European lactase persistence genotype shows evidence of association with increase in body mass index

Johannes Kettunen1,2,3,*, Kaisa Silander2,3, Olli Saarela2, Najaf Amin5, Martina Müller6,7,8, Nicholas Timpson9, Ida Surakka2,3, Samuli Ripatti2,3, Jaana Lahtinen10, Anna-Liisa Hartikainen11, Anneli Pouta11,12, Päivi Lahermo4, Verneuri Anttila1, Satu Männistö2, Antti Jula2, Jarmo Virtamo2, Veikko Salomaa2, Terho Lehtimäki13, Olli Raitakari14, Christian Gieger7, Erich H. Wichmann7,8,15, Cornelia M. Van Duijn5, George Davey Smith6, Mark I. McCarthy16,17, Marjo-Riitta Järvelin18,19,20,21, Markus Perola2,3,22 and Leena Peltonen1,2,3,22,23

1Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1HH, UK, 2Department of Chronic Disease Prevention, National Institute for Health and Welfare FIN-00251, Helsinki, Finland, 3FIMM, Institute for Molecular Medicine Finland and 4Finnish Genome Center, Institute for Molecular Medicine Finland, University of Helsinki, FIN-00251 Helsinki, Finland, 5Department of Epidemiology, Erasmus University Medical Center (MC) Rotterdam, 3000 CA, Rotterdam, The Netherlands, 6Department of Medicine I, Klinikum Grosshadern, 81377 Munich, Germany, 7Helmholtz Zentrum München, German Research Center for Environmental Health, Institute of Epidemiology, 85764 Neuherberg, Germany, 8Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology, Ludwig-Maximilians-Universität, 81377 Munich, Germany, 9MRC Centre for Causal Analyses in Translational Epidemiology, University of Bristol, Bristol BS2 8BN UK, 10Finnish National Institute of Occupational Health, FIN-90220 Oulu, Finland, 11Department of Clinical Sciences/Odontostomatology and Gynecology, University of Oulu, FIN-90014 Oulu, Finland, 12National Public Health Institute and University of Oulu, FIN-90101 Oulu, Finland, 13Department of Clinical Chemistry, Tampere University Hospital and University of Tampere, FIN-33521 Tampere, Finland, 14Department of Clinical Physiology, University of Turku and Turku University Central Hospital, FIN-20521 Turku, Finland, 15Klinikum Grosshadern, Ludwig Maximilians University, 81377 Munich, Germany, 16Oxford Centre for Diabetes, Endocrinology and Metabolism, Headington, Oxford OX3 7LJ, UK, 17The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK, 18Department of Epidemiology and Public Health, Imperial College London, London W2 1PG, UK, 19Institute of Health Sciences, University of Oulu, FIN-90014 Oulu, Finland, 20Department of Child and Adolescent Health, National Public Health Institute, FIN-90101 Oulu, Finland, 21Biocenter Oulu, University of Oulu, FIN-90014 Oulu, Finland, 22Department of Medical Genetics, University of Helsinki, FIN-00014 Helsinki, Finland and 23The Broad Institute of MIT and Harvard, Boston, MA 02142, USA

Received July 2, 2009; Revised December 4, 2009; Accepted December 14, 2009

The global prevalence of obesity has increased significantly in recent decades, mainly due to excess calorie intake and increasingly sedentary lifestyle. Here, we test the association between obesity measured by body mass index (BMI) and one of the best-known genetic variants showing strong selective pressure: the functional variant in the cis-regulatory element of the lactase gene. We tested this variant since it is presumed to provide nutritional advantage in specific physical and cultural environments. We genetically defined

*To whom correspondence should be addressed. Tel: +44 1223834244; Fax: +44 1223496826; Email: johannes@sanger.ac.uk

© The Author 2009. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.5), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
lactase persistence (LP) in 31 720 individuals from eight European population-based studies and one family study by genotyping or imputing the European LP variant (rs4988235). We performed a meta-analysis by pooling the β-coefficient estimates of the relationship between rs4988235 and BMI from the nine studies and found that the carriers of the allele responsible for LP among Europeans showed higher BMI (P = 7.9 × 10⁻⁵). Since this locus has been shown to be prone to population stratification, we paid special attention to reveal any population substructure which might be responsible for the association signal. The best evidence of exclusion of stratification came from the Dutch family sample which is robust for stratification. In this study, we highlight issues in model selection in the genome-wide association studies and problems in imputation of these special genomic regions.

INTRODUCTION

The lactase gene [LCT (MIM 603202)] is expressed in the intestinal epithelial cells and its protein product contributes to carbohydrate metabolism via digestion of the milk sugar lactose. The activity of the lactase enzyme in intestinal cells normally declines during childhood. Individuals with lactase non-persistence show varying degrees of abdominal pain, diarrhoea, bloating and gas formation after consumption of lactose-containing foods, in particular milk and other non-fermented dairy products. Molecular studies of lactose intolerance led to the identification of specific variations in a cis-regulatory element of the LCT gene resulting in lactase persistence (LP) which is an autosomal dominant condition (MIM 223100). The LP single-nucleotide polymorphism (SNP) most common among individuals of European descent (C/T, rs4988235) was first identified by Ennahal et al. (1) in Finnish lactose intolerance families. The T allele correlated perfectly with the LP phenotype of intestinal cells and functional studies have shown that the T allele is associated with the disrupted down-regulation of the LCT gene during childhood (2,3). The enhancer function of this mutation has also been shown in vitro (4,5). Subsequently, this region has yielded a spectrum of variants, some showing population specificity (6–8). These variants have been under strong positive selection regionally and seem to reflect independent events of cattle domestication, as exemplified by the two different allelic backgrounds of the Arabic and European LP alleles (6,9). The European LP allele has been associated with higher milk consumption in Finnish, Estonian and Austrian populations, one of which being the Finnish YF sample used in this study (10–15).

We hypothesized that since lactase non-persistent individuals are on a more restricted diet when compared with the LP individuals, this may have an impact on body mass index (BMI) and/or height. Lactose-containing dairy products are often high in energy content potentially increasing daily calorie intake, and for the lactase non-persistent, they may cause diarrhoea which can prevent the absorption of nutrients. Further, milk is rich in calcium, required for skeletal growth.

BMI is calculated as weight in kilograms divided by the square of height in meters. BMI has been studied extensively due to the health effects of obesity and its relative ease of measurement. Excessive weight increases the risk of type II diabetes, coronary artery disease and hypertension, and is associated with many forms of cancer along with adverse social and psychological consequences (16,17). The need for more effective clinical strategies has become evident with the increasing prevalence of obesity in recent years (18). This has put pressure on genetic studies to find variants responsible for differences in response to positive energy balance. Recent genome-wide association studies (GWASs) have been able to show association between BMI and genetic variants in a total of 13 genomic regions (19–22). In this study, we show the association between genetically defined LP and BMI.

RESULTS

The effect of LP was evaluated in five Finnish representative population cohorts, all non-ascertained for BMI or height: The North Finland Birth Cohort 1966 (NFBC 1966), the Health 2000 Health Examination Survey (Health2000), The Cardiovascular Risk of Young Finns Study (YF), the Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) study and the FINRISK which is a cross-sectional population survey targeting coronary risk factors collected every 5 years. The cohort-specific analyses were performed on corrected phenotype comparing the C T13910 allele homozygotes (lactase non-persistent, CC genotypes) against T -13910 allele carriers (LP, CT and TT genotypes) using linear regression. There was evidence for association with BMI and the LP in the ATBC cohorts (n = 2126, C allele frequency = 0.40, β = −0.12, P = 0.04) and the same direction of association in Health2000 (n = 5320, C allele frequency = 0.42, β = −0.06, P = 0.10), YF (n = 2165, C allele frequency = 0.42, β = −0.04, P = 0.51) and FINRISK (n = 2265, C allele frequency = 0.45, β = −0.09, P = 0.09) cohorts. Table 1 summarizes the cohort-specific analyses.

The CC genotype was associated with decreased BMI compared with CT/TT genotypes in the meta-analysis of 17 374 Finns (β = −0.08, P = 1.5 × 10⁻⁵). The individuals carrying the LP allele had 0.3 kg/m² higher BMI than the non-persistent individuals which corresponds to ~1 kg for an average person. The single study results from linear regression and the meta-analyses are summarized as a forest plot (Fig. 1).

We also performed a sex-specific meta-analysis and found that males (n = 9739, β = −0.09, P = 1.6 × 10⁻⁴) had somewhat higher effect estimate in the model than females (n = 7635, β = −0.06, P = 0.02), but there was no robust statistical evidence of a gender difference (P = 0.6).
We tested for the correlation of BMI between these eigenvectors in both genome-wide SNP data sets. We therefore, we tested for the correlation of BMI between eigenvectors and BMI. No correlation was used linear regression to evaluate evidence of correlation these eigenvectors in both genome-wide SNP data sets. We  

The 5th $P = 0.02$ and 10th $P = 0.01$ principal components showed most evidence of correlation with LP in NFBC 1966, and the 2nd ($P = 0.02$), 7th ($P = 0.008$) and 12th ($P = 0.008$) eigenvectors with LP in Health2000. Including these eigenvectors as covariates in the linear regression did not change association between BMI and LP in the NFBC 1966 data or in the Health2000 subsample. In NFBC 1966, the linear regression $\beta$ changed marginally ($\beta = -0.12$, $P = 0.003$), and in the Health2000 subsample, the effect size remained the same ($\beta = -0.13$, $P = 0.03$). Since the Health2000 subsample was ascertained for metabolic syndrome, we tested for the effect in both cases and controls separately, and found the linear regression $\beta$ to be equal in both groups.

Frayling et al. (19) have shown that the FTO rs9939609 A allele is positively associated with BMI. We had both LCT and FTO genotypes available in NFBC 1966 cohort for 4955 individuals. The difference between FTO AA and TT heterozygotes was 0.5 kg/m$^2$. We compared the proportion of variance in BMI explained by FTO and the proportion explained by LP. The additive model ($n = 4955$, $P = 6.7 \times 10^{-4}$) of the FTO gene association was found to explain 0.2% of the variance in BMI in the NFBC 1966 cohort. Similarly, the dominant effect of LP was found to account for 0.2% of the variance in BMI in this cohort.

We tested for the association with height and LP marker but we did not find evidence of a meaningful association in the meta-analysis in Finns ($n = 17374$, $\beta = -0.004$, $P = 0.41$). We also performed a weight analysis which—as expected, given the association with BMI but lack of association with height—demonstrated robust association ($n = 17374$, $\beta = -0.08$, $P = 1.4 \times 10^{-5}$). Both height and weight analyses were performed in the same way as the BMI analyses.

Since the initial GWASs have not identified this association in the LCT region, we tested the effect of using an additive model in the analyses instead of the correct dominant model and combined the results with meta-analysis. The additive model decreased the association ($P = 0.001$) compared with the dominant model ($P = 1.5 \times 10^{-5}$). We simulated the size of the power decrease resulting from using the incorrect

---

**Table 1. Cohort-specific results**

<table>
<thead>
<tr>
<th>Cohort</th>
<th>n (males/females)</th>
<th>LP prevalence (%)</th>
<th>$\beta$</th>
<th>SE</th>
<th>$P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFBC 1966</td>
<td>5498 (2636/2862)</td>
<td>85</td>
<td>-0.11</td>
<td>0.04</td>
<td>0.002</td>
</tr>
<tr>
<td>ATBC</td>
<td>2126 (2126/0)</td>
<td>84</td>
<td>-0.12</td>
<td>0.06</td>
<td>0.04</td>
</tr>
<tr>
<td>FINRISK</td>
<td>2265 (1555/710)</td>
<td>80</td>
<td>-0.09</td>
<td>0.05</td>
<td>0.09</td>
</tr>
<tr>
<td>Health2000</td>
<td>5320 (2437/2883)</td>
<td>82</td>
<td>-0.06</td>
<td>0.03</td>
<td>0.10</td>
</tr>
<tr>
<td>YF</td>
<td>2165 (985/1180)</td>
<td>83</td>
<td>-0.04</td>
<td>0.06</td>
<td>0.51</td>
</tr>
<tr>
<td>BWHHS</td>
<td>3109 (0/3109)</td>
<td>94</td>
<td>-0.06</td>
<td>0.07</td>
<td>0.41</td>
</tr>
<tr>
<td>ERF</td>
<td>2104 (909/1195)</td>
<td>90</td>
<td>-0.08</td>
<td>0.06</td>
<td>0.16</td>
</tr>
<tr>
<td>Rotterdam</td>
<td>5689 (3320/2369)</td>
<td>91</td>
<td>-0.02</td>
<td>0.05</td>
<td>0.69</td>
</tr>
<tr>
<td>KORA S3</td>
<td>1578 (783/795)</td>
<td>87</td>
<td>-0.001</td>
<td>0.07</td>
<td>0.99</td>
</tr>
<tr>
<td>KORA S4</td>
<td>1755 (859/896)</td>
<td>87</td>
<td>0.08</td>
<td>0.06</td>
<td>0.24</td>
</tr>
</tbody>
</table>

$n$, number of individuals; SE, standard error; LP, lactase persistence determined by the frequency of T_13910 allele carriers, $\beta$ (SE) are from the association analysis between standardized BMI and LP (rs4988235 CC against CT/TT genotypes). The nationwide Finnish cohort analyses of Health2000, YF and FINRISK were corrected for region of residence (Southwest, East and Oulu region), age and sex effects in Health2000 cohort and FINRISK cohort, with sex in NFBC 1966 cohort, with age in ATBC cohort, with age and birth place coordinates in BWHHS, with sex and age in ERF study (family-based sample), with age, sex and first 10 principal components in Rotterdam and KORA studies. Outliers deviating more than three standard deviations were removed from the analyses.

---

Figure 1. The forest plot of the linear regression analyses and meta-analyses.
additive model instead of the correct dominant model through bootstrapping. The results of this power comparison are summarized in Table 2.

We performed a dominance deviation test from additivity in each cohort separately and pooled the \( \beta \)-estimates with meta-analysis. The dominance term suggested that we could reject the additive model (\( P = 0.001, \beta = 0.047 \)).

The linear regression analysis of the British Women’s Heart and Health Study (BWHHS) showed that the difference between the LP and lactase non-persistent was 0.30 kg/m² (\( n = 3109, \beta = 0.06, P = 0.4, \) C allele frequency = 0.25) which is in line with our finding in the Finnish population but underpowered due to smaller sample size. The Dutch analysis was performed with the Erasmus Rucphen Family (ERF) study with extended pedigrees and Rotterdam study. The effect size estimate from the family data was again in line with our finding in the Finnish population but underpowered due to smaller sample size. The Dutch analysis was performed with the Erasmus Rucphen Family (ERF) study with extended pedigrees and Rotterdam study.

Table 2. The results of power calculation for using the additive model instead of correct dominant model where sample size (\( n = 17374 \)), effect size (\( \beta = 0.082 \)) and MAF (0.39) the same as the LCT variant had in combined Finnish sample of this study.

<table>
<thead>
<tr>
<th>Significance level</th>
<th>Power (%)</th>
<th>Dominant model</th>
<th>Additive model</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>98</td>
<td>83</td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>91</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>0.001</td>
<td>73</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>( 1 \times 10^{-4} )</td>
<td>51</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>( 1 \times 10^{-5} )</td>
<td>30</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>( 1 \times 10^{-6} )</td>
<td>16</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>( 1 \times 10^{-7} )</td>
<td>8</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>( 5 \times 10^{-8} )</td>
<td>6</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Although the \( Q \) statistics was not significant (\( P = 0.35 \)). Results of the individual study linear regression analyses and the meta-analyses are summarized as a forest plot (Fig. 1) and in Table 1.

### DISCUSSION

In this study, we present novel evidence of association between genetically defined LP and BMI (\( P = 7.9 \times 10^{-5} \)) in 31720 individuals from four European populations. The \( LCT \) C/T\(_{13910} \) marker frequency shows strong population differences. Campbell et al. (23), sampling from an European American population, tested for an association between height and the C/T\(_{13910} \) marker and found robust association between the marker and height if they did not correct for the ancestral origin of the samples. After the correction for ancestral origin, the association was considerably attenuated, thus indicating that the association was most likely an incidental reflection of population stratification of this variant. Similar generation of a link between the \( LCT \) C/T\(_{13910} \) marker and self-rated health has been shown to be due to population stratification in a British study (24). We tested for the association with height and the LP marker and we did not find this marker to be associated with height. The NFBC 1966 cohort and a subset of the Health2000 cohort had genome-wide SNP data available with which we could assess the effect of population substructure. The first two principal components describe the geographical origin of the individuals in NFBC 1966 cohort, as has been shown by Jakkula et al. (25). We showed that none of the first 14 eigenvectors had a material effect on the association analysis in both the regional NFBC 1966 sample and a country-wide Health2000 subsample, which strongly suggests that the association was not due to population stratification. The best evidence which would exclude the effect of stratification came from the ERF study from the Netherlands. Although the QTDT (26) analyses were not significant, the effect size was in line with the population-based studies which is a good indicator that the population-based results were not due to stratification. All five Finnish study cohorts showed the same trend of increased BMI (heterogeneity \( I^2 = 0 \)), although the association was not conventionally significant in all cohorts individually. However, after combining the cohorts in the meta-analysis, the association was robust (Fig. 1). We showed that the increase in BMI was mediated through the increase in body weight (\( P = 1.4 \times 10^{-5} \)) but not through stature (\( P = 0.41 \)). The non-Finnish population samples did not replicate the association significantly (\( P = 0.24 \)). The marker was genotyped in the British study which showed similar effect size than the Finnish sample. The cohorts which show smaller effect size have all been imputed. In our Finnish sample, we found that the imputation of this region proves to be difficult and the erroneous genotypes concentrate on the phenotype-defining genotype. This effect is shown in Finnish population and it is not possible to assess whether it is the same case in our European replication cohorts. The ERF study is more robust for this kind of error since it benefits from the family structure in the imputation. Should the same phenomenon exist in the imputed European replication samples, the effect size would be
a drastic underestimate and would explain why the European replication cohorts show smaller effect size.

We had FTO genotypes available for the NFBC 1966 cohort and were able to compare the effect size of these two variants. The LP variant had a similar effects size and explained a similar proportion of variance to FTO in this cohort.

A GWAS by Meyre et al. (27) has previously shown association in this region. They showed an association ($P = 9.2 \times 10^{-4}$) with a marker which is in LD (HapMap, $r^2 = 0.58$, $D' = 0.80$, rs2011946) with the LP variant and BMI in 135 obese adults and 794 lean adults of French origin. It is not possible to determine whether this association was due to stratification or not. They could not show association in an independent population-based cohort possibly because the cohort was underpowered ($n = 4417$). The effect in the study by Meyre et al. was in the same direction as in our study in all of the three tested adult cohorts. Any further comparison is not possible since they did not have the same marker genotyped.

There are several possible explanations as to why the LP association with BMI may have remained undetected in rest of the GWA analyses. First, the functional variant is not included in any currently used SNP arrays and the power has been reduced due to insufficient linkage disequilibrium (LD). The recently published GWASs were genotyped either by Affymetrix 5.0 or Illumina array which has less coverage in this area than the HumanHap 650 array (19–22). We found that the strongest LD ($r^2$) with the LCT C/T-13910 and any marker within 500 000 bp in the HapMap CEPH (Utah residents with ancestry from northern and western Europe) sample included in the Illumina HumanHap 650 or Affymetrix 5.0 array was 0.78 (rs309160) (28). Second, the first four GWASs had used an additive model, which given the dominant population-based cohort possibly because the cohort was underpowered ($n = 4417$). The effect in the study by Meyre et al. was in the same direction as in our study in all of the three tested adult cohorts. Any further comparison is not possible since they did not have the same marker genotyped.

The mechanism by which LP affects body composition may be related to the more restricted diet the lactase non-persistent individuals may choose or perhaps the negative symptoms such as diarrhoea after using lactose-containing food products play some role. This finding may also reflect the positive selection the LP allele has been subject to in recent history. Our study highlights the fact that the recent GWASs have not exhaustively revealed the common variants affecting BMI. Although explanation at a functional level is a matter for future studies and finding is not significant in the GWAS established genome-wide level of significance, we propose that the European variant behind LP, and perhaps other variants affecting the regulatory region of LCT, contributes to human obesity.

**MATERIALS AND METHODS**

**Study samples**

Informed consent was obtained from all participants and study was conducted according to the principles expressed in the Declaration of Helsinki and the studies were reviewed and accepted by local ethics committees. Table 3 shows the cohort characteristics. More detailed description of the study cohorts can be found at the National Biobank of Finland website (see Web Resources) and from the University of Turku website (YF cohort, see Web Resources). In brief, the NFBC 1966 includes individuals born in the Northernmost provinces of Finland with expected date of birth in 1966 who underwent a health examination at age 31. The Health2000 was conducted in 2000 to explore the health of the Finnish population. Our Health2000 cohort consists of individuals aged 30 and over. The ATBC study was initially collected to test whether α-tocopherol and β-carotene supplements would reduce the incidence of lung and other cancers. The participants were male smokers aged 50–69 years from South-western Finland. FINRISK is a cross-sectional population survey targeting coronary risk factors collected every 5 years. The current study contains individuals recruited during years 1992 and 1997 and were selected for cardiovascular study. The YF is a population-based prospective cohort study conducted at five university departments of medical schools in Finland (i.e. Turku, Helsinki, Kuopio, Tampere and Oulu), with the aim of studying the levels of cardiovascular risk factors in children and adolescents in different parts of the country. The latest follow-up was conducted in 2007. The present analysis included 2165 subjects with data on height, weight and the gene variant. The study and data collection protocols have been described in detail by Raitakari et al. (30). The Health2000, YF and FINRISK participants were collected across Finland, whereas NFBC 1966 was sampled within the Northern region and the ATBC within the South-western region of Finland. The same analyses were performed in the BWHHS. Full details of the selection of participants and measurements have been reported earlier (31,32). Women aged 60–79 years were randomly selected from primary care lists in 23 British towns. Women were recognized as non-white in the BWHHS and excluded from further analyses.

The ERF study is an extended-pedigree study (n = 2700) focused on the identification of quantitative traits loci related to neuropsychiatric, cardiovascular, endocrinological, ophthalmological and musculoskeletal disorders. The ERF sample has been described in more detail by Aulchenko et al. (33).
The Rotterdam study (34) is an ongoing prospective, population-based cohort study among 7983 persons aged 55 years and older, living in Ommoord, a district of Rotterdam, The Netherlands. The study was designed to investigate the incidence and determinants of chronic disabling diseases. The study is composed of an outbred ethnically homogeneous population of Dutch Caucasian origin. Rationale and design have been described previously. Informed consent was obtained from each participant, and the Medical Ethics Committee of the Erasmus Medical Centre Rotterdam approved the protocols. The rs4988235 marker was imputed to the ERF, Rotterdam and KORA studies. The genetic data for ERF, the Rotterdam study and KORA S4 was imputed using MACH with HAPMAP rel 22 CEU build 36 as a reference (ERF imputation, \( r^2 = 0.94 \), Rotterdam imputation study, \( r^2 = 0.92 \), KORA S4 observed versus expected variance = 0.89), whereas KORA S3 was imputed with MACH based on HAPMAP rel 21 build 35 (observed versus expected variance = 0.91). The genotype frequencies followed HWE in all four non-Finnish cohorts (BWHHS, \( P = 0.88 \); ERF, \( P = 0.23 \); Rotterdam study, \( P = 0.09 \); KORA S3, \( P = 0.75 \); KORA S4, \( P = 0.27 \)). The association analyses were performed in parallel by three collaborating groups, hence NFBC 1966 and Health2000 were analysed with PLINK 1.04, FINRISK, KORA and ATBC cohorts were analysed with R, and YF and Rotterdam study cohort with SPSS, using linear regression. The ERF study was analysed using QTDT. Lahti-Koski et al. (36) have shown that there are no regional differences in BMI within Finland in males, but a slight difference in females between southern and western Finland. Therefore, we corrected the nationwide Health2000, YF and FINRISK analyses for the area of residence (Southwest, East and Northern Finland). The NFBC 1966 cohort was collected from Northern Finland and ATBC from southwest Finland, and therefore, no regional correction was applied. The phenotypes were corrected for age and sex effects in Health2000, YF and FINRISK cohorts, with sex in NFBC 1966 cohort (all 31 years old) and age in ATBC cohort (all men). BMI outliers deviating more than three standard deviations from the mean within a cohort were removed prior to analysis, resulting in approximately normal distribution (kurtosis and skewness both between –1 and 1). The Rotterdam study and KORA were corrected for sex, age and first 10 principal components from the genome-wide data and outliers deviating more than four standard deviations from principal components or three

### Methods

In Finland, the C allele homozygosity of the \( LCT \ C/T_{-13910} \) polymorphism (rs4988235) serves as a diagnostic marker for lactase non-persistence, whereas T homozygotes and heterozygotes can be classified as LP. We tested this SNP in a set of representative population cohorts and measured its potential association with BMI and height. The SNP \( LCT \ C/T_{-13910} \) was genotyped using iPLEX assay on the MassARRAY System (Sequenom, San Diego, CA, USA) using standard protocols in a total of 17 374 Finnish individuals. The YF cohort \( LCT \ C/T_{-13910} \) genotyping was performed by using the 5′-nuclease assay and fluorogenic, allele-specific, TaqMan probes (Applied Biosystems, Foster City, CA, USA) and primers, with the ABI Prism 7000 sequence detection system (Applied Biosystems). The genotype frequencies followed Hardy–Weinberg equilibrium (HWE) in all five cohorts (YF, \( P = 0.814 \); ATBC, \( P = 0.84 \); FR, \( P = 0.77 \); NFBC, \( P = 0.93 \); Health2000, \( P = 0.54 \) and call rate was >95% in all samples. BWHHS genotypes for \( LCT \ C/T_{-13910} \) (rs4988235) were generated using the KASPar chemistry which is a competitive allele-specific PCR SNP genotyping system using FRET quencher cassette oligos using standard protocols. The rs4988235 marker was imputed to the ERF, Rotterdam and KORA studies. The genetic data for ERF, the Rotterdam study and KORA S4 was imputed using MACH with HAPMAP rel 22 CEU build 36 as a reference (ERF imputation, \( r^2 = 0.94 \), Rotterdam imputation study, \( r^2 = 0.92 \), KORA S4 observed versus expected variance = 0.89), whereas KORA S3 was imputed with MACH based on HAPMAP rel 21 build 35 (observed versus expected variance = 0.91). The genotype frequencies followed HWE in all four non-Finnish cohorts (BWHHS, \( P = 0.88 \); ERF, \( P = 0.23 \); Rotterdam study, \( P = 0.09 \); KORA S3, \( P = 0.75 \); KORA S4, \( P = 0.27 \)). The association analyses were performed in parallel by three collaborating groups, hence NFBC 1966 and Health2000 were analysed with PLINK 1.04, FINRISK, KORA and ATBC cohorts were analysed with R, and YF and Rotterdam study cohort with SPSS, using linear regression. The ERF study was analysed using QTDT.
standard deviations from standardized corrected phenotype means were removed. The ERF cohort was corrected for sex and age. The family-based analyses were performed with QTDT (26). The association method was orthogonal association which is very robust to stratification. Dominance effect was included in the analysis model.

The meta-analysis was performed by pooling the β-estimates and standard errors from the individual studies using the inverse variance method.

We had genome-wide SNP data (Illumina HumanHap 370) and phenotype data available for 4911 individuals from the NFBC 1966 cohort. Similarly, we had genome-wide SNP data (Illumina HumanHap 610) available for a subset of the Health2000 cohort (2145 individuals) ascertainment for matched case–control study for metabolic syndrome. We tested for hidden population substructure by performing a principal components analysis using EIGENSOFT (37). The imputation of the marker in these cohorts was performed with MACH using phased HAPMAP rel 22 CEU build 36 as a reference.

The power calculations were performed via bootstrapping. We performed 10 000 simulations where we simulated a sample of 17 374 individuals where the minor allele frequency (MAF = 0.39) and effect size ($\beta = -0.082$) were the same as LCT variant found in this study. Each sample was analysed with both additive and dominant models. The simulations were performed with R.

The dominance deviation test was performed by fitting jointly the additive term and a heterozygote indicator.

The gender interaction analysis was performed by using the meta-analysis summary statistics. The method is described further in a book by Hardy (38).

WEB RESOURCES

National Biobank of Finland: http://www.nationalbiobanks.fi;

ACKNOWLEDGEMENTS

We warmly thank all the Finnish volunteers who participated in the studies. Thanks are due to Siv Knaappila and Minna Suvela for excellent laboratory work.

Conflict of Interest statement. None declared.

FUNDING

Northern Finland Birth Cohort 1966 (NFBC 1966) is supported by the European Commission; contract number QLG1-CT-2000-01643, Biocenter Oulu, University of Oulu, Finland, and the Academy of Finland, and by NHLBI, Stampeed (Genetics of cardiovascular risk factors in large founder population birth cohorts) NIH/NHLBI 1-R01-HL087679-01. V.A. was supported by the Finnish Culture Foundation. V.S. is supported by Sigrid Juelius Foundation and the Finnish Foundation for Cardiovascular Research. L.P. and K.S. are supported by the Academy of Finland Centre of Excellence in Complex Disease Genetics, the Biocentrum Helsinki Foundation, Helsinki, Finland, and the Finnish Foundation for Cardiovascular Research. The ATBC study was supported by US Public Health Service contracts N01-CN-45165, N01-RC-45035 and N01-RC-37004 from the National Cancer Institute, Department of Health and Human Services. The Young Finns cohort was supported by the Emil Aaltonen Foundation (T.L.), the Academy of Finland (grants 53392 and 34316), the Social Insurance Institution of Finland, the Turku University Foundation, the Juho Vainio Foundation, the Finnish Foundation of Cardiovascular Research, the Finnish Cultural Foundation, and the Tampere and Turku University Central Hospital Medical Fund. We would like to acknowledge the International HapMap Consortium for the linkage disequilibrium data. The KORA research platform (KORA: Kooperativer Gesundheitsforschung in der Region Augsburg) was initiated and financed by the Helmholtz Zentrum München—National Research Center for Environmental Health, which is funded by the German Federal Ministry of Education, Science, Research and Technology and by the State of Bavaria. Part of this work was financed by the German National Genome Research Network (NGFN-2 and NGFNPlus: 01GS08223) and within the Munich Center of Health Sciences (MC Health) as part of LMU/innovativ. Funding to pay the Open Access Charge was provided by Wellcome Trust Sanger Institute.

REFERENCES


