ARTICLE

Genome-wide meta-analysis of macronutrient intake of 91,114 European ancestry participants from the cohorts for heart and aging research in genomic epidemiology consortium

Jordi Merino [1,2,3], Hassan S. Dashti [1,2] et al.

Abstract

Macronutrient intake, the proportion of calories consumed from carbohydrate, fat, and protein, is an important risk factor for metabolic diseases with significant familial aggregation. Previous studies have identified two genetic loci for macronutrient intake, but incomplete coverage of genetic variation and modest sample sizes have hindered the discovery of additional loci. Here, we expanded the genetic landscape of macronutrient intake, identifying 12 suggestively significant loci ($P < 1 \times 10^{-6}$) associated with intake of any macronutrient in 91,114 European ancestry participants. Four loci replicated and reached genome-wide significance in a combined meta-analysis including 123,659 European descent participants, unraveling two novel loci; a common variant in *RARB* locus for carbohydrate intake and a rare variant in *DRAM1* locus for protein intake, and corroborating earlier *FGF21* and *FTO* findings. In additional analysis of 144,770 participants from the UK Biobank, all identified associations from the two-stage analysis were confirmed except for *DRAM1*. Identified loci might have implications in brain and adipose tissue biology and have clinical impact in obesity-related phenotypes. Our findings provide new insight into biological functions related to macronutrient intake.

Introduction

Macronutrient intake refers to the proportion of calories consumed from carbohydrate, fat, and protein dietary sources and is an important modifiable risk factor for prevalent diseases such as obesity, type 2 diabetes (T2D), cardiovascular disease (CVD), and cancer [1]. The relevance of macronutrient intake and dietary quality for disease prevention is reflected by related goals across numerous public health guidelines such as the U.S. Department of Health and Human Services' 2015–2020 Dietary Guidelines for Americans [2–5]. Macronutrient

intake, and eating behavior in general, is an excellent example of a complex trait involving the simultaneous interplay among environmental, physiological, and genetic factors [6]. Genetic analyses of eating behavior, including those in family studies, have suggested that the family-based heritability of macronutrient intake ranges between 20–40% [7]. This information has generated interest in pinpointing specific genetic loci that influence macronutrient intake [7, 8].

Previous genome-wide association (GWA) study for macronutrient intake have identified associations between a genetic variant mapping near the fibroblast growth factors 21 gene (*FGF21*) [9]. The *FGF21* locus is associated with diets higher in carbohydrate and alcohol, and lower in fat and protein [9, 10]. Functional characterization studies have linked this locus with regulating food intake, macronutrient preference, and central reward pathways [11]. Earlier GWA investigations have also provided evidence for the association between an obesity and fat-mass associated locus (*FTO*) with protein intake, where individuals carrying the BMI-raising allele reported diets higher in protein [9, 12].

Investigations of other complex traits, where common genetic variants have been shown to exert modest effects, suggest that attaining a larger sample size and improving

---

These authors contributed equally: Jordi Merino, Hassan S. Dashti.

These authors jointly supervised this work: Daniel I. Chasman, Audrey Y. Chu, Toshiko Tanaka.

✉ Toshiko Tanaka
   tanakato@mail.nih.gov

Extended author information available on the last page of the article

genotyping coverage may help identify novel associations [13–15]. Thus, to advance our understanding of the genetic architecture of macronutrient intake, we conducted comprehensive GWA meta-analyses for percentage of total energy intake from carbohydrate, fat, and protein using 1000 Genomes Project-based imputation (minor allele frequency (MAF) in the range of 0.5–5% [14]) in 91,114 European ancestry participants representing 24 cohorts. We performed a two-stage analysis where the suggestive loci from the discovery stage were subsequently examined in a replication meta-analysis of 32,545 additional participants from five independent epidemiologic cohorts. The significant loci from this combined analysis were investigated in additional analysis of 144,770 participants of the UK Biobank. Finally, we applied an array of complementary computational approaches to investigate potential mechanistic functions of the novel loci associated with macronutrient intake.

Methods

Study populations

GWA included 91,114 European ancestry participants from 24 epidemiologic cohorts from the CHARGE Consortium Nutrition Working Group (Supplemental Table S1). In silico replication was conducted in 32,545 additional European ancestry participants from five epidemiologic cohort studies. Participants in discovery and replication analyses for EPIC-Norfolk and Fenland cohort studies did not overlap. Additional verification for replicated genetic variants were conducted in association analyses of up to 144,770 European ancestry participants with genetic and macronutrient intake information of the UK Biobank [16]. Participants provided written informed consent, and each cohort's study protocol was reviewed and approved by their respective institutional review board.

Assessment of macronutrient intake

Assessment tools to estimate habitual dietary intake in the participating cohorts including validated cohort-specific food frequency questionnaires (FFQ), diet history and diet records (Supplemental Table S2). The FFQ used by each cohort was tailored to best capture the dietary habits of the specific population under study. Based on the responses to each dietary assessment tool and study-specific nutrient databases, habitual nutrient consumption was estimated. Daily total energy intake was estimated from the sum of intakes of carbohydrate, fat, protein, and alcohol. The present analysis focused on the percentage of total energy intakes from carbohydrate, fat, and protein. Over-reporters

and under-reporters were excluded by standard cut-offs determined by each study cohort as part of quality control [9].

Genotyping

Genome-wide genotyping was conducted using Affymetrix or Illumina platforms. Each study performed quality control for genotyped variants based on MAF, call rate, and departure from Hardy-Weinberg Equilibrium (Supplemental Table S3). Phased haplotypes from 1000G were used to impute ~38 million autosomal variants using a Hidden Markov Model algorithm implemented in MACH/minimac [17, 18] or SHAPEIT/IMPUTE [19, 20]. Variants with low minor allele count (MAC < 20) and low imputation quality (<0.4) were removed. The number of autosomal genetic variants analyzed in this study was ~11.8 million.

Statistical analysis

Discovery and replication meta-analysis

Study-specific GWA analyses were conducted for each macronutrient using genotyped and imputed genotypes dosages assuming an additive genetic model using continuous allelic dosage values between 0 and 2. The basic model included age and sex for all studies, and study-specific covariates (e.g., study site) and population stratification principal components, where applicable. In a second model, BMI was added to the covariates to decrease variance of the macronutrient phenotypes and to account for genetic effects mediated through body composition. Since each study estimate of macronutrient consumption are comparable, the results from each study were combined in a fixed-effect meta-analysis with inverse variance weighting using METAL (version—released 25 March 2011) software [21]. To address additional inflation due to population stratification, the association results from individual studies as well as meta-analyses were adjusted for genomic control. Following the meta-analysis, genetic variants with low MAF (<0.5%) or those missing data from more than half the samples were removed. Heterogeneity across studies was tested by using Cochran's Q statistic and quantified using the heterogeneity statistic, I^2, and presented as %. Genome-wide significance was considered at the standard genome-wide Bonferroni-corrected threshold of $P < 5 \times 10^{-8}$ given that we studied three partially correlated traits. In addition, we used summary statistics from the discovery basic model GWA analyses to estimate single nucleotide polymorphism (SNP)-based heritability of each macronutrient intake using LD score regression (LDSC) [22].

To confirm the associations of loci from the GWA meta-analyses, an in silico replication of 12 variants with

suggestive significance ($P < 1 \times 10^{-6}$) was conducted in five independent epidemiologic cohort studies. We pursued replication of hits for the strongest corresponding association from the discovery analysis (i.e., macronutrient and BMI-adjusted or unadjusted model). Significant replication was considered at a Bonferroni-corrected threshold of $P_{1\text{-sided}} < 4.17 \times 10^{-3}$ (=0.05/12 loci). The results from the GWA results from the discovery and replication cohort studies were combined using a fixed-effect inverse variance-weighted meta-analysis using METAL software. For this combined analysis, genome-wide significance was also considered at the genome-wide Bonferroni-corrected threshold of $P < 5 \times 10^{-8}$. In addition, we performed follow-up analyses of the replicated genetic variants in the UK Biobank in unrelated subjects of white British ancestry with dietary data using PLINK [23] linear regression and an additive genetic model adjusted for age, sex, 10 PCs, genotyping array, and BMI (if warranted) to determine SNP effects on macronutrient intake. Similarly, we performed meta-analyses including the combined analysis (discovery and replication epidemiologic cohorts) and the UK Biobank.

Biological insights

To determine whether any of our identified genetic variants might be tagging potentially functional variants, we identified all variants within 1 Mb window and in LD ($r^2 \geq 0.8$) with our replicated-index variants. We then annotated all identified tagging variants using ANNOVAR [24]. To predict functional elements likely to be phenotypically relevant we used LINSIGHT, a computational method that combines a generalized linear model for functional genomic data with a probabilistic model of molecular evolution [25]. Next, we aimed to identify a set of 99% credible causal variants for the lead independent variants at novel loci using PAINTOR, a probabilistic framework that integrates association strength with genomic functional annotation data to improve accuracy in selecting plausible causal variants for functional validation [26]. We used regional association plots to define the locus boundaries in each region comprising the lead genetic variant. We identified the outermost variants from the set of variants in $r^2 \geq 0.4$ with the lead genetic variant. We set the maximal number of causal genetic variants in each region to three. Next, we conducted colocalization of genetic variants in regions encompassing the newly associated lead variants based on regional plots with expression quantitative trait loci (eQTL) using genotype-tissue expression (GTEx) database [27]. Finally, we used public available data from an atlas of the human long non-coding RNAs (lncRNAs), a comprehensive atlas with substantially improved gene models that integrates new data from gene expression, evolutionary conservation and genetic studies models allowing to better assess the diversity and functionality of these RNAs [28].

Cross-phenotype associations and causal inference analysis

To understand the pathways which new loci might be related to macronutrient intake, we examined the associations of the two new and two known macronutrient intake regions with a wide range of risk factors, molecular traits and clinical disorders, using Phenoscanner [29], which encompasses 137 genotype-phenotype datasets from the NHGRI-EBI GWAS catalog and other databases. We set a Bonferroni-corrected threshold for significance at $P < 3.6 \times 10^{-4}$ (=0.05/137 phenotypes). Since the top hit at *DRAM1* was only available for inflammatory diseases in Phenoscanner, we looked for association with other metabolic traits in the T2D Knowledge Portal and set a nominal *p*-value for significance [30]. We also used LD score regression [31] to estimate the genetic correlations between macronutrient intake and a range of disease outcomes and intermediate traits relating to food choice and eating behaviors (psychiatric traits, eating disorders, and used years of education as a surrogate for socioeconomic status) and cardiometabolic traits (BMI, glycemic traits, T2D, blood lipids, and coronary artery disease). Bidirectional Mendelian randomization (MR) was subsequently used to examine causality between traits found to have a significant genetic correlation. In MR analyses, our genetic instrument comprised of genetic variants associated with the proportion of protein intake that achieved genome-wide significance in the combined discovery + replication + UK Biobank meta-analyses. The genetic proxy for protein intake thus included the *FGF21* locus. Since the same *FGF21* genetic variant is a genetic proxy for fat and carbohydrate intake, we additionally evaluated whether the genetic effect on fat or carbohydrate intake raised BMI. BMI effect sizes were extracted from the largest published GWA meta-analysis for BMI of predominantly European ancestry participants [32], and supplemented with data from additional UK Biobank participants. We performed fixed-effect inverse variance-weighted meta-analysis [33], median and weighted median [34], and MR Egger approaches where we identified potential pleiotropy.
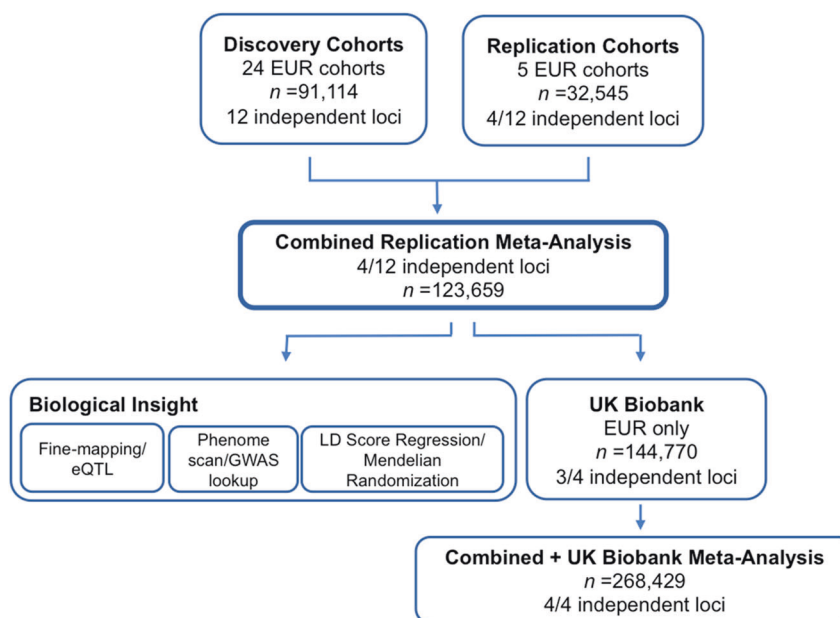
Data availability

Summary statistics of all analyses are available in dbGaP (accession number phs000930).

Results

Discovery GWA meta-analyses of percentage of total energy intake from carbohydrate, fat, and protein were

Discovery GWA meta-analysis

In discovery analyses, we tested the association of ~11.8 million genetic variants (MAF > 0.5%, imputation quality >0.4) in 91,114 participants of European ancestry with macronutrient intake. Twelve independent loci, three at genome-wide significance ($P < 5 \times 10^{-8}$) and nine at sub-genome wide significance ($P < 1 \times 10^{-6}$), showed associations with macronutrient intake either with or without BMI adjustment (Table 1; Supplemental Figure S1; Supplemental Figure S2). Estimated genetic effects sizes per each copy of the minor allele are detailed in Table 1 and Supplemental Table S5. Estimated SNP-based heritability was 3.9, 3.3, and 3.2% for carbohydrate, fat, and protein, respectively.

Replication meta-analysis

In silico replication was conducted for the set of 12 independent loci identified in the discovery meta-analyses. Replication included 32,545 additional participants of European ancestry from five epidemiologic cohort studies.

conducted with participants from 24 epidemiologic cohort studies of the CHARGE Consortium. General characteristics of participating cohort studies are presented in Supplemental Table S4. Mean macronutrient intake distribution was 48.5% ± 8.4, 32.1% ± 6.7, and 17.8% ± 3.6 for carbohydrate, fat and protein intake, respectively, and was consistent with earlier estimates of macronutrient intake distribution [9, 12] including those from population-based survey studies [35]. Figure 1 represents a schematic of the study design and main findings.

The distribution of macronutrient intake in the replication studies was consistent with the discovery studies (Supplemental Table S4). In total, four independent loci including two novel hits in Retinoic Acid Receptor Beta (*RARB*) locus, and DNA Damage Regulated Autophagy Modulator 1, (*DRAM1*) locus, and two previously known (*FGF21* and *FTO*) were confirmed in the subsequent two-stage replication of discovery findings (Table 2; Supplemental Table S6). Specifically, we replicated the association between rs7619139 in *RARB* locus and higher carbohydrate intake ($\beta = 0.20\%$ per each copy of the minor allele, SE = 0.052, $P_{\text{replication}} = 4.0 \times 10^{-5}$), achieving genome-wide significance in the combined meta-analysis including GWA results from the discovery and replication cohort studies ($\beta = 0.20\%$, SE = 0.031, $P_{\text{combined}} = 4.13 \times 10^{-11}$) (Table 2). Similarly, the association between rs77694286 in *DRAM1* locus and higher protein intake was significant in the replication analysis ($\beta = 0.55\%$ per each copy of the minor allele, SE = 0.194, $P_{\text{replication}} = 2 \times 10^{-3}$) and also achieved genome-wide significance in the combined meta-analysis ($\beta = 0.56\%$, SE = 0.092, $P_{\text{combined}} = 1.90 \times 10^{-9}$). In addition, we confirmed the previously reported associations between the *FGF21* locus (rs838133) and intake for all macronutrients, and the *FTO* locus (rs1421085) and higher protein intake (Table 2).

UK Biobank analysis

In analysis of the UK Biobank, three of the four loci achieved the Bonferroni-corrected threshold for significance. We noted similar effect sizes and directionality for *RARB* locus and higher carbohydrate intake ($\beta = 0.17\%$ per each copy of the minor allele, SE = 0.049, $P = 4.60 \times 10^{-4}$). The association

Table 1 Meta-analyses result for the discovery samples with or without BMI adjustment

SNP	Nearest gene	Chr. position[a]	Alleles[b]	MAF	Macronutrient	BMI unadjusted ($n = 91{,}114$)				BMI adjusted ($n = 90{,}407$)			
						β	SE	P	I^2	β	SE	P	I^2
rs7619139	*RARB*	3:25110415	T/A	0.41	Carbohydrate	**0.203**	**0.038**	**1.3×10^{-7}**	**3.24**	0.193	0.038	5.3×10^{-7}	0
					Fat	−0.093	0.031	2.5×10^{-3}	0	−0.087	0.031	4.6×10^{-3}	0
					Protein	−0.074	0.016	3.6×10^{-6}	2.9	−0.069	0.016	1.6×10^{-5}	3.78
rs77694286	*DRAM1*	12:102305219	G/A	0.01	Carbohydrate	−0.997	0.118	0.13	0	−0.278	0.255	0.28	0
					Fat	0.043	0.203	0.83	0	0.037	0.203	0.86	0
					Protein	**0.556**	**0.105**	**1.2×10^{-7}**	**13.17**	0.509	0.108	2.7×10^{-6}	7.64
rs1421085	*FTO*	16:53800954	C/T	0.41	Carbohydrate	−0.151	0.038	7.5×10^{-5}	8.29	−0.095	0.038	1.3×10^{-2}	2.16
					Fat	0.100	0.030	1.1×10^{-3}	2.59	0.041	0.030	0.18	0
					Protein	**0.097**	**0.016**	**8.6×10^{-10}**	**25.82**	0.073	0.015	4.0×10^{-6}	17.02
rs838133	*FGF21*	19:49259529	A/G	0.42	Carbohydrate	0.244	0.043	1.5×10^{-8}	0	**0.245**	**0.043**	**1.4×10^{-8}**	**0**
					Fat	**−0.192**	**0.035**	**3.2×10^{-8}**	**0**	−0.188	0.035	5.2×10^{-8}	0
					Protein	−0.102	0.018	1.1×10^{-8}	37.65	**−0.109**	**0.017**	**1.3×10^{-9}**	**34.13**

Increasing beta indicate % of total energy higher intake of macronutrient. I^2 represents the heterogeneity statistic, presented as %.

Bold represents the more significant result and model for the subsequent two-stage replication

Chr chromosome, *MAF* minor allele frequency, *N* total sample size

[a]Position based on the Genome Reference Consortium assembly version 37 (GRCh37 or hg19)

[b]Minor/major alleles.

between *FGF21* and *FTO* loci and macronutrient intake was also confirmed with effect sizes similar to the combined analyses (Table 2). We were unable to confirm the association between the lead *DRAM1* signal and higher protein intake in the UK Biobank ($\beta = -0.17\%$ per each copy of the minor allele, SE = 0.12, $P = 0.16$) or any other genetic variants in LD with the lead signal (results not shown). A meta-analysis including discovery cohorts, replication cohorts and the UK Biobank ($n = 268{,}429$) showed consistent evidence for *RARB*, *FTO,* and *FGF21* loci (Table 2). The *DRAM1* association with protein intake in this meta-analysis showed significant evidence of substantial heterogeneity ($I^2 = 91\%$) likely a result of the 1% MAF.

Biological insights

Figure 2 summarizes biological insights for the two variants identified for macronutrient intake. The lead genetic variant in *RARB* locus (rs7619139) is located in a lncRNA (*AC133680.1*), while the lead *DRAM1* signal (rs77694286) is an intronic variant in *DRAM1* gene. Using ANNOVAR, we did not find evidence for coding variants close to (<1 Mb) and in linkage disequilibrium (LD) ($r^2 > 0.8$) with our two index variants. We next applied LINSIGHT and showed that the lead variant at *RARB* locus is a highly constrained variant (median LINSIGHT score of 0.958), indicating a 95.8% probability of fitness consequences due to mutations at this nucleotide site (Supplemental Figure S3). No evidence of a constrained variant was detected for the *DRAM1* signal.

To identify 99% credible sets of causal variants for each lead variant at the novel loci, we identified the outermost variants from the set of variants in $r^2 \geq 0.4$ with the lead

variant using regional association plots to define the locus boundaries. The 99% credible sets included 102 and 127 variants for the *RARB* and *DRAM1* loci, respectively. The lead variant at *RARB* was the best-ranked variant in the region (posterior probability = 0.99 and Z-score = 4.63). The variant is annotated as a functional variant with a high CADD score [Combined Annotation Dependent Depletion (CADD) score = 21.1], a score that integrates multiple annotations into one metric by contrasting variants that survived natural selection with simulated mutations [36], and was predicted to be deleterious (Supplemental Figure 4A). For *DRAM1*, the GWA signal as well as two other variants (rs58512731 and rs78927281) in perfect LD with the lead variant are the most likely causal variants in the region. These variants lie in super-enhancers of *DRAM1* (super-enhancer 32592) (Supplemental Figure 4B). In an eQTL analysis of regions encompassing the newly associated lead variants, we did not detect any significant eQTL for *RARB* within the high LD window. However, the allele associated with higher protein intake in *DRAM1* region was associated with lower expression of *DRAM1* in several tissues including subcutaneous adipose tissue ($P = 5.1 \times 10^{-8}$), artery ($P = 2.9 \times 10^{-7}$), esophagus ($P = 6.5 \times 10^{-10}$), left ventricle ($P = 1.6 \times 10^{-8}$), skin ($P = 5.6 \times 10^{-8}$) and tibial nerve ($P = 3.1 \times 10^{-8}$) (Supplemental Table S7). Finally, we integrated data from the Atlas of the Human lncRNAs to gain insights from the lead GWA hit at *RARB* locus lying in the lncRNA *AC133680.1*. We observed that the lncRNA is differentially expressed in several brain regions [the strongest being in the caudate nucleus (39.3 fold-change increase), and plays a role in H1-neuronal progenitor cells differentiation (3.7 fold-change increase, false discovery rate

Table 2 Results from replication and meta-analysis of discovery and (A) replication and (B) UK Biobank

A.

SNP	Nearest gene	Chr. position[a]	Alleles[b]	Macronutrient	MAF	Replication ($n = 32,545$)				Discovery + replication ($n = 123,659$)			
						β	SE	P	I^2	β	SE	P	I^2
rs7619139	*RARB*	3:25110415	T/A	Carbohydrate	0.41	0.203	0.052	4.0×10^{-5}	40.8	0.203	0.031	4.1×10^{-11}	0
rs77694286	*DRAM1*	12:102305219	G/A	Protein	0.01	0.551	0.194	2.0×10^{-3}	0	0.555	0.092	1.9×10^{-9}	0
rs1421085	*FTO*	16:53800954	C/T	Protein	0.41	0.092	0.024	5.9×10^{-5}	0	0.095	0.013	4.7×10^{-13}	0
rs838133	*FGF21*	19:49259529	A/G	Carbohydrate[c]	0.43	0.264	0.054	5.2×10^{-7}	19.8	0.252	0.034	7.7×10^{-14}	0
				Fat	0.43	−0.152	0.050	1.0×10^{-3}	0	−0.179	0.029	3.4×10^{-10}	0
				Protein[c]	0.43	−0.093	0.026	1.3×10^{-4}	21.7	−0.104	0.015	1.6×10^{-12}	0

B.

SNP	Nearest gene	Chr. position[a]	Alleles[b]	Macronutrient	MAF	UK Biobank ($n = 144,770$)				Combined + UK Biobank ($n = 268,429$)			
						β	SE	P		β	SE	P	I^2
rs7619139	*RARB*	3:25110415	T/A	Carbohydrate	0.41	0.138	0.021	1.1×10^{-5}		0.159	0.017	5.9×10^{-20}	34.7
rs77694286	*DRAM1*	12:102305219	G/A	Protein	0.01	−0.172	0.123	0.16		0.293	0.074	7.5×10^{-5}	91.0
rs1421085	*FTO*	16:53800954	C/T	Protein	0.40	0.062	0.014	5.0×10^{-6}		0.079	0.010	1.0×10^{-16}	34.2
rs838133	*FGF21*	19:49259529	A/G	Carbohydrate[c]	0.45	0.212	0.033	1.1×10^{-10}		0.232	0.024	9.2×10^{-23}	0
				Fat	0.45	−0.176	0.027	6.1×10^{-11}		−0.177	0.020	1.4×10^{-19}	0
				Protein[c]	0.45	−0.118	0.014	1.1×10^{-16}		−0.111	0.010	5.0×10^{-28}	0

Increasing beta indicate % of total energy higher intake of macronutrient. I^2 represents the heterogeneity statistic, presented as %

Chr chromosome, *MAF* minor allele frequency, *N* total sample size

[a]Position based on the Genome Reference Consortium assembly version 37 (GRCh37 or hg19)

[b]Minor(effect)/major alleles

[c]BMI-adjusted

(FDR) at 5% $P = 1.03 \times 10^{-3}$), as well as cardiomyocyte (5.5 fold-change increase, FDR at 5% $P = 3.87 \times 10^{-4}$), and melanocyte differentiation (2.9 fold-change increase, FDR at 5% $P = 3.22 \times 10^{-3}$)] (Supplemental Table S8, Supplemental Figure S5).

Cross-phenotype associations of significant loci

To investigate the clinical importance of the macronutrient intake loci, we examined the associations of the identified loci with a range of disease risk factors, molecular traits, and clinical disorders. The *RARB* rs7619139 T-allele associated with higher carbohydrate intake was also associated with lower BMI ($\beta = -0.019$, SE = 0.004, $P = 3.32 \times 10^{-6}$) (Supplemental Table S9). The *DRAM1* rs77694286 G-allele associated with higher protein intake displayed significant association with higher T2D risk (OR = 1.89, (95%CI: 1.36–2.54); $P = 0.019$). Associations for the *FGF21* and *FTO* loci are also listed in Supplemental Table S9.

Genetic correlation and causal inference analysis

We examined the genetic correlation between macronutrient intake and a range of disease outcomes and intermediate traits using LDSC. We found an inverse genetic correlation between the intake of carbohydrate and fat ($r = -0.78$; $P < 0.001$), carbohydrate and protein ($r = -0.33$; $P < 0.001$), but not fat and protein. (Supplemental Figure S6). We found a moderate concordant genetic correlation between protein intake and BMI ($r_g = 0.23$, $P = 4 \times 10^{-4}$; Supplemental Figure S6). Also, we found an inverse genetic correlation between dietary fat intake and years of education ($r_g = -0.24$, $P = 5 \times 10^{-4}$; Supplemental Figure S6). Upon identifying a significant genetic correlation between higher protein intake and higher BMI, we used a bi-directional MR approach to investigate whether genetically driven protein intake has a causal role for BMI and vice versa. We found that genetically raised protein intake (per 1% of total energy intake) was associated with higher BMI ($\beta = 0.09 \, \text{kg/m}^2$, SE = 0.03, $P = 6.92 \times 10^{-4}$) (Table 3). Given that the lead variant at *FGF21* was associated with all three macronutrients, it is possible that any one of the macronutrients might genetically raise BMI. Conversely, we also noted that genetically determined higher BMI increased the amount of protein intake, but not carbohydrate or fat intake (Table 3). A one standard deviation (1-SD) increase in BMI due to a 93-variant polygenic risk score (excluding *FTO*) was associated with
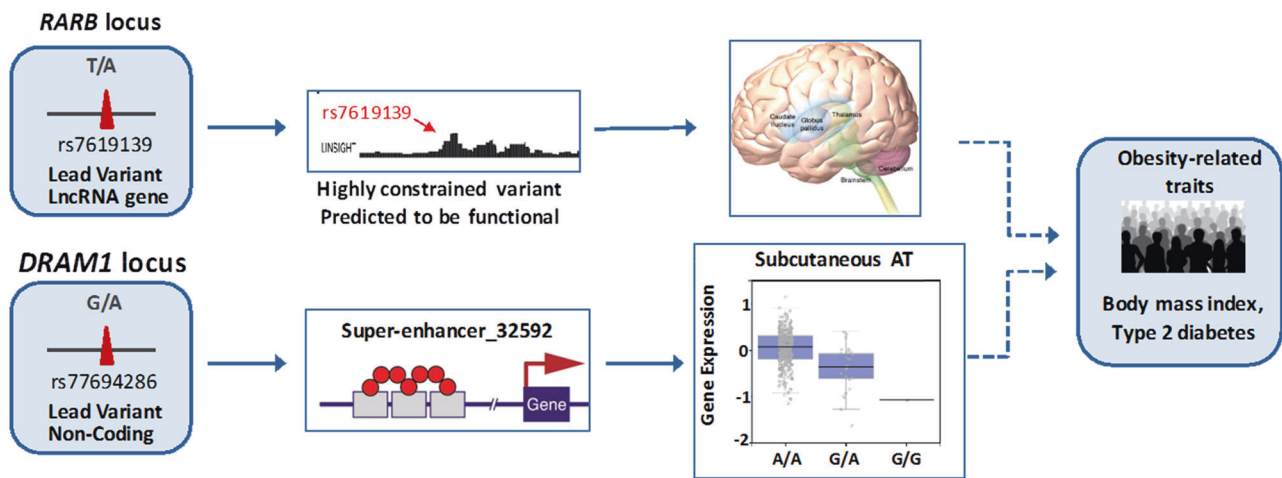
Fig. 2 Biological insight of the two replicated variants in CHARGE. Data generated from an array of complementary approaches including genetic fine-mapping, epigenetic regulation and expression data to investigate the biological relevant of novel variants. Through genetic fine mapping, we showed that the lead variant at *RARB* locus associated with CHO intake was the best-ranked variant in the region. This variant is placed in a lncRNA, is a highly constrained variant and predicted to be deleterious. We showed that the lncRNA gene is differently expressed in several brain regions and is involved in neuronal progenitor cells differentiation. For *DRAM1*, the GWA hit is predicted to be the most likely causal variant in the region and lie in a super-enhancer of *DRAM1* (super-enhancer 32592). Gene-tissue expression showed that the allele associated with higher protein intake was associated with lower expression (rank normalized expression) of *DRAM1* in subcutaneous adipose tissue (A/A; Homozygote reference, $n = 359$. G/A; Heterozygote, $n = 29$, G/G; Homozygote alternate, $n = 1$). The clinical transcendence of the two identified loci in CHARGE is supported by the association with obesity-related traits. *DRAM1* findings did not replicate when using data from the UK Biobank data alone

0.58% higher protein intake (SE = 0.08, $P = 9.88 \times 10^{-13}$) (Table 3, Supplemental Figure S7).

Discussion

In this study including data from up to 91,114 participants from European ancestry, we identified 12 suggestively significant loci ($P < 1 \times 10^{-6}$) associated with macronutrient intake including four genome-wide significant loci in combined meta-analysis from discovery and replication cohort studies. Meta-analysis including up to 123,659 individuals supported a novel common variant in *RARB* locus associated with 0.20% higher carbohydrate intake, a novel rare variant in the *DRAM1* locus (MAF = 1% associated with 0.55% higher protein intake, and corroborated previous findings between *FGF21* with higher carbohydrate intake and lower fat and protein intake, and *FTO* with higher protein intake [9, 12]. In additional analysis of 144,770 participants from the UK Biobank, all identified associations from the two-stage analysis were confirmed except for *DRAM1*, which warrants further investigation given its 1% MAF. The identified loci are predicted to be relevant regulatory regions mainly functional in brain and subcutaneous adipose tissues. The clinical translation of these variants is supported by the associations with obesity related-traits.

Suboptimal diets represent a major driving force behind escalating obesity epidemic worldwide and their associated risk of T2D, CVD, and cancer [37]. Ecologic studies suggest that increasing intake of carbohydrates, especially added sugars, is most strongly linked to these trends [38, 39]. The present analysis suggests that a regulatory common genetic variant in the *RARB* locus, situated in the lncRNA *AC133680.1*, is associated with increased carbohydrate intake. The *RARB* locus has been identified as a novel obesity locus in a recent GWA meta-analysis for BMI [32]. Fine-mapping confirmed that the identified variant in this study, in high LD with the BMI reported variant (rs6804842, LD = 0.89), is likely to be the causal variant in the region. We showed that the identified variant is differentially expressed in several brain regions, where the strongest association was seen for caudate nucleus (39.3 fold-change increase). In humans, ingestion of high-density and palatable food, such as added sugar foods, has been shown to release dopamine in the caudate and putamen regions [40]. Still, the potential functional overlap between the lncRNA *AC133680.1* and other relevant genes in the region supports further explorations of biological implications. Our findings may serve as preliminary evidence for the design and implementation of in vitro and in vivo experimentations investigating how this genetic variant might contribute to food selection in humans. In this regard, the identification and characterization of relevant macronutrient intake genes, such as *FGF21*, has contributed evidence that help create the framework to develop an *FGF21* analogue that has been shown to suppress sugar intake, sweet taste preference, and decrease central reward

Table 3 Estimated associations of each genetically predicted exposure on outcome[1]

Exposure	Outcome	Scaling of OR	SNP, genetic locus	MR-IVW β (se)	MR-IVW P	MR-Egger β (se)	MR-Egger P	MR-WM β (se)	MR-WM P
Protein intake (adj BMI)*	BMI[a]	1%TEI from protein	rs838133 (FGF21)	0.09 (0.03)	6.92×10^{-4}				
Fat intake (unadj BMI)*	BMI[a]	1%TEI from fat	rs838133 (FGF21)	0.05 (0.02)	6.92×10^{-4}				
Carbohydrate (unadj BMI)*	BMI[a]	1%TEI from CHO	rs838133 (FGF21), rs7619139 (RARB)	−0.06 (0.01)	4.8×10^{-09}				
BMI	Protein intake[a]	1 SD of BMI	94 BMI SNPs excluding FTO	0.58 (0.08)	9.88×10^{-13}	0.92 (0.23)	6.93×10^{-5}	0.66 (0.11)	2.41×10^{-9}
BMI	Protein intake[b]	1 SD of BMI	94 BMI SNPs excluding FTO	0.35 (0.08)	1.8×10^{-5}	0.64 (0.23)	6×10^{-3}	0.39 (0.11)	6.23×10^{-4}
BMI	Fat intake[a]	1 SD of BMI	94 BMI SNPs excluding FTO	0.13 (0.19)	0.48	−0.78 (0.52)	0.13	−0.10 (0.22)	0.65
BMI	Fat intake[b]	1 SD of BMI	94 BMI SNPs excluding FTO	−0.36 (0.19)	0.06	1.45 (0.53)	6×10^{-3}	−0.49 (0.23)	0.03
BMI	CHO intake[a]	1 SD of BMI	94 BMI SNPs excluding FTO	−0.43 (0.24)	0.07	0.30 (0.68)	0.66	−0.50 (0.29)	0.09
BMI	CHO intake[b]	1 SD of BMI	94 BMI SNPs excluding FTO	0.17 (0.25)	0.48	0.86 (0.70)	0.22	−0.01 (0.29)	0.96

IVW inverse variance weighted method, *WM* weighted median, *CHO* carbohydrate, *N* sample size, *OR* odds ratio, *SNP* single nucleotide polymorphism, *MR* Mendelian randomization

[a]BMI-unadjusted

[b]BMI-adjusted

[1]Bidirectional Mendelian randomization analysis on the effect of genetically driven macronutrient intake on BMI obtained from estimates derived from this study and publicly-available data from GIANT Consortium and the UK Biobank

*Based on combined GWA results for respective macronutrient intake. Modeling is fixed

pathways when administered in monkeys with obesity and humans with obesity and T2D [11].

In this study, we used 1000G imputation reference panel which allowed us to identify a rare variant in *DRAM1* locus for the association with protein intake. *DRAM1* encodes the DNA damage regulated autophagy modulator, a lysosomal protein that is required for induction of autophagy by the p53 pathway [41]. Through genetic fine mapping, we showed that the GWA hit is the most likely causal variant in the region together with other two variants in perfect LD with the lead variant. These variants lie in a super-enhancer region of *DRAM1* (super-enhancer 32,592), and the protein increasing G-allele in rs77694286 is associated with higher T2D risk and lower expression of *DRAM1* in subcutaneous adipose tissue. The association between *DRAM1* and protein intake was not verified in additional analysis of 144,770 subjects from the UK Biobank. A potential explanation for the lack of confirmation of the *DRAM1* findings in the UK Biobank may be due the MAF of this variant. Therefore, our observation linking *DRAM1* with protein intake requires further evaluation.

In our discovery analysis, worth noting is a novel genome-wide signal in the *ABO* locus for protein intake that did not replicate in subsequent analyses. This variant is in perfect LD with an intronic genetic polymorphism in the *ABO* gene (rs651007) and associates with a host of cardiometabolic traits including higher fasting glucose levels and moderate increases in T2D risk [42]. A recent study has also identified that the minor allele at this polymorphism significantly interacts with higher dietary fat intake to exacerbate BMI [43]. Pending replication of the interaction observation and the present *ABO* association with protein intake, these observations may be utilized for future studies to better understand the genetic architecture of macronutrient intake and related metabolic outcomes.

The observation of a moderate concordant genetic correlation between protein intake and BMI, but not other obesity-related traits, suggests that genetic effects of higher protein intake are shared with greater BMI genome-wide. First, we observed reasonably clear evidence to support causality for BMI with protein intake. However, since our GWA were with and without adjustment for BMI, we cannot exclude the potential for BMI-related bias (i.e. reporting bias in those who are overweight or obese) to account for our current MR observations [44]. A previous report highlighted that the *FTO* locus was associated with higher protein intake [45], a finding we now extend to other BMI-raising alleles, suggest that higher BMI is associated with higher reported protein intake not specific to the effect of *FTO* (Supplemental Figure S6). Second, for our MR analysis, the lead variant at *FGF21* associated with percentage of energy intake from protein was also associated with percentage of energy from carbohydrate and fat intake,

therefore we cannot ascribe causality between one specific macronutrient group with BMI or whether this reflects substitution of macronutrients as a proportion of total energy intake. A meta-analysis of randomized controlled trials demonstrated that diets with any macronutrient composition result in weight loss [46]. However, given that the current macronutrient intake genetic predisposition is limited to a few number of variants, future research using more genetically increased macronutrient consumption is needed to confirm these findings.

In the U.S., dietary factors are estimated to account for >650,000 deaths per year and 14% of all disability-adjusted life-years lost [47]. The nutritional shift towards increased consumption of ultra-processed foods has been a consequence of globalization and rapid economic development during the last few decades [48]. Family-based heritability estimates for macronutrient intake ranges between 20–40 %, whereas SNP-based heritability reported previously [9] and here indicate more modest heritability estimates. Although the small estimates indicate environmental variation plays a major role in explaining the remainder of the trait, genes still have a role in explaining a significant proportion of macronutrient intake heritability. Our analysis indicates shared genetic correlations across macronutrients, particularly carbohydrate with both fat and protein. Understanding the biological basis of dietary intake can help guide future studies and shape public health initiatives. Our results may be used to assess dietary pattern recommendations based on genetic risk profile, or may be used by recall-by-genotype studies to evaluate whether a dietary pattern tailored to an individual's genetic risk will lead to more desirable health outcomes. Nevertheless, there are several challenges in identifying and validating genetic associations for macronutrient intake worth noting. Variability in dietary intake across geographical locations, error in dietary assessment, differences in allele frequencies across studies, imperfect imputation within studies, and ancestry-specific LD patterns may hinder discovery and replication of genetic associations and could potentially induce false positive findings. In addition, noting the variability in dietary habits across populations and the use of dissimilar dietary assessment tools across studies is particularly relevant for meta-analyses of lifestyle traits [49]. Although the present investigation was limited to individuals of European ancestry for the purpose of reducing ancestry-specific LD patterns, we cannot account for differences in other intrinsic factors across studies. In addition, the present findings require further validation in individuals of other ancestries. Finally, because we studied three partially correlated traits, we set the p-value for statistical significance at the genome-wide Bonferroni-corrected threshold of $P < 5 \times 10^{-8}$, and whether a more stringent threshold is more appropriate for partially correlated traits is unclear.

In summary, our results provide compelling novel evidence of the genetic architecture of macronutrient intake and contribute biological insights relating dietary intake to the central nervous system and adipose tissue biology. Our findings add to the current understanding of macronutrient intake and are hypothesis-generating for future studies.

Authors contributions D.I.C., T.T., and A.Y.C. conceived and designed the study. A.E.J., M.G., C.P., G.V.D., R.N.L., M.K.W., S.M., J.S.N., M.K., T.S.A., N.P., D.K.H., C.B., T.H., M.O., A.C.F.-W., D.O.M.-K., L.P., C.E.P., P.S.V., T.V., O.L., S.K., L.M.R., T.L., J.H.Z., M.F.F., J.L., N.M.M., J.A.S., T.H., N.E., M.A.N., T.R., J.H., D.G.H., C.-A.S., A.M., R.L.-G., M.-C.V., C.A.W., F.J.A.R., J.S., I.P.K., F.D., P.M.R., M.K., D.S.S., C.L., W.Z., A.A., P.K., M.G., D.C.R., Q.Q., L.F., U.E., J.B., A.H., Z.P., V.M., N.J.W., S.L.R.K., O. P., A.J., J.E.C., M.C.Z., J.S.V., N.G.F., J.M.O., J.C.L., H.R., A.G.U., O.T.R., J.C.K.-d.J., J.D., J.I.R., K.W.N., R.A.S., M.A.P., M.P., L.A. C., S.T.T., T.I.A.S., V.K., Y.L., Y.S., L.Q., S.B., S.S.R., R.M., A.T., W.H.O., O.H.F., T.P., J.C.F., P.D., L.-P.L., contributed to data collection. J.M., H.S.D., D.I.C., T.T., and A.Y.C. performed quality control and meta-analyses. J.M., A.E.J., C.S., S.X.L. performed additional analyses including look-up in the N.H.G.R.I.-E.B.I. G.W.A.S. Catalog, fine-mapping analyses, colocalization analyses for Expression Quantitative Trait loci and Human Atlas of long non-coding R.N.A., L.D. score regression analyses and Mendelian randomization. J.M., H. S.D., S.X.L., A.E.J., M.G., C.P., C.E.S., G.V.D., D.I.C., T.T., and A.Y.C. wrote the manuscript. All authors approved the final version of the manuscript.

Compliance with ethical standards

References

1. Ezzati M, Riboli E. Behavioral and dietary risk factors for non-communicable diseases. N Engl J Med. 2013;369:954–64.
2. U.S. Department of Health and Human Services and U.S. Department of Agriculture. *2015 – 2020 Dietary Guidelines for Americans*. 8th edn. December 2015. https://health.gov/dietaryguidelines/2015/guidelines/.
3. The Eatwell Guide -GOV.UK. https://www.gov.uk/government/publications/the-eatwell-guide
4. Montagnese C, Santarpia L, Buonifacio M, Nardelli A, Caldara AR, Silvestri E, et al. European food-based dietary guidelines: a comparison and update. Nutrition. 2015;31:908–15.

5. Food based dietary guidelines in the WHO European Region Nutrition and Food Security Programme WHO Regional Office for Europe Scherfigsvej 8, 2100 Copenhagen Denmark; http://www.euro.who.int/__data/assets/pdf_file/0017/150083/E79832.pdf

6. de Castro JM. The control of food intake of free-living humans: putting the pieces back together. Physiol Behav. 2010;100:446–53.

7. Rankinen T, Bouchard C. Genetics of food intake and eating behavior phenotypes in humans. Annu Rev Nutr. 2006;26:413–34.

8. Teucher B, Skinner J, Skidmore PML, Cassidy A, Fairweather-Tait SJ, Hooper L, et al. Dietary patterns and heritability of food choice in a UK female twin cohort. Twin Res Hum Genet. 2007;10:734–48.

9. Chu AY, Workalemahu T, Paynter NP, Rose LM, Giulianini F, Tanaka T, et al. Novel locus including FGF21 is associated with dietary macronutrient intake. Hum Mol Genet. 2013;22:1895–902.

10. Søberg S, Sandholt CH, Jespersen NZ, Toft U, Madsen AL, von Holstein-Rathlou S, et al. FGF21 Is a sugar-induced hormone associated with sweet intake and preference in humans. Cell Metab. 2017;25:1045–53.e6.

11. Potthoff MJ. FGF21 and metabolic disease in 2016: a new frontier in FGF21 biology. Nat Rev Endocrinol. 2017;13:74–6.

12. Tanaka T, Ngwa JS, van Rooij FJA, Zillikens MC, Wojczynski MK, Frazier-Wood AC, et al. Genome-wide meta-analysis of observational studies shows common genetic variants associated with macronutrient intake. Am J Clin Nutr. 2013;97:1395–402.

13. Spencer CCA, Su Z, Donnelly P, Marchini J. Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. PLoS Genet. 2009;5:e1000477.

14. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491:56–65.

15. de Vries PS, Sabater-Lleal M, Chasman DI, Trompet S, Ahluwalia TS, Teumer A, et al. Comparison of HapMap and 1000 genomes reference panels in a large-scale genome-wide association study. Yao Y-G, editor. PLoS ONE. 2017;12:e0167742.

16. Manolio TA, Weis BK, Cowie CC, Hoover RN, Hudson K, Kramer BS, et al. New models for large prospective studies: is there a better way? Am J Epidemiol. 2012;175:859–66.

17. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol. 2010;34:816–34.

18. Fuchsberger C, Abecasis GR, Hinds DA. minimac2: faster genotype imputation. Bioinformatics. 2015;31:782–4.

19. Delaneau O, Marchini J, 1000 Genomes Project Consortium GA, 1000 Genomes Project Consortium P, Lunter G, Marchini JL, et al. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. Nat Commun. 2014;5:3934.

20. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet. 2012;44:955–9.

21. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26:2190–1.

22. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium N. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015;47:291–5.

23. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association

24. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nat Protoc. 2015;10:1556–66.

25. Huang Y-F, Gulko B, Siepel A. Fast, scalable prediction of deleterious noncoding variants from functional and population genomic data. Nat Genet. 2017;49:618–24.

26. Kichaev G, Yang W-Y, Lindstrom S, Hormozdiari F, Eskin E, Price AL, et al. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. Di Rienzo A, editor. PLoS Genet. 2014;10:e1004722.

27. Ardlie KG, Deluca DS, Segre AV, Sullivan TJ, Young TR, Gelfand ET, et al. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015;348:648–60.

28. Hon C-C, Ramilowski JA, Harshbarger J, Bertin N, Rackham OJL, Gough J, et al. An atlas of human long non-coding RNAs with accurate 5' ends. Nature. 2017;543:199–204.

29. Staley JR, Blackshaw J, Kamat MA, Ellis S, Surendran P, Sun BB, et al. PhenoScanner: a database of human genotype-phenotype associations. Bioinformatics. 2016;32:3207–9.

30. T2D-GENES Consortium, GoT2D Consortium, DIAGRAM Consortium. 2017.

31. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015;47:1236–41.

32. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015;518:197–206.

33. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013;37:658–65.

34. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genet Epidemiol. 2016;40:304–14.

35. Ford ES, Dietz WH. Trends in energy intake among adults in the United States: findings from NHANES. Am J Clin Nutr. 2013;97:848–53.

36. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46:310–5.

37. Abajobir AA, Abate KH, Abbafati C, Abbas KM, Abd-Allah F, Abdulle AM, et al. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017;390:1345–422.

38. Basu S, Yoffe P, Hills N, Lustig RH. The relationship of sugar to population-level diabetes prevalence: an econometric analysis of repeated cross-sectional data. PLoS ONE. 2013;8:e57873.

39. Siervo M, Montagnese C, Mathers JC, Soroka KR, Stephan BCM, Wells JCK. Sugar consumption and global prevalence of obesity and hypertension: an ecological analysis. Public Health Nutr. 2014;17:587–96.

40. Volkow ND, Wang G-J, Baler RD. Reward, dopamine and the control of food intake: implications for obesity. Trends Cogn Sci. 2011;15:37–46.

41. Crighton D, Wilkinson S, O'Prey J, Syed N, Smith P, Harrison PR, et al. DRAM, a p53-induced modulator of autophagy, is critical for apoptosis. Cell. 2006;126:121–34.

42. Wessel J, Chu AY, Willems SM, Wang S, Yaghootkar H, Brody JA, et al. Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. Nat Commun. 2015;6:5897.

and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75.

43. Almgren P, Lindqvist A, Krus U, Hakaste L, Ottosson-Laakso E, Asplund O, et al. Genetic determinants of circulating GIP and GLP-1 concentrations. JCI insight. 2017;2:e93306.

44. Heitmann BL, Lissner L. Dietary underreporting by obese individuals--is it specific or non-specific? BMJ. 1995;311: 986–9.

45. Qi Q, Kilpeläinen TO, Downer MK, Tanaka T, Smith CE, Sluijs I, et al. FTO genetic variants, dietary intake, and body mass index: insights from 177,330 individuals. Hum Mol Genet. 2014;23:6961–72.

46. Johnston BC, Kanters S, Bandayrel K, Wu P, Naji F, Siemieniuk Ra, et al. Comparison of weight loss among named diet programs in overweight and obese adults. JAMA. 2014;312:923–33.

47. Murray CJL, Atkinson C, Bhalla K, Birbeck G, Burstein R, Chou D, et al. The state of US health, 1990-2010: burden of diseases, injuries, and risk factors. JAMA. 2013;310:591–608.

48. Hu FB, Satija A, Manson JE. Curbing the diabetes pandemic: the need for global policy solutions. JAMA. 2015;313:2319–20.

49. Barnard ND, Willett WC, Ding EL. The misuse of meta-analysis in nutrition research. JAMA. 2017;318:1435.

Affiliations

Jordi Merino [1,2,3] · Hassan S. Dashti [1,2] · Sherly X. Li[4] · Chloé Sarnowski[5] · Anne E. Justice[6,7] · Misa Graff[7] · Constantina Papoutsakis[8] · Caren E. Smith[9] · George V. Dedoussis[10] · Rozenn N. Lemaitre[11] · Mary K. Wojczynski[12] · Satu Männistö[13] · Julius S. Ngwa[5,14] · Minjung Kho[15] · Tarunveer S. Ahluwalia [16] · Natalia Pervjakova[13,17,18,19] · Denise K. Houston[20] · Claude Bouchard[21] · Tao Huang[22] · Marju Orho-Melander[23] · Alexis C. Frazier-Wood[24] · Dennis O. Mook-Kanamori[25,26] · Louis Pérusse [27,28] · Craig E. Pennell[29] · Paul S. de Vries[30] · Trudy Voortman [31] · Olivia Li[32] · Stavroula Kanoni[33] · Lynda M. Rose[34] · Terho Lehtimäki[35] · Jing Hua Zhao[4] · Mary F. Feitosa[12] · Jian'an Luan[4] · Nicola M. McKeown[36] · Jennifer A. Smith [15] · Torben Hansen [16] · Niina Eklund[13] · Mike A. Nalls[37,38] · Tuomo Rankinen[21] · Jinyan Huang[39] · Dena G. Hernandez[38] · Christina-Alexandra Schulz[23] · Ani Manichaikul[40] · Ruifang Li-Gao[25] · Marie-Claude Vohl[28,41] · Carol A. Wang [29] · Frank J. A. van Rooij[31] · Jean Shin[32] · Ioanna P. Kalafati[10] · Felix Day [4] · Paul M. Ridker[34,42] · Mika Kähönen[35] · David S. Siscovick[43] · Claudia Langenberg[4] · Wei Zhao[15] · Arne Astrup [44] · Paul Knekt[13] · Melissa Garcia[45] · D. C. Rao[46] · Qibin Qi[47] · Luigi Ferrucci[48] · Ulrika Ericson[23] · John Blangero[49] · Albert Hofman[31,50] · Zdenka Pausova[32,51] · Vera Mikkilä[52] · Nick J. Wareham[4] · Sharon L. R Kardia[15] · Oluf Pedersen[16] · Antti Jula[13] · Joanne E. Curran[49] · M. Carola Zillikens[53] · Jorma S. Viikari[54,55] · Nita G. Forouhi[4] · José M. Ordovás[9,56,57] · John C. Lieske[58] · Harri Rissanen[13] · André G. Uitterlinden[31,53] · Olli T. Raitakari[59,60] · Jessica C. Kiefte-de Jong[31,61] · Josée Dupuis[5] · Jerome I. Rotter[62,63] · Kari E. North[7,64] · Robert A. Scott[4] · Michael A. Province[12] · Markus Perola[13] · L. Adrienne Cupples[5,65] · Stephen T. Turner[58] · Thorkild I. A. Sørensen[16,66] · Veikko Salomaa[13] · Yongmei Liu[67] · Yun J. Sung[46] · Lu Qi[68] · Stefania Bandinelli[69] · Stephen S. Rich[70] · Renée de Mutsert[25] · Angelo Tremblay[27] · Wendy H. Oddy[71,72] · Oscar H. Franco[31] · Tomas Paus[73,74,75] · Jose C. Florez[1,2,3] · Panos Deloukas [33,76] · Leo-Pekka Lyytikäinen [35] · Daniel I. Chasman[35,77] · Audrey Y. Chu[34] · Toshiko Tanaka[48]

1   Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

2   Programs in Metabolism and Medical & Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA

3   Diabetes Unit, Massachusetts General Hospital, Boston, MA, USA

4   MRC Epidemiology Unit, University of Cambridge, Cambridge, UK

5   Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA

6   Biomedical and Translational Informatics Institute, Geisinger Health Weis Center for Research, Danvilla, PA, USA

7   Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

8   Research International and Scientific Affairs, Academy of Nutrition and Dietetics, Chicago, IL, USA

9   Nutrition and Genomics, JM-USDA-HNRCA at Tufts University, Boston, MA, USA

10 Department of Dietetics and Nutritional Science, School of Health Science and Education, Harokopio University, Athens, Greece

11 Departments of Medicine, University of Washington, Seattle, USA

12 Department of Genetics, Washington University School of Medicine, Saint Louis, MO, USA

13 National Institute for Health and Welfare, Helsinki, Finland

14 Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

15 Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA

16 Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

17 Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia

18 Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia

19 Genomics of Common Disease, Division of Diabetes, Endocrinology and Metabolism, Department of Medicine,

Imperial College London, London, UK

[20] Department of Internal Medicine, Wake Forest School of Medicine, Winston Salem, NC, USA

[21] Human Genomics Laboratory, Pennington Biomedical Research Center, Baton Rouge, LA, USA

[22] Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Peking, China

[23] Department of Clinical Sciences, Lund University Diabetes Center, Lund University, Malmö, Sweden

[24] USDA/ARS Children's Nutrition Research Center, Houston, TX, USA

[25] Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

[26] Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, The Netherlands

[27] Department of Kinesiology, Faculty of Medicine, Laval University, Québec, Canada

[28] Institute of Nutrition and Functional Foods, Québec, Canada

[29] School of Medicine and Public Health, Faculty of Medicine and Health, The University of Newcastle, Callaghan, Australia

[30] Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX, USA

[31] Department of Epidemiology, Erasmus MC University Hospital, Rotterdam, The Netherlands

[32] The Hospital for Sick Children, Translational Medicine, University of Toronto, Toronto, Canada

[33] William Harvey Research Institute, Queen Mary University of London, London, UK

[34] Division of Preventive Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

[35] Department of Clinical Chemistry, Fimlab Laboratories, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland

[36] Nutritional Epidemiology, JM-USDA-HNRCA at Tufts University, Boston, MA, USA

[37] Data Tecnica International, Glen Echo, MD, USA

[38] Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA

[39] Department of Bioinformatics, Shanghai Institute of Hematology, Shanghai, China

[40] Center for Public Health Genomics, Division of Biostatistics and Epidemiology, University of Virginia, Charlottesville, VA, USA

[41] School of Nutrition, Laval University, Québec, Canada

[42] Division of Cardiovascular Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

[43] New York Academy of Medicine, New York, NY, USA

[44] Department of Nutrition, Exercise, and Sports, Faculty of Science,

University of Copenhagen, Copenhagen, Denmark

[45] Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Bethesda, MD, USA

[46] Division of Biostatistics, Washington University School of Medicine, St. Louis, MO, USA

[47] Department of Epidemiology & Population Health, Albert Einstein College of Medicine, Bronx, NY, USA

[48] Translational Gerontology Branch, National Institute on Aging, Baltimore, MD, USA

[49] South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley, Brownsville, TX, USA

[50] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

[51] Departments of Physiology and Nutritional Sciences, University of Toronto, Toronto, Canada

[52] Department of Food and Environmental Sciences, Division of Nutrition, Helsinki, Finland

[53] Department of Internal Medicine, Erasmus MC University Hospital, Rotterdam, The Netherlands

[54] Division of Medicine, Turku University Hospital, Turku, Finland

[55] Department of Medicine, University of Turku, Turku, Finland

[56] Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid, Spain

[57] IMDEA Food Institute, CEI UAM + CSIC, Madrid, Spain

[58] Division of Nephrology and Hypertension, Mayo Clinic, Rochester, MN, USA

[59] Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland

[60] Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland

[61] Global Public Health, Leiden University College, The Hague, The Netherlands

[62] Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA

[63] Departments of Pediatrics and Medicine, Division of Genomic Outcomes, Harbor-UCLA Medical Center, Torrance, CA, USA

[64] Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill, NC, USA

[65] NHLBI Framingham Heart Study, Framingham, MA, USA

[66] Department of Public Health, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

[67] Department of Epidemiology and Prevention, Wake Forest School of Medicine, Winston Salem, NC, USA

[68] Department of Epidemiology, Tulane University, New Orleans, LA, USA

[69] Geriatric Unit, Azienda Sanitaria Firenze-ASF, Florence, Italy

[70] Department of Public Health Sciences, Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA

[71] Telethon Kids Institute, The University of Western Australia, Perth, Western Australia, Australia

[72] Menzies Institute for Medical Research, University of Tasmania, Tasmania, Australia

[73] Rotman Research Institute, University of Toronto, Toronto, Canada

[74] Department of Psychiatry and Psychology, University of Toronto, Toronto, Canada

[75] Child Mind Institute, New York, NY, USA

[76] Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, Saudi Arabia

[77] Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA