

## RESEARCH ARTICLE

## Genome-wide association study of fish oil supplementation on lipid traits in 81,246 individuals reveals new gene-diet interaction loci

Michael Francis<sup>1</sup>, Changwei Li<sup>2</sup>, Yitang Sun<sup>3</sup>, Jingqi Zhou<sup>3</sup>, Xiang Li<sup>2</sup>, J. Thomas Brenna<sup>4,5,6</sup>, Kaixiong Ye<sup>1,3\*</sup>

**1** Institute of Bioinformatics, University of Georgia, Athens, Georgia, United States of America, **2** Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, Louisiana, United States of America, **3** Department of Genetics, University of Georgia, Athens, Georgia, United States of America, **4** Division of Nutritional Sciences, Cornell University, Ithaca, New York, United States of America, **5** Department of Pediatrics, Dell Pediatric Research Institute, University of Texas at Austin, Austin, Texas, United States of America, **6** Departments of Nutrition and Chemistry, University of Texas at Austin, Austin, Texas, United States of America

\* [kaixiong.ye@uga.edu](mailto:kaixiong.ye@uga.edu)

## OPEN ACCESS

**Citation:** Francis M, Li C, Sun Y, Zhou J, Li X, Brenna JT, et al. (2021) Genome-wide association study of fish oil supplementation on lipid traits in 81,246 individuals reveals new gene-diet interaction loci. *PLoS Genet* 17(3): e1009431. <https://doi.org/10.1371/journal.pgen.1009431>

**Editor:** Heather J Cordell, Newcastle University, UNITED KINGDOM

**Received:** June 9, 2020

**Accepted:** February 16, 2021

**Published:** March 24, 2021

**Copyright:** © 2021 Francis et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Individual-level genetic and phenotypic data cannot be shared publicly because of participant privacy. Data are available from the UK Biobank Institutional Data Access / Ethics Committee (<https://www.ukbiobank.ac.uk/register-apply/>) with applications. All summary statistics for Gene-Fish Oil Interactions are publicly available on figshare (<https://doi.org/10.6084/m9.figshare.14171069.v1>). All other relevant data are within the manuscript and its [Supporting information files](#). Key computational scripts are available here:

## Abstract

Fish oil supplementation is widely used for reducing serum triglycerides (TAGs) but has mixed effects on other circulating cardiovascular biomarkers. Many genetic polymorphisms have been associated with blood lipids, including high- and low-density-lipoprotein cholesterol (HDL-C, LDL-C), total cholesterol, and TAGs. Here, the gene-diet interaction effects of fish oil supplementation on these lipids were analyzed in a discovery cohort of up to 73,962 UK Biobank participants, using a 1-degree-of-freedom (1df) test for interaction effects and a 2-degrees-of-freedom (2df) test to jointly analyze interaction and main effects. Associations with  $P < 1 \times 10^{-6}$  in either test (26,157; 18,300 unique variants) were advanced to replication in up to 7,284 participants from the Atherosclerosis Risk in Communities (ARIC) Study. Replicated associations reaching 1df  $P < 0.05$  (2,175; 1,763 unique variants) were used in meta-analyses. We found 13 replicated and 159 non-replicated (UK Biobank only) loci with significant 2df joint tests that were predominantly driven by main effects and have been previously reported. Four novel interaction loci were identified with 1df  $P < 5 \times 10^{-8}$  in meta-analysis. The lead variant in the *GJB6-GJB2-GJA3* gene cluster, rs112803755 (A>G; minor allele frequency = 0.041), shows exclusively interaction effects. The minor allele is significantly associated with decreased TAGs in individuals with fish oil supplementation, but with increased TAGs in those without supplementation. This locus is significantly associated with higher *GJB2* expression of connexin 26 in adipose tissue; connexin activity is known to change upon exposure to omega-3 fatty acids. Significant interaction effects were also found in three other loci in the genes *SLC12A3* (HDL-C), *ABCA6* (LDL-C), and *MLXIPL* (LDL-C), but highly significant main effects are also present. Our study identifies novel gene-diet interaction effects for four genetic loci, whose effects on blood lipids are modified

<https://github.com/michaelofrancis/FishOil-Lipid-Interaction>.

**Funding:** KY is supported by the University of Georgia Research Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

by fish oil supplementation. These findings highlight the need and possibility for personalized nutrition.

### Author summary

We utilized the unprecedentedly large genotype and phenotype dataset in the UK Biobank to perform a genome-wide association study (GWAS) which accounts for the interplay between genotype and dietary intake. We examined the interaction effects of fish oil supplementation on levels of blood lipids (LDL-C, HDL-C, TAGs, and total cholesterol). Our findings were replicated in the Atherosclerosis Risk in Communities (ARIC) Study. We found that at the genetic variant rs112803755 (A>G), the minor allele (G) is associated with a decrease in TAGs among individuals with fish oil supplementation, but is associated with an increase in TAGs among those without supplementation. In other words, only individuals carrying the minor allele benefit from fish oil supplementation in reducing TAG levels. We further analyzed rs112803755 with functional genomics data from the Genotype-Tissue Expression (GTEx) project to identify potential target genes, and found a connexin coding gene which has been previously reported to respond to cellular omega-3 levels. This research suggests that inter-personal variation in TAG response to fish oil supplementation is in part explained by genotype, and that fish oil dose adjustment based on genotype should be investigated as a means to protect against cardiovascular disease risk.

### Introduction

Dyslipidemia, characterized by imbalances in low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), and triglycerides (TAGs), is a common predictive factor for metabolic conditions such as cardiovascular disease and type 2 diabetes [1,2]. Use of dietary supplements in lieu of xenobiotic pharmaceuticals for the management of dyslipidemia may produce comparable benefits with fewer side effects [1,2]. In particular, the omega-3 long-chain polyunsaturated fatty acids (n-3 LCPUFAs) eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) supplied by fish oil supplements are an effective treatment for hypertriglyceridemia, though results are mixed for LDL-C and HDL-C [3–5]. Genetic polymorphisms have been consistently associated with intra- and inter-population differences in levels of LDL-C, HDL-C, total cholesterol, and TAGs [6,7]. Gene-environment interactions (GEIs; specifically, gene-diet interactions) between n-3 LCPUFAs and genetic variants have been reported, though few have been replicated, likely due to small sample sizes and inconsistencies in study designs such as study length and supplement dosage [8]. Studies of GEIs may reveal novel genetic loci that are otherwise obscure in conventional main-effect-only association studies, and may identify genetic loci whose phenotypic effects are modifiable by specific environmental exposures. Further identification of these GEIs may help explain both missing heritability in lipid biomarker traits [9], and heterogeneity of individual lipid response to fish oil supplementation [8,10–13].

To identify genomic factors which interact with n-3 LCPUFAs supplementation to affect levels of blood lipids, we performed a genome-wide association study (GWAS) among participants of the large UK Biobank cohort [14]. We used only participants whose genetic ethnic grouping is Caucasian, the largest sample available, to avoid population stratification

[15]. The focus on single-ancestry groups is particularly important in studies related to LCPUFAs because their metabolic genes have been shown to undergo genetic adaptation to local diets in multiple geographical regions, and exhibit population-specific allele frequency patterns [16,17]. We used both the traditional 1-degree-of-freedom (1df) interaction test and a 2-degrees-of-freedom (2df) joint test to evaluate interactions between genetic variants and fish oil supplementation on blood lipid phenotypes. The 2df joint test evaluates single nucleotide polymorphism (SNP) main effects and interaction effects jointly, and therefore has higher power to detect SNPs with moderate main effects and moderate interaction effects that would otherwise be missed in the 1df test [18–20]. This method has recently been employed to examine the GEIs of these lipid traits with smoking [21], and sleep duration [22]. We further confirmed promising UK Biobank findings in a US cohort, the Atherosclerosis Risk in Communities Study (ARIC). Replicated SNPs were utilized in a meta-analysis of these studies to reveal new gene-diet interaction loci.

## Results

### Cohort demographics

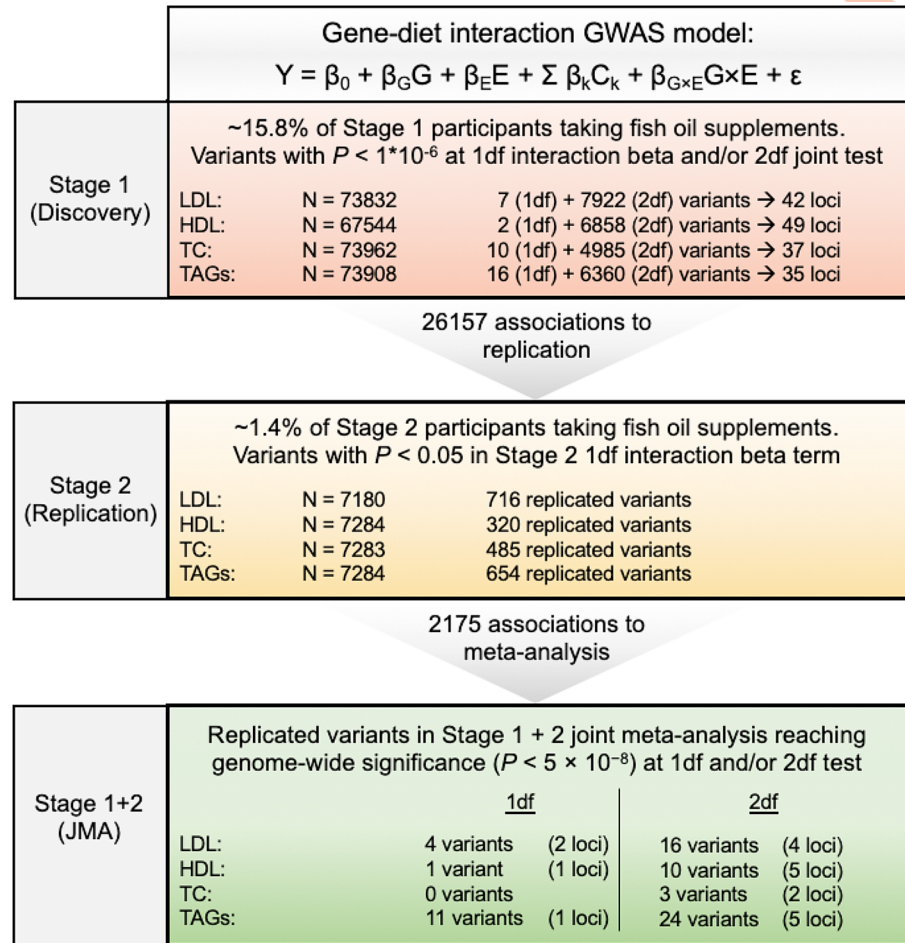
Stage 1 discovery analyses were performed in up to 73,962 genetically Caucasian UK Biobank participants (S1 Table). Approximately 15.8% of these participants answered yes to taking fish oil supplements on dietary questionnaires at two time points taken between one and five years apart. The percentage male, mean age and BMI of UK Biobank participants were ~46.6%,  $55.6 \pm 7.9$  ( $\pm 1$  SD) years old, and  $27.0 \pm 4.6$  kg/m<sup>2</sup>, respectively. Stage 2 (replication) analyses were performed in up to 7,284 white participants in the ARIC cohort study. Approximately 1.4% of these participants answered yes to taking fish oil at one time point during their primary assessment. The percentage male, mean age and BMI of ARIC participants were ~47.0%,  $54.3 \pm 5.7$  years old, and  $26.9 \pm 4.7$  kg/m<sup>2</sup>, respectively.

### Gene-diet interaction GWAS

A three-stage discovery, replication, and meta-analysis approach for identification of significant GWAS loci was adopted for the blood lipid phenotypes LDL-C, HDL-C, total cholesterol, and TAGs (Fig 1). Genomic control (GC) correction was applied during Stage 1 2df *P*-value calculation and Stage 3 2df *P*-value calculation; lambda ( $\lambda$ ) values after GC correction were 1. GC values of 1df *P*-values for Stage 1 and Stage 3 were  $< 1$ , therefore GC correction was not necessary.

Variants with 1df or 2df  $P < 1 \times 10^{-6}$  in a gene-fish-oil interaction GWAS model (Eq (1)) were selected for replication (S2 Table). For the four lipid traits, LDL-C, HDL-C, total cholesterol, and TAGs, 26,157 associations (18,300 unique variants) met this criterion (S1 and S2 Figs).

Stage 2 replication analyses were performed in up to 7,284 white participants in the ARIC cohort (S1 Table). A gene-fish-oil interaction GWAS model was performed. Variants passed from Stage 1 with 1df  $P < 0.05$  were considered as replicated. Of the 26,157 associations from Stage 1, a total of 2,175 associations (1763 unique variants) for the lipid traits were replicated (S3 Table) and passed to the meta-analysis step. There were also 17,259 associations (12,440 unique variants, 85 unique loci; S4 and S5 Tables) which reached genome-wide significance in Stage 1 ( $P < 5 \times 10^{-8}$ ), but were not replicated in Stage 2, and therefore not sent to meta-analysis. All of the 85 lead variants had significant 2df joint test *P*-values, and none of their 1df interaction *P*-values approached significance, suggesting these variants influence lipid traits predominantly through main effects.



**Fig 1. Overview of the analysis performed in this study.** A three-stage discovery, replication, and meta-analysis process was used to identify significant variants. Stage 1 revealed 26,157 associations with 1df and/or 2df  $P < 1 \times 10^{-6}$  in a cohort of up to 73,962 participants. Of these associations, 2,175 were replicated in a cohort of up to 7,284 participants. In meta-analysis, 4 1df loci (Table 1) and 16 2df loci (13 additional loci, Table 2) reached the genome-wide significance of  $P < 5 \times 10^{-8}$ . TC, total cholesterol; TAGs, triglycerides.

<https://doi.org/10.1371/journal.pgen.1009431.g001>

Meta-analysis of Stage 1 and Stage 2 results for both 1df and 2df tests were performed for each blood lipid phenotype. Significant variants were defined in meta-analysis as those meeting the genome-wide significance threshold in their 1df or 2df tests ( $P < 5 \times 10^{-8}$ ). This revealed 16 novel and significant 1df associations (4 unique loci; Table 1) and 53 significant 2df associations (11 unique loci; Table 2 and S6 Table). One variant, rs112803755 (*GJB6*; A>G; minor allele frequency (MAF) = 0.0410) had a significant 1df interaction term and no significant 2df or main effects terms. In the discovery cohort, the minor allele of SNP rs112803755 is associated with a strong decrease in TAGs among those taking fish oil supplements ( $\beta_{G(E=1)} = -0.12$  mmol/L,  $P = 5.59 \times 10^{-5}$ ), but is suggestively associated with a mild increase in TAGs for those without supplementation ( $\beta_{G(E=0)} = 0.030$  mmol/L,  $P = 0.024$ ), resulting in a significant interaction effect (1df  $P = 1.95 \times 10^{-7}$ ). There is no association between the SNP main effects and TAGs in the UK Biobank ( $\beta_G = 0.0063$  mmol/L,  $P = 0.60$ ) if not considering the interaction effect. Meta-analysis revealed that the interaction effect at this SNP reaches genome-wide significance (1df  $P = 5.65 \times 10^{-10}$ ). Three additional variants have both significant 1df interaction and 2df joint test  $P$ -values in the meta-analysis: rs799157 (*MLXIPL*;

**Table 1. Loci with significant interaction between fish oil supplementation and blood lipid levels.** Listed variants represent the lead association within a 1 Mb region for 1df tests of variant × fish oil interaction after meta-analysis. The name of the nearest gene is listed with each lead variant. Bold P-values indicate meeting the genome-wide significance threshold of  $P < 5 \times 10^{-8}$ . Main effect P-values are calculated using Stage 1 (UK Biobank) participants only, and without interaction (Eq (2)); stratified for exposure groups as in Eq (3). Effect, beta coefficient of the minor allele dose term ( $\beta_G$  in Eq (1)); MAF, minor allele frequency; SE, standard error; Int effect, beta coefficient of the interaction term ( $\beta_{G \times E}$  in Eq (1)). Lipid traits were measured in mmol/L.

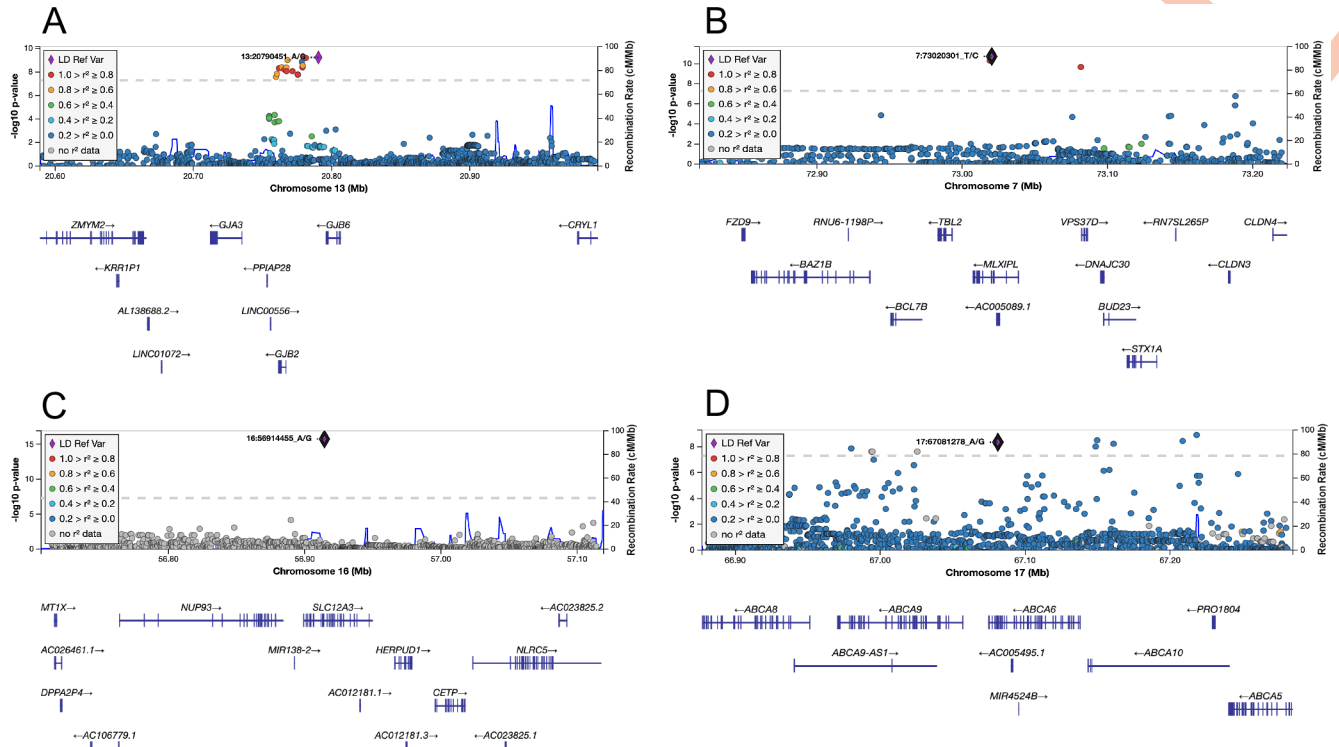
hg19 chr: pos	Stage 1 + 2 Meta-analysis genome-wide significant interaction loci										Stage 1					
	rsID (nearest gene)	Minor (effect) allele (avg freq)	Reference allele	Stage1 MAF / Stage 2 MAF	Lipid trait	n	Effect	SE	Int Effect	Int SE	1df interaction P-value	2df joint P-value	n	Adj. Main effect P-value	Main effect P-value (E = 0)	Main effect P-value (E = 1)
13:20790451	rs112803755 (GJB6: -5650 bp downstream)	g (0.0410)	a	0.0416 / 0.0352	TAGs	81192	0.04272	0.0447	-0.3084	0.1048	<b>5.65E-10</b>	0.01247	73908	0.6007	2.42E-02	5.59E-05
7:73020301	rs799157 (MLXIPL; Synonymous Variant)	t (0.0407)	c	0.0435 / 0.0112	LDL	81012	0.07162	0.05678	1.459	0.1371	<b>1.92E-11</b>	<b>1.93E-33</b>	73832	<b>3.33E-08</b>	5.36E-06	8.14E-04
17:67081278	rs77542162 (ABCA6; Missense Variant)	g (0.0218)	a	0.0226 / 0.0134	LDL	81012	0.1741	0.06495	-1.900	0.1224	<b>4.48E-09</b>	<b>6.58E-63</b>	73832	<b>5.40E-23</b>	<b>4.24E-21</b>	1.55E-03
16:56914455	rs148931404 (SLC12A3; Intron Variant)	a (0.0221)	g	0.0226 / 0.0172	HDL	74824	0.05594	0.02328	0.7734	0.04719	<b>1.82E-16</b>	<b>5.04E-91</b>	67544	<b>2.70E-16</b>	<b>5.67E-12</b>	3.49E-06

<https://doi.org/10.1371/journal.pgen.1009431.t001>

**Table 2. Loci with significant 2df joint test between fish oil supplementation and blood lipid levels.** Listed variants represent the lead association within a 1-Mb region for 2df tests of variant × fish oil interaction after meta-analysis. The name of the nearest gene is listed with each lead variant. Bold *P*-values indicate meeting the genome-wide significance threshold of  $P < 5 \times 10^{-8}$ . Main effect *P*-values are calculated using Stage 1 (UK Biobank) participants only, and without interaction (Eq (2)); stratified for exposure groups as in Eq (3)). Effect, beta coefficient of the minor allele dose term ( $\beta_G$  in Eq (1)); MAF, minor allele frequency; SE, standard error; Int effect, beta coefficient of the interaction term ( $\beta_{G \times E}$  in Eq (1)). Lipid traits were measured in mmol/L.

Stage 1 + 2 Meta-analysis replicated genome-wide significant loci												
hg19 chr:pos	rsID (nearest gene)	Minor (effect) allele (avg freq)	Reference allele	Stage1 MAF / Stage 2 MAF	Lipid trait	n	Effect	SE	Int Effect	Int SE	1df interaction <i>P</i> -value	2df joint <i>P</i> -value
6:34094919	rs115675705 (GRAM4; Intron Variant)	g (0.0357)	a	0.0367 / 0.0268	HDL	74824	-0.03882	0.01795	-0.3348	0.04479	0.00873	1.23E-19
8:19722204	rs117860853 (JPL; ~217kb upstream)	a (0.0135)	g	0.0137 / 0.0116	HDL	74824	-0.08003	0.02691	-0.4139	0.05311	1.48E-02	5.46E-28
11:116916060	rs144018203 (SIK3; Intron Variant)	c (0.0102)	g	0.0107 / 0.0052	HDL	74824	-0.05472	0.03791	0.4387	0.05912	0.04122	1.03E-16
17:42061277	rs147438979 (PYY; Intron Variant)	c (0.011)	g	0.0116 / 0.006	HDL	74824	-0.06327	0.02952	-0.3336	0.05753	4.18E-02	2.75E-16
1:55583210	rs530804537 (USP24; Intron Variant)	a (0.011)	g	0.011 / 0.0012	LDL	81012	-0.07453	0.08869	-1.133	0.1360	0.1325	2.68E-31
19:45198060	rs112952132 (LOC107985305; Intron Variant)	t (0.01)	c	0.01 / 0.0101	LDL	81012	-0.2051	0.08381	-1.055	0.1329	2.77E-02	7.58E-34
7:72921771	rs117788606 (BAZ1B; Intron Variant)	c (0.0112)	t	0.0120 / 0.0032	TAGs	81192	-0.06165	0.09183	1.012	0.1464	2.67E-04	3.93E-16
8:19768150	rs142084074 (LOC107986921; Intron Variant)	a (0.015)	g	0.0150 / 0.0148	TAGs	81192	-0.1182	0.07241	0.911	0.1319	2.72E-02	3.153E-12
11:116916060	rs144018203 (SIK3; Intron Variant)	c (0.0103)	g	0.0108 / 0.0052	TAGs	81192	0.2531	0.1359	-0.9079	0.1735	0.008127	2.48E-09
15:43820717	rs55707100 (MAPIA; Missense Variant)	t (0.0249)	c	0.0247 / 0.0273	TAGs	81192	0.1201	0.05043	0.6116	0.1122	0.01044	2.36E-13
19:19365178	rs141844019 (HAPLN4; 500B Downstream Variant)	t (0.0104)	c	0.0109 / 0.0054	TAGs	81192	-0.06922	0.09157	1.637	0.1366	1.64E-06	4.27E-53
1:55583210	rs530804537 (USP24; Intron Variant)	a (0.0109)	g	0.0113 / 0.0073	Tot. Chol.	81245	-0.05886	0.1084	-1.057	0.1603	0.1002	2.24E-20
17:67081278	rs77542162 (ABCA6; Missense Variant)	g (0.0218)	a	0.0226 / 0.0135	Tot. Chol.	81245	0.1599	0.06831	-1.587	0.1261	4.58E-07	1.80E-41

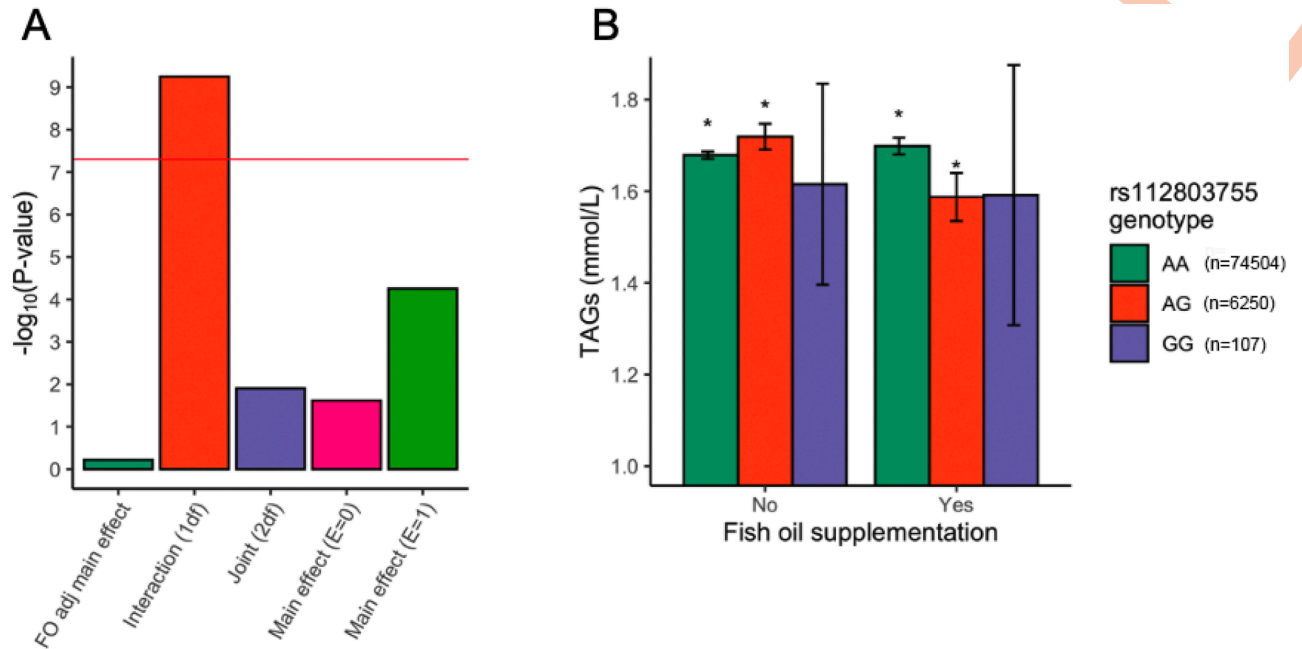
<https://doi.org/10.1371/journal.pgen.1009431.t002>



**Fig 2. LocusZoom for genome-wide significant ( $P < 5 \times 10^{-8}$ ) replicated gene-fish oil interaction loci.** (A) rs112803755  $\times$  fish oil and TAGs, stage 1 + 2 1df tests ( $n = 81,192$ ). (B) rs799157  $\times$  fish oil and LDL-C, stage 1 + 2 1df tests ( $n = 81,012$ ). (C) rs148931404  $\times$  fish oil and HDL-C, stage 1 + 2 1df tests ( $n = 74,824$ ). (D) rs77542162  $\times$  fish oil and LDL-C, stage 1 + 2 1df tests ( $n = 81,012$ ).

<https://doi.org/10.1371/journal.pgen.1009431.g002>

C>T; MAF = 0.0407) with LDL-C, rs77542162 (*ABCA6*; A>G; MAF = 0.0218) with LDL-C, and rs148931404 (*SLC12A3*; G>A; MAF = 0.0221) with HDL-C (Table 1 and Fig 2). In the discovery cohort, the minor allele of SNP rs799157 is associated with an increase in LDL-C ( $\beta_G = 0.057$  mmol/L,  $P = 3.33 \times 10^{-8}$ ) after adjusting for fish oil supplementation status and other covariates. Less significant associations were observed in the stratified groups with fish oil supplementation ( $\beta_{G(E=1)} = 0.087$  mmol/L,  $P = 8.14 \times 10^{-4}$ ) and in those without ( $\beta_{G(E=0)} = 0.052$  mmol/L,  $P = 5.36 \times 10^{-6}$ ). Meta-analysis confirmed the presence of main effect and revealed an interaction effect (1df  $P = 1.92 \times 10^{-11}$ , 2df  $P = 1.93 \times 10^{-33}$ ). Similarly, SNP rs77542162 is associated with an increase in LDL-C in the overall discovery cohort ( $\beta_G = 1.41$  mmol/L,  $P = 5.40 \times 10^{-23}$ ), in those without ( $\beta_{G(E=0)} = 1.50$ ,  $P = 4.24 \times 10^{-21}$ ) and with ( $\beta_{G(E=1)} = 1.11$ ,  $p = 1.55 \times 10^{-3}$ ) fish oil supplementation. Meta-analysis revealed genome-wide significance in both tests (1df  $P = 4.48 \times 10^{-9}$ , 2df  $P = 6.58 \times 10^{-63}$ ). For HDL-C, there is only one SNP, rs148931404, that reaches genome-wide significant 1df P-value ( $1.82 \times 10^{-16}$ ) in the meta-analysis. It is associated with an increase in HDL-C in the overall discovery cohort ( $\beta_G = 0.049$  mmol/L,  $P = 2.70 \times 10^{-16}$ ), in those without ( $\beta_{G(E=0)} = 0.045$  mmol/L,  $P = 5.67 \times 10^{-12}$ ) and with ( $\beta_{G(E=1)} = 0.071$  mmol/L,  $P = 3.49 \times 10^{-6}$ ) fish oil supplementation. The three variants with both significant 1df and 2df P-values are mainly driven by main effects, as reflected by the much more significant 2df P-values and the consistent associations across subgroups in UK Biobank. All four loci have been previously found to be associated with the corresponding lipid. Overall, we unraveled novel gene-fish oil interaction effects for four previously known lipid-associated genetic loci.



**Fig 3. Significant results for the replicated interaction locus with lead SNP rs112803755.** (A) rs112803755  $P$ -values in five regression models. The red line is the negative  $\log_{10}$ -transformed genome-wide significance of  $5 \times 10^{-8}$ . (B) Triglyceride lowering effect of fish oil supplementation on rs112803755 heterozygotes. Levels of TAGs stratified by genotypes at rs112803755 and fish oil supplementation status. Error bars show 95% confidence intervals. Exact numbers and sample sizes can be found in S7 Table.

<https://doi.org/10.1371/journal.pgen.1009431.g003>

There are 11 unique genetic loci whose 2df joint test  $P$ -values reached the genome-wide significance cutoff ( $P < 5 \times 10^{-8}$ ) but their 1df interaction test  $P$ -values did not. For instance, a SNP upstream of *LPL* rs117860853 is associated with a decrease in HDL-C in the overall discovery cohort ( $\beta_G = -0.078$  mmol/L,  $P = 5.72 \times 10^{-24}$ ), in those without ( $\beta_{G(E=0)} = -0.081$  mmol/L,  $P = 3.47 \times 10^{-22}$ ) and with fish oil supplementation ( $\beta_{G(E=1)} = -0.063$  mmol/L,  $P = 3.00 \times 10^{-3}$ ). Meta-analysis revealed that this SNP has a significant main effect but no interaction effect (1df  $P = 0.015$ , 2df  $P = 5.46 \times 10^{-28}$ ). Notably, two loci have 1df interaction test  $P$ -values that are close to the genome-wide significance level. SNP rs141844019, downstream of *HAPLN4*, has a suggestive interaction effect on TAGs ( $\beta_{G \times E} = 1.64$  mmol/L,  $P = 1.64 \times 10^{-6}$ ), while SNP rs77542162, a missense variant of *ABCA6*, may have an interaction effect on total cholesterol ( $\beta_{G \times E} = -1.59$  mmol/L,  $P = 4.58 \times 10^{-7}$ ). All these significant 2df replicated loci (Tables 1 and 2) were within 1 Mb of one or more previously reported loci associated with the same blood lipid phenotype and are therefore not reported as novel.

### rs112803755 modifies the effect of fish oil on TAGs

Using TAG levels as a phenotype, the locus of 11 significant variants whose lead SNP is rs112803755 (*GJB6*: 5650 bp downstream; A>G; MAF = 0.0410) has a significant 1df interaction  $P$ -value ( $5.65 \times 10^{-10}$ ), while its 2df joint  $P$ -value is not significant ( $P = 0.0124$ ) (Fig 3A). Its fish-oil adjusted main effects model SNP term is not significant ( $P = 0.600$ ), and in a stratified analysis the  $P$ -value is lower in the fish-oil supplementation exposure group ( $P = 5.59 \times 10^{-5}$ ) than the non-supplementing group ( $P = 2.42 \times 10^{-2}$ ) (Fig 3A). This evidence suggests that this locus is involved predominantly with interaction effects but not main effects.

The rs112803755 locus has significant TAG-lowering effect in those who supplement fish oil versus those who do not when considering AA vs. AG genotypes (Fig 3B and



[S7 Table](#)). Since this variant has low MAF (~4.1%), homozygous individuals of GG genotype are rare. TAG levels were significantly higher in AG heterozygotes who did not take fish oil ( $\Delta\bar{x}_{TAGs} = +0.04$  mmol/L) versus those who did ( $\Delta\bar{x}_{TAGs} = -0.111$  mmol/L). However, with respect to rs112803755, while fish oil supplementation is associated with lower TAGs in heterozygous individuals, it has a slight opposite effect in AA homozygotes ( $\Delta\bar{x}_{TAGs} = +0.0197$  mmol/L;  $P = 0.0258$ ).

### rs112803755 eQTL mapping

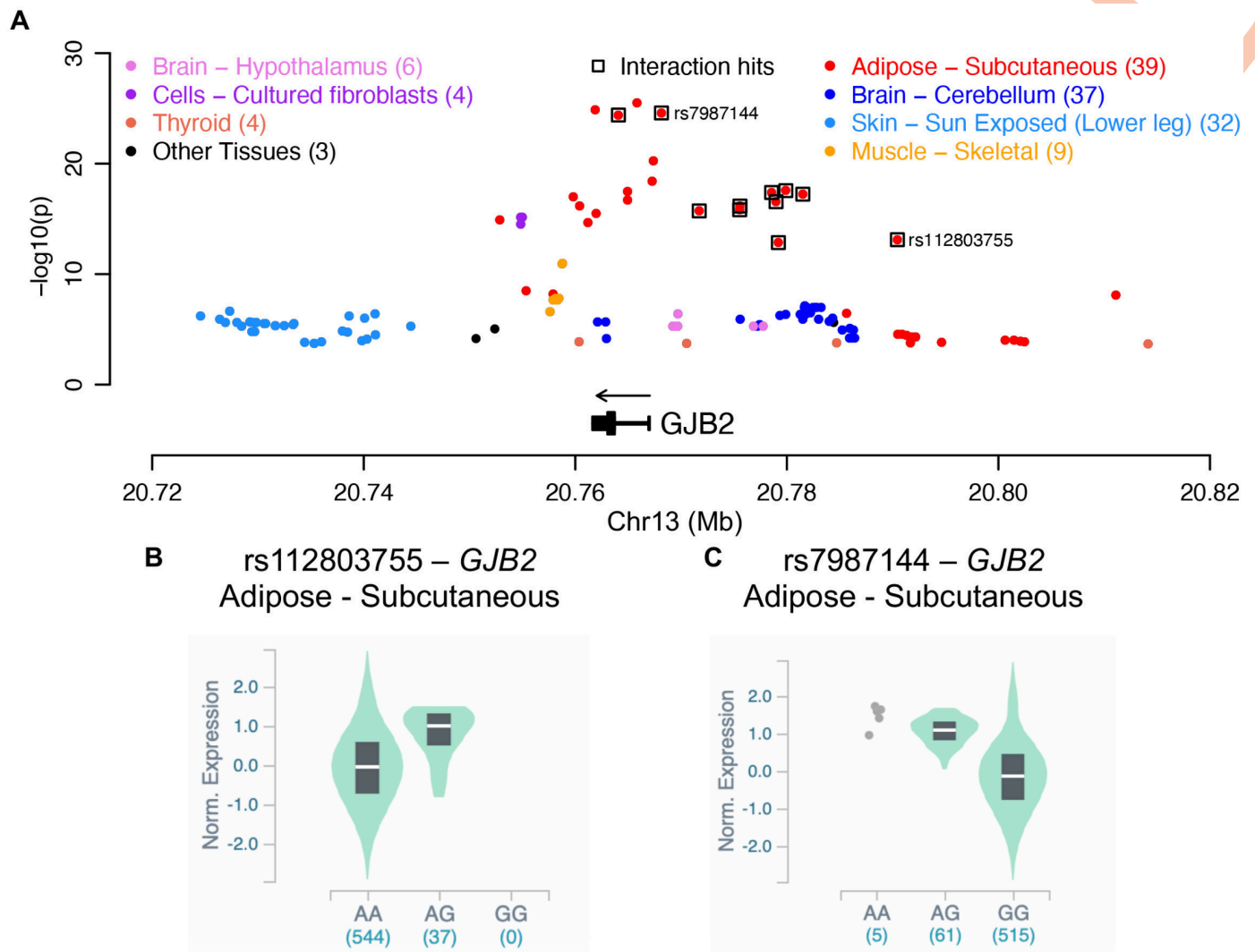
To evaluate if regulation of gene expression is an underlying molecular mechanism for the interaction locus whose lead SNP is rs112803755 ([Table 1](#)), we interrogated the association of these genetic markers with expression levels of nearby genes using data from the Genotype-Tissue Expression (GTEx) project. For the 11 genetic markers in this locus with genome-wide significance of interaction with fish oil, all of them are exclusively associated with the expression of *GJB2*. Expression quantitative trait loci (eQTLs) for *GJB2* were found in multiple tissues but the strongest signals were observed in subcutaneous adipose, which overlap with the significant interaction signals ([Fig 4A](#)). rs112803755 is associated with *GJB2* expression in subcutaneous adipose ( $P = 7.7 \times 10^{-14}$ ; [Fig 4B](#)), while another interaction SNP in this locus, rs7987144 (G>A; MAF = 0.0375), has an even stronger association ( $P = 2.6 \times 10^{-25}$ ; [Fig 4C](#)). Both of these SNPs show increased *GJB2* expression with increased minor allele dosage. These eQTLs results indicate that regulatory variants of *GJB2* are likely responsible for the interaction signals at this locus.

### Discussion

In this gene-diet interaction GWAS, we identified and replicated novel interaction loci, in which fish oil supplementation affected levels of continuous lipid traits in a large Caucasian cohort. We found one locus, rs112803755, with a significant interaction effect but a non-significant main effect, suggesting that the presence of minor alleles at this locus can enhance the TAG-lowering effects of fish oil supplementation. We found three additional new significant interaction loci related to LDL-C and HDL-C levels, though these appear predominantly influenced by main effects ([Table 1](#)).

rs112803755 is found 5.65 kb downstream from *GJB6*, or alternatively 23.3 kb upstream from *GJB2*. It is also in high LD with variants found in the other genes in the *GJB6-GJB2-GJA3* gene cluster at 13q12.11 ([Fig 2A](#)). *GJB6*, *GJB2*, and *GJA3* are connexin (Cx) gap junction protein-coding genes that encode Cx30, Cx26, and Cx46, respectively. Cxs are responsible for forming hemichannels across gap junctions to enable the exchange of messenger molecules between adjacent cells. An n-6 LCPUFA, linoleic acid, has been shown to increase hemichannel activity of Cx26 in HeLa cells [23], and n-3 LCPUFAs lowered the expression of another connexin, Cx43, in rats with hypertriglyceridemia [24]. Genetic polymorphisms in another Cx gene are associated with protective effects on cardiovascular disease [25]. It is therefore plausible that changes in n-3 LCPUFA status induced by fish oil supplementation could interact with one Cx in this cluster to affect TAG levels. Although our analysis supports the likely presence of a regulatory variant, we also cannot rule out the existence of a causal coding variant.

rs799157 is a synonymous variant in exon 6 of *MLXIPL*, whose gene product is known as Carbohydrate-responsive element-binding protein (ChREBP). We found this variant has a significant interaction effect of fish oil on LDL-C. Variants in *MLXIPL* have previously been associated with changes in LDL-C and TAGs [26,27]. Intracellular levels of PUFAs are known to suppress ChREBP transactivity, though the molecular basis for this is not defined [28,29]. GTEx reveals that SNPs in this locus are significantly associated with increased expression of



**Fig 4. The gene-fish oil interaction locus with lead SNP rs112803755 overlaps eQTLs of *GJB2*.** (A) Genetic variants significantly associated with the expression of *GJB2* as detected in the GTEx project. Colors indicate the tissues or cells. For variants with significant association in more than one tissues, the most significant p value is shown. The association of (B) rs112803755 and (C) rs7987144 with the expression of *GJB2* in subcutaneous adipose tissues.

<https://doi.org/10.1371/journal.pgen.1009431.g004>

*TYW1B*. Specifically, increased minor allele dosage at rs799157, which we found to be associated with lower LDL-C levels, is most significantly associated with higher *TYW1B* expression in subcutaneous adipose tissue. *TYW1B* is a tRNA-yW synthesizing protein coding pseudogene involved in wybutosine synthesis, whose characteristics are not well-studied. This evidence suggests biological support for the ChREBP coding variant, while the regulatory variant for *TYW1B* is unlikely to be the causal variant.

rs148931404 is an intron variant of *SLC12A3* which we found to be associated with lower HDL-C levels. This gene has previously been associated with HDL-C in a large multi-ethnic GWAS [6]. *SLC12A3* encodes the sodium-chloride symporter protein. We did not find any plausible underlying biological mechanism for this variant. rs77542162 is a missense variant in *ABCA6* that we found to be associated with LDL-C. This variant has been reported in several GWAS studies in relation to LDL-C levels [6,7] and also in a 2df test joint GEI test with alcohol

consumption [30]. Therefore, it is likely that this SNP is driven by main effects as *ABCA6* is thought to be regulated with macrophage lipid homeostasis [31].

Fish oil supplementation for treatment of hypertriglyceridemia has long been recognized [32]. Recent studies suggest that EPA and DHA have differential effects on HDL-C subfractions, but their overall effects on cardiometabolic lipid risk markers remain unresolved [33], despite dozens of human trials. Nearly all studies to date ignored genetic variants and focused on random cross sections of the population. Our unbiased study identified a variant modulating TAG levels, the only one of the lipid biomarker traits examined that is known to be clearly related to fish oil intake. Further, we identified variants modulating HDL-C and LDL-C, though these effects require further study. Overall, our study found no strong variants that may modulate LDL-C or HDL-C differentially between individuals based on fish oil supplementation status, thus supporting the hypothesis that EPA and DHA effects on these biomarkers are well represented by clinical trials that do not consider interaction with genotype. Our findings emphasize that a one-size-fit-all recommendation of fish oil supplementation to reduce TAG may not be appropriate. While individuals who are heterozygous (AG) at SNP rs112803755 experience a reduction in blood TAG when taking fish oil supplements, homozygotes of AA actually experience an increase. Based on the strong relationship between TAG and cardiovascular diseases, it is natural to hypothesize that the same genetic locus at *GJB2* might interact with n-3 LCPUFAs intakes to have differing effects on the risk of cardiovascular diseases. This is a promising hypothesis calling for direct tests in future studies.

Our study has several strengths and weaknesses. One strength granted by the UK Biobank is a large sample size with two data points taken several years apart for fish oil supplementation. This makes our discovery dataset quite robust and reduces the measurement error of our environmental exposure, which is an important consideration for GEI studies [34]. The ARIC data is less reliable, with only one fish oil data point. A weakness that we recognize is that other dietary quantities of n-3 and n-6 PUFAs are difficult to ascertain, and may interfere with the effects of fish oil. Another limitation of this study is that the ratio of samples in the discovery and replication cohorts is about 10:1. Currently, datasets which provide participant genotype data, fish oil supplementation use, and blood lipid measurements, are rare. Despite the difference in sample size between the UK Biobank and ARIC datasets, each is sufficiently powered to identify significant variants, with the exception of those which are rare or have low effect sizes. Previous gene-diet interaction studies of fish oil have had participants in the hundreds [8], and this is the largest fish oil interaction GWAS to date. One additional weakness is that there may be heterogeneity in the dosage of n-3 LCPUFAs provided by fish oil supplements. These limitations of exact nutrient quantification are present in most nutritional studies which rely on food frequency questionnaires and/or 24-hour recall surveys. Lastly, as in any other association study, ours is associative in nature and could not pinpoint the causal environmental exposure or the genetic variant [35]. We only examine one environmental exposure in this study, fish oil supplementation, which is correlated with many other lifestyle factors [36]. It is possible that other unexamined but correlated environmental factors drive the observed interaction effects, highlighting the need to perform interaction analysis with more environmental factors. However, our novel results make biological sense and many can be placed in a plausible mechanistic context. Finding significant interactions associated with the genes *GJB2* and *MLXIPL*, which have been shown to be regulated by PUFAs, is a validation of our approach.

Our study unravels novel gene-diet interaction effects for four genetic loci, whose effects on blood lipids are modified by fish oil supplementation. Such results lend further support to the practice of precision nutrition to catalyze nutrition science into meaningful and clinically relevant dietary suggestions [37]. Personalizing and optimizing fish oil supplementation recommendations based on a person's unique genetic composition can improve our understanding

of nutrition, and lead to significant improvements in human health and well-being. Once validated, these variants in *GJB2*, *SLC12A3*, *ABCA6*, and *MLXIPL*, will contribute to our understanding of how accounting for genetic differences can allow every person to implement their optimal nutrient intake. Accounting for interaction effects can also help us better understand biological processes leading to disease, and improve the accuracy of future risk prediction models.

## Methods

### Ethics statement

Use of participant data was approved by the University of Georgia Institutional Review Board, UK Biobank (Project ID 48818), and the National Center for Biotechnology Information. Participants of UK Biobank and the Atherosclerosis Risk in Communities Study (ARIC) have signed written consent forms authorizing the use of their medical and genetic data for use in research studies. All methods were performed securely and in accordance with ethical guidelines and regulations.

### Participants

UK Biobank is a prospective cohort study which recruited > 500,000 volunteer participants between 2006 and 2010 in England, Scotland and Wales. Biochemical, clinical, and genotype data were collected. ARIC is a prospective cohort study conducted in four U.S. communities, which began in 1987 and continued to 2007. ARIC participants were randomly selected from pre-defined populations to have medical, social, and demographic data collected. All participants were 40 to 70 years of age at the time of assessment. Participant characteristics can be found in [S1 Table](#).

Participants were quality controlled on the following criteria: genetic ethnicity is Caucasian, used in PCA analysis, not an outlier for heterogeneity and missing genotype rate, no sex chromosome aneuploidy, does not have high degree of genetic kinship (ten or more third-degree relatives identified), and self-reported sex matches genetic sex. Additionally, we removed the minimum number of participants to eliminate all related pairs.

### Phenotypes

All continuous blood lipid measures are reported and analyzed in mmol/L. For stage 1 participants, lipid measures were collected during the UK Biobank Assessment Centres initial assessment from 2006–2010. Blood lipids were analyzed by direct aliquot assays in UK Biobank participants using a Beckman Coulter AU5800. LDL-C was measured by enzymatic protective selection analysis; HDL-C was measured by enzyme immunoinhibition analysis; total cholesterol was measured by CHO-POD analysis; TAGs were measured by GPO-POD analysis.

For ARIC participants, plasma was ultracentrifuged to obtain VLDL-free infranate. LDL-C was precipitated by addition of dextran sulfate and  $Mg^{2+}$  to separate an HDL-C supernate. HDL-C was re-precipitated with dextran sulfate and  $Mg^{2+}$ , and separated by centrifugation. LDL-C levels were calculated using the Friedewald equation. TAGs and total cholesterol were processed and their levels measured by spectrophotometry as described in the ARIC manual for Lipid and Lipoprotein Determinations [38].

LDL-C was adjusted for those who self-reported the use of statins or lipid-lowering drugs as described in [19]; this adjustment was performed in 9,951 UK Biobank participants and 316 ARIC participants. No adjustments were made for other lipids.

### Fish oil supplementation status

Blood LCPUFA levels were not taken in UK Biobank or ARIC cohort studies. Because omega-3 content in dietary intake can vary significantly depending on animal feed quality (e.g. egg laying hens fed an omega-3 rich diet), as well as source (e.g. wild or farmed raised salmon) [39–42], and since neither dietary questionnaire specifies these details, we use fish oil consumption as a minimally confounded contributor to EPA and DHA consumption [43].

Dietary intake data for UK Biobank participants was taken at two time points approximately 3–4 years apart. Participants were asked of their supplement use, including fish oil, in their health and medical history questionnaire at the initial assessment, "Do you regularly take any of the following? (You can select more than one answer)" (f.6179). An online follow-up assessment which included the Oxford WebQ, a digital 24-hour dietary recall questionnaire, was completed by UK Biobank participants on a voluntary basis between 2011–2012 [44,45]. Participants self-reported their use of dietary supplements from the preceding 24 hours (f.20084). Those who answered yes to fish oil supplementation at both time points were coded as 1, those who answered no at both points were coded as 0, and those with different answers were excluded from our analysis (S3 Fig).

ARIC participants indicated their fish oil supplementation status at one time point during their primary assessment. Participants were asked "Do you regularly take fish oil? (Including omega-3 fatty acids, EPA, cod liver oil)." in the "Vitamin Survey Form" at the date of their primary assessment between 1985–2007.

### Covariates

Covariates used in our association analyses were age, sex, body mass index (BMI), weekly servings of oily fish, socioeconomic status measured by Townsend deprivation index, and the first ten genetic principal components. BMI is measured in  $\text{kg}/\text{m}^2$ , and was transformed using ordered quantile normalization for ARIC participants. Weekly servings of oily fish were converted to ordinal variables ranging from 0 (none) to 5 (more than one serving per day). Genetic principal components were provided in the original genotype data of both cohorts.

### Genotype data

The first 50,000 UK Biobank participants of the full study cohort were genotyped using the Affymetrix UK BiLEVE Axiom array, and the remaining 450,000 participants were genotyped using the Affymetrix UK Biobank Axiom array; the two arrays are more than 95% similar in their variant content. Imputation and initial quality control of UK Biobank SNPs were performed by a collaborative group headed by the Wellcome Trust Centre for Human Genetics. We excluded autosomal SNPs with imputation quality score  $< 0.5$ , minor allele frequency (MAF)  $< 1\%$ , missing genotype per individual  $> 5\%$ , missing genotype per variant  $> 2\%$ , or Hardy-Weinberg equilibrium (HWE)  $P < 1 \times 10^{-6}$ . After quality control, a total of 7,954,107 autosomal variants among 73,962 participants were included in the analyses. Our quality control and genotype file format conversions were performed using PLINK2 alpha-v2.3 [46–48].

ARIC participants were genotyped using the Affymetrix GeneChip SNP Array 6.0. Before imputation, quality control removed variants with missing rate  $> 10\%$ , or MAF  $< 1\%$ , and individuals with missing genotype rate  $> 80\%$ . After quality control, genotypes were imputed to the ALL ancestry panel of the 1000 Genome Phase III integrate Release Version 5 [49] using MiniMac software [50]. After imputation, SNPs with  $r^2 < 0.50$ , MAF  $< 1\%$ , or HWE  $P < 1 \times 10^{-6}$  were removed.

### Stage 1 analysis

Stage 1 analysis included up to 73,962 UK Biobank participants and up to 7,954,107 variants after quality control (S1 Table).

Interaction regression was performed for each variant using QuickTest (v1.2) according to the following fixed effects GWAS interaction model:

$$Y = \beta_0 + \beta_G G + \beta_E E + \sum \beta_{C_k} C_k + \beta_{G \times E} G \times E + \varepsilon \quad (1)$$

where  $Y$  is a measure of lipid traits (LDL-C, HDL-C, total cholesterol, and TAGs),  $G$  is the effect variant count (0/1/2),  $E$  is a binary variable representing fish oil supplementation status (0/1),  $C_k$  are covariates, and  $G \times E$  is the GEI term (S4 Fig). Regression coefficients and  $P$ -values were calculated using QuickTest normal mean method for expected genotype dosages; this method is implemented to reduce false positives [51]. Robust Huber sandwich estimates of the variance-covariance matrix were generated.

Main effects adjusted by  $E$  were calculated according to the fixed effects model:

$$Y = \beta_0 + \beta_G G + \beta_E E + \sum \beta_{C_k} C_k + \varepsilon \quad (2)$$

Main additive variant effects, and variant effects stratified by ( $E$ ) were also calculated using the generalized fixed effects model:

$$Y = \beta_0 + \beta_G G + \sum \beta_{C_k} C_k + \varepsilon \quad (3)$$

These main effects models were performed using the same QuickTest normal mean method.

Joint  $P$ -values of main and interaction effects ( $\beta_G$  and  $\beta_{G \times E}$ ) were calculated according to a 2df  $\chi^2$  distribution which corrected for the determinant of the covariance matrix between these two terms [18]. Genomic control was applied to Stage 1 2df joint  $P$ -values for each lipid phenotype. Variants reaching  $P < 1 \times 10^{-6}$  in either the 1df interaction test or 2df joint test were advanced to replication in Stage 2.

### Stage 2 analysis

Stage 2 analysis included up to 7,284 ARIC participants, and 48,608,505 variants (S1 Table). Participants were filtered on the basis of their ethnicity (white) only. Additional quality control on samples and genomic data (as in Stage 1) was not conducted, because these filters are meant to reduce the rate of false positives, which was not relevant for Stage 2 replication. Regression coefficients and  $P$ -values were calculated using QuickTest normal mean method. Variants advanced from Stage 1 which also had a  $P < 0.05$  in the 1df interaction term in the ARIC cohort were advanced to joint meta-analysis between the two cohorts in Stage 1+2.

### Meta-analysis of stage 1+2

METAL meta-analysis software (2010-02-08) [52] was used to perform a meta-analysis of those associations with  $P < 1 \times 10^{-6}$  in 1df interaction and/or 2df joint tests in Stage 1, and  $P < 0.05$  in 1df interaction test in Stage 2 (patch provided by A. Manning to enable 2df GEI testing [18]; genome.sph.umich.edu/wiki/Meta\_Analysis\_of\_SNPxEnvironment\_Interaction). Stage 1+2 meta-analyses were performed using a weighted  $z$ -statistic by sample size [52]. Genomic control was applied to all meta-analyses as implemented by METAL. Associations exceeding the genome-wide significance threshold of  $P < 5 \times 10^{-8}$  were passed to FUMA to identify the lead SNP for each locus.

### Identifying lead SNPs

Variants exceeding the genome-wide significance threshold of  $P < 5 \times 10^{-8}$  were inputted to FUMA to identify independent loci and their lead SNPs [53]. Lead SNPs are defined as the SNP within a locus having the lowest  $P$ -value. UK Biobank release 2b 10k White British was used as the reference panel population. The maximum  $P$ -value cutoff was set to 0.05, and a first threshold of  $r^2 \geq 0.6$  and second threshold of  $r^2 \geq 0.1$  were used to define independent significant SNPs. The maximum distance between LD blocks to merge into a locus was  $< 1$ Mb.

### Identifying novel variants

For replicated and non-replicated variants with joint meta-analysis  $P < 5 \times 10^{-8}$ , GWAS Catalog [54] was used to identify novel variants. Gene-fish-oil interaction variants were checked in a literature search for their novelty. Variants within 1Mb from previously published variants associated with the same trait were considered to be non-novel.

### Additional analyses

The R package qqman v 0.1.4 was used to generate Manhattan plots and QQ plots [55]. Regional loci plots were made using LocusZoom [56]. Data analysis was conducted in R v3.6.1 [57]. The Genotype-Tissue Expression Project (GTEx) data used were obtained from the GTEx Portal on 04/29/20 [58].

### Supporting information

**S1 Fig. Manhattan plots for Stage 1 1df interaction term  $P$ -values and 2df joint test  $P$ -values for lipid traits.** Plots show post-genomic control values.

(TIF)

**S2 Fig. QQ plots for Stage 1 1df interaction term  $P$ -values and 2df joint test  $P$ -values for lipid traits.** Plots show post-genomic control values.

(TIF)

**S3 Fig. Fish oil supplementation taken at two time points.** The number of UK Biobank participants who responded yes/yes, no/no, yes/no, and no/yes to the two dietary assessment time points at the initial assessment and in the 24-hour follow-up questionnaire are shown. Numbers reflect the total number of participants who answered in both assessments, but not the number of participants used in this study after quality control.

(TIF)

**S4 Fig. Visualization of the  $G \times E$  interaction regression model.**  $Y = \beta_0 + \beta_G G + \beta_E E + \sum \beta_k C_k + \beta_{G \times E} G \times E + \epsilon$ , where  $Y$  = phenotype,  $G$  = minor variant dosage (0/1/2 coding),  $E$  = environmental exposure,  $C_k$  = covariates, and  $G \times E$  = interaction term. In this study,  $Y$  is a continuous lipid trait, and  $E$  is a binary variable representing the presence or absence of self-reported dietary fish oil supplementation.

(TIF)

**S1 Table. Participant characteristics.** Participant characteristics, by blood lipid phenotype, for those included in GEI analyses for Stage 1 (UK Biobank) and Stage 2 (ARIC). Mean and standard deviation values are shown for blood lipid phenotypes and for applicable covariates.

(XLSX)

**S2 Table. Numbers of stage 1 significant variants.** Variants which passed a significance threshold of  $P < 1e-06$  in Stage 1 (UK Biobank) are counted here. Significance was assessed for both 1df interaction terms and 2df joint terms. Variant count and number of independent loci are shown, as well as unique variants with 1df and 2df tests.  
(XLSX)

**S3 Table. Numbers of replicated variants.** Variants which reached Stage 1  $P < 1e-06$  (in either 1df or 2df) and were found to have 1df  $P < 0.05$  in Stage 2 interaction models.  
(XLSX)

**S4 Table. Numbers of genome-wide significance loci in only Stage 1.** Counts of variants and loci which met the significance threshold of  $P < 5e-08$  in Stage 1 (in either 1df or 2df) but which were not replicated in Stage 2. Note that no Stage 1 1df  $P$ -values reached this threshold so all variants in this table refer to their 2df joint test  $P$ -values.  
(XLSX)

**S5 Table. Non-replicated genome-wide significant Stage 1 variants.** Full details for the loci counted in Table S4. Effect, beta coefficient of the minor allele dose term ( $\beta_G$  in Eq (1)); MAF, minor allele frequency; SE, standard error; Int effect, beta coefficient of the interaction term ( $\beta_{G \times E}$  in Eq (1)). Lipid traits were measured in mmol/L. All  $P$ -values are calculated using Stage 1 (UK Biobank) participants only.  
(XLSX)

**S6 Table. Numbers of genome-wide significance loci after meta-analyses.** Counts of replicated results reaching genome-wide significance ( $P < 5e-08$ ) in Stage 1+2 meta-analyses. Significant variants determined by 1df  $P$ -values (top) and 2df  $P$ -values (bottom).  
(XLSX)

**S7 Table. Data used in Fig 3B.** Fish oil status, number of G alleles at rs112803755, mean triglycerides, sample size, standard deviation of triglycerides, and 95% confidence interval for combined participants from Stage 1 and Stage 2.  
(XLSX)

## Acknowledgments

The authors would like to thank all UK Biobank participants and administrators for data access. We also thank all Ye lab members for helpful discussions.

## Author Contributions

**Conceptualization:** Michael Francis, Kaixiong Ye.

**Data curation:** Michael Francis, Kaixiong Ye.

**Formal analysis:** Michael Francis, Kaixiong Ye.

**Funding acquisition:** Kaixiong Ye.

**Investigation:** Michael Francis, Yitang Sun, Jingqi Zhou, Xiang Li, J. Thomas Brenna.

**Methodology:** Michael Francis, Changwei Li, Kaixiong Ye.

**Project administration:** Michael Francis, Kaixiong Ye.

**Resources:** Kaixiong Ye.

**Software:** Michael Francis, Kaixiong Ye.



**Supervision:** Kaixiong Ye.

**Validation:** Michael Francis, Yitang Sun, Kaixiong Ye.

**Visualization:** Michael Francis, Kaixiong Ye.

**Writing – original draft:** Michael Francis, Kaixiong Ye.

**Writing – review & editing:** Michael Francis, Changwei Li, Yitang Sun, Jingqi Zhou, Xiang Li, J. Thomas Brenna, Kaixiong Ye.

## References

1. Thaipitakwong T, Aramwit P. A Review of the Efficacy, Safety, and Clinical Implications of Naturally Derived Dietary Supplements for Dyslipidemia. *Am J Cardiovasc Drug*. 2017; 17(1):27–35. <https://doi.org/10.1007/s40256-016-0191-2> PMID: 27637494
2. Scicchitano P, Cameli M, Maiello M, Modesti PA, Muesan ML, Novo S, et al. Nutraceuticals and dyslipidaemia: Beyond the common therapeutics. *J Funct Food*. 2014; 6:11–32. <https://doi.org/10.1016/j.jff.2013.12.006>.
3. Eslick GD, Howe PRC, Smith C, Priest R, Bensoussan A. Benefits of fish oil supplementation in hyperlipidemia: a systematic review and meta-analysis. *Int J Cardiol*. 2009; 136(1):4–16. <https://doi.org/10.1016/j.ijcard.2008.03.092> PMID: 18774613
4. Lombardo YB, Chicco AG. Effects of dietary polyunsaturated n-3 fatty acids on dyslipidemia and insulin resistance in rodents and humans. A review. *J Nutr Biochem*. 2006; 17(1):1–13. <https://doi.org/10.1016/j.jnutbio.2005.08.002> PMID: 16214332
5. Goldberg RB, Sabharwal AK. Fish oil in the treatment of dyslipidemia. *Curr Opin Endocrinol, Diabetes and Obesity*. 2008; 15(2):167–74. <https://doi.org/10.1097/MED.0b013e3282f76728> PMID: 18316953
6. Hoffmann TJ, Theusch E, Haldar T, Ranatunga DK, Jorgenson E, Medina MW, et al. A large electronic-health-record-based genome-wide study of serum lipids. *Nat Genet*. 2018; 50(3):401–13. Epub 2018/03/05. <https://doi.org/10.1038/s41588-018-0064-5> PMID: 29507422
7. Klarin D, Damrauer SM, Cho K, Sun YV, Teslovich TM, Honerlaw J, et al. Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. *Nat Genet*. 2018; 50(11):1514–23. Epub 2018/10/03. <https://doi.org/10.1038/s41588-018-0222-9> PMID: 30275531
8. Madden J, Williams CM, Calder PC, Lietz G, Miles EA, Cordell H, et al. The Impact of Common Gene Variants on the Response of Biomarkers of Cardiovascular Disease (CVD) Risk to Increased Fish Oil Fatty Acids Intakes. *Annu Rev Nutr*. 2011; 31(1):203–34. <https://doi.org/10.1146/annurev-nutr-010411-095239> PMID: 21568708
9. Klarin D, Damrauer SM, Cho K, Sun YV, Teslovich TM, Honerlaw J, et al. Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. *Nat Genet*. 2018; 50(11):1514–23. Epub 2018/10/01. <https://doi.org/10.1038/s41588-018-0222-9> PMID: 30275531
10. Aung T, Halsey J, Kromhout D, Gerstein HC, Marchioli R, Tavazzi L, et al. Associations of Omega-3 Fatty Acid Supplement Use With Cardiovascular Disease Risks: Meta-analysis of 10 Trials Involving 77 917 Individuals Meta-analysis of Associations of Omega-3 Fatty Acids and Cardiovascular Risk Meta-analysis of Associations of Omega-3 Fatty Acids and Cardiovascular Risk. *JAMA Cardiol*. 2018; 3(3):225–33.
11. Zheng J, Huang T, Yu Y, Hu X, Yang B, Li D. Fish consumption and CHD mortality: an updated meta-analysis of seventeen cohort studies. *Public Health Nutr*. 2012; 15(4):725–37. Epub 2011/09/15. <https://doi.org/10.1017/S1368980011002254> PMID: 21914258
12. Martinelli N, Girelli D, Malerba G, Guarini P, Illig T, Trabetti E, et al. FADS genotypes and desaturase activity estimated by the ratio of arachidonic acid to linoleic acid are associated with inflammation and coronary artery disease. *Am J Clin Nutr*. 2008; 88(4):941–9. Epub 2008/10/10. <https://doi.org/10.1093/ajcn/88.4.941> PMID: 18842780
13. Bokor S, Dumont J, Spinneker A, Gonzalez-Gross M, Nova E, Widhalm K, et al. Single nucleotide polymorphisms in the FADS gene cluster are associated with delta-5 and delta-6 desaturase activities estimated by serum fatty acid ratios. *J Lipid Res*. 2010; 51(8):2325–33. Epub 2010/04/30. <https://doi.org/10.1194/jlr.M006205> PMID: 20427696
14. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018; 562(7726):203–9. <https://doi.org/10.1038/s41586-018-0579-z> PMID: 30305743

15. Bush WS, Moore JH. Chapter 11: Genome-wide association studies. *PLOS Comput Biol*. 2012; 8(12): e1002822. <https://doi.org/10.1371/journal.pcbi.1002822> PMID: 23300413
16. Ye K, Gao F, Wang D, Bar-Yosef O, Keinan A. Dietary adaptation of FADS genes in Europe varied across time and geography. *Nat Ecol Evol*. 2017; 1(7):0167. <https://doi.org/10.1038/s41559-017-0167> PMID: 29094686
17. Kothapalli KSD, Ye K, Gadgil MS, Carlson SE, O'Brien KO, Zhang JY, et al. Positive Selection on a Regulatory Insertion-Deletion Polymorphism in FADS2 Influences Apparent Endogenous Synthesis of Arachidonic Acid. *Mol Biol Evol*. 2016; 33(7):1726–39. Epub 2016/03/29. <https://doi.org/10.1093/molbev/msw049> PMID: 27188529
18. Manning AK, LaValley M, Liu C-T, Rice K, An P, Liu Y, et al. Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP  $\times$  environment regression coefficients. *Genet Epidemiol*. 2011; 35(1):11–8. <https://doi.org/10.1002/gepi.20546> PMID: 21181894
19. Rao DC, Sung YJ, Winkler TW, Schwander K, Borecki I, Cupples LA, et al. Multiancestry Study of Gene-Lifestyle Interactions for Cardiovascular Traits in 610 475 Individuals From 124 Cohorts: Design and Rationale. *Circ Cardiovasc Genet*. 2017; 10(3). Epub 2017/06/18. <https://doi.org/10.1161/CIRCGENETICS.116.001649> PMID: 28620071
20. Psaty BM, O'Donnell CJ, Gudnason V, Lunetta KL, Folsom AR, Rotter JI, et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet*. 2009; 2(1):73–80. Epub 2009/12/25. <https://doi.org/10.1161/CIRCGENETICS.108.829747> PMID: 20031568
21. Sung YJ, de las Fuentes L, Winkler TW, Chasman DI, Bentley AR, Kraja AT, et al. A multi-ancestry genome-wide study incorporating gene-smoking interactions identifies multiple new loci for pulse pressure and mean arterial pressure. *Hum Mol Genet*. 2019; 28(15):2615–33. <https://doi.org/10.1093/hmg/ddz070> PMID: 31127295
22. Noordam R, Bos MM, Wang H, Winkler TW, Bentley AR, Kilpeläinen TO, et al. Multi-ancestry sleep-by-SNP interaction analysis in 126,929 individuals reveals lipid loci stratified by sleep duration. *Nat Comm*. 2019; 10:5121.
23. Figueroa V, Saez PJ, Salas JD, Salas D, Jara O, Martinez AD, et al. Linoleic acid induces opening of connexin26 hemichannels through a PI3K/Akt/Ca(2+)-dependent pathway. *Biochim Biophys Acta*. 2013; 1828(3):1169–79. Epub 2012/12/25. <https://doi.org/10.1016/j.bbame.2012.12.006> PMID: 23261389
24. Dlugosova K, Weismann P, Bernatova I, Sotnikova R, Slezak J, Okruhlicova L. Omega-3 fatty acids and atorvastatin affect connexin 43 expression in the aorta of hereditary hypertriglyceridemic rats. *Can J Physiol Pharmacol*. 2009; 87(12):1074–82. Epub 2009/12/24. <https://doi.org/10.1139/Y09-104> PMID: 20029544
25. Brisset AC, Isakson BE, Kwak BR. Connexins in vascular physiology and pathology. *Antioxid Redox Signal*. 2009; 11(2):267–82. <https://doi.org/10.1089/ars.2008.2115> PMID: 18834327
26. Kooner JS, Chambers JC, Aguilar-Salinas CA, Hinds DA, Hyde CL, Warnes GR, et al. Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides. *Nat Genet*. 2008; 40(2):149–51. <https://doi.org/10.1038/ng.2007.61> PMID: 18193046
27. Zeng X-N, Yin R-X, Huang P, Huang K-K, Wu J, Guo T, et al. Association of the MLXIPL/TBL2 rs17145738 SNP and serum lipid levels in the Guangxi Mulao and Han populations. *Lipids Health Dis*. 2013; 12(1):156. <https://doi.org/10.1186/1476-511X-12-156> PMID: 24160749
28. Iizuka K. The transcription factor carbohydrate-response element-binding protein (ChREBP): A possible link between metabolic disease and cancer. *BBA-Mol Basis Dis*. 2017; 1863(2):474–85. <https://doi.org/10.1016/j.bbadis.2016.11.029> PMID: 27919710
29. Jump DB, Tripathy S, Depner CM. Fatty acid-regulated transcription factors in the liver. *Annu Rev Nutr*. 2013; 33:249–69. Epub 2013/03/22. <https://doi.org/10.1146/annurev-nutr-071812-161139> PMID: 23528177
30. de Vries PS, Brown MR, Bentley AR, Sung YJ, Winkler TW, Ntalla I, et al. Multiancestry Genome-Wide Association Study of Lipid Levels Incorporating Gene-Alcohol Interactions. *Am J Epidemiol*. 2019; 188(6):1033–54. <https://doi.org/10.1093/aje/kwz005> PMID: 30698716
31. Kaminski WE, Wenzel JJ, Piehler A, Langmann T, Schmitz G. ABCA6, a novel subclass ABC transporter. *Biochem Biophys Res Commun*. 2001; 285(5):1295–301. Epub 2001/08/02. <https://doi.org/10.1006/bbrc.2001.5326> PMID: 11478798
32. Harris WS. Fish oils and plasma lipid and lipoprotein metabolism in humans: a critical review. *J Lipid Res*. 1989; 30(6):785–807. Epub 1989/06/01. PMID: 2677200
33. Innes JK, Calder PC. The Differential Effects of Eicosapentaenoic Acid and Docosahexaenoic Acid on Cardiometabolic Risk Factors: A Systematic Review. *Int J Mol Sci*. 2018; 19(2). Epub 2018/02/10. <https://doi.org/10.3390/ijms19020532> PMID: 29425187

34. Greenwood DC, Gilthorpe MS, Cade JE. The impact of imprecisely measured covariates on estimating gene-environment interactions. *BMC Med Res Methodol*. 2006 4; 6:21. <https://doi.org/10.1186/1471-2288-6-21> PMID: 16674808
35. Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet*. 2018; 19(8):491–504. <https://doi.org/10.1038/s41576-018-0016-z> PMID: 29844615
36. Cundiff DK, Lanou AJ, Nigg CR. Relation of omega-3 Fatty Acid intake to other dietary factors known to reduce coronary heart disease risk. *Am J Cardiol*. 2007 1; 99(9):1230–3. <https://doi.org/10.1016/j.amjcard.2006.12.032> PMID: 17478148
37. Rodgers GP, Collins FS. Precision Nutrition—the Answer to “What to Eat to Stay Healthy”. *JAMA*. 2020; 324(8):735–736. <https://doi.org/10.1001/jama.2020.13601> PMID: 32766768
38. The National Heart, Lung, and Blood Institute (NHLBI) of the National Institutes of Health. Manual 8: Lipid and Lipoprotein Determinations. ARIC Protocol. 1987.
39. Cladis DP, Kleiner AC, Freiser HH, Santerre CR. Fatty Acid Profiles of Commercially Available Finfish Fillets in the United States. *Lipids*. 2014; 49(10):1005–18. <https://doi.org/10.1007/s11745-014-3932-5> PMID: 25108414
40. Bostock J, McAndrew B, Richards R, Jauncey K, Telfer T, Lorenzen K, et al. Aquaculture: global status and trends. *Philos T R Soc B*. 2010; 365(1554):2897–912. <https://doi.org/10.1098/rstb.2010.0170> PMID: 20713392
41. Henriques J, Dick JR, Tocher DR, Bell JG. Nutritional quality of salmon products available from major retailers in the UK: content and composition of n-3 long-chain PUFA. *Brit J Nutr*. 2014; 112(6):964–75. Epub 2014/07/14. <https://doi.org/10.1017/S0007114514001603> PMID: 25017007
42. Sprague M, Dick JR, Tocher DR. Impact of sustainable feeds on omega-3 long-chain fatty acid levels in farmed Atlantic salmon, 2006–2015. *Sci Rep-UK*. 2016; 6(1):21892. <https://doi.org/10.1038/srep21892> PMID: 26899924
43. Tur JA, Bibiloni MM, Sureda A, Pons A. Dietary sources of omega 3 fatty acids: public health risks and benefits. *Brit J Nutr*. 2012; 107 Suppl 2:S23–52. Epub 2012/05/25. <https://doi.org/10.1017/S0007114512001456> PMID: 22591897
44. Bradbury KE, Young HJ, Guo W, Key TJ. Dietary assessment in UK Biobank: an evaluation of the performance of the touchscreen dietary questionnaire. *J Nutr Sci*. 2018; 7:e6–e. <https://doi.org/10.1017/jns.2017.66> PMID: 29430297
45. Liu B, Young H, Crowe FL, Benson VS, Spencer EA, Key TJ, et al. Development and evaluation of the Oxford WebQ, a low-cost, web-based method for assessment of previous 24 h dietary intakes in large-scale prospective studies. *Public Health Nutr*. 2011; 14(11):1998–2005. Epub 2011/07/07. <https://doi.org/10.1017/S1368980011000942> PMID: 21729481
46. Purcell S, Chang CC. PLINK 1.9-beta3. [www.cog-genomics.org/plink/1.9/](http://www.cog-genomics.org/plink/1.9/).
47. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015; 4(1).
48. Purcell S, Chang CC. PLINK 2.3 alpha. 2020. [www.cog-genomics.org/plink/2.0/](http://www.cog-genomics.org/plink/2.0/).
49. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. *Nature*. 2015; 526(7571):68–74. <https://doi.org/10.1038/nature15393> PMID: 26432245
50. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016; 48(10):1284–7. Epub 2016/08/29. <https://doi.org/10.1038/ng.3656> PMID: 27571263
51. Kutalik Z, Johnson T, Bochud M, Mooser V, Vollenweider P, Waeber G, et al. Methods for testing association between uncertain genotypes and quantitative traits. *Biostatistics*. 2010; 12(1):1–17. <https://doi.org/10.1093/biostatistics/kxq039> PMID: 20543033
52. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010; 26(17):2190–1. <https://doi.org/10.1093/bioinformatics/btq340> PMID: 20616382
53. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017; 8(1):1826. Epub 2017/12/01. <https://doi.org/10.1038/s41467-017-01261-5> PMID: 29184056
54. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. 2019; 47(D1):D1005–d12. Epub 2018/11/18. <https://doi.org/10.1093/nar/gky1120> PMID: 30445434

55. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv*. 2014:005165. <https://doi.org/10.1101/005165>
56. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010; 26(18):2336–7. Epub 2010/07/17. <https://doi.org/10.1093/bioinformatics/btq419> PMID: 20634204
57. Team R Consortium. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2019.
58. Battle A, Brown CD, Engelhardt BE, Montgomery SB. Genetic effects on gene expression across human tissues. *Nature*. 2017; 550(7675):204–13. Epub 2017/10/13. <https://doi.org/10.1038/nature24277> PMID: 29022597