

## Article

# Genome-Wide Association Analysis of Over 170,000 Individuals from the UK Biobank Identifies Seven Loci Associated with Dietary Approaches to Stop Hypertension (DASH) Diet

Olatz Mompeo <sup>1</sup>, Maxim B. Freidin <sup>1</sup>, Rachel Gibson <sup>2</sup>, Pirro G. Hysi <sup>1</sup>, Paraskevi Christofidou <sup>1</sup>, Eran Segal <sup>3</sup>, Ana M. Valdes <sup>1,4</sup>, Tim D. Spector <sup>1</sup>, Cristina Menni <sup>1</sup> and Massimo Mangino <sup>1,5,\*</sup>

<sup>1</sup> Department of Twin Research and Genetic Epidemiology, King's College London, London SE1 7EH, UK

<sup>2</sup> Department of Nutritional Sciences, King's College London, London SE1 9NH, UK

<sup>3</sup> Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 7610001, Israel

<sup>4</sup> Academic Rheumatology Clinical Sciences Building, Nottingham City Hospital, University of Nottingham, Nottingham NG5 1PB, UK

<sup>5</sup> NIHR Biomedical Research Centre at Guy's and St Thomas' Foundation Trust, London SE1 9RT, UK

\* Correspondence: massimo.mangino@kcl.ac.uk; Tel.: +49-(0)-207-188-86727



**Citation:** Mompeo, O.; Freidin, M.B.; Gibson, R.; Hysi, P.G.; Christofidou, P.; Segal, E.; Valdes, A.M.; Spector, T.D.; Menni, C.; Mangino, M. Genome-Wide Association Analysis of Over 170,000 Individuals from the UK Biobank Identifies Seven Loci Associated with Dietary Approaches to Stop Hypertension (DASH) Diet. *Nutrients* **2022**, *14*, 4431. <https://doi.org/10.3390/nu14204431>

Academic Editor: Clive J. Petry

Received: 27 September 2022

Accepted: 17 October 2022

Published: 21 October 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Diet is a modifiable risk factor for common chronic diseases and mental health disorders, and its effects are under partial genetic control. To estimate the impact of diet on individual health, most epidemiological and genetic studies have focused on individual aspects of dietary intake. However, analysing individual food groups in isolation does not capture the complexity of the whole diet pattern. Dietary indices enable a holistic estimation of diet and account for the intercorrelations between food and nutrients. In this study we performed the first ever genome-wide association study (GWA) including 173,701 individuals from the UK Biobank to identify genetic variants associated with the Dietary Approaches to Stop Hypertension (DASH) diet. DASH was calculated using the 24 h-recall questionnaire collected by UK Biobank. The GWA was performed using a linear mixed model implemented in BOLT-LMM. We identified seven independent single-nucleotide polymorphisms (SNPs) associated with DASH. Significant genetic correlations were observed between DASH and several educational traits with a significant enrichment for genes involved in the AMP-dependent protein kinase (AMPK) activation that controls the appetite by regulating the signalling in the hypothalamus. The colocalization analysis implicates genes involved in body mass index (BMI)/obesity and neuroticism (*ARPP21*, *RP11-62H7.2*, *MFHAS1*, *RHEBL1*). The Mendelian randomisation analysis suggested that increased DASH score, which reflect a healthy diet style, is causal of lower glucose, and insulin levels. These findings further our knowledge of the pathways underlying the relationship between diet and health outcomes. They may have significant implications for global public health and provide future dietary recommendations for the prevention of common chronic diseases.

**Keywords:** DASH; genetic; GWA

## 1. Introduction

In 2016, the World Health Organization (WHO) estimated that more than 1.9 billion adults were overweight or obese with alarming projections indicating that by 2030, nearly 60% of the worldwide population could be either overweight or obese [1]. Obesity is a common metabolic disease and a major risk factor for other common chronic diseases, including cardiovascular disease (CAD), type 2 diabetes (T2D), metabolic syndrome, and cancer [2,3]. Moreover, obesity has been linked to common mental disorders [4]. Since both chronic diseases and mental disorder present an enormous economic burden to society, understanding the relationships between nutrition, lifestyle, and individual health has become one of the highest priorities for public health organizations [5]. Indeed, diet

interventions combined with physical activity have been shown to prevent or mitigate the risk of developing common chronic diseases and mental health disorders [5–7].

To estimate the impact of diet on individual health, epidemiological and genetic studies have so far focused on different aspects of dietary intake such as macronutrient composition, curated measures of single food intake, multivariate dietary patterns described by principal component analysis (PCA) and food liking. Five genome-wide association studies (GWAs) have been conducted on macronutrient intake [8–11], two GWAs were performed on the PCA derived from the average consumption of defined diet components (i.e., meat, fish, vegetables) [12,13] and, more recently, a large scale GWA of food liking assessed food preferences over 139 specific foods [14]. Moreover, the Neale Lab (<http://www.nealelab.is/uk-biobank/> accessed on 1 April 2022) conducted GWAs of thousands single food intakes (i.e., wholemeal bread vs. all food intakes) in 361,000 unrelated individuals in UK Biobank (UKB).

However, analysing individual food groups in isolation presents some limitations since the complexity of the diet pattern as a whole is not considered [15]. Indeed, consumption of certain food groups (e.g., fruit often with vegetables, or fat with sugar) is often correlated [16] and the impacts of macronutrient composition on weight change have been particularly controversial [17].

Over recent decades, the research has shifted towards the analysis of dietary patterns (dietary indices). Dietary indices enable a holistic estimation of diet and account for the intercorrelations between nutrients or foods and for the possible synergic effects of nutrients [15]. We have recently reported that adherence to the Dietary Approaches to Stop Hypertension (DASH) diet decreased cardiovascular risk phenotypes [18] and we found that DASH is under strong genetic influence, with heritability estimates of 37% [19].

Following up on these findings [18,19], here we report the results of what, to our knowledge, is the first GWA that seeks to identify putative genetic determinants of DASH. For this purpose, we used available food questionnaire data and genotypes from 173,701 subjects of European ancestry participating in the UKB collection [20]. We applied *in silico* functional analyses on the identified loci to gain insights into the biological processes that potentially regulate dietary intake and conducted a Mendelian randomization analysis to provide evidence of the causal relationship between DASH and both cardiovascular risk factors and mental health disorders.

## 2. Materials and Methods

### 2.1. Study Population

We carried out a GWA analysis for DASH index that was calculated based on the UK Biobank 24 h dietary recall questionnaire [21]. In the UK Biobank, dietary intake was based on the average of a 24 h dietary recall of the previous day over 5 instances between 2009 and 2012 (Table 1) [21]. At the first instance, the questionnaire was completed on a touch screen. For the remaining 4 instances, an on-line questionnaire was sent to the participants. The average period between recalls was 6 months. Participants were questioned about whether they had eaten or drunk any of the approximately 300 commonly consumed foods and beverages in the previous 24 h, along with their amount and portion sizes.

A total of 203,581 individuals replied to one or more questionnaires. DASH diet scores were computed using the individual food intakes for each food group included in the DASH index [22] (Supplementary Materials Table S1). Scores range from 8 to 40 a higher score indicating closer adherence to a DASH dietary pattern. A total of 105,930 (60%) individuals replied to more than one questionnaire (Table 1).

For these subjects we computed the averages of their scores. We excluded from the analysis subjects who fell in each of the following category: (i) incomplete diet questionnaire, (ii) nutritional data not credible (UKB field 100,026) and (iii) low energy intake reported (energy intake  $< 1.1 \times$  resting metabolic rate (RMR), with the RMR calculated by Mifflin St Jeor equation [23]). After the exclusions, we included in the GWA analysis a total of 173,701 individuals with both questionnaire and genotype data available (Table 1).

**Table 1.** Baseline characteristics of the UK Biobank participants included in the GWA analysis.

	<b>Total Participants (n = 173,701)</b>
Age (years) (Mean (SD))	56.420 (7.864)
Gender	
Female	94,721 (54.5%)
Male	78,980 (45.5%)
BMI (kg/m <sup>2</sup> ) (Mean (SD))	26.462 (4.633)
Townsend Index (Mean (SD))	−1.519 (2.424)
Smoking	
Yes	98,378 (56.6%)
No	75,323 (43.4%)
Alcohol (g/day per week) (Mean (SD))	13.971 (21.439)
Instances of 24 h FFQ answered	
1	67,771 (39%)
2	39,797 (22.9%)
3	35,567 (20.5%)
4	25,713 (14.8%)
5	4853 (2.8%)
Energy intake (kcal) (Mean (SD))	2106.596 (595.421)
Fruit (Mean Servings (SD))	3.16 (2.52)
Vegetables (Mean Servings (SD))	3.49 (3.06)
Nuts and legumes (Mean Servings (SD))	0.9 (1.06)
Whole grains (Mean Servings (SD))	4.05 (3.12)
Low-fat dairy (Mean Servings (SD))	0.66 (0.84)
Sodium (mg) (Mean (SD))	2998.9 (3046.66)
Red and processed meat (Mean Servings (SD))	1.66 (1.75)
Sweetened beverages (Mean Servings (SD))	0.47 (0.89)
DASH (Mean (SD))	24 (4.24)

SD, standard deviation; BMI, body mass index; FFQ, food frequency questionnaire; DASH, Dietary Approaches to Stop Hypertension; GWA, genome-wide association study.

## 2.2. Genome Wide Association Analysis

UK Biobank is a large prospective cohort including genome wide genotyping, deep phenotyping and molecular data on over 500,000 individuals recruited throughout the UK between 2006–2010 [20]. Full details on genotyping, imputation and initial quality control of the genetic dataset have been described previously [24]. Briefly, UK Biobank participants were genotyped using two similar and mutually compatible SNP arrays platforms: the Affymetrix UK Biobank Axiom array and the UK BiLEVE array [24] and the imputation was performed combining Haplotype Reference Consortium (HRC) [25] and UK10K [26] reference panels. We also excluded from further analyses individuals based on (1) high SNP missingness (>2%); (2) extreme heterozygosity ( $\pm 3$  SD from mean heterozygosity rate); (3) withdrew their consent at the time of analysis; and/or (4) were not of European ancestry (flagged from the UKBB principal component analysis). Furthermore, we also excluded SNPs because they had: (1) minor allele frequency (MAF)  $\leq 5\%$  and/or (2) info score (imputation quality)  $\leq 0.8$ .

The genome wide association analysis was performed in BOLT-LMM [27] using a linear mixed model, in order to provide additional corrections for population structure and cryptic relatedness. DASH diet score was used as the outcome variable in the regression model, under the assumption of an additive model for allelic effects. In order to account for confounding effects, the regression model was adjusted for age, sex, BMI, energy intake, smoking status, age, alcohol intake as a categorical (0–5 g/d, 5–15 g/d, or >15 g/d) and the first five principal components. We also included in the association model the Townsend deprivation Index [28] (a composite measure of socioeconomic deprivation and household income) to account for the socio-economic status of the participants. Sex chromosomes were not included in the analysis.

### 2.3. Mapping and Conditional Analysis

We used the Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA) web-based application (<https://fuma.ctglab.nl/>) [29] to identify single-nucleotide polymorphisms (SNPs) associated with DASH at genome-wide significant  $p$ -value ( $p < 5 \times 10^{-8}$ ) that are in approximate linkage disequilibrium with each other at  $r^2 < 0.1$ . The independence of these signals was directly assessed using the imputed genotype dosage for the lead SNPs as a covariate along with the other covariates from the primary GWA using a linear mixed model in BOLT-LMM (conditional analysis). We defined a novel locus if all the variants and the genes mapping in the identified loci were not previously reported associated with DASH by querying NHGRI-EBI GWAS Catalog [30], PhenoScanner (v2.0) [31] and Open Target Genetics [32].

### 2.4. Pathway and Colocalization Analysis

We used Multi-marker Analysis of GenoMic Annotation (MAGMA) (Version 1.8), implemented in FUMA web-based application, applying standard settings to identify the most likely causal genes [33]. The statistical threshold for the most credible gene was defined at  $p < 2.86 \times 10^{-6}$  (0.05/17,460 analysed genes). To validate MAGMA results, we also conducted gene analysis and gene set analysis using VEGAS2 (version 0.2) [34] using the default options. In this case, the statistical threshold set at  $p < 2.55 \times 10^{-6}$  (0.05/19,640 analysed genes). For each GWA locus, we performed a colocalization analysis with the “coloc” R package (Version 3.2-1, R Core Team, Vienna, Austria) [35].

We performed the “coloc” analysis using the publicly available genome-wide expression quantitative trait locus (eQTL) data from 31,684 whole-blood samples deposited on the eQTLgen portal [36,37]. We included in the analysis all cis-eQTLs (false discovery rate (FDR) < 0.05) present in both DASH GWA and eQTLgen results and mapping 1 Mb across the lead SNP of each locus. Analyses were performed using the recommended defaults prior probabilities (PP) (PP for association in the GWA (P1):  $1 \times 10^{-4}$ ; PP for association in the eQTL (P2):  $1 \times 10^{-4}$ ; and PP for association in both dataset (P12):  $1 \times 10^{-5}$ ).

To investigate the SNP functional relevance on DASH, we applied the Summary-data-based Mendelian Randomization (SMR) (Version 1.03) approach [38], integrating the summary results from eQTLGen Consortium [36,37] and DASH GWA. SMR applies the principles of Mendelian randomization (MR) to test the association between gene eQTLs and a trait using the most associated SNP as a genetic instrument [38]. A significant SMR test indicates that a functional variant determines both gene expression and the trait of interest via causality or pleiotropy.

The heterogeneity in dependent instruments (HEIDI) test evaluates the existence of linkage disequilibrium (LD) in the observed association. Rejection of the null hypothesis ( $P_{\text{HEIDI}} < 0.05$ ) indicates that the association might be due to two distinct variants in high LD. We performed SMR using the recommended default options and utilised the genotypes from 3601 independent TwinsUK samples to estimate the LD structures. To account for multiple testing, SMR  $p$ -values were adjusted using the Benjamini and Hochberg method [39]. Association tests with  $p_{\text{SMR}_{\text{FDR}}} < 0.01$  were considered statistically significant, while  $p_{\text{SMR}_{\text{FDR}}} < 0.05$  and  $\geq 0.01$  were considered “suggestive”.

### 2.5. Shared Genetic Architecture with Disease

Genetic correlations ( $r_g$ ) between DASH and complex traits were estimated using linkage disequilibrium score regression (LDSC) through LD-hub (Version 1.9.3) software performed using the LD-Hub online portal (<http://ldsc.broadinstitute.org/>) [40] that automates the computation of  $r_g$  between one phenotypic trait of interest and 855 diseases or other phenotypic traits whose summary-level GWAs results are deposited in the database. To account for multiple testing, the significance for the LDSC analysis was set at  $p_{\text{LDSC}_{\text{FDR}}} < 0.01$ .

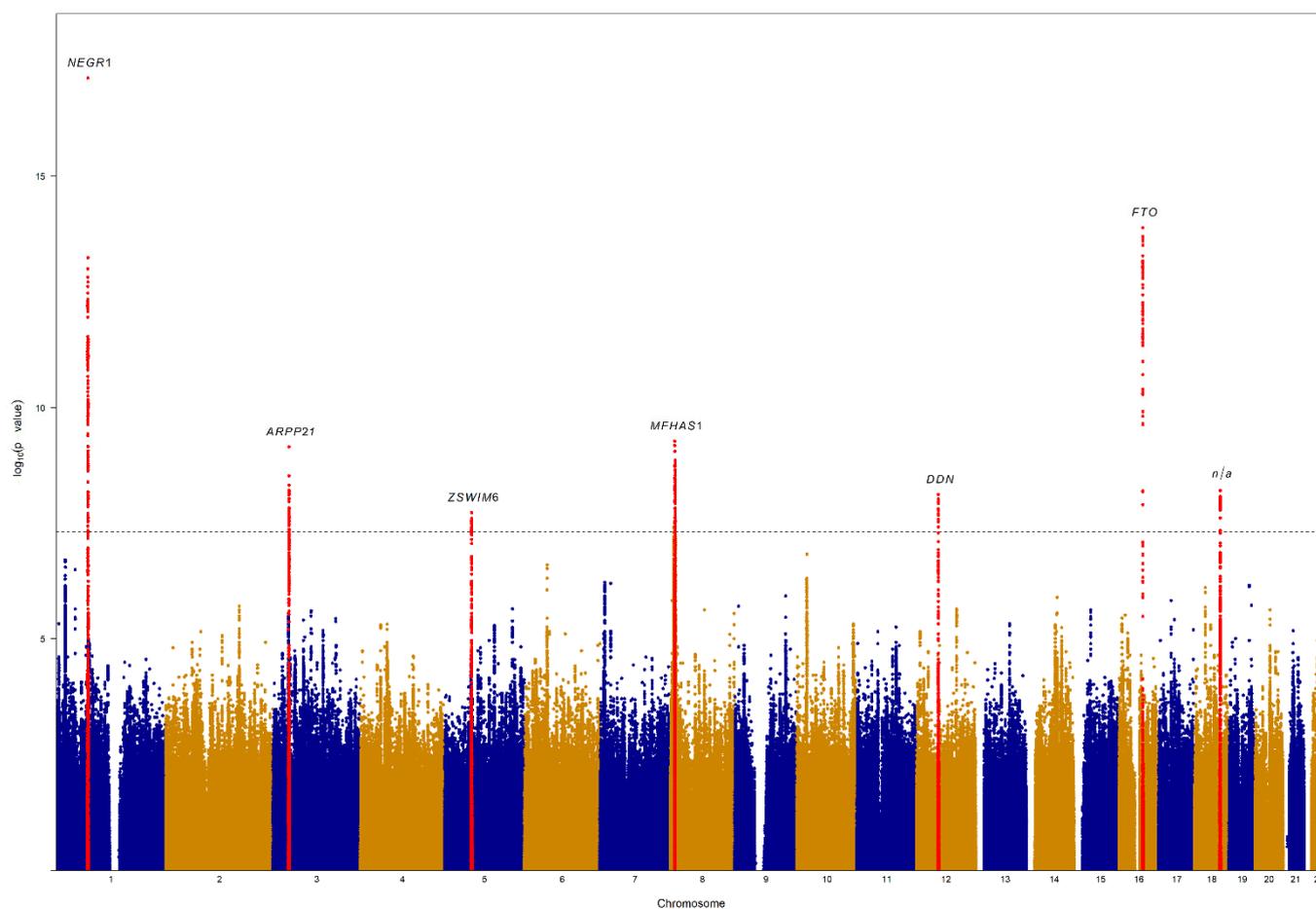
### 2.6. Mendelian Randomization

We performed a two-sample bidirectional Mendelian randomization analyses utilising the Generalised Summary-data-based Mendelian Randomization (GSMR) (Version 1.0.9) [41] implemented in the Genome-wide Complex Trait Analysis (GCTA) (Version 1.93.2) suite [42]. The GSMR estimate the effect and its standard error from multiple SNPs associated with the analysed traits at a genome-wide significance level. To perform the GSMR analysis we utilised the following summary statistics from genetic studies not overlapping UK Biobank: Body mass index (BMI) [43], high-density lipoprotein (HDL), low-density lipoprotein (LDL), triglycerides (TAG), total cholesterol (TC) [44], Glucose, Insulin [45], Homeostatic Model Assessment-Insulin Resistance (HOMA-IR) [46], coronary artery disease (CAD) [47] and Body Fat percentage [48]. In light of the results observed in the LDSC and the gene-enrichment analyses, we also included neuroticism [49] and educational attainment (years spent in formal education) [50] in the GSMR analysis. We included in the GSMR analysis data from the most recent meta-analyses (not including UK Biobank individuals). The summary statistics were downloaded either from the original consortia or from NHGRI-EBI GWAS Catalog [30] webpages and harmonized utilising the “snp\_match” command implemented in the bigsnpr (Version 1.6.1) [51] package in R (R Core Team, Vienna, Austria). We only used SNPs on autosomal chromosomes and available in the HRC reference panel, which allowed us to estimate the linkage disequilibrium among the instrument SNPs and prune them.

The HEIDI test implemented in GSMR was used to detect and remove variants showing independent effects on both exposure and outcome (i.e., horizontal pleiotropy), because they do not satisfy the assumptions for valid instruments. The HEIDI test is more conservative than excluding SNPs that have an outlying association likely driven by locus-specific pleiotropy. GSMR is more powerful than other MR methods (i.e. inverse-weighted MR (IVW-MR) and MR-Egger) because it takes account of the sampling variation of both the exposure and outcome effects [41]. GSMR also accounts for LD between the clumped SNPs. We used the genome-wide significant p-value threshold ( $p < 1 \times 10^{-8}$ ) to select a minimum number of instrument SNPs ( $n > 5$ ) to perform the GSMR analysis. Genotypes of unrelated TwinsUK cohort participants were used as reference to estimate the LD structures. To further validate the GSMR results we also conducted IVW-MR and MR-Egger analyses [52] using the TwoSampleMR (Version 0.5.5) [53] package in R (R Core Team, Vienna, Austria).

## 3. Results

We accessed the 24 h dietary recall records for 203,581 individuals from UKBiobank (Supplementary Materials Figure S1) and, after the quality controls exclusions, we calculated a DASH score for 173,701 subjects with genotype available. Full characteristics of the study population are reported in Table 1. The participants included in the final analysis had an average age of  $56.4 \pm 7.9$ , were overweight (BMI mean =  $26.5 \text{ kg/m}^2 \pm 4.6 \text{ kg/m}^2$ ), had average daily energy intake of  $2106.6 \text{ kcal} \pm 595.4 \text{ kcal}$  and a mean DASH score of  $24 \pm 4.2$ . Half of the samples were females (54.5%) and smokers (56.6%) (Table 1). Some of the values for the food consumption reported in Table 1 reflect the typical skewed distribution observed when collecting food group data. However, the skewed distribution did not affect the DASH score because, using a point score based on the quintiles of the distribution of each food component, it has been specifically designed to address this issue [22]. We tested over 12 million autosomal SNPs for association with DASH. A genomic inflation factor [54] of  $\lambda_{\text{DASH}} = 1.2$  and the linkage disequilibrium score regression (LDSC) intercept of  $0.997 \pm 0.0099$  is consistent with the expectations of polygenicity and large sample sizes [55] and indicate adequate population structure control. We observed 641 genome-wide significant associations ( $p < 5 \times 10^{-8}$ ) (Figure 1), clustered within 7 distinct genomic regions (Table 2 and Supplementary Materials Figure S2).



**Figure 1.** Manhattan plot of the DASH genome wide association study in UK Biobank. The x axis represents the position on each chromosome (represented with alternate colours), while the y axis represents the  $-\log_{10}(P)$  of SNPs. The dotted line indicates the genome-wide significance threshold ( $p < 5 \times 10^{-8}$ ). Only SNPs with  $p < 0.1$  are represented in the figure. Independent genome-wide significant variants (in red) are annotated with the genes associated with DASH based on both MAGMA and VEGAS2 analyses. No genes were associated with DASH on chromosome 18q21.32. DASH, Dietary Approaches to Stop Hypertension; SNPs, single-nucleotide polymorphisms; MAGMA, Multi-marker Analysis of GenoMic Annotation; n/a, no genes present.

To investigate the presence of multiple independent sources of associations within the respective associated regions, we performed conditional analyses on each of the seven regions, adjusting for the effect of the lead SNPs that were included in models as covariates. We did not detect any additional independent SNP with either significant ( $p < 5 \times 10^{-8}$ ) or suggestive ( $p < 1 \times 10^{-7}$ )  $p$ -value.

Most associated loci point towards a shared genetic background between DASH and behaviour, food related or metabolic traits. In particular, the strongest association was detected within a 445 Kbp region on chromosome 1p31.1 (rs66495454,  $p = 7.6 \times 10^{-18}$ ) (Supplementary Materials Figure S2A). The 1p31.1 locus harbours the neuronal growth regulator 1 (*NEGR1*) gene, which has previously been associated with psychiatric [56], behavioural [57], nutritional [13] and metabolic disorders [43]. We also identified one locus on chromosome 16q12.2 (rs56094641,  $p = 1.3 \times 10^{-14}$ ) harbouring *FTO* (Supplementary Materials Figure S1F), best known genes influencing both nutrition and obesity [44]. Another variant (rs56331918,  $p = 6.9 \times 10^{-10}$ ) mapping on chromosome 3p22.3 in the intronic region of CAMP Regulated Phosphoprotein 21 (*ARPP21*) gene (Supplementary Materials Figure S2B) which has been associated with neuroticism [58] and BMI [59]. Our results showed a significant association between DASH and a ~3 Mb region (Chr8: 8,088,230–11,463,015;

rs73195303,  $p = 5.3 \times 10^{-10}$ ) on chromosome 8p23.1 (Supplementary Materials Figure S1D). Chromosome 8p23.1 locus comprises numerous genes performing functions important to the nervous system and associated with cancer and developmental neuropsychiatric disorders [60]. Finally, three additional loci were identified on chromosome 5q12.1 (rs544711163,  $p = 1.9 \times 10^{-8}$ ), 12q13.12 (rs1054442,  $p = 7.7 \times 10^{-9}$ ) and 18q21.32 (rs35614134,  $p = 6.3 \times 10^{-9}$ ), (Supplementary Materials Figures S2C, S2E, S2F and S2G, respectively).

Using MAGMA [33], we performed a gene-based analysis including the complete GWA results. This analysis identified nineteen genes associated ( $p < 2.66 \times 10^{-6}$ ) with DASH. While most of these genes (15 out of 19) mapped within the identified loci, MAGMA analysis also found four genes which are physically distant (>250 kb) from the lead SNPs. Supplementary Materials Table S2 lists the fifteen most likely causal genes (reported by MAGMA analysis) in or near the lead SNP at each locus. Similar results were obtained when the analysis was performed using VEGAS2 (Supplementary Materials Table S3).

In order to annotate these genes in a biological context, we used the function GENE2FUNC included in FUMA online platform [29]. We observed that our GWA results were enriched for genes participating in the Activation of AMPK downstream of N-methyl D-aspartate receptors (NMDARs) pathway, even after Bonferroni multiple testing correction ( $p_{\text{adjusted}} = 2.8 \times 10^{-2}$ ) (Supplementary Materials Table S4). We also find significant enrichment for genes participating in another 40 gene-sets selected from the GWA catalogue ( $p_{\text{adjusted}}$  ranging from  $2.7 \times 10^{-17}$  to  $4.4 \times 10^{-2}$ ), in particular, the “general factor of neuroticism” which was one of the most significantly enriched entries ( $p_{\text{adjusted}} = 2.7 \times 10^{-17}$ ) (Supplementary Materials Table S4 and Figure S3).

Next, we proceeded with a functional characterisation of genomic variants within or near genes harboured by the identified loci. We performed a Bayesian test [35] to examine whether GWA loci co-localize with the gene eQTLs in blood. To this end, we utilised publicly available expression data from the large eQTLgen consortium meta-analysis [37]. We identified 111 genes mapping in the 8 identified loci. Seventy-two genes were excluded from the analysis because they were not available in the eQTLgen dataset (Supplementary Materials Table S5). Our COLOC-estimated posterior probabilities (PP) [35] suggested that two eQTL effects (*ARPP21* and *RHEBL1*) were probably sharing the same common causal variant ( $PP_{H4}$  ranging from 1 to 0.80) with the associated loci (Supplementary Materials Table S5). Twenty-six eQTLs overlapped with their corresponding locus, without necessarily sharing the same causal variant ( $PP_{H3}$  ranging from 1 to 0.80). The COLOC-PP for *NEGR1* ( $PP_{H4} = 0.68$ ), *PRKAG1* ( $PP_{H4} = 0.65$ ) and *FTO* ( $PP_{H4} = 0.74$ ) eQTLs showed a suggestive probability for a causal variant shared with chromosome 1p31.1, 8p23.1 and 16q12.2 loci, respectively. Finally, the analysis of eight eQTL transcripts showed no colocalization ( $PP_{H0}$  or  $PP_{H1}$  ranging from 1 to 0.80) or failed to support any tested hypotheses (All PPs < 0.80) (Supplementary Materials Table S5).

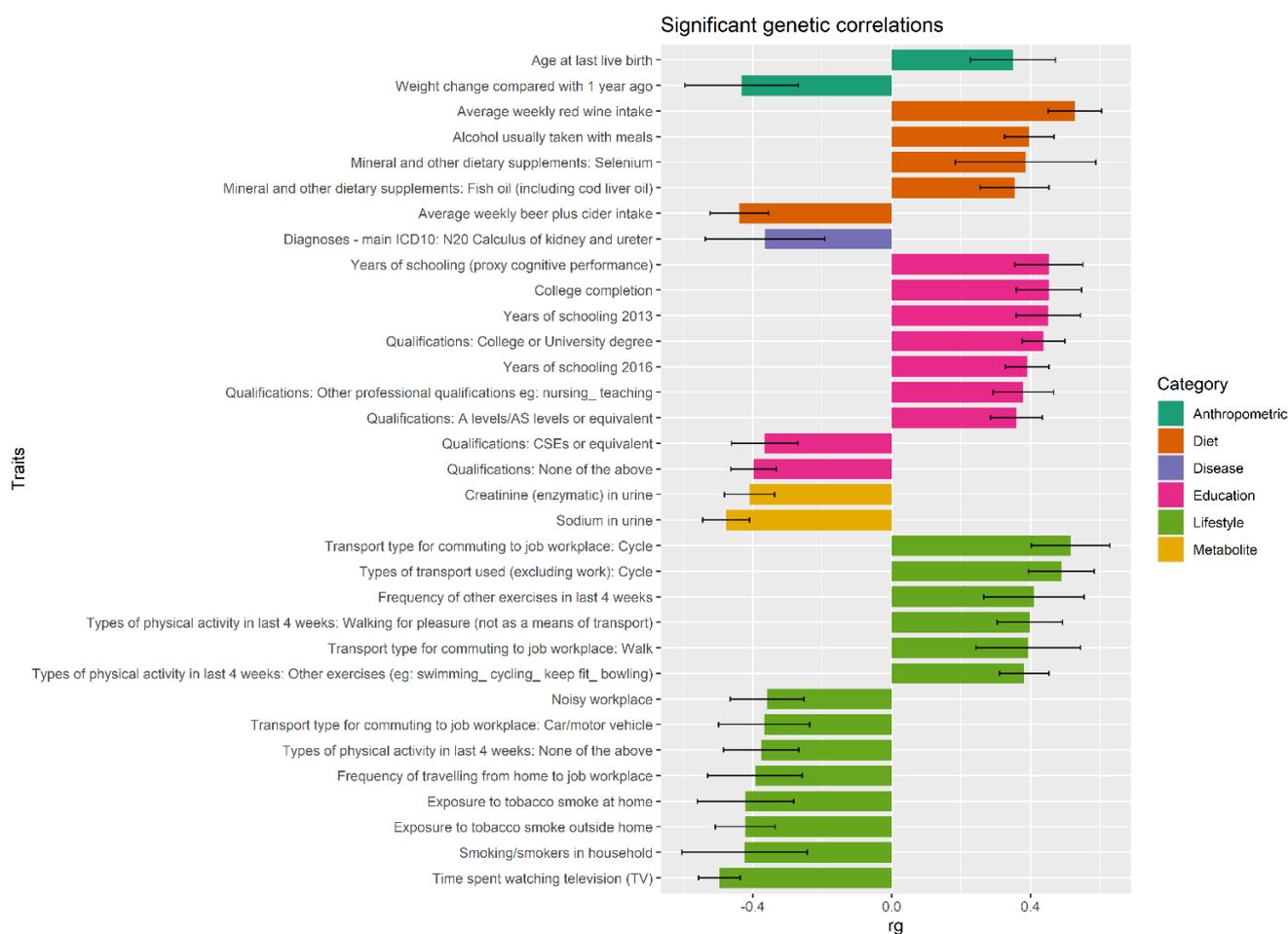
To further evaluate whether any of the cis-eQTLs mediated the association between genetic variants and DASH, we applied SMR [38] on the GWA results. Nine genes did not have any significant eQTL SNP ( $p < 5 \times 10^{-8}$ ) to be utilised as a genetic instrument and therefore were excluded from the analysis (Supplementary Materials Table S6). We observed statistically significant associations ( $p_{\text{SMR}_{\text{FDR}}} < 0.01$ ) for fifteen genes ( $p_{\text{SMR}_{\text{FDR}}}$  ranging from  $1.03 \times 10^{-20}$  to  $2.06 \times 10^{-3}$ ) with eight genes showing a suggestive SMR ( $0.05 < p_{\text{SMR}_{\text{FDR}}} \leq 1 \times 10^{-2}$ ) (Table S6). The SMR test was not significant for seven genes (Supplementary Materials Table S6). We obtained similar results when using a SMR multi-SNP approach ( $p_{\text{SMR}_{\text{FDR}}}$  ranging from  $3.15 \times 10^{-19}$  to  $3.34 \times 10^{-3}$ ) (Supplementary Materials Table S6). Next, we performed the HEIDI test to distinguish pleiotropy/causality from linkage. The HEIDI results suggested that four genes (*ARPP21*, *RP11-62H7.2*, *MFHAS1*, *RHEBL1*) were consistent with either pleiotropy or causality (Supplementary Materials Table S6) while for eleven genes it was not possible to distinguish between pleiotropy/causality and linkage disequilibrium (HEIDI  $p$  ranging from  $6.8 \times 10^{-13}$  to  $4.4 \times 10^{-2}$ ).

**Table 2.** GWA summary results. Seven independent genomic regions associated with DASH at genome-wide significance ( $p$  Value  $< 5 \times 10^{-8}$ ).

Locus	Chromosome	Locus Starts	Locus End	Top SNP	Top SNP Position	Effect Allele	Non-Effect Allele	Effect Allele Frequency	Beta	Standard Error	$p$ Value
1	1	72,511,514	72,956,535	rs66495454	72,748,567	GTCCT	G	0.38	0.126	0.01	$7.60 \times 10^{-18}$
2	3	35,778,773	35,913,342	rs56331918	35,801,168	G	C	0.28	-0.1	0.02	$6.90 \times 10^{-10}$
3	5	60,613,826	60,844,213	rs544711163	60,775,743	CT	C	0.38	-0.082	0.01	$1.90 \times 10^{-8}$
4	8	8,088,230	11,463,015	rs73195303	10,200,253	T	C	0.23	-0.105	0.02	$5.30 \times 10^{-10}$
5	12	49,385,679	49,479,968	rs1054442	49,389,320	C	A	0.37	0.085	0.01	$7.70 \times 10^{-9}$
6	16	53,797,908	53,845,487	rs56094641	53,806,453	G	A	0.40	0.111	0.01	$1.30 \times 10^{-14}$
7	18	57,732,418	57,912,226	rs35614134	57,832,856	AC	A	0.24	0.097	0.02	$6.30 \times 10^{-9}$

SNPs, single-nucleotide polymorphisms.

To compute the amount of shared genetic correlation ( $r_g$ ) between DASH and other complex traits, we performed LD score regression (LDSC) analyses on 855 other phenotypic traits or diseases [40] and observed significant genetic correlations ( $pLDSC_{FDR} < 0.01$ ) with 193 of them (Supplementary Materials Table S7). Among traits with strong positive genetic correlations with DASH the most significant was educational attainment (“Qualifications: College or University degree”,  $r_g = 0.44$ ,  $pLDSC_{FDR} = 4.7 \times 10^{-41}$ ). We also found significant positive genetic correlations with other measures of educational attainment (Supplementary Materials Table S7 and Figure 2). “Average weekly red wine intake” was the trait with the strongest genetic correlation ( $r_g = 0.53$ ,  $pLDSC_{FDR} = 2.6 \times 10^{-39}$ ). Among the strongest negative correlations ( $r_g < -0.35$ ), “Time spent watching television (TV)” was the trait showing the strongest and most significant genetic correlation with DASH ( $r_g = -0.50$ ,  $pLDSC_{FDR} = 2.3 \times 10^{-56}$ ) (Supplementary Materials Table S7 and Figure 2)



**Figure 2.** Genetic correlations ( $r_g$ ). Pairwise genome-wide genetic correlations between DASH and 855 other phenotypic traits or diseases were estimated using LD score regression (LDSC). In the figure are represented only the strongest genetic correlations ( $r_g < -0.35$  and  $r_g > 0.35$ ) (see ESI Supplementary Materials Table S7 for the report of the full results). Error bars show 95% confidence intervals, while colours represent the different categories. LD, linkage disequilibrium; ICD10, International Statistical Classification of Diseases and Related Health Problems 10th Revision; A, Advanced; AS, Advanced Subsidiary; CSE, Certificate of Secondary Education.

We previously described the association between DASH and cardiometabolic traits [18] and next sought to explore the evidence of causality versus pleiotropy between DASH and these traits using bidirectional a GSMR [41]. In light of the newly observed LDSC and gene-enrichment evidence implicating neuroticism and educational attainment, both phenotypes were also included in the GSMR analyses.

Bidirectional GSMR strongly suggested that high DASH scores causally lower insulin ( $\text{Beta}_{\text{DASH} \rightarrow \text{Insulin}}$  (standard error (SE)) =  $-0.041$  (0.01);  $p_{\text{GSMR}} = 3.86 \times 10^{-4}$ ) and glucose levels ( $\text{Beta}_{\text{DASH} \rightarrow \text{Glucose}}$  (SE) =  $-0.036$  (0.01);  $p_{\text{GSMR}} = 8.82 \times 10^{-3}$ ) (Supplementary Materials Table S8 and Figure S4A,B). Our results also provided evidence of a causality of high DASH score on increased educational attainment ( $\text{Beta}_{\text{DASH} \rightarrow \text{Educational Attainment}}$  (standard error (SE)) =  $0.101$  (0.02);  $p_{\text{GSMR}} = 2.15 \times 10^{-8}$ ) (Supplementary Materials Table S8 and Figure S4C). We did not find evidence for a causal role for insulin ( $p_{\text{GSMR}} = 7.82 \times 10^{-2}$ ) and glucose ( $p_{\text{GSMR}} = 0.821$ ) levels on DASH score (Supplementary Materials Table S8). We could not test reverse causality of educational attainment on DASH because there were not enough independent instrument variants ( $n < 5$ ) to perform the GSMR analyses. For CAD, neuroticism score, body fat percentage, HOMA-IR, triglycerides, LDL, HDL and total cholesterol levels the GSMR analyses were not significant when considering DASH as exposure ( $p_{\text{GSMR}} = 0.123$ ,  $p_{\text{GSMR}} = 0.451$ ,  $p_{\text{GSMR}} = 0.634$ ,  $p_{\text{GSMR}} = 0.127$ ,  $p_{\text{GSMR}} = 0.854$ ,  $p_{\text{GSMR}} = 0.397$ ,  $p_{\text{GSMR}} = 0.517$  and  $p_{\text{GSMR}} = 0.059$ , respectively) (Supplementary Materials Table S8). We also observed a causal relationship between lower HDL levels ( $\text{Beta}_{\text{HDL} \rightarrow \text{DASH}}$  (SE) =  $-0.066$  (0.03)  $p_{\text{GSMR}} = 2.32 \times 10^{-2}$ ), high LDL levels ( $\text{Beta}_{\text{LDL} \rightarrow \text{DASH}}$  (SE) =  $0.112$  (0.03)  $p_{\text{GSMR}} = 1.75 \times 10^{-5}$ ), increased levels of total cholesterol ( $\text{Beta}_{\text{TC} \rightarrow \text{DASH}}$  (SE) =  $0.071$  (0.03)  $p_{\text{GSMR}} = 8.65 \times 10^{-3}$ ), increased CAD risk ( $\text{Beta}_{\text{CAD} \rightarrow \text{DASH}}$  (SE) =  $0.081$  (0.03);  $p_{\text{GSMR}} = 2.57 \times 10^{-3}$ ) and increased body fat percentage ( $\text{Beta}_{\text{BODY FAT\%} \rightarrow \text{DASH}}$  (SE) =  $0.426$  (0.15);  $p_{\text{GSMR}} = 4.83 \times 10^{-3}$ ) when considering DASH as outcome (Supplementary Materials Table S8 and Figure S4D–G). We could not test reverse causality of neuroticism and HOMA-IR on DASH because there were not enough independent instrument variants ( $n < 5$ ) to perform the GSMR analyses. Finally, for BMI, we found significant bidirectional effect with DASH (Supplementary Materials Table S8). This result may reflect the presence of shared biological pathways (vertical pleiotropy). The GSMR results are reported in full in Supplementary Materials Table S8. We obtained qualitatively similar results with other MR methods implemented in the two-sample MR R library (Supplementary Materials Table S8). As different MR methods rely on different assumptions and models of horizontal pleiotropy, the consistency of the results across different methods builds confidence in the obtained estimates.

#### 4. Discussion

In this first GWA, we investigated the genetic influences on DASH on over 170,000 subjects of European ancestry and identified seven associated loci, which provide new insights into the genetic basis of this dietary pattern. By leveraging these genetic findings, we performed Mendelian randomization analyses to assess the causal relationship between DASH and health outcomes. Our results indicate that a healthy diet style may causally lead to reduced levels of glucose and insulin. The gene-based analysis revealed nineteen genes associated with DASH. We identified four genes (*ARPP21*, *RP11-62H7.2*, *MFHAS1* and *RHEBL1*) whose expressions were potentially associated with DASH due to causality or pleiotropy. Interestingly, most of these genes have been consistently associated with cardiometabolic diseases [59,61], educational attainment [50], cognitive abilities [62], neuroticism [58] and major depressive disorder [60]. These findings support the hypothesis that genes influencing dietary choice may also influence the liability to psychiatric and cardiometabolic disorder [13].

A recent UK Biobank study defined two independent diet component (DC) intakes based on the principal component (PC) of UK Biobank generic diet questionnaire and identified a number of genetic loci associated with either DC1 (a meat-related diet) or DC2 (a fish/plant-related diet) [13]. Moreover, utilising the same UK Biobank generic diet questionnaire, May-Wilson et al. identified the genetic determinants of food liking [14]. Two markers reported in our study overlap with the variants associated with either DC1 (rs66495454 on chromosome 1p31.1) or DC2 (rs56094641 on chromosome 16q12.2). Specifically, the rs66495454 A allele that is associated with a lower DASH score, indicating a

propensity to lower diet quality also increased processed meat intake [13] as well as red meat and beef steak liking [14] (all negative components of the DASH diet).

Similarly, the variant (rs56094641) on chromosome 16q12.2 is associated with lower DASH score as well as lower non-oily fish intake/liking [13,14]. Similar results were also reported by Cole et al.'s analyses [12] of measures of single food intake (FI) in UK Biobank (1p31.1 and 16q12.2, harbouring *NEGR1* and *FTO* genes, respectively). These two genes have been consistently associated with BMI [43,63] obesity and cardiometabolic diseases [43]. Although we were not able to distinguish between causal effect and linkage disequilibrium, the SMR analysis on chromosome 1p31.1 locus showed that decreased *NEGR1* expression levels are associated with lower DASH score. Decreased expression level of *Negr1* in murine periventricular hypothalamic areas lead to an increase in body weight [64]. Altogether, these results are consistent with previous observation that increased red/processed meat consumption is mostly responsible for the association between DASH and increased cardiometabolic disease risk [18] and may indicate a very complex genetic relationship between DASH and obesity.

Gene-set enrichment analyses provide evidence that the genes annotated to the variants associated with DASH participate in the "Activation of AMPK downstream of NMDARs" pathway. The AMP-dependent protein kinase (AMPK) is highly expressed in the hippocampus [65] and is activated when AMP and ADP levels in the cells rise due to a variety of physiological stresses, such as increased ghrelin levels, glucose deprivation and exercise [66]. AMPK is one of the signalling components of the Neuropeptide Y (NPY) network, which is the master regulator of the appetite signal in the arcuate nucleus-paraventricular nucleus (ARC-PVN) of the hypothalamus [67]. Intracerebroventricular administration of a pharmacological AMPK activator (AICAR) in murine experiments stimulates food intake and weight gain [68].

Our study should be interpreted in the context of the following limitations. First, our SMR analyses is based on cis-eQTL effects estimated in peripheral blood because the currently available brain eQTL studies have very limited statistical power due to their small sample sizes. However, Ting Qi et al. demonstrated that, when the genes are expressed in both brain and blood, then using cis-eQTL effects estimated in blood as proxies of those in brain increase the power to identify putative functional genes for brain-related complex traits and diseases [69]. Additionally, similar to other UK Biobank dietary studies [12–14], we calculated the DASH score using the self-reported questionnaire data. Single 24-hr recalls are unlikely to capture episodic food of some items included in the DASH score (nuts, legumes). However, more than 60% of the participants included in this study answered to two or more questionnaire. Therefore, by averaging multiple recalls from participants, it is more likely that they represent the individual habitual intake. Finally, large biobanks are well powered to discover common variant associations. However, replicating their findings is one of the main issue that has been recently discussed [70]. Indeed, while using data derived from the same biobank, studies analysing similar traits (same phenotype but different modelling and phenotype definition) have solved this problem adopting different solutions [71,72]. As suggested by Huffman [70] we performed a range of secondary analyses (functional annotation, pathway analysis, eQTL and colocalization) utilizing publicly available datasets. We presented a number of orthogonal biological evidences which may be considered in the same vein as statistical validation [70]. Using different phenotype definition and analysis model, some of the findings described in this study overlap with previous observations [12–14]. The consistency of the results across different studies builds confidence in our findings and may represent a form of validation [70]. However, although our secondary analyses pointed towards plausible genes/pathways and some of the loci identified in this study overlap with previous observations in UK Biobank, our findings need to be further tested in functional and interventional studies in animal models and humans to fully determine the biological mechanisms underlying DASH.

## 5. Conclusions

In conclusion, this study provides novel insights into the genetic architecture of DASH and highlighted its putative causal relationship with health outcomes. These findings extend our knowledge of the genetic pathways underlying DASH and may have significant implications for global public health providing future dietary recommendations for the prevention of common chronic diseases and mental health disorders.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/nu14204431/s1>, Figure S1: Study design and analysis workflow; Figure S2: Regional associations plots; Figure S3: Gene enrichment analysis results; Figure S4: Mendelian randomization analyses (GSMR); Table S1: Food groups included in the computation of DASH score; Table S2: Genes significant in MAGMA gene-wise analysis; Table S3: Genes significant in Vegas2 gene-wise analysis; Table S4: Gene-set Enrichment analysis (GENE2FUNC); Table S5: Co-localization analysis; Table S6: SMR results; Table S7: Linkage Disequilibrium Score Correlation analysis; Table S8: Mendelian randomization results.

**Author Contributions:** M.M. and C.M.; methodology, M.M., M.B.F., R.G. and P.G.H.; software, M.M., M.B.F., P.G.H. and P.C.; formal analysis, O.M. and M.M.; resources, E.S., A.M.V. and T.D.S.; writing—original draft preparation, M.M., O.M. and C.M.; funding acquisition, T.D.S. and E.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** The research has been conducted using the UK Biobank Resource (project # 28784). UK Biobank was established by the Wellcome Trust medical charity, Medical Research Council, Department of Health, Scottish Government and Northwest Regional Development Agency. It also had funding from the Welsh Assembly Government, British Heart Foundation and Diabetes UK. The authors are grateful to the UK Biobank participants for making such research possible. OM and PL are supported by Chronic Disease Research Foundation (CDRF). PC is supported by the European Union (H2020 contract #733100). AMV is supported by the National Institute for Health Research Nottingham Biomedical Research Centre. CM is funded by the Chronic Disease Research Foundation and by the Medical Research Council (MRC)/British Heart Foundation Ancestry and Biological Informative Markers for Stratification of Hypertension (AIMHY; MR/M016560/1). MM is supported by the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London.

**Institutional Review Board Statement:** This study was approved by the ethics committees of the UK Biobank consortium providers (UKBB-28784 approved on 13 May 2016).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The UKBB phenotypic data and GWA data analyzed during the current study are available from UK Biobank.

**Acknowledgments:** We thank the UK Biobank (UKBB) for their assistance in preparing and access to the data.

**Conflicts of Interest:** T.D.S. is co-founder and stakeholder of Zoe Global Ltd. A.M.V. is a consultant for Zoe Global Ltd. All other authors declare no conflict of interest.

## References

1. Abarca-Gómez, L.; Abdeen, Z.A.; Hamid, Z.A.; Abu-Rmeileh, N.M.; Acosta-Cazares, B.; Acuin, C.; Adams, R.J.; Aekplakorn, W.; Afsana, K.; Aguilar-Salinas, C.A.; et al. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: A pooled analysis of 2416 population-based measurement studies in 128.9 million children, adolescents, and adults. *Lancet* **2017**, *390*, 2627–2642. [[CrossRef](#)]
2. Kompaniyets, L.; Goodman, A.B.; Belay, B.; Freedman, D.S.; Sucusky, M.S.; Lange, S.J.; Gundlapalli, A.V.; Boehmer, T.K.; Blanck, H.M. Body Mass Index and Risk for COVID-19-Related Hospitalization, Intensive Care Unit Admission, Invasive Mechanical Ventilation, and Death—United States, March–December 2020. *MMWR Morb. Mortal. Wkly. Rep.* **2021**, *70*, 355–361. [[CrossRef](#)]
3. Roberts, C.K.; Barnard, R.J. Effects of exercise and diet on chronic disease. *J. Appl. Physiol.* **2005**, *98*, 3–30. [[CrossRef](#)]
4. Kivimäki, M.; Lawlor, D.A.; Singh-Manoux, A.; Batty, G.D.; Ferrie, J.E.; Shipley, M.J.; Nabi, H.; Sabia, S.; Marmot, M.G.; Jokela, M. Common mental disorder and obesity: Insight from four repeat measures over 19 years: Prospective Whitehall II cohort study. *BMJ* **2009**, *339*, b3765. [[CrossRef](#)] [[PubMed](#)]

5. Mozaffarian, D.; Rosenberg, I.; Uauy, R. History of modern nutrition science-implications for current research, dietary guidelines, and food policy. *BMJ* **2018**, *361*, k2392. [[CrossRef](#)] [[PubMed](#)]
6. Gómez-Pinilla, F. Brain foods: The effects of nutrients on brain function. *Nat. Rev. Neurosci.* **2008**, *9*, 568–578. [[CrossRef](#)]
7. Deane, K.H.O.; Jimoh, O.F.; Biswas, P.; O'Brien, A.; Hanson, S.; Abdelhamid, A.S.; Fox, C.; Hooper, L. Omega-3 and polyunsaturated fat for prevention of depression and anxiety symptoms: Systematic review and meta-analysis of randomised trials. *Br. J. Psychiatry J. Ment. Sci.* **2021**, *218*, 135–142. [[CrossRef](#)]
8. Chu, A.Y.; Workalemahu, T.; Paynter, N.P.; Rose, L.M.; Giulianini, F.; Tanaka, T.; Ngwa, J.S.; Qi, Q.; Curhan, G.C.; Rimm, E.B.; et al. Novel locus including FGF21 is associated with dietary macronutrient intake. *Hum. Mol. Genet.* **2013**, *22*, 1895–1902. [[CrossRef](#)]
9. Meddens, S.F.W.; de Vlaming, R.; Bowers, P.; Burik, C.A.P.; Linnér, R.K.; Lee, C.; Okbay, A.; Turley, P.; Rietveld, C.A.; Fontana, M.A.; et al. Genomic analysis of diet composition finds novel loci and associations with health and lifestyle. *Mol. Psychiatry* **2020**, *26*, 2056–2069. [[CrossRef](#)]
10. Merino, J.; Dashti, H.S.; Li, S.X.; Sarnowski, C.; Justice, A.E.; Graff, M.; Papoutsakis, C.; Smith, C.E.; Dedoussis, G.V.; Lemaitre, R.N.; et al. 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. *Mol. Psychiatry* **2019**, *24*, 1920–1932. [[CrossRef](#)]
11. Tanaka, T.; Ngwa, J.S.; van Rooij, F.J.; Zillikens, M.C.; Wojczynski, M.K.; Frazier-Wood, A.C.; Houston, D.K.; Kanoni, S.; Lemaitre, R.N.; Luan, J.; et al. Genome-wide meta-analysis of observational studies shows common genetic variants associated with macronutrient intake. *Am. J. Clin. Nutr.* **2013**, *97*, 1395–1402. [[CrossRef](#)] [[PubMed](#)]
12. Cole, J.B.; Florez, J.C.; Hirschhorn, J.N. Comprehensive genomic analysis of dietary habits in UK Biobank identifies hundreds of genetic associations. *Nat. Commun.* **2020**, *11*, 1467. [[CrossRef](#)] [[PubMed](#)]
13. Niarchou, M.; Byrne, E.M.; Trzaskowski, M.; Sidorenko, J.; Kemper, K.E.; McGrath, J.J.; MC, O.D.; Owen, M.J.; Wray, N.R. Genome-wide association study of dietary intake in the UK biobank study and its associations with schizophrenia and other traits. *Transl. Psychiatry* **2020**, *10*, 51. [[CrossRef](#)] [[PubMed](#)]
14. May-Wilson, S.; Matoba, N.; Wade, K.H.; Hottenga, J.-J.; Concas, M.P.; Mangino, M.; Grzeszkowiak, E.J.; Menni, C.; Gasparini, P.; Timpson, N.J.; et al. Large-scale GWAS of food liking reveals genetic determinants and genetic correlations with distinct neurophysiological traits. *Nat. Commun.* **2022**, *13*, 2743. [[CrossRef](#)] [[PubMed](#)]
15. Hu, F.B. Dietary pattern analysis: A new direction in nutritional epidemiology. *Curr. Opin. Lipidol.* **2002**, *13*, 3–9. [[CrossRef](#)] [[PubMed](#)]
16. Waijers, P.M.; Feskens, E.J.; Ocké, M.C. A critical review of predefined diet quality scores. *Br. J. Nutr.* **2007**, *97*, 219–231. [[CrossRef](#)]
17. Vergnaud, A.-C.; Norat, T.; Mouw, T.; Romaguera, D.; May, A.M.; Bueno-de-Mesquita, H.B.; van der A, D.; Agudo, A.; Wareham, N.; Khaw, K.-T.; et al. Macronutrient composition of the diet and prospective weight change in participants of the EPIC-PANACEA study. *PLoS ONE* **2013**, *8*, e57300. [[CrossRef](#)]
18. Mompeo, O.; Berry, S.E.; Spector, T.D.; Menni, C.; Mangino, M.; Gibson, R. Differential associations between a priori diet quality scores and markers of cardiovascular health in women: Cross-sectional analyses from TwinsUK. *Br. J. Nutr.* **2020**, *126*, 1017–1027. [[CrossRef](#)]
19. Mompeo, O.; Gibson, R.; Christofidou, P.; Spector, T.D.; Menni, C.; Mangino, M. Genetic and Environmental Influences of Dietary Indices in a UK Female Twin Cohort. *Twin Res. Hum. Genet. Off. J. Int. Soc. Twin Stud.* **2020**, *23*, 330–337. [[CrossRef](#)]
20. Sudlow, C.; Gallacher, J.; Allen, N.; Beral, V.; Burton, P.; Danesh, J.; Downey, P.; Elliott, P.; Green, J.; Landray, M.; et al. UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **2015**, *12*, e1001779. [[CrossRef](#)]
21. Allen, N.E.; Sudlow, C.; Peakman, T.; Collins, R. UK biobank data: Come and get it. *Sci. Transl. Med.* **2014**, *6*, 224ed224. [[CrossRef](#)] [[PubMed](#)]
22. Fung, T.T.; Chiuve, S.E.; McCullough, M.L.; Rexrode, K.M.; Logroscino, G.; Hu, F.B. Adherence to a DASH-style diet and risk of coronary heart disease and stroke in women. *Arch. Intern. Med.* **2008**, *168*, 713–720. [[CrossRef](#)] [[PubMed](#)]
23. Mifflin, M.D.; St Jeor, S.T.; Hill, L.A.; Scott, B.J.; Daugherty, S.A.; Koh, Y.O. A new predictive equation for resting energy expenditure in healthy individuals. *Am. J. Clin. Nutr.* **1990**, *51*, 241–247. [[CrossRef](#)]
24. Bycroft, C.; Freeman, C.; Petkova, D.; Band, G.; Elliott, L.T.; Sharp, K.; Motyer, A.; Vukcevic, D.; Delaneau, O.; O'Connell, J.; et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **2018**, *562*, 203–209. [[CrossRef](#)] [[PubMed](#)]
25. McCarthy, S.; Das, S.; Kretzschmar, W.; Delaneau, O.; Wood, A.R.; Teumer, A.; Kang, H.M.; Fuchsberger, C.; Danecek, P.; Sharp, K.; et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **2016**, *48*, 1279–1283. [[CrossRef](#)]
26. Huang, J.; Howie, B.; McCarthy, S.; Memari, Y.; Walter, K.; Min, J.L.; Danecek, P.; Malerba, G.; Trabetti, E.; Zheng, H.F.; et al. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat. Commun.* **2015**, *6*, 8111. [[CrossRef](#)]
27. Loh, P.R.; Tucker, G.; Bulik-Sullivan, B.K.; Vilhjálmsson, B.J.; Finucane, H.K.; Salem, R.M.; Chasman, D.I.; Ridker, P.M.; Neale, B.M.; Berger, B.; et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **2015**, *47*, 284–290. [[CrossRef](#)]
28. Townsend, P.; Phillimore, P.; Beattie, A. *Health and Deprivation: Inequality and the North*; Routledge: London, UK, 1988.
29. Watanabe, K.; Taskesen, E.; van Bochoven, A.; Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **2017**, *8*, 1826. [[CrossRef](#)]

30. Buniello, A.; MacArthur, J.A.L.; Cerezo, M.; Harris, L.W.; Hayhurst, J.; Malangone, C.; McMahon, A.; Morales, J.; Mountjoy, E.; Sollis, E.; et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **2018**, *47*, D1005–D1012. [[CrossRef](#)]
31. Kamat, M.A.; Blackshaw, J.A.; Young, R.; Surendran, P.; Burgess, S.; Danesh, J.; Butterworth, A.S.; Staley, J.R. PhenoScanner V2: An expanded tool for searching human genotype–phenotype associations. *Bioinformatics* **2019**, *35*, 4851–4853. [[CrossRef](#)]
32. Ghousaini, M.; Mountjoy, E.; Carmona, M.; Peat, G.; Schmidt, E.M.; Hercules, A.; Fumis, L.; Miranda, A.; Carvalho-Silva, D.; Buniello, A.; et al. Open Targets Genetics: Systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res.* **2020**, *49*, D1311–D1320. [[CrossRef](#)] [[PubMed](#)]
33. de Leeuw, C.A.; Mooij, J.M.; Heskes, T.; Posthuma, D. MAGMA: Generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **2015**, *11*, e1004219. [[CrossRef](#)] [[PubMed](#)]
34. Mishra, A.; Macgregor, S. VEGAS2: Software for More Flexible Gene-Based Testing. *Twin Res. Hum. Genet. Off. J. Int. Soc. Twin Stud.* **2015**, *18*, 86–91. [[CrossRef](#)]
35. Giambartolomei, C.; Vukcevic, D.; Schadt, E.E.; Franke, L.; Hingorani, A.D.; Wallace, C.; Plagnol, V. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **2014**, *10*, e1004383. [[CrossRef](#)] [[PubMed](#)]
36. Vosa, U.; Claringbould, A.; Westra, H.J.; Bonder, M.J.; Deelen, P.; Zeng, B.; Kirsten, H.; Saha, A.; Kreuzhuber, R.; Yazar, S.; et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.* **2021**, *53*, 1300–1310. [[CrossRef](#)]
37. Vōsa, U.; Claringbould, A.; Westra, H.-J.; Bonder, M.J.; Deelen, P.; Zeng, B.; Kirsten, H.; Saha, A.; Kreuzhuber, R.; Kasela, S.; et al. Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv* **2018**, 447367. [[CrossRef](#)]
38. Zhu, Z.; Zhang, F.; Hu, H.; Bakshi, A.; Robinson, M.R.; Powell, J.E.; Montgomery, G.W.; Goddard, M.E.; Wray, N.R.; Visscher, P.M.; et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **2016**, *48*, 481–487. [[CrossRef](#)]
39. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **1995**, *57*, 289–300. [[CrossRef](#)]
40. Zheng, J.; Erzurumluoglu, A.M.; Elsworth, B.L.; Kemp, J.P.; Howe, L.; Haycock, P.C.; Hemani, G.; Tansey, K.; Laurin, C.; Pourcain, B.S.; et al. LD Hub: A centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **2017**, *33*, 272–279. [[CrossRef](#)]
41. Zhu, Z.; Zheng, Z.; Zhang, F.; Wu, Y.; Trzaskowski, M.; Maier, R.; Robinson, M.R.; McGrath, J.J.; Visscher, P.M.; Wray, N.R.; et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* **2018**, *9*, 224. [[CrossRef](#)]
42. Yang, J.; Lee, S.H.; Goddard, M.E.; Visscher, P.M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **2011**, *88*, 76–82. [[CrossRef](#)] [[PubMed](#)]
43. Locke, A.E.; Kahali, B.; Berndt, S.I.; Justice, A.E.; Pers, T.H.; Day, F.R.; Powell, C.; Vedantam, S.; Buchkovich, M.L.; Yang, J.; et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **2015**, *518*, 197–206. [[CrossRef](#)]
44. Willer, C.J.; Schmidt, E.M.; Sengupta, S.; Peloso, G.M.; Gustafsson, S.; Kanoni, S.; Ganna, A.; Chen, J.; Buchkovich, M.L.; Mora, S.; et al. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **2013**, *45*, 1274–1283. [[CrossRef](#)] [[PubMed](#)]
45. Manning, A.K.; Hivert, M.F.; Scott, R.A.; Grimsby, J.L.; Bouatia-Naji, N.; Chen, H.; Rybin, D.; Liu, C.T.; Bielak, L.F.; Prokopenko, I.; et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat. Genet.* **2012**, *44*, 659–669. [[CrossRef](#)]
46. Dupuis, J.; Langenberg, C.; Prokopenko, I.; Saxena, R.; Soranzo, N.; Jackson, A.U.; Wheeler, E.; Glazer, N.L.; Bouatia-Naji, N.; Gloyn, A.L.; et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* **2010**, *42*, 105–116. [[CrossRef](#)] [[PubMed](#)]
47. Nikpay, M.; Goel, A.; Won, H.H.; Hall, L.M.; Willenborg, C.; Kanoni, S.; Saleheen, D.; Kyriakou, T.; Nelson, C.P.; Hopewell, J.C.; et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **2015**, *47*, 1121–1130. [[CrossRef](#)] [[PubMed](#)]
48. Lu, Y.; Day, F.R.; Gustafsson, S.; Buchkovich, M.L.; Na, J.; Bataille, V.; Cousminer, D.L.; Dastani, Z.; Drong, A.W.; Esko, T.; et al. New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk. *Nat. Commun.* **2016**, *7*, 10495. [[CrossRef](#)]
49. de Moor, M.H.; van den Berg, S.M.; Verweij, K.J.; Krueger, R.F.; Luciano, M.; Arias Vasquez, A.; Matteson, L.K.; Derringer, J.; Esko, T.; Amin, N.; et al. Meta-analysis of Genome-wide Association Studies for Neuroticism, and the Polygenic Association With Major Depressive Disorder. *JAMA Psychiatry* **2015**, *72*, 642–650. [[CrossRef](#)]
50. Rietveld, C.A.; Medland, S.E.; Derringer, J.; Yang, J.; Esko, T.; Martin, N.W.; Westra, H.J.; Shakhbazov, K.; Abdellaoui, A.; Agrawal, A.; et al. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **2013**, *340*, 1467–1471. [[CrossRef](#)]
51. Privé, F.; Aschard, H.; Ziyatdinov, A.; Blum, M.G.B. Efficient analysis of large-scale genome-wide data with two R packages: Bigstatsr and bigsnpr. *Bioinformatics* **2018**, *34*, 2781–2787. [[CrossRef](#)]
52. Bowden, J.; Davey Smith, G.; Burgess, S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **2015**, *44*, 512–525. [[CrossRef](#)] [[PubMed](#)]

53. Hemani, G.; Zheng, J.; Elsworth, B.; Wade, K.H.; Haberland, V.; Baird, D.; Laurin, C.; Burgess, S.; Bowden, J.; Langdon, R.; et al. The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **2018**, *7*, e34408. [[CrossRef](#)] [[PubMed](#)]
54. Devlin, B.; Roeder, K. Genomic control for association studies. *Biometrics* **1999**, *55*, 997–1004. [[CrossRef](#)] [[PubMed](#)]
55. Bulik-Sullivan, B.K.; Loh, P.R.; Finucane, H.K.; Ripke, S.; Yang, J.; Patterson, N.; Daly, M.J.; Price, A.L.; Neale, B.M. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **2015**, *47*, 291–295. [[CrossRef](#)] [[PubMed](#)]
56. Howard, D.M.; Adams, M.J.; Clarke, T.K.; Hafferty, J.D.; Gibson, J.; Shirali, M.; Coleman, J.R.I.; Hagenars, S.P.; Ward, J.; Wigmore, E.M.; et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **2019**, *22*, 343–352. [[CrossRef](#)]
57. Karlsson Linnér, R.; Biroli, P.; Kong, E.; Meddens, S.F.W.; Wedow, R.; Fontana, M.A.; Lebreton, M.; Tino, S.P.; Abdellaoui, A.; Hammerschlag, A.R.; et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat. Genet.* **2019**, *51*, 245–257. [[CrossRef](#)]
58. Nagel, M.; Jansen, P.R.; Stringer, S.; Watanabe, K.; de Leeuw, C.A.; Bryois, J.; Savage, J.E.; Hammerschlag, A.R.; Skene, N.G.; Muñoz-Manchado, A.B.; et al. Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nat. Genet.* **2018**, *50*, 920–927. [[CrossRef](#)]
59. Pulit, S.L.; Stoneman, C.; Morris, A.P.; Wood, A.R.; Glastonbury, C.A.; Tyrrell, J.; Yengo, L.; Ferreira, T.; Marouli, E.; Ji, Y.; et al. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum. Mol. Genet.* **2019**, *28*, 166–174. [[CrossRef](#)]
60. Tabarés-Seisdedos, R.; Rubenstein, J.L. Chromosome 8p as a potential hub for developmental neuropsychiatric disorders: Implications for schizophrenia, autism and cancer. *Mol. Psychiatry* **2009**, *14*, 563–589. [[CrossRef](#)]
61. Feitosa, M.F.; Kraja, A.T.; Chasman, D.I.; Sung, Y.J.; Winkler, T.W.; Ntalla, I.; Guo, X.; Franceschini, N.; Cheng, C.Y.; Sim, X.; et al. Novel genetic associations for blood pressure identified via gene-alcohol interaction in up to 570K individuals across multiple ancestries. *PLoS ONE* **2018**, *13*, e0198166. [[CrossRef](#)]
62. Savage, J.E.; Jansen, P.R.; Stringer, S.; Watanabe, K.; Bryois, J.; de Leeuw, C.A.; Nagel, M.; Awasthi, S.; Barr, P.B.; Coleman, J.R.I.; et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.* **2018**, *50*, 912–919. [[CrossRef](#)] [[PubMed](#)]
63. Park, S.L.; Cheng, I.; Pendergrass, S.A.; Kucharska-Newton, A.M.; Lim, U.; Ambite, J.L.; Caberto, C.P.; Monroe, K.R.; Schumacher, F.; Hindorff, L.A.; et al. Association of the FTO obesity risk variant rs8050136 with percentage of energy intake from fat in multiple racial/ethnic populations: The PAGE study. *Am. J. Epidemiol.* **2013**, *178*, 780–790. [[CrossRef](#)] [[PubMed](#)]
64. Boender, A.J.; van Gestel, M.A.; Garner, K.M.; Luijendijk, M.C.; Adan, R.A. The obesity-associated gene *Negr1* regulates aspects of energy balance in rat hypothalamic areas. *Physiol. Rep.* **2014**, *2*, e12083. [[CrossRef](#)] [[PubMed](#)]
65. Culmsee, C.; Monnig, J.; Kemp, B.E.; Mattson, M.P. AMP-activated protein kinase is highly expressed in neurons in the developing rat brain and promotes neuronal survival following glucose deprivation. *J. Mol. Neurosci. MN* **2001**, *17*, 45–58. [[CrossRef](#)]
66. Mihaylova, M.M.; Shaw, R.J. The AMPK signalling pathway coordinates cell growth, autophagy and metabolism. *Nat. Cell Biol.* **2011**, *13*, 1016–1023. [[CrossRef](#)] [[PubMed](#)]
67. Kalra, S.P.; Kalra, P.S. NPY and cohorts in regulating appetite, obesity and metabolic syndrome: Beneficial effects of gene therapy. *Neuropeptides* **2004**, *38*, 201–211. [[CrossRef](#)]
68. Anderson, K.A.; Means, R.L.; Huang, Q.H.; Kemp, B.E.; Goldstein, E.G.; Selbert, M.A.; Edelman, A.M.; Freneau, R.T.; Means, A.R. Components of a calmodulin-dependent protein kinase cascade. Molecular cloning, functional characterization and cellular localization of Ca<sup>2+</sup>/calmodulin-dependent protein kinase kinase beta. *J. Biol. Chem.* **1998**, *273*, 31880–31889. [[CrossRef](#)]
69. Qi, T.; Wu, Y.; Zeng, J.; Zhang, F.; Xue, A.; Jiang, L.; Zhu, Z.; Kemper, K.; Yengo, L.; Zheng, Z.; et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat. Commun.* **2018**, *9*, 2282. [[CrossRef](#)]
70. Huffman, J.E. Examining the current standards for genetic discovery and replication in the era of mega-biobanks. *Nat. Commun.* **2018**, *9*, 5054. [[CrossRef](#)]
71. Ramírez, J.; Duijvenboden, S.V.; Ntalla, I.; Mifsud, B.; Warren, H.R.; Tzanis, E.; Orini, M.; Tinker, A.; Lambiase, P.D.; Munroe, P.B. Thirty loci identified for heart rate response to exercise and recovery implicate autonomic nervous system. *Nat. Commun.* **2018**, *9*, 1947. [[CrossRef](#)]
72. Verweij, N.; van de Vegte, Y.J.; van der Harst, P. Genetic study links components of the autonomous nervous system to heart-rate profile during exercise. *Nat. Commun.* **2018**, *9*, 898. [[CrossRef](#)] [[PubMed](#)]