5.2 (4.8)
AV/AV (nontasters)	9.1 (11.2)
Firenze	
PA/PA, AV/PA, or AA/PA (tasters)	9.6 (15.2)
AV/AV (nontasters)	10.8 (14.0)
Ragusa	
PA/PA, AV/PA, or AA/PA (tasters)	17.0 (21.2)
AV/AV (nontasters)	23.8 (33.6)

* Numbers in parentheses, standard deviation.

study centers also had an uneven distribution of alleles, particularly for *LCT*: Twenty-nine (78.3 percent) subjects born in the south had the CC genotype, while eight (21.6 percent) had the CT genotype and none had the TT genotype; corresponding numbers for persons born in the north were 98 (57.6 percent), 51 (30.0 percent), and 21 (12.3 percent), respectively ($p = 0.024$). For *TAS2R38* haplotypes, the distributions were 29 (78.3 percent) tasters and eight (21.6 percent) nontasters for those born in the south and 107 tasters (64.8 percent) and 58 nontasters (35.1 percent) for those born in the north ($p = 0.11$).

Tables 3 and 4 show mean and median intakes of relevant foods by genetic variant. A clear association was evident for ice cream and *LCT* variants ($p = 0.004$); less so for milk intake. As expected, no difference was found for yogurt and cheese, since lactose is hydrolyzed in these products. An association was also evident for intake of cruciferous vegetables and *TAS2R38* haplotypes ($p = 0.02$).

The association between dietary intakes and genotypes was not explained by population stratification or other characteristics of local populations. As tables 5 and 6 show, intakes of the relevant foods differed between genotypes at each of the study centers involved. For *LCT*, apparently only the TT genotype, and not the heterozygous genotype, was associated with different intakes of dairy products.

DISCUSSION

Our study confirms the existence of a considerable difference between Northern and Southern Europe for the frequency of the T allele of the *LCT* gene, also suggesting a gradient for the *LCT* variant within Italy. In addition, it confirms that persons with the *LCT* CC genotype have a lower intake of dairy products, particularly ice cream. The latter observation has the possible explanation that ice cream intake might be less susceptible to classification error than milk intake. As predicted, no association was found for cheese and yogurt. Most studies have found that lactase persistence is related to milk consumption (18–24), although other studies, generally with small samples, have observed no association (25–27).

The data showed an association of the *TAS2R38* variants with intake of cruciferous vegetables, as hypothesized. The direction of the association was clear, since people carrying the haplotypes associated with the inability to taste bitter compounds (PA/PA, AV/PA, and AA/PA), which make cruciferous vegetables more palatable, had significantly greater mean and median intakes of cruciferous vegetables. A study of food preferences in children found that tasters selected fewer vegetables to consume than nontasters (28). In a study of 4,286 British women, Timpson et al. (29) found no association between taster status and consumption of green vegetables. Timpson et al. suggested that the lack of association could have been related to a British ''debittering'' approach to cooking, with long boiling periods and the addition of salt, sugar, or fat that reduces the bitter taste of vegetables and makes them more palatable (29).

In conclusion, our study suggests that the *LCT* and *TAS2R38* gene variants have a geographic gradient and are associated with the expected phenotypes. Therefore, they are good candidates for Mendelian randomization studies of cancer and other health outcomes. However, the relatively weak associations between the variants and the exposures of interest mean that large study samples will be required when relating these variants to disease outcomes and imputing causal effects of the proxy exposures. In addition, given the geographic distribution of the variants across Italy (and other areas), investigators will need to take population stratification into account when relating the variants to disease outcomes (30).

ACKNOWLEDGMENTS

This work was made possible by grants from the Compagnia di San Paolo to the ISI Foundation and from the Associazione Italiana per la Ricerche sul Cancro to Dr. Giuseppe Matullo.

Conflict of interest: none declared.

REFERENCES

1. Ganmaa D, Li XM, Wang J, et al. Incidence and mortality of testicular and prostatic cancers in relation to world dietary practices. Int J Cancer 2002;98:262–7.
2. Skowronski RJ, Peehl DM, Feldman D. Vitamin D and prostate cancer: 1,25 dihydroxyvitamin D3 receptors and actions in human prostate cancer cell lines. Endocrinology 1993;132:1952–60.
3. Ganmaa D, Wang PY, Qin LQ, et al. Is milk responsible for male reproductive disorders? Med Hypotheses 2001;57:510–14.
4. Murillo G, Mehta RG. Cruciferous vegetables and cancer prevention. Nutr Cancer 2001;41:17–28.
5. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 2003;32:1–22.
6. Smith GD, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. Int J Epidemiol 2004;33:30–42.
7. Gugatschka M, Dobnig H, Fahrleitner-Pammer A, et al. Molecularly-defined lactose malabsorption, milk consumption and anthropometric differences in adult males. QJM 2005;98:857–63.
8. Swallow DM. Genetics of lactase persistence and lactose intolerance. Annu Rev Genet 2003;37:197–219.
9. Weiss KM. The unkindest cup. Lancet 2004;363:1489–90.
10. Grand RJ, Montgomery RK, Chitkara DK, et al. Changing genes; losing lactase. Gut 2003;52:617–19.
11. Enattah NS, Sahi T, Savilahti E, et al. Identification of a variant associated with adult-type hypolactasia. Nat Genet 2002;30:233–7.
12. Tishkoff SA, Reed FA, Ranciaro A, et al. Convergent adaptation of human lactase persistence in Africa and Europe. Nat Genet 2007;39:31–40.
13. Kim UK, Jorgenson E, Coon H, et al. Positional cloning of the human quantitative trait locus underlying taste sensitivity to phenylthiocarbamide. Science 2003;299:1221–5.

14. Bufe B, Breslin PA, Kuhn C, et al. The molecular basis of individual differences in phenylthiocarbamide and propylthiouracil bitterness perception. Curr Biol 2005;15:322–7.
15. Bufe B, Meyerhof W. Taste receptors and their variants. In: Brigelius-Flohé R, Joost H-G, eds. Nutritional genomics: impact on health and disease. Weinheim, Germany: Wiley-VCH Verlag GmbH & Company, 2006:386–411.
16. 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.
17. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 2001;68:978–89.
18. Obermayer-Pietsch BM, Bonelli CM, Walter DE, 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.
19. Newcomer AD, Gordon H, Thomas PJ, et al. Family studies of lactase deficiency in the American Indian. Gastroenterology 1977;73:985–8.
20. Honkanen R, Pulkkinen P, Jarvinen R, et al. Does lactose intolerance predispose to low bone density? A population-based study of perimenopausal Finnish women. Bone 1996;19:23–8.
21. Horowitz M, Wishart J, Mundy L, et al. Lactose and calcium absorption in postmenopausal osteoporosis. Arch Intern Med 1987;147:534–6.
22. Newcomer AD, Hodgson SF, McGill DB, et al. Lactase deficiency: prevalence in osteoporosis. Ann Intern Med 1978;89:218–20.
23. Rasinpera H, Savilahti E, Enattah NS, et al. A genetic test which can be used to diagnose adult-type hypolactasia in children. Gut 2004;53:1571–6.
24. Mainguet P, Faille I, Destrebecq L, et al. Lactose intolerance, calcium intake, and osteopenia. Lancet 1991;338:1156–7.
25. Stephenson LS, Latham MC. Lactose intolerance and milk consumption: the relation of tolerance to symptoms. Am J Clin Nutr 1974;27:296–303.
26. Corazza GR, Benati G, Di Sario A, et al. Lactose intolerance and bone mass in postmenopausal Italian women. Br J Nutr 1995;73:479–87.
27. Harma M, Alhava E. Is lactose malabsorption a risk factor in fractures of the elderly? Ann Chir Gynaecol 1988;77:180–3.
28. Bell KI, Tepper BJ. Short-term vegetable intake by young children classified by 6-n-propylthoiuracil bitter-taste phenotype. Am J Clin Nutr 2006;84:245–51.
29. Timpson NJ, Christensen M, Lawlor DA, et al. *TAS2R38* (phenylthiocarbamide) haplotypes, coronary heart disease traits, and eating behavior in the British Women's Heart and Health Study. Am J Clin Nutr 2005;81:1005–11.
30. Campbell CD, Ogburn EL, Lunetta KL, et al. Demonstrating stratification in a European American population. Nat Genet 2005;37:868–72.