Novel locus including *FGF21* is associated with dietary macronutrient intake

Audrey Y. Chu[1], Tsegaselassie Workalemahu[2,6], Nina P. Paynter[1], Lynda M. Rose[1], Franco Giulianini[1], Toshiko Tanaka[7] and Julius S. Ngwa[8] on behalf of the CHARGE Nutrition Working Group, Qibin Qi[2,6], Gary C. Curhan[2], Eric B. Rimm[2,6], David J. Hunter[2,6], Louis R. Pasquale[3,9], Paul M. Ridker[1,4], Frank B. Hu[2,6], Daniel I. Chasman[1,5,†] and Lu Qi[2,6,*,†] on behalf of the DietGen Consortium

[1]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA [2]The Channing Laboratory and the Department of Medicine, [3]Department of Ophthalmology, [4]Division of Cardiology, and [5]Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA [6]The Departments of Nutrition and Epidemiology, Harvard School of Public Health, Boston, MA, USA [7]Clinical Research Branch, National Institute on Aging, Baltimore, MD, USA [8]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA and [9]The Mass Eye and Ear Infirmary, Harvard Medical School, Boston, MA, USA

Dietary intake of macronutrients (carbohydrate, protein, and fat) has been associated with risk of chronic conditions such as obesity and diabetes. Family studies have reported a moderate contribution of genetics to variation in macronutrient intake. In a genome-wide meta-analysis of a population-based discovery cohort ($n = 33\,533$), rs838133 in *FGF21* (19q13.33), rs197273 near TRAF family member-associated NF-kappa-B activator (*TANK*) (2p24.2), and rs10163409 in *FTO* (16q12.2) were among the top associations ($P < 10^{-5}$) for percentage of total caloric intake from protein and carbohydrate. rs838133 was replicated *in silico* in an independent sample from the Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium (CHARGE) Nutrition Working Group ($n = 38\,360$) and attained genome-wide significance in combined analysis ($P_{\text{joint}} = 7.9 \times 10^{-9}$). A cytokine involved in cellular metabolism, *FGF21* is a potential susceptibility gene for obesity and type 2 diabetes. Our results highlight the potential of genetic variation for determining dietary macronutrient intake.

INTRODUCTION

Dietary intake is tightly related to human health and excess intake of energy has been associated with obesity, diabetes and cardiovascular disease, as well as many other chronic diseases in epidemiology studies (1). Total energy intake can be split into energy derived from three primary macronutrients: fat, protein, and carbohydrate. In spite of wide variety in food ingredients and preparation according to culture and geography, mean proportions of dietary macronutrient intake are relatively constant across many populations from industrialized nations providing ∼40–50, 10–20 and 25–40% of total caloric intake from carbohydrate, protein and fat, respectively (1). Inter-individual variation in macronutrient intake remains substantial however, and is heritable as judged by family and twin studies, although the range of heritability estimates is broad at 11–65% (2). Prior approaches to discovering genetic influences on macronutrient intake have focused on candidate hypotheses related to metabolic function (3) and taste (4), and have suggested some genetic links between dietary habits, obesity and diabetes. However, there have been no genome-wide association studies (GWASs) investigating macronutrient intake to date. To gain additional insights into genetic control of macronutrient intake, we conducted a

*To whom correspondence should be addressed at: Department of Nutrition and Epidemiology, Harvard School of Public Health, Building 2, Room 323, 655 Huntington Ave, Boston, MA 02115, USA. Tel: +1 6174324116; Fax: +1 6174322435; Email: luqi@hsph.harvard.edu
†These authors contributed equally to this manuscript.

GWAS by performing a genome-wide meta-analysis among population-based cohorts of European ancestry.

RESULTS

We performed a genome-wide meta-analysis of 33 355 male and female participants of European ancestry in DietGen, a consortium consisting of three large population-based studies investigating the genetics of dietary intake and nutrition (the Women's Genome Health Study: WGHS, $n = 22\,691$; the Health Professionals Follow-up Study: HPFS, $n = 4077$ among three nested case–control studies and the Nurses' Health Study: NHS, $n = 6765$ among four nested case–control studies; Supplementary Material, Table S1). The proportion of total energy derived from each of the three macronutrients as a percentage was estimated from similarly designed, self-administered, semi-quantitative and highly validated food frequency questionnaires (FFQs) (5) in all the three cohorts. Discovery-stage genome-wide association scans and meta-analyses across the three cohorts were performed for all three macronutrient phenotypes adjusted for age, location, sub-population stratification, with and without adjustment for body mass index (BMI). BMI adjustments were performed to decrease variance of the macronutrient phenotypes and to account for genetic effects mediated through body composition.

A genome-wide significant ($P < 5 \times 10^{-8}$) association was identified for SNPs in an intron of the *FTO* gene at 16q12.2 with percentage of total caloric intake from carbohydrate in models with or without BMI adjustment (index SNP rs10163409, $P = 3.4 \times 10^{-8}$ and $P = 7.4 \times 10^{-9}$, respectively; Table 1), although this association appeared to be driven by the WGHS cohort ($P = 5.7 \times 10^{-9}$ and $P = 1.2 \times 10^{-9}$ with and without adjustment for BMI; see Supplementary Material, Figure S1). A total of 22 independent SNPs (see Methods) were associated with macronutrient intake with or without BMI adjustment at sub-genome-wide significance ($5 \times 10^{-8} < P < 1 \times 10^{-5}$); 13 for percentage carbohydrate intake, 7 for percentage protein intake, and 2 for percentage fat intake (Table 1, Supplementary Material, Table S2 and Figure S2).

The most significantly associated variant at each locus was further evaluated through *in silico* replication from a parallel genome-wide meta-analysis of 38 360 samples among 12 cohorts from the Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium (CHARGE) Nutrition Working Group (6). The sub-genome-wide association between rs838133 in the *FGF21* gene at 19q13.33 and percentage protein intake adjusted for BMI from DietGen was replicated in a model adjusted for BMI after correcting for multiple hypothesis testing ($P_{\text{one-sided}} = 7.2 \times 10^{-4}$; $\alpha_{\text{protein}} = 7.1 \times 10^{-3} = 0.05/7$ protein loci), and attained a genome-wide significant P-value in a joint inverse-variance weighted meta-analysis combining the discovery cohorts with the replication cohorts ($P_{\text{joint}} = 7.9 \times 10^{-9}$). In the replication cohorts, each copy of the minor allele was associated with a decrease of 0.098% (SE 0.030%, $P = 7.3 \times 10^{-4}$) protein intake (Table 1). The sub-genome-wide association of percentage carbohydrate intake adjusted for BMI with

rs197273 $\sim$100 kb to the 5′ end of the *TANK* gene on chromosome 2q24 was replicated in a model adjusting for BMI after correction for multiple hypothesis testing ($P_{\text{one-sided}} = 2.9 \times 10^{-3}$; $\alpha_{\text{carbohydrate}} = 3.8 \times 10^{-3} = 0.05/13$ carbohydrate loci), but did not attain genome-wide significance in the joint analysis ($P_{\text{joint}} = 9.6 \times 10^{-8}$). The genome-wide significant association between the *FTO* variant rs10163409 and percentage carbohydrate intake was not replicated by CHARGE either with or without BMI adjustment ($P = 0.34$ and $P_{\text{BMI adjusted}} = 0.23$, respectively) (Table 1). For this SNP, there was evidence of moderate heterogeneity in CHARGE ($I^2 = 34.8$ and 32.3%, unadjusted and adjusted for BMI, respectively, Supplementary Material, Table S2).

Together, the three macronutrients account for the majority of caloric intake, raising the potential for simultaneous associations between the lead SNPs and more than one macronutrient. In discovery, the variant rs838133 at the *FGF21* locus, which was identified in association with decreased protein intake, was also associated with increased carbohydrate intake (beta[se]: 0.23% [0.07%], $P = 5.9 \times 10^{-4}$) and decreased fat intake (beta[se]: -0.21% [0.05%], $P = 7.1 \times 10^{-5}$) in models adjusted for BMI (Fig. 1). The variant rs10163409 in *FTO*, which was discovered in association with increased percentage carbohydrate intake in models adjusted for BMI, was associated with decreased fat intake (beta[se]: $-0.22\%[0.05\%]$, $P = 1.6 \times 10^{-4}$), but not associated with protein intake (beta[se]: $-0.05\%[0.03\%]$, $P = 0.08$; Fig. 1). Variant rs197273 near the *TANK* gene, in addition to a sub-genome-wide association with increased carbohydrate intake, was significantly associated with decreased fat intake (beta[se]: $-0.17\%[0.05\%]$, $P = 0.78 \times 10^{-4}$) and marginally associated with decreased protein intake (beta[se]: -0.05% [0.02%], $P = 0.068$) in models adjusted for BMI. Associations at all three loci were similar in the absence of BMI adjustment (Supplementary Material, Table S3).

To find evidence for additional sub-genome-wide associations, we estimated the proportion of variance in macronutrient intake in the WGHS using GCTA (7) (see Methods). This procedure found that 6.6% (SE 2.1%, $P = 9.1 \times 10^{-5}$), 8.0% (SE 2.2%, $P = 2.7 \times 10^{-5}$) and 7.3% (SE 2.2%, $P = 3.8 \times 10^{-5}$) of the variance in total energy from carbohydrate, protein and fat, respectively, could be explained by the common WGHS tag-SNPs. By comparison, SNP rs838133 in *FGF21* explained 0.052, 0.062, and 0.057% of the variance in carbohydrate, protein, and fat intake, respectively, in the WGHS. Adjustment for age, location and population substructure as measured by 10 eigenvectors did not substantially affect the estimated genetic contribution to the variance, nor did additional adjustment for BMI (Supplementary Material, Table S4). Partitioning the contribution to the variance according to MAF suggested that the estimates were mostly due to the SNPs with MAF > 5% compared with less common SNPs with MAF between 1 and 5% (Supplementary Material, Table S4).

DISCUSSION

We performed a two-stage genome-wide meta-analysis of percentage of total energy intake from carbohydrate, protein and

Table 1. Genome-wide association results from meta-analysis for percentage of total caloric intake from macronutrients (carbohydrate and protein) in the discovery (DietGen), replication (CHARGE) and combined (Joint) cohort analyses

Model	SNP	Gene	Chr	BP	EA/nonEA	EAF	DietGen Beta	SE	P(2GC)	CHARGE Beta	SE	P(2GC)	one-sided P	Joint Analysis Beta	SE	P	N
Carbohydrate																	
Model1[a]	rs10163409	*FTO*	16	52 696 381	A/T	0.69	0.44	0.08	7.4E−09	−0.03	0.07	6.9E−01	3.4E−01	0.19	0.05	2.2E−04	71 885
Model2[b]	rs197273	*TANK*	2	161 602 909	A/G	0.48	0.31	0.06	2.0E−06	0.17	0.06	5.7E−03	2.9E−03	0.23	0.04	9.6E−08	71 328
Model2[b]	rs10163409	*FTO*	16	52 696 381	A/T	0.69	0.42	0.07	3.4E−08	−0.05	0.07	4.6E−01	2.3E−01	0.17	0.05	1.1E−03	71 326
Protein																	
Model2[b]	rs838133	*FGF21*	19	53 951 341	A/G	0.45	−0.12	0.03	2.3E−06	−0.10	0.03	1.5E−03	7.3E−04	−0.11	0.02	7.9E−09	68 968

CHARGE, Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium; SNP, single-nucleotide polymorphism; Chr, chromosome; BP, base pair position—build 36; EA, effect allele; nonEA, non-effect allele; EAF, effect allele frequency; beta, beta-coefficient; SE, standard error; P(2GC), P-value corrected for genomic control; one-sided P, one-sided P-value; P, P-value; N, sample size.
[a]Model 1 adjusts for age, sex (in CHARGE), location and sub-population stratification.
[b]Model 2 adjusts for model 1 covariates in addition to BMI (kg/m^2).

fat among 33 533 participants with independent replication in a sample of equivalent size. The replication yielded a locus at 19q33 (rs838133) associated with protein intake at genome-wide significance, and provided nominal support for a second locus at 2q24 (rs197273) associated with carbohydrate intake. In addition, heritability analyses in a subset of the discovery cohort showed that common SNPs explained a modest but statistically significant proportion of genetic variance for carbohydrate, protein and fat intake (6–8%).

Variant rs838133 maps to exon 1 of the *FGF21* gene, but also maps in the vicinity of a gene-dense region on chromosome 19 that includes other genes *FUT1, FUT2, IZUMO1* and *RASIP1*—all within 100 kb. While we cannot exclude the possibility this variant could be influencing other genes in this region, *FGF21* is a strong candidate due to its biology and location. Additional potential SNP associations in the region were fully attenuated by conditional meta-analysis of rs838133, reinforcing choice of *FGF21* as the candidate gene. rs838133 is a synonymous SNP located in the first exon of the *FGF21* gene encoding fibroblast growth factor 21. Secreted by hepatocytes and adipocytes, FGF21 promotes glucose uptake in adipocytes and has other functions in cellular metabolism, such as regulation of carbohydrate and lipid metabolism (8). FGF21 expression can be induced through fasting or feeding in animal models (9). Serum levels of FGF21 have been associated with increased odds of type 2 diabetes (10,11) and obesity(12), and there is suggestive evidence that human FGF21 infusion into rodents may regulate food intake(13,14), although the effects were not consistently observed (15). And FGF21 protein is being investigated as a potential agent for pharmacologic intervention targeting these conditions (15).

A different variant at the *FGF21* locus was identified from the CHARGE Nutrition Working Group meta-analysis (6) in association with carbohydrate and fat intake at genome-wide significance, whereas the *FGF21* locus was associated with protein intake in the DietGen analysis. The *FGF21* variants identified in the DietGen and CHARGE analyses are in moderate linkage disequilibrium ($r^2 = 0.7$, 1000 Genomes) and likely represent the same signal. Because macronutrient intake is expressed as a proportion of the total caloric intake, a SNP association with one macronutrient could imply association with one or both of the others. The *FGF21* variant identified in DietGen in association with percentage protein intake was also associated with carbohydrate and fat intake at a lower significance level in our analyses. Which macronutrient is identified with the greatest significance in one study or another may reflect differences in instruments used for nutritional assessment, true differences in the dietary habits of the underlying populations, or genetic influence on the composition of macronutrient intake rather than individual macronutrient intake, among other explanations. All of these possibilities still support the association of the *FGF21* variant with macronutrient intake.

However, there are some published findings that may be inconsistent with the identification of *FGF21* as the causal gene. For example, rs838133 was mapped as an eQTL for *FUT2*, *RASIP1* and *NTN5* ($P < 3.8 \times 10^{-5}$) in skin tissue, but not adipose tissue or lymphoblasts (16, 17). *RASIP1* and *NTN5* have not been associated with dietary intake, anthropometric
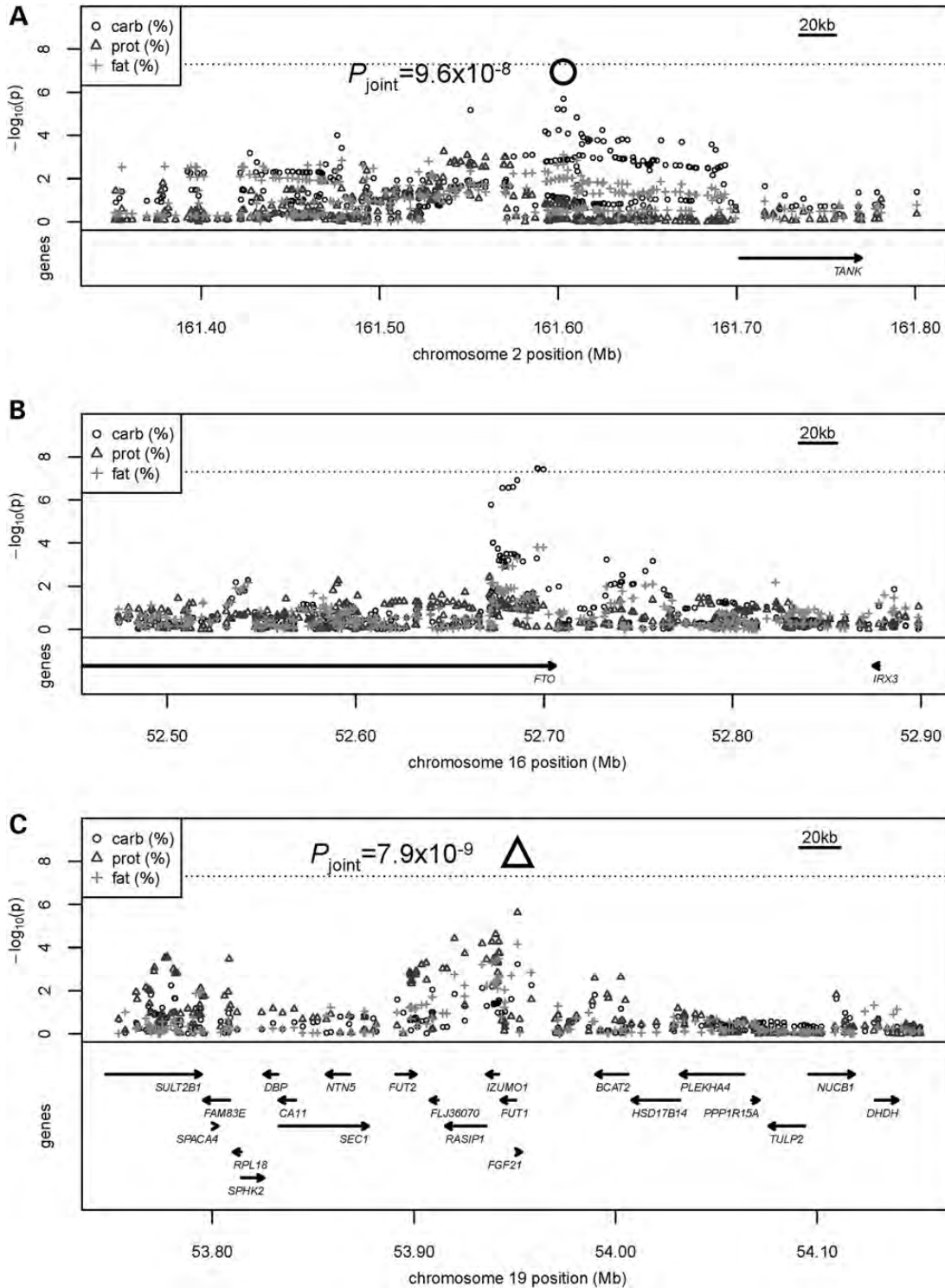
Figure 1. Regional association plots showing $-\log_{10}$ P-values for percentage of total energy intake from macronutrients. Closest genes and -log10 P-values for SNPs at (**A**) chromosome 2q24, (**B**) chromosome16q12 and (**C**) chromosome 19q13 in association with percentage of total energy intake from carbohydrate (circles), protein (triangles) and fat (crosses) adjusted for BMI in the discovery cohort. The P-values from joint meta-analysis of DietGen and CHARGE Nutrition are indicated for association between (**A**) rs197273 and carbohydrate intake by the bolded circle and (**C**) rs838133 and protein intake by the bolded triangle. The threshold for genome-wide significance ($P < 5 \times 10^{-8}$) is indicated by the horizontal dotted line.

measures or metabolic disorders, but variants in *FUT2* have been associated with plasma vitamin B-12 levels (18), and the *FUT2* polymorphism rs601338 ($r^2 = 0.39$ with

rs838133) is known to determine glycoprotein and glycolipid secretor phenotype (19). Taken together the evidence suggests that there remains some ambiguity to the causality of *FGF21*,

and indicates that while *FGF21* is a top candidate for macronutrient intake at this locus, other genes in the region cannot be entirely excluded.

It is unlikely that the association between the *FGF21* locus and protein intake in our meta-analysis could be mediated or confounded by BMI, or other related characteristics. We did not observe any association between rs838133 and BMI within the individual DietGen cohorts, and the variant is only nominally associated with BMI in the publically available summary statistics from the most recent GIANT meta-analysis ($P = 0.027$, $n = 123\,702$) (20). Currently available GIANT results do not allow investigation of whether this weak effect is mediated by dietary intake. In addition, no *FGF21* candidate variants reached genome-wide significance in the recent DIAGRAM+ meta-analysis of 34 840 type 2 diabetes cases and 114 981 controls (21). Furthermore, the DietGen meta-analysis results for *FGF21* were stronger with the inclusion of BMI than without (beta[se]: $-0.124[0.026]$, $P = 2.3 \times 10^{-6}$; beta[se]: $-0.115[0.026]$, $P = 1.1 \times 10^{-5}$, respectively). Sensitivity analyses from the WGHS demonstrated that this association was robust to adjustment for BMI, baseline diabetes status, physical activity, current smoking and daily alcohol intake (data not shown). Residual confounding, if any, is unlikely expected to be mediated through clinical characteristics typically thought to influence the diet.

TANK, the nearest gene to rs197273 on chromosome 2 encodes a signaling protein that inhibits tumor necrosis factor receptor-associated factor (TRAF) function. The TRAF family of proteins plays a role in inflammation and immune response, and has not been previously linked to regulation of dietary macronutrient intake.

Although the *FTO* locus at 16q12 is strongly and consistently associated with BMI and adiposity, and there is evidence that it influences dietary intake and energy balance, the genome-wide significant association of rs10163409 with percentage carbohydrate was not replicated by the independent meta-analysis from the CHARGE Nutrition Working Group. Heterogeneity for rs10163409 was estimated to be qualitatively larger among the CHARGE cohorts ($I^2 = 34.8\%$ and $I^2_{BMI\ adjusted} = 32.3\%$) compared with the DietGen cohorts ($I^2 = 8.7\%$, and $I^2_{BMI\ adjusted} = 19.9\%$). It is possible that this heterogeneity may have partially influenced the lack of replication at this locus. Furthermore, rs10163409 is not in LD with SNPs at *FTO* that are strongly associated with BMI (rs1558902, $r^2 = 0.07$; 1000 Genomes), including in the DietGen cohorts (20), nor is it associated with BMI in the individual DietGen cohorts (data not shown). These results suggest that there may be additional effects of *FTO* on dietary composition and intake unrelated to its effect on the body size and composition.

Unhealthy habitual dietary behaviors especially recent changes in the proportion of macronutrients consumed underlie the current epidemic of obesity and its associated risk of metabolic diseases, including cardiovascular disease and type 2 diabetes (22). The present work, together with the study by *Tanaka et al.* (6) on behalf of the CHARGE Nutrition Working Group, represents the first and largest genetic investigations of macronutrient intake on a genome-wide scale. Our results demonstrate associations with macronutrient intake at two loci, one at genome-wide significance, while using heritability analysis to infer that additional genetic associations with macronutrient intake remain to be discovered. The proportion of variance explained in the GCTA analysis likely underestimates the actual heritability due to residual imprecision in measuring macronutrient intake from FFQs and to intrinsic limitations in using tag rather than causal SNPs to assess the correlation between relatedness and macronutrient intake (23). In sum, our results provide a compelling impetus for learning more about the biological function of *FGF21* and for performing additional analysis to further delineate the genetic determinants of macronutrient intake.

MATERIALS AND METHODS

Study populations

Discovery cohorts

DietGen consortium. DietGen is composed of three US population-based cohorts: HPFS, NHS and WGHS investigating the genetics of dietary intake and nutrition.

Health professionals follow-up study (HPFS). The study population was derived from the HPFS, a prospective cohort of 51 529 male health care professionals between the ages of 40–75 years at enrollment in 1986 (24). Since the inception of the study, data on lifestyle and medical history have been ascertained through a self-administered questionnaire, including a semi-quantitative FFQ every 4 years. For this analysis, we included participants who completed the 1986 FFQ. Participants of validated self-reported European ancestry from three sets of the GWAS from nested case-control studies were included in this analysis: type 2 diabetes (25) (HPFS T2D, $n = 2316$), coronary heart disease (26) (HPFS CHD, $n = 1212$), and kidney stones (27) (HPFS KS, $n = 549$). The study was approved by the Institutional Review Board (IRB) of the Harvard School of Public Health.

Nurses' health study (NHS). The study population was derived from the NHS, a prospective cohort of 121 700 female registered nurses between the ages of 30–55 years at enrollment in 1976 (28). Since the inception of the NHS, data on lifestyle and medical history have been ascertained through a self-administered questionnaire, including a semi-quantitative FFQ every 2–4 years. For this analysis, we included participants who responded to the 1986 FFQ. Participants of validated self-reported European ancestry from four sets of the GWAS from nested case–control studies were included in this analysis: type 2 diabetes (25) (NHS T2D, $n = 3219$), coronary heart disease (26) (NHS CHD, $n = 1013$), kidney stones (27) (NHS KS, $n = 490$) and breast cancer (29) (NHS BC, $n = 2043$). The study was approved by the IRB of the Harvard School of Public Health.

Women's genome health study (WGHS). The study population was derived from the WGHS (30), a prospective cohort of women from the Women's Health Study aged 45 years or older with no history of cardiovascular disease, cancer or major chronic illness, and who provided blood samples at baseline from which genomic DNA was extracted. Individuals

of confirmed self-reported European ancestry and had genotyping information available were included in this analysis ($n = 23\,294$). Among these participants, 22 691 provided FFQ data at baseline. The study was approved by the IRB of Brigham and Women's Hospital.

Replication cohorts

Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium (CHARGE) (31). A subset of CHARGE, the CHARGE Nutrition Working Group, is composed of 12 cohorts from the United States and Europe with a total sample size of 38 360 (6) and includes the Atherosclerosis Risk in Communities Study (ARIC), Cardiovascular Health Study (CHS), The European Prospective Investigation of Cancer-Norfolk (EPIC-Norfolk), Family Heart Study, Fenland study, Framingham Heart Study, Genetics of Lipid Lowering Drugs and Diet Network study (GOLDN), Health Aging and Body Composition study (Health ABC), the InCHIANTI study, Multi-Ethnic Study of Atherosclerosis (MESA), the Rotterdam study and the Young Finns Study.

Assessment of percentage of caloric intake from macronutrients

In the NHS and HPFS, dietary intake was assessed every 2–4 years using the Willett semi-quantitative food frequency questionnaire containing 118-food items (32,33). The FFQ from 1986 was used for this analysis. In the WGHS, dietary intake was assessed at baseline between 1992 and 1994 using the Willett semi-quantitative FFQ derived from the NHS containing 131-food items (34). For all cohorts, participants reported an average frequency of consumption and portion sizes of all food items listed in the FFQ during the previous year. Average daily intake per food item was calculated by multiplying the frequency of consumption by portion size (1). Absolute macronutrient intake was calculated by multiplying the average daily intake by the nutrient content derived from food composition tables from the Harvard School of Public Health.

In the replication cohorts, dietary intake was assessed using similar but slightly varied FFQs. Average daily intake of food and nutrients was calculated using the frequency of consumption, portion sizes and study-specific nutrient and food composition tables.

Genotyping

Genotyping for each of the discovery cohort studies was performed using either the Illumina HumanHap 300 Duo+ (WGHS), Illumina 550 K (NHS BC), Illumina 610Q (NHS/HPFS KS cohort), or the Affymetrix 6.0 (NHS/HPFS T2D and CHD) array; each panel was imputed up to ~2.6 million SNPs using MaCH software (reference panel—HapMap release 22/NCBI build 36)(35,36). Imputed genotypes were expressed as allelic dosage values between 0 and 2.

Similarly, the replication cohorts were genotyped with either Affymetrix- or Illumina-based genotyping arrays and imputed to HapMap release 21 or 22/NCBI build 35 or 36.

Statistical analysis

Three primary outcomes were examined in this study: percentage of total energy intake from (i) carbohydrate, (ii) protein, and (iii) fat. Percentage of total energy intake from each of the three macronutrients was calculated by multiplying the absolute macronutrient intake in grams by the caloric content of each type of macronutrient (4 kcal/g for carbohydrate and protein, 9 kcal/g for fat) (1), and dividing by the total energy intake. Absolute macronutrient intake and total energy intake were computed by multiplying the relative frequency of consumption by the nutrient content using United States Department of Agriculture food composition sources.

Study-specific genome-wide association analyses of percentage macronutrient intake were performed using linear regression in ProbABEL (37) for ~2.6 million SNPs, adjusted for age, location, measures of population stratification, with and without adjustment for BMI. Distributions of the residuals from each of the macronutrient intake analyses are provided in the Supplementary material, Figure S3. If the genomic inflation factor was > 1.0, the analysis was corrected before the meta-analysis.

We conducted a two-stage genome-wide meta-analysis for macronutrient intake. In the initial stage, meta-analyses for each phenotype–model combination were performed in METAL (38). Fixed-effects meta-analyses using the inverse-variance weighting scheme were performed for each macronutrient and model combination (total of six meta-analyses). The meta-analysis results were corrected if the genomic inflation factor was > 1.0. Heterogeneity was estimated using the I^2 statistic. Meta-analyses were limited to 2.1 million SNPs with minor allele frequencies (MAF) ≥ 0.05 and imputation quality scores $(R^2) \geq 0.30$. The genome-wide significance threshold was set at $P < 5 \times 10^{-8}$, and the sub-genome-wide significance threshold was set at $1 \times 10^{-5} > P > 5 \times 10^{-8}$. Lead SNPs were deemed independent if they had the most significant association at a locus, and were > 500 kb from a lead SNP at an adjacent locus. Conditional meta-analysis in GCTA(7) on the lead SNP at each locus did not reveal any remaining SNP associations with $P < 1 \times 10^{-5}$ at the candidate loci.

In the second stage, we approached the CHARGE consortium and requested to investigate the association of our index genome-wide and sub-genome-wide SNPs in their parallel genetic analysis of macronutrient intake (6). The Bonferroni method was used to correct for multiple hypothesis testing by dividing the $\alpha(=0.05)$ by the number of loci brought forward for replication per macronutrient phenotype. The P-values from one-tailed z-distributions were used to determine the significance in replication. Genome-wide significance was declared if the P-value from the joint analysis was $< 5 \times 10^{-8}$.

GCTA estimation of heritability

We randomly selected 15 000 participants (out of 23 294) from the WGHS to estimate heritability of macronutrient intake using the GCTA software (7); we were unable to include all WGHS participants in the heritability estimate due to computational restraints. We restricted the analysis to 339 596 genotyped SNPs with 98.5% complete genotyping and MAF > 0.01.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

REFERENCES

1. Willett, W. (1998) *Nutritional Epidemiology*. New York: Oxford University Press.
2. Rankinen, T. and Bouchard, C. (2006) Genetics of food intake and eating behavior phenotypes in humans. *Annu. Rev. Nutr.*, **26**, 413–434.
3. Qi, L., Kraft, P., Hunter, D.J. and Hu, F.B. (2008) The common obesity variant near MC4R gene is associated with higher intakes of total energy and dietary fat, weight change and diabetes risk in women. *Hum. Mol. Genet.*, **17**, 3502–3508.
4. Reed, D.R., Bachmanov, A.A., Beauchamp, G.K., Tordoff, M.G. and Price, R.A. (1997) Heritable variation in food preferences and their contribution to obesity. *Behav. Genet.*, **27**, 373–387.
5. Willett, W.C., Reynolds, R.D., Cottrell-Hoehner, S., Sampson, L. and Browne, M.L. (1987) Validation of a semi-quantitative food frequency questionnaire: comparison with a 1-year diet record. *J. Am. Diet Assoc.*, **87**, 43–47.
6. Tanaka, T., Ngwa, J.S., van Rooij, F.J.A., Zillikens, M.C., Wojczynski, M.K., Frazier-Wood, A.C., Houston, D.K., Kanoni, S., Lemaitre, R.N., Luan, J. *et al.* (2013) Genome-wide meta-analysis of observational studies reveals common genetic variants associated with macronutrient intake. *Am. J. Clin. Nutr.* (in press).
7. Yang, J., Lee, S.H., Goddard, M.E. and Visscher, P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.*, **88**, 76–82.
8. Berglund, E.D., Li, C.Y., Bina, H.A., Lynes, S.E., Michael, M.D., Shanafelt, A.B., Kharitonenkov, A. and Wasserman, D.H. (2009) Fibroblast growth factor 21 controls glycemia via regulation of hepatic glucose flux and insulin sensitivity. *Endocrinology*, **150**, 4084–4093.
9. Uebanso, T., Taketani, Y., Yamamoto, H., Amo, K., Ominami, H., Arai, H., Takei, Y., Masuda, M., Tanimura, A., Harada, N. *et al.* (2011) Paradoxical regulation of human FGF21 by both fasting and feeding signals: is FGF21 a nutritional adaptation factor? *PLoS. One*, **6**, e22976.
10. Chen, W.W., Li, L., Yang, G.Y., Li, K., Qi, X.Y., Zhu, W., Tang, Y., Liu, H. and Boden, G. (2008) Circulating FGF-21 levels in normal subjects and in newly diagnose patients with Type 2 diabetes mellitus. *Exp. Clin. Endocrinol Diabetes*, **116**, 65–68.
11. Bobbert, T., Schwarz, F., Fischer-Rosinsky, A., Pfeiffer, A.F., Mohlig, M., Mai, K. and Spranger, J. (2013) Fibroblast growth factor 21 predicts the metabolic syndrome and type 2 diabetes in Caucasians. *Diabetes Care*, **36**, 145–149.
12. Zhang, X., Yeung, D.C., Karpisek, M., Stejskal, D., Zhou, Z.G., Liu, F., Wong, R.L., Chow, W.S., Tso, A.W., Lam, K.S. *et al.* (2008) Serum FGF21 levels are increased in obesity and are independently associated with the metabolic syndrome in humans. *Diabetes*, **57**, 1246–1253.

13. Coskun, T., Bina, H.A., Schneider, M.A., Dunbar, J.D., Hu, C.C., Chen, Y., Moller, D.E. and Kharitonenkov, A. (2008) Fibroblast growth factor 21 corrects obesity in mice. *Endocrinology*, **149**, 6018–6027.
14. Sarruf, D.A., Thaler, J.P., Morton, G.J., German, J., Fischer, J.D., Ogimoto, K. and Schwartz, M.W. (2010) Fibroblast growth factor 21 action in the brain increases energy expenditure and insulin sensitivity in obese rats. *Diabetes*, **59**, 1817–1824.
15. Dutchak, P.A., Katafuchi, T., Bookout, A.L., Choi, J.H., Yu, R.T., Mangelsdorf, D.J. and Kliewer, S.A. (2012) Fibroblast growth factor-21 regulates PPARgamma activity and the antidiabetic actions of thiazolidinediones. *Cell*, **148**, 556–567.
16. Yang, T.P., Beazley, C., Montgomery, S.B., Dimas, A.S., Gutierrez-Arcelus, M., Stranger, B.E., Deloukas, P. and Dermitzakis, E.T. (2010) Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics*, **26**, 2474–2476.
17. Grundberg, E., Small, K.S., Hedman, A.K., Nica, A.C., Buil, A., Keildson, S., Bell, J.T., Yang, T.P., Meduri, E., Barrett, A. *et al.* (2012) Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat. Genet.*, **44**, 1084–1089.
18. Hazra, A., Kraft, P., Selhub, J., Giovannucci, E.L., Thomas, G., Hoover, R.N., Chanock, S.J. and Hunter, D.J. (2008) Common variants of FUT2 are associated with plasma vitamin B12 levels. *Nat. Genet.*, **40**, 1160–1162.
19. Rouquier, S., Lowe, J.B., Kelly, R.J., Fertitta, A.L., Lennon, G.G. and Giorgi, D. (1995) Molecular cloning of a human genomic region containing the H blood group alpha(1,2)fucosyltransferase gene and two H locus-related DNA restriction fragments. Isolation of a candidate for the human secretor blood group locus. *J. Biol. Chem.*, **270**, 4632–4639.
20. Speliotes, E.K., Willer, C.J., Berndt, S.I., Monda, K.L., Thorleifsson, G., Jackson, A.U., Allen, H.L., Lindgren, C.M., Luan, J., Magi, R. *et al.* (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.*, **42**, 937–948.
21. Morris, A.P., Voight, B.F., Teslovich, T.M., Ferreira, T., Segre, A.V., Steinthorsdottir, V., Strawbridge, R.J., Khan, H., Grallert, H., Mahajan, A. *et al.* (2012) Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.*, **44**, 981–990.
22. Popkin, B.M., Adair, L.S. and Ng, S.W. (2012) Global nutrition transition and the pandemic of obesity in developing countries. *Nutr. Rev.*, **70**, 3–21.
23. Visscher, P.M., Yang, J. and Goddard, M.E. (2011) A commentary on 'common SNPs explain a large proportion of the heritability for human height' by Yang et al. (2010). *Twin Res. Hum. Genet.*, **13**, 517–524.
24. Rimm, E.B., Giovannucci, E.L., Willett, W.C., Colditz, G.A., Ascherio, A., Rosner, B. and Stampfer, M.J. (1991) Prospective study of alcohol consumption and risk of coronary disease in men. *Lancet*, **338**, 464–468.
25. Qi, L., Cornelis, M.C., Kraft, P., Stanya, K.J., Linda Kao, W.H., Pankow, J.S., Dupuis, J., Florez, J.C., Fox, C.S., Pare, G. *et al.* (2010) Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes. *Hum. Mol. Genet.*, **19**, 2706–2715.
26. Jensen, M.K., Pers, T.H., Dworzynski, P., Girman, C.J., Brunak, S. and Rimm, E.B. (2011) Protein interaction-based genome-wide analysis of incident coronary heart disease. *Circ. Cardiovasc. Genet.*, **4**, 549–556.
27. Curhan, G.C. and Taylor, E.N. (2008) 24-h uric acid excretion and the risk of kidney stones. *Kidney Int.*, **73**, 489–496.
28. Colditz, G.A. and Hankinson, S.E. (2005) The Nurses' Health Study: lifestyle and health among women. *Nat. Rev. Cancer*, **5**, 388–396.
29. Hunter, D.J., Kraft, P., Jacobs, K.B., Cox, D.G., Yeager, M., Hankinson, S.E., Wacholder, S., Wang, Z., Welch, R., Hutchinson, A. *et al.* (2007) A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat. Genet.*, **39**, 870–874.
30. Ridker, P.M., Chasman, D.I., Zee, R.Y., Parker, A., Rose, L., Cook, N.R. and Buring, J.E. (2008) Rationale, design, and methodology of the Women's Genome Health Study: a genome-wide association study of more than 25,000 initially healthy American women. *Clin. Chem.*, **54**, 249–255.
31. Psaty, B.M., O'Donnell, C.J., Gudnason, V., Lunetta, K.L., Folsom, A.R., Rotter, J.I., Uitterlinden, A.G., Harris, T.B., Witteman, J.C. and Boerwinkle, E. (2009) Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ. Cardiovasc. Genet.*, **2**, 73–80.

32. Willett, W.C., Sampson, L., Stampfer, M.J., Rosner, B., Bain, C., Witschi, J., Hennekens, C.H. and Speizer, F.E. (1985) Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am. J. Epidemiol.*, **122**, 51–65.

33. Rimm, E.B., Giovannucci, E.L., Stampfer, M.J., Colditz, G.A., Litin, L.B. and Willett, W.C. (1992) Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. *Am. J. Epidemiol.*, **135**, 1114–1126; discussion 1127–1136.

34. Liu, S., Manson, J.E., Lee, I.M., Cole, S.R., Hennekens, C.H., Willett, W.C. and Buring, J.E. (2000) Fruit and vegetable intake and risk of cardiovascular disease: the Women's Health Study. *Am. J. Clin. Nutr.*, **72**, 922–928.

35. Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M. *et al.* (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature*, **449**, 851–861.

36. Li, Y., Willer, C.J., Ding, J., Scheet, P. and Abecasis, G.R. (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.*, **34**, 816–834.

37. Aulchenko, Y.S., Struchalin, M.V. and van Duijn, C.M. (2010) ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics*, **11**, 134.

38. Willer, C.J., Li, Y. and Abecasis, G.R. (2010) METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics*, **26**, 2190–2191.