medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Fifty-one novel, replicated loci identified in genome-wide association study of polyunsaturated and monounsaturated fatty acids in 124,024 European individuals

Michael Francis[1], Yitang Sun[2], Huifang Xu[2], J. Thomas Brenna[3,4], Kaixiong Ye[1,2*]

[1] Institute of Bioinformatics, University of Georgia, Athens, Georgia, US

[2] Department of Genetics, University of Georgia, Athens, Georgia, US

[3] Division of Nutritional Sciences, Cornell University, Ithaca, NY, US

[4] Dell Pediatric Research Institute and the Depts of Pediatrics, of Nutrition, and of Chemistry University of Texas at Austin, Austin, TX, US

*Corresponding Author

Email: kaixiong.ye@uga.edu

Abstract

Circulating polyunsaturated and monounsaturated fatty acid (PUFA and MUFA) levels, whose imbalances co-occur with human metabolic diseases, have strong heritable components. We performed the largest genome-wide association study (GWAS) to-date on fourteen PUFA and MUFA phenotypes, measured by nuclear magnetic resonance in plasma. We identified 612 significant loci-phenotype associations (115 unique loci; $P < 1.678*10^{-8}$) in a European cohort from UK Biobank (UKB-EUR; n=101,729). Replication of five phenotypes (omega-3, omega-6, DHA, LA, MUFAs) was conducted in two external European studies: FinMetSeq (n=8,751) and a meta-analysis by Kettunnen *et al.* (n=3,644-13,544). Meta-analysis of these three studies yielded 254 significant loci-phenotype associations (109 unique loci; $P < 2.439*10^{-8}$); we identified 87 novel loci, 51 of which were replicated. A transcriptome-wide association study of the UKB-EUR cohort revealed an additional twelve novel loci. This study improves our understanding of the genetic architecture of unsaturated fatty acids, and can inform future gene-based dietary interventions.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Acknowledgements

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institute of Health under award numbers T32GM007103 (MF) and R35GM143060 (KY). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Special thanks to the UGA GACRC staff for facilitating our data analyses.

Competing interests

We have no conflicts of interest to disclose.

Author contributions

KY conceived and supervised the study. KY and MF designed the analyses. MF led the data analyses, with assistance from YS and HX. MF, KY, and JTB. interpreted the results. MF and KY wrote the manuscript. MF created data visualizations. All authors reviewed, revised, and approved the final paper.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Introduction

Polyunsaturated fatty acids (PUFAs) are dietary fats containing two to six double bonds along linear carbon chains from 14 to 22 carbons in length. Imbalance of tissue PUFAs are involved in the pathophysiology of a broad array of diseases, including cardiovascular disease, cancer, depression, and dementia [1-3]. Omega-3 long-chain PUFAs (n-3 LCPUFAs) have been consistently shown to improve aspects of metabolic syndrome that are related to cardiovascular disease and obesity-related risk factors such as insulin resistance, hypertension, and dyslipidemia [1, 4, 5], while omega-6 (n-6) LCPUFAs have been associated with both positive and negative health outcomes [6]. Excess n-6 linoleic acid (LA) suppresses tissue and circulating n-3 LCPUFAs, due to common enzymes operating on both PUFA families; balance in dietary n-6 and n-3 is necessary to avoid suppression of the functional n-3 LCPUFAs, EPA and DHA [7]. Modest overall dietary PUFAs and an n-6/n-3 ratio up to 4/1 have been recommended, while the typical industrialized diet has a ratio of 15/1 or more [8, 9].

Dietary PUFA intake strongly influences circulating fatty acid levels [10]. Heritability analyses in twin studies and large cohorts have indicated that substantial genetic components contribute to determining circulating PUFA levels [11-13]. Genome-wide association studies (GWASs) have identified 37 unique genomic loci related to PUFAs and monounsaturated fatty acids (MUFAs; Supplementary Table 1). However, they collectively only explain a small fraction of the phenotypic variance [12, 14], suggesting more loci may be found in large-sample GWAS. High-throughput nuclear magnetic resonance spectroscopy (NMR) enables rapid large-scale quantification of metabolic biomarkers [15]. The UK Biobank (UKB) has recently released NMR data of plasma fatty acid levels for over 110,000 participants, presenting an opportunity to identify novel genetic loci for PUFA levels.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Here, we perform a linear mixed model (LMM) GWAS to identify the genetic variants associated with the fourteen available NMR-derived plasma unsaturated fatty acid phenotypes in UKB participants: n-3 LCPUFAs, n-6 LCPUFAs, docosahexaenoic acid (DHA), linoleic acid (LA), total PUFAs, total MUFAs, the percentages of each of these fatty acid groups per the total amount of fatty acids, as well as the ratios of n-6/n-3 and PUFA/MUFA. (These PUFA- and MUFA-related phenotypes are collectively referred to as "PUFA traits" for simplicity's sake.) Our discovery cohort consists of UKB participants with NMR PUFA trait data, who were determined to be genetically European (EUR) by the Pan-UK Biobank project (n=101,729) [16]. Additional multi-ancestry analyses were performed in Central and South Asian (CSA), East Asian (EAS), and African (AFR) UKB participants (n=4,400). Two EUR studies were used in our replication and meta-analysis of the UKB-EUR cohort: the Finnish Metabolic Sequencing (FinMetSeq) study [11], which itself consists of exome sequencing derived from two cohorts, FINRISK [17] and Metabolic Syndrome in Men (METSIM) [18], and a meta-analysis of 14 genotyped datasets derived from ten EUR studies by Kettunnen *et al*. [14] (n=8,751 and 3,644-13,544, respectively).

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Methods

Ethics

Participant data use was approved by UK Biobank (UKB; Project ID 48818). UKB participants have consented to and authorized the use of their medical and genetic data for use in research studies. This research was performed on a University of Georgia (UGA) data server using strict data protection protocols and two-factor authentication. The UGA Institutional Review Board (IRB) approved the use of human subject data in this study. Additional datasets use publicly available summary statistics from previous GWAS; approval was not required.

Participants

The full UKB cohort consists of > 500,000 volunteer participants between ages 40 and 70 that were recruited between 2006 and 2010 in England, Scotland, and Wales. Approximately 120,000 participants had metabolic traits measured using NMR between June 2019 and April 2020, from plasma samples taken at recruitment. Participants were removed on the following criteria: withdrawn consent, mismatches between self-reported and genetic sex, poor quality genotyping as flagged by UKB, sex chromosome aneuploidy, or poor-quality NMR measurement flagged by UKB. After quality control (QC) and stratification by ancestry using Pan UKBB designations [16], counts of participants included in our analyses were: 101,729 European (EUR); 1,564 African (AFR); 2,203 Central South Asian (CSA); and 633 East Asian (EAS). UKB participant characteristics can be found in Supplementary Table 2.

For replication and meta-analysis, we included two external EUR studies. First, the Finnish Metabolic Sequencing (FinMetSeq) study [11], which consists of a combination of FINRISK and METSIM cohorts. METSIM participants were 10,197 men from Kuopio, Eastern Finland, aged 45 to 73 years during initial examinations from 2005 to 2010. FINRISK

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

participants were recruited every five years from 1972 to 2012, and consisted of random population samples of men and women aged 30-59 years; FinMetSeq uses 10,192 participants from 1992-2007 FINRISK surveys who had a residence in northeastern Finland. Of the approximately 19,000 participants in FinMetSeq, 8,751 had NMR metabolomics data available for the fatty acid phenotypes of interest. Pregnant women, type 1 and 2 diabetics, and those fasting less than eight hours were excluded from this cohort.

We also utilized a meta-analysis consisting of fourteen genotyped datasets derived from ten EUR studies performed by Kettunnen *et al.* in 2016 [14] in our replication and meta-analysis. The number of participants contributed by the Kettunnen *et al.* summary statistics in this analysis ranges from 3,644-13,544 participants (from six to ten studies), depending on the variant; this participant range is consistent across the five traits meta-analyzed. There is an overlap of 225 participants between the FinMetSeq and Kettunnen *et al.* cohorts, which we determined would not affect type I error in a meaningful way [19].

Fatty acid phenotypes

UKB EDTA (ethylenediaminetetraacetic acid) plasma samples were taken at the baseline recruitment timepoint and measured between June 2019 and April 2020 by the metabolic biomarker profiling platform of Nightingale Health Ltd., as described previously [20, 21]. We analyzed fourteen quantitative PUFA and MUFA phenotypes in UKB cohorts, specifically, n-3 LCPUFAs, n-6 LCPUFAs, docosahexaenoic acid (DHA), linoleic acid (LA), total PUFAs, and total monounsaturated fatty acids (MUFAs), all reported in mmol/L, the percentages of n-3 LCPUFAs, n-6 LCPUFAs, DHA, LA, MUFAs, and PUFAs out of the total amount of fatty acids, as well as n-6/n-3 ratio, and PUFA/MUFA ratio. (These PUFA- and MUFA-related phenotypes are collectively referred to as “PUFA traits” for simplicity’s sake.) We regressed all

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

PUFA traits on selected covariates in each model (described below) and applied rank-based inverse normal transformation to the residuals for use in all analyses [22]; this was consistent with transformations performed in our external GWAS replication studies.

Genotype data

Genotype data was initially QCed and imputed with Haplotype Reference Consortium (HRC) and 1000 Genomes variants by UKB (v3) as previously described [23]. For discovery analyses, we excluded variants with imputation quality score < 0.3, minor allele frequency (MAF) < 0.1%, missing genotype per individual > 5%, missing genotype per variant > 2%, or Hardy-Weinberg equilibrium (HWE) $P < 1\times10^{-8}$. After quality control, a total of 15,587,898 variants among 101,729 participants were included in the UKB-EUR discovery cohort. QC and genotype file format conversions were performed using PLINK2 alpha-v2.3 [24, 25]. All genomic positions in this study refer to the Genome Reference Consortium Human Build 37 (GRCh37), also known as hg19.

A pruned set of variants

A pruned set of variants were inputted as PLINK-format genotypes to BOLT-LMM for model-fitting in the UKB-EUR discovery GWAS [26]. After filtering for only participants included in this analysis, the exclusion criteria for variants in the pruned set were imputation quality score < 0.8, MAF < 1%, missing genotype per variant > 1%, or Hardy-Weinberg equilibrium (HWE) $P < 1\times10^{-8}$. A hard-call threshold of 0.1 was applied to the remaining variants. The lactase locus on chromosome 2, the major histocompatibility complex (MHC) on chromosome 6, and inversions on chromosomes 8 and 17 were excluded. Linkage disequilibrium (LD) pruning was performed at $r^2 = 0.2$, (plink2 --indep-pairwise 50 5 0.2). After pruning, 821,405 variants remained.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Discovery model selection

Two sensitivity models were used in this analysis. Model 1 (M1) included the covariates sex, age, age^2, genotyping array, and assessment center. Model 2 (M2) included the M1 covariates, plus body mass index (BMI), lipid medication usage, and socioeconomic status as measured by Townsend deprivation index (Supplementary Table 2). Additionally, the first twenty principal components for study participants were included as covariates in both models as calculated by PLINK2 (randomized algorithm) [24, 25]. For meta-analysis cohorts, FinMetSeq summary statistics as provided were adjusted by age, age^2, sex, cohort year, BMI, sex hormones, and lipid medications. Kettunnen *et al.* adjusted for sex, age and ten genetic principal components in their analyses. We compared our summary statistics from M1 and M2 with the publicly available results from OpenGWAS (batch met-d) [27], as well as to each other, using Spearman's rank correlation.

Identification of significant GWAS signals

BOLT-LMM v2.3.6 [26] was used to perform linear mixed-effects model association analyses on fourteen PUFA traits in the UKB-EUR discovery cohort. The provided 1000G European LD scores [28] were used to calibrate the BOLT-LMM statistic. Covariates and pruned variant sets were included as described above. Non-infinitesimal BOLT-LMM *P*-values (P_BOLT_LMM) were used in all reporting and downstream analyses. Because these PUFA traits were highly correlated, we calculated the effective number of traits to use for Bonferroni multiple testing correction. Eigenvalues (λ) for the fourteen PUFA traits were used to calculate the number of effective traits as: $\left(\sum_{k=1}^{14}\lambda_k\right)^2 \Big/ \sum_{k=1}^{14}\lambda_k^2$ = 2.98 [29]. The threshold of $P < (5*10^{-8}/2.98) = 1.678*10^{-8}$ was used to designate significant variant-trait associations.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Transcriptome-wide association analysis

Transcriptome-wide association analysis (TWAS) was performed on UKB-EUR discovery cohort summary statistics using S-PrediXcan [30]. This method used Genotype-Tissue Expression (GTEx) v8 [31] expression quantitative trait loci (eQTL) data across 49 available tissues. Summary statistics were harmonized and imputed to GTEx models. A total of 601,176 gene-tissue pairs were analyzed across fourteen PUFA traits, using a Bonferroni corrected significance threshold of $P < 2.791*10^{-8}$ (0.05/(601,176*2.98)). S-MultiXcan was used to integrate tissue-level associations and improve association detection. The cutoff condition number of eigenvalues was set to 30 for truncating singular-value decomposition components. S-MultiXcan was run across 21,846 genes, using a Bonferroni corrected significance threshold of $P < 7.68 *10^{-7}$ (0.05/(21,846*2.98)).

Replication studies

Variant-trait associations from the discovery UKB-EUR cohort for five traits were sent to replication and meta-analysis steps: omega-3 fatty acids (n-3 LCPUFAs), omega-6 fatty acids (n-6 LCPUFAs), docosahexaenoic acid (DHA), linoleic acid (LA), and total monounsaturated fatty acids (MUFAs). Summary statistics were obtained from the publicly available Finnish Metabolic Sequencing (FinMetSeq) study [11] and the meta-analysis of 10 studies by Kettunen *et al.* [14]. Discovery cohort variants were considered replicated in these external cohorts on a per-variant basis at $P < 0.05$.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Meta-analysis

Meta-analyses were performed using the METAL [32] software using the STDERR scheme, which weights effect size estimates using the inverse of the corresponding standard errors. The meta-analysis of this study consists of the three EUR cohorts: UKB-EUR+ FinMetSeq+ Kettunen *et al.* datasets (N=114,470 to 124,370). MungeSumstats [33] was used in pre-processing to harmonize effect alleles from separate cohorts to the reference genome. UKB-EUR was used to estimate the number of effective traits for the three EUR meta-analysis cohorts. Number of effective traits was calculated from phenotype eigenvalues as 2.05 using the formula shown above, and the multiple-testing corrected threshold of $P < (5*10^{-8} /2.05) = 2.439 *10^{-8}$ was used to designate significant meta-analysis variant-trait associations.

Heritability and LD score regression

Restricted maximum likelihood (REML) estimates for genetic correlation and multi-trait heritability were calculated using BOLT-REML [26]. The six phenotypes measured in absolute concentration units (mmol/L) from the discovery UKB-EUR cohort (n-3, n-6, DHA, LA, PUFAs, MUFAs) were inputted with the covariates age, age^2, sex, assessment center, genotype batch, and the first twenty PCs. The provided 1000G European LD scores [28] were used to calibrate the BOLT statistic. The refinement step was skipped (--remlNoRefine) to increase computational efficiency.

LD score regression (LDSC) [34] was used to calculate non-partitioned LD Score regression, SNP-based h^2, and genomic control (lambda). The 1000 Genomes European set was used as the LD reference panel [28]. MungeSumstats [33] was used to harmonize alleles and convert summary statistics to LDSC format for this and subsequent steps. Additionally, pairwise

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

genetic correlations between PUFA traits were computed with summary statistics using LDSC [35].

Identifying genetic loci

Lead variants for each independent genomic risk loci were defined in the discovery (UKB-EUR) and meta-analysis cohorts (UKB-EUR + Kettunen *et al.* + FinMetSeq) by inputting summary statistics to FUMA web server [36]. The UKB release2b 10k European set was used as the LD reference panel. 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 loci and lead SNPs. SNPs not available in the GWAS input but contained in the reference panel were included. The maximum distance between LD blocks to merge into a locus was < 250 Kb. *P*-values were set with a lower cap of P = 1e-300 to resolve FUMA processing errors related to minimum Python float size limit. Variants from meta-analysis were annotated to genes with SNP2GENE using positional mapping (maximum distance 10 Kb), eQTL mapping from GTEx v8 liver tissue, and chromatin interaction mapping using built-in data from Hi-C (GSE87112) liver tissue. All other FUMA mapping settings were kept as default.

Identifying novel loci

A table of previously reported ("known") PUFA- and MUFA- associated lead variants was prepared from full summary statistics (where available) and significance tables found within previous PUFA-GWAS publications (Supplementary Table 1). All reported genomic coordinates were set to hg19 using liftOver [37]. Genomic risk loci coordinates were identified in each study by *P*-values of reported variants using FUMA (setting LD reference by ancestry of study, otherwise default settings). These loci were then grouped together within a ± 500 Kb window when checking for novelty of our results, regardless of the ancestry of the study cohorts. Trait

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

names were harmonized across studies. LDtrait of the LDlinkR tools [38] was used to cross-check our novelty table and no additional "known" loci were found with this method.

Multi-ancestry replication in non-EUR UKB Cohorts

GCTA-MLMA [39, 40] was used to perform mixed-model GWAS analyses in the smaller UKB African (AFR), Central/South Asian (CSA), and East Asian (EAS) cohorts (Supplementary Tables 2, 14). A genetic relatedness matrix was generated for each population using GCTA-GRM [39, 41]. Covariates used in these models were age, age^2, sex, and the first ten principal components. Genotype removal parameters were imputation quality score < 0.3, MAF < 1 %, missing genotype per individual > 5 %, missing genotype per variant > 5%, or Hardy-Weinberg equilibrium (HWE) $P < 1\times10^{-8}$.

Gene sets and pathway analysis

FUMA GENE2FUNC [36] was performed on genes mapped from SNP2GENE using parameters described above including all background gene-sets in hypergeometric tests, and using expression data from all GTEx v8 datasets. Benjamini-Hochberg (FDR) was used as the gene set enrichment multiple test correction method. Gene Ontology (GO) was used to categorize sets of genes in downstream analyses [42].

Conditional and joint association analysis

Conditional and joint association analysis (COJO) with stepwise model selection to identify conditionally independent variants was performed with GCTA-COJO (--cojo-slct) [43] for discovery and meta-analysis summary statistics. A random set of 20,000 unrelated EUR UKB participants were used as the LD reference (--bfile). Variants with MAF < 1% were removed.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

COJO was run per chromosome with *P*-value settings based on effective trait Bonferroni correction, using default settings for collinearity and window size. For METAL meta-analysis summary statistics, SE values inputted to COJO were re-calculated based on beta and P-values to ensure accuracy of GCTA-COJO output.

Visualizing results

CMplot [44] was used to generate the circular Manhattan plot in Figure 2. The qqman R package [45] was used to generate standard Manhattan and QQ plots in Supplementary Figure 4. S-MultiXcan Manhattan plots in Supplementary Figure 5 generated using the Manhattan R package [46]. Regional Manhattan plots in Supplementary Figure 9 were generated in FUMA [36]. Bar plots and scatterplots were generated using ggplot2 [47] in R v4.1.0. Correlation plot was generated using corrplot [48].

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Results

Discovery analysis

We performed a GWAS of NMR-measured PUFA traits in European (EUR) individuals. Our three-stage discovery, replication, and meta-analysis approach is shown in Figure 1. First, we performed a GWAS discovery analysis in UK Biobank (UKB). After applying quality control, the UKB-EUR discovery cohort (UKB-EUR) was composed of 101,729 participants with NMR data who were designated as genetically EUR by the Pan UKBB project [16]. Mean age of participants was 56.8 yr old and 45.98% were male (full participant characteristics found in Supplementary Table 2, Supplementary Figure 1). All subsequent analyses of PUFA traits used rank-based inverse normal transformed values. Two covariate sensitivity models were evaluated. Model 1 included the covariates sex, age, age^2, genotyping array, and assessment center, while Model 2 additionally included BMI, lipid medication usage, and socioeconomic status. No residual confounding was observed with either model. The *P*-values were highly correlated between the two models and also with the summary statistics released on OpenGWAS (batch met-d) [27], derived from a generalized GWAS of all metabolic traits in the metabolomics panel (Supplementary Table 3). The range of LDSC intercepts is 1.00 - 1.041 for Model 1 and 0.99 - 1.045 for Model 2 (Supplementary Table 4). The statistical significance in Model 2 was always comparable to or more significant than that of Model 1 and OpenGWAS (Supplementary Figure 2); we infer this is because the inclusion of relevant covariates reduces residual variability and enhances the statistical power for variant discovery [49]. The primary results in the main text are therefore based on Model 2.

A total of 15,578,593 variants were tested for associations with all fourteen available PUFA traits in discovery analysis (Supplementary Table 5). Linear mixed model GWAS found a

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

total of 146,982 significant associations (35,869 unique variants across traits; Supplementary Table 5) at a significance threshold corrected for the effective number of traits ($P < 1.678*10^{-8}$). Conditional and joint analysis (COJO) identified 711 independent significant associations in this cohort (328 unique across traits; Supplementary Tables 5 and 6). We grouped significant associations into LD blocks and merged loci < 250Kb apart; this yielded 612 genomic risk loci (115 unique loci across traits; Supplementary Table 7).

Replication and meta-analysis

Replication of discovery cohort GWAS associations utilized two external EUR PUFA-GWAS studies: FinMetSeq and Kettunen *et al.* (Supplementary Table 13). These studies contained five out of the fourteen traits analyzed in the discovery stage. After munging the three EUR studies to ensure high quality alleles and to harmonize alleles to the reference genome, ~8.7 million variants overlapped between Kettunen *et al.* and UKB-EUR; 209,509 variants overlapped between FinMetSeq and UKB-EUR. Across the five available phenotypes, there were 19,929 UKB-EUR associations (8,543 unique variants) replicated at $P < 0.05$ in one of the two external EUR studies. Of these, 615 associations (266 unique variants) were replicated in both studies (Supplementary Table 13).

We also performed replication analyses across UKB multi-ancestry groups with adequate sample sizes, to evaluate the reproducibility of associations found in the UKB-EUR cohort. No significant phenotypic differences were found between UKB ancestry groups in any of fourteen PUFA traits (Supplementary Table 2, Supplementary Figure 1). Counts of UKB-EUR associations replicated ($P < 0.05$) were: UKB-AFR 5,327 (2,358 unique variants), UKB-CSA 16,560 (5,179 unique variants), and UKB-EAS 5,466 (2,113 unique variants) (Supplementary Table 14). Out of the 612 significant loci for UKB-EUR associations, 170 were replicated in ≥ 1

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

UKB multi ancestry group (Supplementary Table 14). Despite having a smaller sample size, UKB-EAS had more replicated UKB-EUR loci than UKB-AFR (46 vs. 31).

Meta-analysis was performed on variants which appeared in at least two EUR studies; a total of ~10,200,000 variants were tested. Across the five PUFA traits there were 38,344 significant associations ($P < 2.439*10^{-8}$; 17,301 unique variants; Supplementary Table 15). LDSC intercepts for the five meta-analyzed traits ranged from 1.013 (SE = 0.0079) for DHA to 1.04 (0.011) for MUFAs, indicating that there was no inflation or residual confounding (Supplementary Table 16). Conditional and joint analysis revealed 402 independent significant variants (258 unique; Supplementary Table 17). Of these 402 significant COJO associations, 265 were replicated in at least one external EUR study ($P < 0.05$). When grouping all significant meta-analysis variants into loci, we found 254 significant loci (109 unique; Supplementary table 18). Of the 254 grouped meta-analysis loci, 171 were replicated in at least one external EUR study (Supplementary Table 18). The loci identified across phenotypes from COJO independent variants and FUMA were nearly identical. Four loci found in FUMA were not in COJO, and a separate four loci found in COJO were not in FUMA.

We searched the literature and found 210 previously reported PUFA and MUFA GWAS associations (106 unique variants, 37 loci based on 1Mb grouping; Supplementary Table 1). We compared these known loci with our significant meta-analysis loci; of the 254 meta-analysis loci, 173 were novel (87 unique across phenotypes), and 102 of these novel loci were replicated in at least one external EUR study (51 unique).

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Notable associated genes

Among the 109 unique genomic loci identified in our meta-analysis of five traits, thirteen loci were associated with all five traits (Figure 2, Supplementary Table S18), nine of which have been previously identified. These spanned genes that are well-known in lipid metabolism, including the apolipoprotein gene clusters at chr11q23 (i.e., *APOA5*, *APOA4*, *APOC3*, and *APOA1*) and chr19q13 (i.e., *APOE, APOC1*, *APOC4*, and *APOC2*), *APOB*, LDL receptor adaptor protein 1 (*LDLRAP1)*, LDL receptor (*LDLR*), lipase C (*LIPC*), and lysophosphatidic acid receptor 2 (*LPAR2*). Another notable gene is glucokinase regulator (*GCKR*), which has been previously associated with DPA and palmitoleic acid [50-52]. Of the four novel loci associated with all five traits, the locus of chr18q21 covers a candidate gene of lipase G (*LIPG*). The candidate genes at loci chr1p13 and chr8q24 include *PSRC1*, *SORT1*, *TRIB1*, and *SQLE*. Both *PSRC1* and *TRIB1* have been previously associated with familial hypercholesterolemia [53]. *SQLE* encodes squalene epoxidase, a rate-limiting enzyme catalyzing the first oxygenation step in sterol biosynthesis.

When considering omega-3, DHA, omega-6, and LA, there are three loci associated with all these four traits, located at chr2q21 (the *LCT* locus), chr4q13, and chr7p22; all three loci are novel. The chr4q13 encompasses multiple members of the UDP-glycosyltransferase family (i.e., *UGT2B17; UGT2B10; UGT2B11; UGT2A1; UGT2A2;* and *UGT2A1*), which play important roles in the bile acid (BA) detoxification by catalyzing the glucuronidation of BA substrates and impact dietary lipids absorption [54]. One candidate gene at chr7p22 is *CYP2W1*, a member of the cytochrome P450 superfamily, which encodes monooxygenases and oxidizes steroids, fatty acids, and xenobiotics [55].

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Nine genomic loci are only associated with omega-3, DHA, or both, eight of which are novel. The known locus has at least two candidate genes, choline kinase alpha (*CHKA*) and carnitine palmitoyltransferase 1A (*CPT1A*). *CHKA* encodes the initial enzyme to catalyze the phosphorylation of ethanolamine in the CDP-choline pathway for phosphatidylcholine biosynthesis. *CPT1A* catalyzes the transfer of the acyl group of LCPUFA-CoA conjugates onto carnitine, enabling the mitochondrial update of LCPUFAs and the subsequent beta-oxidation. One novel locus at chr10q23 has a cluster of genes in the cytochrome P450 superfamily (i.e., *CYP2C18*, *CYP2C19*, *CYP2C9*, and *CYP2C8*). Another notable novel locus at chr11q24 has a candidate gene of *ST3GAL4*, which is involved in the terminal sialylation of glycolipids. There are 35 loci associated with only omega-6, LA, or both; 33 of these are novel. Multiple novel candidate genes are implicated in lipid metabolism, such as LDL receptor related protein 2 (*LRP2*), NPC1 like intracellular cholesterol transporter 1 (*NPC1L1*), scavenger receptor class B member 1 (*SCARB1*), phospholipase C gamma 1 (*PLCG1*), and lipin 3 (*LPIN3*). Three additional novel loci carry cytochrome P450 genes, including chr2q33 (*CYP20A1*), chr8q12 (*CYP7A1)*, and chr19q13 (*CYP2A6*). Another two notable candidate genes are arachidonate 5-lipoxygenase (*ALOX5*) and peroxisome proliferator activated receptor delta (*PPARD*).

The two key sets of genes which catalyze LCPUFA biosynthesis are fatty acid desaturase (*FADS*) and elongase protein family genes (*ELOVL*) [7]. We have re-confirmed the primary importance of *FADS* genes in n-3 LCPUFA genetics, as these genes had the most significant *P*-values in our meta-analysis at lead SNPs rs174528 (DHA, $P < 1E\text{-}300$; MAF = 0.39) and rs509360 (omega-3, $P < 1E\text{-}300$; MAF =0.33). Both of these variants were mapped to *FADS1*, *FADS2*, and *FADS3*. The most significant independent association in COJO analysis, however, was rs964184 (*ZPR1*; *p*-value from joint analysis (*p*J) = 6.08E-207) with MUFAs. *ZPR1* and

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

rs964184 have been repeatedly associated with PUFA and MUFA traits in previous GWAS (Supplementary Table 1). Of the seven *ELOVL* genes, only *ELOVL2* (chr6:10,980,992-11,044,547) has been associated with PUFAs in previous GWAS studies, specifically with omega-3s: DHA, eicosapentaenoic acid (EPA), and *cis*-7,10,13,16,19-docosapentaenoic acid (DPA) (Supplementary Table 1, locus 13). We found an association in UKB-EUR with the *ELOVL2* locus surpassing the suggestive significance threshold (P < 5e-05) for DHA (rs9380082, P=2.6e-05), but did not find this locus associated at genome-wide significance. However, we did find a novel, non-replicated association with MUFAs to total fatty acids percentage at *ELOVL6*. The lead variant rs189866430 (T>A; chr4:110,578,226; $MAF_{UKB\text{-}EUR}$ = 0.01) is ~389Kbp downstream from *ELOVL6*. This association is consistent with previous findings that demonstrate the role of *ELOVL6* gene product in elongating MUFAs [56].

ACSL6 is part of the Acyl-CoA synthetase (ACS) family of enzymes which catalyze the formation of acyl-CoAs from free fatty acids [57]. We identified two novel and externally replicated associations (Supplementary Table 18): omega-3 fatty acids with rs273913 (T>C; $MAF_{EUR\text{-}meta\text{-}analysis}$ = 0.3867; locus start=chr5:131,407,493; Supplementary Figure 9a), and DHA with rs166635 (G>A; $MAF_{EUR\text{-}meta\text{-}analysis}$ = 0.3079; locus start=chr5:131,590,114; Supplementary Figure 9b), that are ~60Kbp and ~242Kbp upstream from the *ACSL6* gene (chr5:131,142,683-131,347,936; reverse strand). It should be noted with regard to novelty, this locus was previously associated with arachidonic acid at rs274559 by Guan *et al.* [6], but their P-value at 3.81e-06 did not reach genome-wide significance. *ACSL6* expression has been previously linked to DHA enrichment in the brain. Our finding of significant associations with only omega-3 and DHA, not omega-6, LA, or MUFAs, is consistent with previous reports [58, 59].

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Heritability and genetic correlation

In the fourteen phenotypes tested in the discovery cohort, Linkage Disequilibrium Score (LDSC) found h^2 ranged from 0.12 (SE = 0.018) for LA percentage to 0.20 (0.032) for MUFA percentage (Supplementary Table 4). Using individual-level genotype data in the discovery cohort, BOLT-REML found h^2 of six PUFA traits measured in absolute concentration units (omega-3, omega-6, DHA, LA, PUFAs, MUFAs) to range from 0.16 (0.0065) for LA to 0.22 (0.0066) for MUFAs (Supplementary Table 8). The lower range of LDSC when compared to BOLT-REML is consistent with our expectation that LDSC reports the lower bound of heritability estimates. LDSC-estimated h^2 was similar in meta-analysis, ranging from 0.12 (0.022) for DHA to 0.16 (0.023) for MUFAs (Supplementary Table 16).

We evaluated the shared genetic basis across PUFA traits using genetic correlations (r_g). For the fourteen phenotypes in the discovery analysis, the levels of phenotypic (r_p) and genetic correlations are broadly consistent across all 91 trait-pairs, with slightly stronger genetic correlations (linear regression $\beta = 1.089$; $P < 2.2e\text{-}16$; Supplementary Figure 3; Supplementary Table 4). Among the 78 trait-pairs that have both nominally significant phenotypic and genetic correlations ($P < 0.05$), 57 have stronger correlations at the genetic level (binomial test $p = 2.79e\text{-}5$). Of note, the genetic correlations between the absolute concentrations and their relative percentages of total fatty acids are not always high, ranging from 0.89 for MUFAs, 0.84 for omega-3, 0.73 for DHA, -0.56 for omega-6, -0.32 for PUFAs, to 0.017 for LA. These medium to low genetic correlations emphasize the need to perform separate GWAS for absolute concentrations and relative percentages. Moreover, the correlation between omega-3 and omega-6 is medium with the absolute concentrations ($r_g = 0.67$; $r_p = 0.45$), and is low with the relative

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

percentages ($r_g = -0.11$; $r_p = -0.12$), indicating that there is substantial unique genetic basis for either trait.

Transcriptome-wide association analyses

S-PrediXcan was used to integrate GTEx (v8) eQTL (expression quantitative trait loci) data from 49 tissues and UKB-EUR cohort GWAS summary statistics to identify genes whose expression is associated with PUFA traits. Across fourteen PUFA phenotypes in the discovery analysis, 24,666 Bonferroni-corrected significant gene-trait associations comprised of 527 unique genes were identified (Supplementary Table 10). We then used S-MultiXcan to find joint effects of gene expression correlation across tissues. We found 2,818 associations (601 unique genes), of which 392 unique genes have not been in previous PUFA GWAS and are thus considered as novel for PUFA traits (Supplementary Tables 11, 12).

Since there was a high degree of overlap between TWAS and GWAS results, we searched for novel gene-trait associations in S-MultiXcan that had not been found in our discovery or meta-analysis GWAS analyses. We found 55 genes, spanning 12 loci, that were novel and identified exclusively in TWAS. Many of these genes (44) are found in a cluster at 6p21. These 44 genes are significantly enriched for immune system process (GO: GO:0002376; 17 genes; FDR = 7.41E-04) and regulation of immune system process (GO:0002682; 17 genes; FDR = 2.09E-06). Three other genes from the 55 novel genes, *F2* (MUFAs), *WDR81* (LA), and *PTK2* (PUFAs), are involved in the regulation of lipid kinase activity (GO:0043550).

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Gene set enrichment analysis

MAGMA tissue expression analysis for sets of positionally mapped genes from each of the five meta-analyzed PUFA traits revealed that liver was exclusively the significantly enriched tissue type (Supplementary Figure 6). Because of this, we sent genes mapped from GTEx liver eQTLs and HiC liver chromatin data, in addition to positionally mapped genes, to GENE2FUNC for gene set enrichment (Supplementary Table 19).

Across the five traits, the most significant gene sets in the categories of Curated gene set, Positional gene set, Gene Ontology (GO): Biological process, GO: Cellular component, GO: Molecular function, Cancer modules, Canonical pathways, Computational gene sets, KEGG pathways, and TF targets, were all driven by genes in the major histocompatibility complex (MHC). The MHC region is the most polymorphic in the human genome, associated with the most disease traits, and determining causal variants in this region is highly prone to confounding [60]. After MHC, the next most significant positional gene set in all five PUFA traits corresponds to rs1260326 in *GCKR*; this gene locus has been identified in several published PUFA GWAS analyses (Supplementary Table 1). In the gene sets defined by GWAS catalog, the most significant enrichments are in genes that have previously been associated with blood lipids, including total cholesterol, LDL cholesterol, and triglycerides.

To identify new relationships between novel genes associated with PUFA traits and previously reported traits in GWAS Catalog, we stratified GENE2FUNC output based on novelty (Supplementary Table 20). The second most significant gene set enrichment of non-lipid GWAS catalog traits for omega-3s and DHA (after “Handedness”) was “Alcohol use disorder (total score)” (Omega-3: adjP_{novel} = 5.95E-10, adjP_{known} = 0.0052, adjP_{all} = 3.38E-09; DHA: adjP_{novel} = 1.57E-10, adjP_{known}=0.075, adjP_{all} = 1.73E-10; Supplementary Figure 8). Alcohol-

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

related traits have been experimentally linked to PUFAs in multiple studies [61-65] (more in discussion).

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Discussion

Here we report the largest GWAS to-date of PUFA and MUFA phenotypes in Europeans (EUR; N_{EUR}=114,133 to 124,024), including fourteen traits in the discovery analysis, five of which were replicated and meta-analyzed in external EUR cohorts. The discovery cohort, UK Biobank (UKB), is the largest publicly available human dataset with measures of these traits in genotyped participants ($N_{UKB\text{-}EUR}$=101,729 after QC). We have identified 51 novel and replicated loci, as well as 36 loci that were not replicated, but have not been reported in previous GWAS studies (Supplementary Tables 1, 18). Considering that only 37 genomic risk loci were previously reported in relation to these traits, this study greatly increases our scope of understanding the genetic architecture of PUFAs and MUFAs. Of the 37 previously reported loci, we have replicated 23 loci in our discovery analysis (UKB-EUR) and 22 loci in our EUR meta-analysis (Supplementary Table 1).

In our follow-up analysis of genes mapped to novel and known loci in the meta-analysis, we found a notable difference in *P*-values and proportion of gene set enrichment for the GWAS catalog trait "alcohol use disorder (total score)" (AUD) (Supplementary Table 20, Supplementary Figure 8). Across the phenotypes omega-3, omega-6, DHA, and LA, there are two novel clusters of PUFA-gene associations that have previously been associated with AUD. These genes are *PLEKHM1*, *CRHR1*, *SPPL2C*, *MAPT*, *STH*, and *KANSL1*, *NSF*, and *WNT3* at chr17q21.31, mapped to omega-3 and DHA; and *FUT2*, *MAMSTR*, *RASIP1*, and *IZUMO1* at chr19q13.33, mapped to omega-6 and LA. The inversion at chr17q21.31 has recently been associated with alcohol intake in a GWAS of ~127,000 European participants from the Million Veterans Program cohort [61]. The association of the gene cluster at chr19q13.33 with AUD was reported as a novel association in an analysis of ~435,000 European participants of UKB [62].

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

In addition to the shared genetic variants between PUFAs and AUD, such as variants in *SNX17* and *GCKR*, variability in PUFA levels has been associated directly with AUD. The direction of causality between these traits has not been clearly disentangled. DHA has a neuroprotective effect against binge alcohol drinking, and is depleted with alcohol exposure [66]. In the opposite causal direction, high alcohol consumption was associated with lower fatty acid intake measured by 24 hour recall in the 2001-2002 National Health and Nutrition Examination Survey in 4,168 adults [63]. Omega-3 deficiencies are associated with bipolar disorder [67], which can lead to higher cravings for alcohol. Additionally, alcohol abuse has been characterized by an increase in oleic acid / LA ratio and Teubert *et al.* demonstrated a shift back to higher LA during alcohol detoxification in a small study of 45 alcoholic patients [65]. Overall, the data on this topic are sparse, and more research should be done to elucidate this relationship.

Along with our novel findings, our study has several limitations. First, the PUFA traits that we were able to investigate are limited to those reported by UKB. We cannot resolve, for instance, differences in specific PUFAs that are often reported with higher resolution metabolite analyses, such as the difference in effects associated with DHA and other omega-3s, notably eicosapentaenoic acid (EPA). Second, UKB is known to have volunteer bias, which can skew results, as has previously been shown [68].

Next, it is unfortunate that our analysis is mostly limited to determining the genetic associations of PUFA and MUFA traits in EUR populations. We recognize that an overwhelming number of genomic analyses to date have been conducted on EUR populations [69], to the detriment of understanding other ancestry groups. Further, our replication analysis shows that of 115 discovery loci in UKB-EUR, only 47 were replicated at $P < 0.05$ in one or more of the AFR, CSA, or EAS multi-ancestry groups. While power calculation shows that the difference was

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

mainly driven by small sample sizes of non-EUR samples, it is possible that there is a distinct set of variants associated with PUFA traits in non-EUR groups. We hope that the results of this study can be meta-analyzed with ancestrally diverse participant groups in future analyses.

Another limitation is possibly introduced by the quantification of PUFA and MUFA traits in our three meta-analyzed studies using nuclear magnetic resonance spectroscopy (NMR). NMR has pros and cons as compared to the gold standard methods for quantitative fatty acid analysis, specifically high-resolution capillary gas chromatography coupled to flame ionization detection (FID) or mass spectrometry (GC-MS), or alternatively, single-stage mass spectrometry (LC-MS) [70]. First, the speed and cost advantages of NMR over GC or MS are advantageous in biobank-scale sample quantification [20]. NMR is also a non-destructive technique, meaning samples can be stored and re-measured in the future. However, NMR is of reduced sensitivity and selectivity compared to GC-based techniques. GC resolves all fatty acids at nanogram levels, compared to NMR which operates at minimum on milligram scale [71]. GC-MS is also able to perform more selective, targeted analyses, and accurately separate analytes, whereas NMR is limited in resolution and inherently will produce confounding overlapping signals, particularly in complex mixtures [71]. Nevertheless, as discussed above, our most significant results are congruent with biochemical expectations and with previous GWAS studies, including those studies which used MS-based quantification (Supplementary Table 1). This adds confidence to the usage of NMR quantification methods and strengthens our novel findings.

Finally, as with any GWAS, associations are candidates for mechanistic processes that, when altered, will have a reproducible influence on traits and ultimately human health. Replication of the associations and detailed investigation in experimental models and in human randomized control trials are required to lead to clinical and precision nutrition applications. This

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

study adds to a growing body of genomics literature that may help realize these applications in relation to PUFA and MUFA traits [72].

Data availability

Full GWAS summary statistics can be found on GWAS Catalog, using the accession codes provided in Supplementary Table 21. Scripts for this analysis can be found at https://github.com/michaelofrancis/PUFA_GWAS.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

References

1. Harris, W.S., et al., *Blood n-3 fatty acid levels and total and cause-specific mortality from 17 prospective studies.* Nature Communications, 2021. **12**(1): p. 2329.
2. Lin, P.-Y., et al., *A meta-analytic review of polyunsaturated fatty acid compositions in dementia.* The Journal of clinical psychiatry, 2012. **73**(9): p. 0-0.
3. Grosso, G., et al., *Dietary n-3 PUFA, fish consumption and depression: A systematic review and meta-analysis of observational studies.* Journal of Affective Disorders, 2016. **205**: p. 269-281.
4. Innes, J.K. and P.C. Calder, *Marine Omega-3 (N-3) Fatty Acids for Cardiovascular Health: An Update for 2020.* International journal of molecular sciences, 2020. **21**(4): p. 1362.
5. Lorente-Cebrián, S., et al., *Role of omega-3 fatty acids in obesity, metabolic syndrome, and cardiovascular diseases: a review of the evidence.* Journal of physiology and biochemistry, 2013. **69**(3): p. 633-651.
6. Guan, W., et al., *Genome-wide association study of plasma N6 polyunsaturated fatty acids within the cohorts for heart and aging research in genomic epidemiology consortium.* Circulation: Cardiovascular Genetics, 2014. **7**(3): p. 321-331.
7. Brenna, J.T. and K.S.D. Kothapalli, *New understandings of the pathway of long-chain polyunsaturated fatty acid biosynthesis.* Curr Opin Clin Nutr Metab Care, 2022. **25**(2): p. 60-66.
8. Kothapalli, K.S.D., H.G. Park, and J.T. Brenna, *Polyunsaturated fatty acid biosynthesis pathway and genetics. implications for interindividual variability in prothrombotic, inflammatory conditions such as COVID-19.* Prostaglandins, Leukotrienes and Essential Fatty Acids, 2020. **162**.
9. Sinclair, A.J., *High Linoleic Acid in the Food Supply Worldwide-What are the Consequences?* Science and Technology of Cereals, Oils and Foods, 2022. **30**(3).
10. Hodson, L., C.M. Skeaff, and B.A. Fielding, *Fatty acid composition of adipose tissue and blood in humans and its use as a biomarker of dietary intake.* Progress in lipid research, 2008. **47**(5): p. 348-380.
11. Locke, A.E., et al., *Exome sequencing of Finnish isolates enhances rare-variant association power.* Nature, 2019. **572**(7769): p. 323-328.
12. Shin, S.Y., et al., *An atlas of genetic influences on human blood metabolites.* Nat Genet, 2014. **46**(6): p. 543-550.
13. Kettunen, J., et al., *Genome-wide association study identifies multiple loci influencing human serum metabolite levels.* Nat Genet, 2012. **44**(3): p. 269-76.
14. Kettunen, J., et al., *Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA.* Nature Communications, 2016. **7**(1): p. 11122.
15. Würtz, P., et al., *Quantitative serum nuclear magnetic resonance metabolomics in large-scale epidemiology: a primer on-omic technologies.* American journal of epidemiology, 2017. **186**(9): p. 1084-1096.
16. team, P.-U. 2020; Available from: https://pan.ukbb.broadinstitute.org.
17. Borodulin, K., et al., *Cohort Profile: The National FINRISK Study.* International Journal of Epidemiology, 2018. **47**(3): p. 696-696i.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

18. Stančáková, A., et al., *Changes in Insulin Sensitivity and Insulin Release in Relation to Glycemia and Glucose Tolerance in 6,414 Finnish Men.* Diabetes, 2009. **58**(5): p. 1212.
19. Lin, D.-Y. and P.F. Sullivan, *Meta-analysis of genome-wide association studies with overlapping subjects.* American journal of human genetics, 2009. **85**(6): p. 862-872.
20. Soininen, P., et al., *Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Cardiovascular Epidemiology and Genetics.* Circulation: Cardiovascular Genetics, 2015. **8**(1): p. 192-206.
21. Ahola-Olli, A.V., et al., *Circulating metabolites and the risk of type 2 diabetes: a prospective study of 11,896 young adults from four Finnish cohorts.* Diabetologia, 2019. **62**(12): p. 2298-2309.
22. McCaw, Z.R., et al., *Operating characteristics of the rank-based inverse normal transformation for quantitative trait analysis in genome-wide association studies.* Biometrics, 2020. **76**(4): p. 1262-1272.
23. Bycroft, C., et al., *The UK Biobank resource with deep phenotyping and genomic data.* Nature, 2018. **562**(7726): p. 203-209.
24. Chang, C.C., et al., *Second-generation PLINK: rising to the challenge of larger and richer datasets.* Gigascience, 2015. **4**: p. 7.
25. Purcell, S.C.C., *PLINK* 2020: www.cog-genomics.org/plink/2.0/.
26. Loh, P.-R., et al., *Efficient Bayesian mixed-model analysis increases association power in large cohorts.* Nature genetics, 2015. **47**(3): p. 284-290.
27. Elsworth, B., et al., *The MRC IEU OpenGWAS data infrastructure.* bioRxiv, 2020: p. 2020.08.10.244293.
28. Consortium, G.P., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68.
29. Bretherton, C.S., et al., *The effective number of spatial degrees of freedom of a time-varying field.* Journal of climate, 1999. **12**(7): p. 1990-2009.
30. Barbeira, A.N., et al., *Integrating predicted transcriptome from multiple tissues improves association detection.* PLOS Genetics, 2019. **15**(1): p. e1007889.
31. Consortium, G., *The GTEx Consortium atlas of genetic regulatory effects across human tissues.* Science, 2020. **369**(6509): p. 1318-1330.
32. Willer, C.J., Y. Li, and G.R. Abecasis, *METAL: fast and efficient meta-analysis of genomewide association scans.* Bioinformatics, 2010. **26**(17): p. 2190-2191.
33. Murphy, A.E., B.M. Schilder, and N.G. Skene, *MungeSumstats: a Bioconductor package for the standardization and quality control of many GWAS summary statistics.* Bioinformatics, 2021. **37**(23): p. 4593-4596.
34. Bulik-Sullivan, B.K., et al., *LD Score regression distinguishes confounding from polygenicity in genome-wide association studies.* Nature Genetics, 2015. **47**(3): p. 291-295.
35. Bulik-Sullivan, B., et al., *An atlas of genetic correlations across human diseases and traits.* Nature Genetics, 2015. **47**(11): p. 1236-1241.
36. Watanabe, K., et al., *Functional mapping and annotation of genetic associations with FUMA.* Nature communications, 2017. **8**(1): p. 1-11.
37. Hinrichs, A.S., et al., *The UCSC genome browser database: update 2006.* Nucleic acids research, 2006. **34**(suppl_1): p. D590-D598.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

38. Myers, T.A., S.J. Chanock, and M.J. Machiela, *LDlinkR: An R Package for Rapidly Calculating Linkage Disequilibrium Statistics in Diverse Populations.* Front Genet, 2020. **11**: p. 157.
39. Yang, J., et al., *GCTA: a tool for genome-wide complex trait analysis.* Am J Hum Genet, 2011. **88**(1): p. 76-82.
40. Yang, J., et al., *Advantages and pitfalls in the application of mixed-model association methods.* Nature genetics, 2014. **46**(2): p. 100-106.
41. Yang, J., et al., *Common SNPs explain a large proportion of the heritability for human height.* Nature genetics, 2010. **42**(7): p. 565-569.
42. *The Gene Ontology resource: enriching a GOld mine.* Nucleic Acids Res, 2021. **49**(D1): p. D325-d334.
43. Yang, J., et al., *Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits.* Nat Genet, 2012. **44**(4): p. 369-75, s1-3.
44. Yin, L., et al., *rMVP: A Memory-efficient, Visualization-enhanced, and Parallel-accelerated tool for Genome-Wide Association Study.* Genomics, Proteomics & Bioinformatics, 2021.
45. Turner, S.D., *qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots.* bioRxiv, 2014: p. 005165.
46. Liu, B., et al., *Abundant associations with gene expression complicate GWAS follow-up.* Nature Genetics, 2019. **51**(5): p. 768-769.
47. Wickham, H., *ggplot2: elegant graphics for data analysis*. 2016: springer.
48. Wei, T. and V. Simko, *package "corrplot": Visualization of a Correlation Matrix; 2017.* Version 0.92, 2021.
49. Mefford, J. and J.S. Witte, *The Covariate's Dilemma.* PLOS Genetics, 2012. **8**(11): p. e1003096.
50. Lemaitre, R.N., et al., *Genetic loci associated with plasma phospholipid n-3 fatty acids: a meta-analysis of genome-wide association studies from the CHARGE Consortium.* PLoS Genet, 2011. **7**(7): p. e1002193.
51. Hu, Y., et al., *Genome-wide meta-analyses identify novel loci associated with n-3 and n-6 polyunsaturated fatty acid levels in Chinese and European-ancestry populations.* Hum Mol Genet, 2016. **25**(6): p. 1215-24.
52. Wu, J.H., et al., *Genome-wide association study identifies novel loci associated with concentrations of four plasma phospholipid fatty acids in the de novo lipogenesis pathway: results from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium.* Circ Cardiovasc Genet, 2013. **6**(2): p. 171-83.
53. Sánchez Muñoz-Torrero, J.F., et al., *Multivariate analysis for coronary heart disease in heterozygote familial hypercholesterolemia patients.* Per Med, 2018. **15**(2): p. 87-92.
54. Perreault, M., et al., *The Human UDP-glucuronosyltransferase UGT2A1 and UGT2A2 enzymes are highly active in bile acid glucuronidation.* Drug Metab Dispos, 2013. **41**(9): p. 1616-20.
55. Arnold, C., et al., *Cytochrome P450-dependent metabolism of omega-6 and omega-3 long-chain polyunsaturated fatty acids.* Pharmacol Rep, 2010. **62**(3): p. 536-47.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

56. Wang, Z., et al., *Fatty acid desaturase 2 (FADS2) but not FADS1 desaturates branched chain and odd chain saturated fatty acids.* Biochim Biophys Acta Mol Cell Biol Lipids, 2020. **1865**(3): p. 158572.
57. Fernandez Regina, F., et al., *Acyl-CoA synthetase 6 enriches the neuroprotective omega-3 fatty acid DHA in the brain.* Proceedings of the National Academy of Sciences, 2018. **115**(49): p. 12525-12530.
58. Marszalek, J.R., et al., *Long-chain acyl-CoA synthetase 6 preferentially promotes DHA metabolism.* Journal of Biological Chemistry, 2005. **280**(11): p. 10817-10826.
59. Fernandez, R.F., et al., *Acyl-CoA synthetase 6 is required for brain docosahexaenoic acid retention and neuroprotection during aging.* JCI Insight, 2021. **6**(11).
60. Trowsdale, J. and J.C. Knight, *Major histocompatibility complex genomics and human disease.* Annual review of genomics and human genetics, 2013. **14**: p. 301-323.
61. Gelernter, J., et al., *Genome-wide Association Study of Maximum Habitual Alcohol Intake in >140,000 U.S. European and African American Veterans Yields Novel Risk Loci.* Biological Psychiatry, 2019. **86**(5): p. 365-376.
62. Zhou, H., et al., *Genome-wide meta-analysis of problematic alcohol use in 435,563 individuals yields insights into biology and relationships with other traits.* Nature Neuroscience, 2020. **23**(7): p. 809-818.
63. Kim, S.Y., et al., *Alcohol consumption and fatty acid intakes in the 2001–2002 National Health and Nutrition Examination Survey.* Alcoholism: Clinical and Experimental Research, 2007. **31**(8): p. 1407-1414.
64. Fogaça, M.N., et al., *The effects of polyunsaturated fatty acids in alcohol dependence treatment - a double-blind, placebo-controlled pilot study.* BMC Clinical Pharmacology, 2011. **11**(1): p. 10.
65. Teubert, A., et al., *Elevated oleic acid serum concentrations in patients suffering from alcohol dependence.* Journal of Molecular Psychiatry, 2013. **1**(1): p. 13.
66. Collins, M., *Alcohol abuse and docosahexaenoic acid: Effects on cerebral circulation and neurosurvival.* Brain Circulation, 2015. **1**(1): p. 63-68.
67. Balanza-Martinez, V., et al., *Therapeutic use of omega-3 fatty acids in bipolar disorder.* Expert review of neurotherapeutics, 2011. **11**(7): p. 1029-1047.
68. Munafò, M.R., et al., *Collider scope: when selection bias can substantially influence observed associations.* International Journal of Epidemiology, 2018. **47**(1): p. 226-235.
69. Popejoy, A.B. and S.M. Fullerton, *Genomics is failing on diversity.* Nature, 2016. **538**(7624): p. 161-164.
70. Brenna, J.T., et al., *Best practices for the design, laboratory analysis, and reporting of trials involving fatty acids.* Am J Clin Nutr, 2018. **108**(2): p. 211-227.
71. Emwas, A.-H., et al., *NMR Spectroscopy for Metabolomics Research.* Metabolites, 2019. **9**(7).
72. Francis, M., et al., *Genome-wide association study of fish oil supplementation on lipid traits in 81,246 individuals reveals new gene-diet interaction loci.* PLOS Genetics, 2021. **17**(3): p. e1009431.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Figure / Table captions

Main Figures

Figure 1. Overview of analyses. Fourteen PUFA traits were analyzed in the UK Biobank (UKB) European discovery cohort. Significant associations were sent to replication for five traits available in external European cohorts FinMetSeq and Kettunnen et al. These three studies were meta-analyzed, resulting in 22,334 significant and replicated associations across the 5 traits, comprised of 51 unique, novel, significant replicated loci. Replication was also performed in UKB multi-ancestry cohorts. Additional software analyses are shown in blue. FUMA: functional mapping and association; COJO: conditional and joint analyses; LDSC: linkage disequilibrium score regression; TWAS: transcriptome-wide association analysis.

Figure 2. Circular Manhattan plot of five meta-analyzed PUFA traits. Plots show the −log10 *P*-values of variant-based GWAS of PUFA traits. Red triangles designate the lead variant of a novel locus associated with the trait in our analysis. Red dotted lines at $P < 2.439*10^{-8}$ indicate the genome-wide significance threshold Bonferroni corrected for number of effective traits. Alternating color shades within each trait ring designate breaks between chromosomes. Genes corresponding to loci with $P < 1e-20$ are labeled. All *P*-values are restrained to an upper limit of 1e-100 for visualization. Rings from outer to inner: omega-3 fatty acids, docosahexaenoic acid, omega-6 fatty acids, linoleic acid, and monounsaturated fatty acids.

Supplementary Figures

S1. Participant characteristics for PUFA traits of four ancestries in UK Biobank. These plots correspond to the values in **S2 Table**. (A) Mean and standard deviation (sd) of PUFA traits measured in absolute concentration units (mmol/L). (B) Left: average percentage and sd of each trait per total amount of fatty acids. Right: Ratio of a:b and sd of these values.

S2. Correlation plot comparing *P*-values of our Models 1 and 2 versus OpenGWAS. These *P*-values were compared because of different QC and model covariate selection procedures. Each dot represents one variant in the UK Biobank. Spearman's Rho (R) and correlation *P*-values shown.

S3. Genetic and phenotypic correlations between fourteen UK Biobank PUFA traits. Above diagonal (circles) are genetic correlations between traits calculated from UKB GWAS summary statistics using LDSC. Below diagonal (squares) are phenotypic correlations by Pearson correlation coefficient. Color and shape size correspond to direction and strength of correlation.

S4. Manhattan and QQ plots of UK Biobank discovery (EUR) dataset. Left: Manhattan plot showing the −log10 *P*-values across 22 chromosomes for associations in each of fourteen PUFA traits in the UK Biobank European discovery cohort. Red line at $1.678*10^{-8}$ shows genome-wide significance threshold Bonferroni corrected for number of effective traits. Right: Quantile-quantile (QQ) plots showing observed versus expected distributions of association *P*-values for each trait.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

S5. SMultiXcan results from UK Biobank discovery (EUR) summary statistics. Manhattan plots showing −log10 *P*-values of significant associations between gene expression levels and fourteen PUFA traits in the UKB-EUR cohort. The red line is the Bonferonni corrected significance threshold at $P < 7.68 * 10^{-7}$. The most significant genes for each 5Mb window of significant associations are labeled. Alternating color shades designate breaks between chromosomes.

S6. MAGMA tissue expression analysis for meta-analyzed phenotypes in European cohorts. Significant tissue-expression specificity by tissue type for five meta-analyzed PUFA traits in meta-analysis of UKB-EUR, FinMetSeq, and Kettunnen et al. studies. Analysis and plots performed by MAGMA as implemented in the FUMA software. Liver is the only significant tissue type identified in the five meta-analyzed traits.

S7. Manhattan and QQ plots of three UK Biobank multi-ancestry cohorts. Manhattan plots showing the −log10 *P*-values across 22 chromosomes for associations in each of fourteen PUFA traits in three UK Biobank multi-ancestry cohorts: African (AFR); Central and South Asian (CSA); and East Asian (EAS). Red line at $P < 1.678*10^{-8}$ shows genome-wide significance threshold Bonferroni corrected for number of effective traits. Right: Quantile-quantile (QQ) plots showing observed versus expected distributions of association *P*-values for each trait.

S8. Enrichment of gene sets for "alcohol use disorder (total score)." Enrichment −log10 (adj*P*)-values of gene sets with GWAS catalog reported genes associated with "Alcohol use disorder (total score)." Genes in query gene sets mapped from significant meta-analysis PUFA trait variant associations by position, GTEx liver eQTLs, and HiC liver chromatin data using GENE2FUNC implemented by FUMA.

S9. LocusZoom regional Manhattan plots. ACLS6 novel nearby associations with omega-3 fatty acids (rs273913, top) and DHA (rs166635, bottom).

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Supplementary Tables

S1. Known PUFA loci. Previously reported PUFA- and MUFA-associated lead variants. Each row represents a genomic risk locus identified by inputting summary statistics from previous publications into FUMA SNP2GENE. Phenotype abbreviations: AA: arachidonic acid; AdrA: adrenic acid; ALA: alpha-linolenic acid; DGLA: dihomo-gamma-linolenic acid; DHA: docosahexaenoic acid; DPA: *cis*-7,10,13,16,19-docosapentaenoic acid; DPAn6: *cis*-4,7,10,13,16-docosapentaenoic acid; EDA: eicosadienoic acid; EPA: eicosapentaenoic acid; FAw3: omega-3 fatty acids; FAw6: omega-6 fatty acids; FAw67: omega-6 and -7 fatty acids; GLA: gamma-linolenic acid; LA: linoleic acid; MUFA: monounsaturated fatty acids; OA: oleic acid; otPUFA: polyunsaturated fatty acids (other than 18:2); POA: palmitoleic acid; PUFA: polyunsaturated fatty acids.

S2. Participant characteristics table for UK Biobank cohorts. Phenotype and covariate data for UKB cohorts (EUR, AFR, CSA, EAS). Continuous variables are represented as: mean (standard deviation). BMI: body mass index.

S3. Comparison of discovery models one and two with each other and with OpenGWAS. Compare Spearman correlation coefficient between *P*-values for variants in UKB-EUR discovery analysis M1 and M2, as well as between M1, M2, and OpenGWAS. All correlation coefficients were highly significant ($P < 2.2e-16$). See Supplementary Figure 2 for plots.

S4. LDSC results for discovery cohort. Number of SNPs in Linkage disequilibrium score regression (LDSC) analysis, heritability (h^2), genomic control (GC; lambda), and LDSC intercept for fourteen PUFA traits in models 1 and 2 of discovery UKB-EUR cohort.

S5. Discovery cohort GWAS summary. Number of significant variants, independent significant variants (from COJO), and significant loci for each of fourteen PUFA traits tested in the UKB-EUR discovery cohort. Novel loci for each trait and unique novel loci are also shown.

S6. Discovery GCTA-COJO results. Conditional and joint (COJO) analysis using summary statistics from fourteen PUFA traits in the UKB-EUR discovery cohort. RefA: effect allele; freq: frequency of the effect allele in the original data; b: effect size; se: standard error; p: *p*-value from original GWAS; n: estimated effective sample size; freq_geno: frequency of the effect allele in the reference sample; bJ: effect size from joint analysis of selected SNPs; bJ_se: standard error from joint analysis of selected SNPs; pJ: *p*-value from joint analysis of selected SNPs; LD_r: LD correlation between the SNP i and SNP i + 1 for the SNPs on the list.

S7. Discovery cohort significant loci. FUMA genomic risk loci from UKB-EUR discovery cohort summary statistics. Corresponding summary statistics from UKB multi-ancestry African (AFR), Central and South Asian (CSA), and East Asian (EAS) cohorts are also provided.

S8. BOLT-REML multi-trait heritability correlations. Output table from BOLT-REML multi-trait heritability correlations for the six traits in the discovery UKB-EUR cohort measured in absolute concentration units (mmol/L). The diagonal represents heritability explained by genotyped SNPs, other values are genetic correlations between traits.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

S9. Genetic and phenotypic correlations heatmap data. Raw data for correlations heatmap Supplementary Figure 3 (upper) and corresponding *P*-values (lower). Above diagonal are genetic correlations calculated using LDSC using UKB GWAS summary statistics. Below diagonal are phenotypic correlations by Pearson correlation coefficient.

S10. S-PrediXcan results. Gene-trait associations by tissue type from S-PrediXcan, passing the Bonferroni corrected significance threshold $P < 2.791*10^{-8}$ (0.05/(601,176*2.98). Gene: gene ID; gene_name: HUGO gene name; Zscore: S-PrediXcan association result for the gene; Pvalue: P-value of Zscore; var_g: variance of the gene expression, calculated as W' * G * W (where W is the vector of SNP weights in a gene's model, W' is its transpose, and G is the covariance matrix); n_snps_used: number of SNPs in the covariance matrix; n_snps_in_model: number of SNPs in the model.

S11. S-MultiXcan results. Significant genes across tissue types from S-MultiXcan, passing the Bonferroni corrected significance threshold of $P < 7.68 *10^{-7}$ (0.05/(21,846*2.98). Gene: gene ID; gene_name: HUGO gene name; pvalue: significance p-value of S-MultiXcan association; n: number of "tissues" available for this gene; n_indep: number of independent components of variation kept among the tissues' predictions. (Synthetic independent tissues); p_i_best: best p-value of single-tissue S-PrediXcan association; t_i_best: name of best single-tissue S-PrediXcan association; p_i_worst: worst p-value of single-tissue S-PrediXcan association; t_i_worst: name of worst single-tissue S-PrediXcan association.

S12. Summarize S-MultiXcan results. Number of significant associations from S-PrediXcan and S-MultiXcan results for fourteen PUFA traits in UKB-EUR discovery cohort. Novelty of results also shown.

S13. UKB-EUR external EUR replication. Number of variants in common between UKB-EUR discovery cohort and external EUR cohorts FinMetSeq and Kettunnen et al. after munging. Counting variants from UKB-EUR that were replicated at $P < 0.05$.

S14. UKB-EUR multi-ancestry replication. Number of variants in common between UKB-EUR discovery cohort and UKB multi-ancestry African (AFR), Central and South Asian (CSA), and East Asian (EAS) cohorts after QC protocol. Counting variants from UKB-EUR that were replicated at $P < 0.05$.

S15. Meta-analysis summary. Number of significant variants, independent significant variants (from COJO), and significant loci for each of five PUFA traits tested in meta-analysis. Novel loci for each trait and unique novel loci are also shown.

S16. LDSC results for meta-analysis. Number of SNPs in Linkage disequilibrium score regression (LDSC) analysis, heritability (h^2), genomic control (GC; lambda), and LDSC intercept for five meta-analyzed PUFA traits.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.27.22275343; this version posted May 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

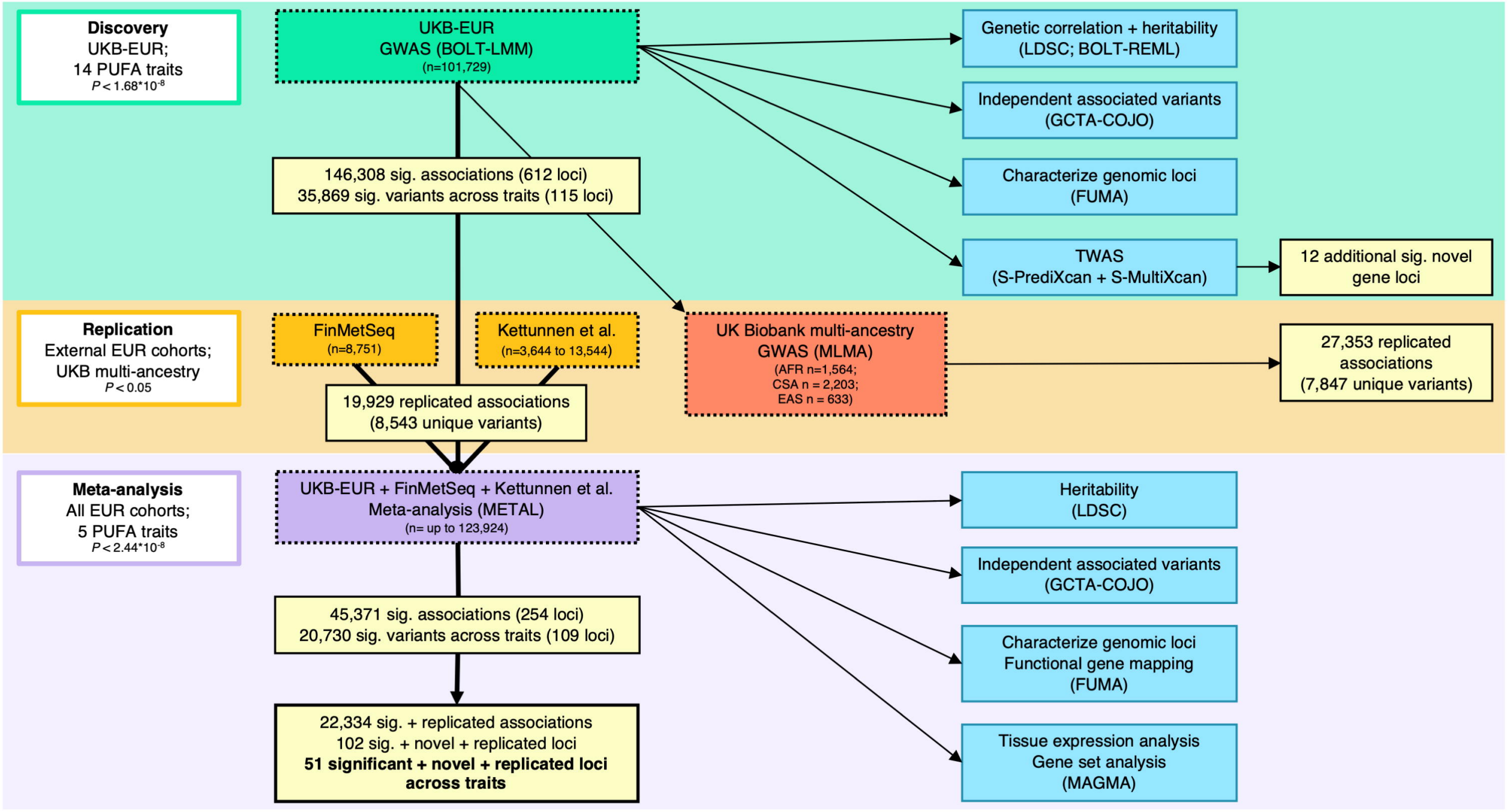
S17. Meta-analysis GCTA-COJO results. Conditional and joint (COJO) analysis using summary statistics from five meta-analyzed PUFA traits. RefA: effect allele; freq: frequency of the effect allele in the original data; b: effect size; se: standard error; p: *p*-value from original GWAS; n: estimated effective sample size; freq_geno: frequency of the effect allele in the reference sample; bJ: effect size from joint analysis of selected SNPs; bJ_se: standard error from joint analysis of selected SNPs; pJ: *p*-value from joint analysis of selected SNPs; LD_r: LD correlation between the SNP i and SNP i + 1 for the SNPs on the list.

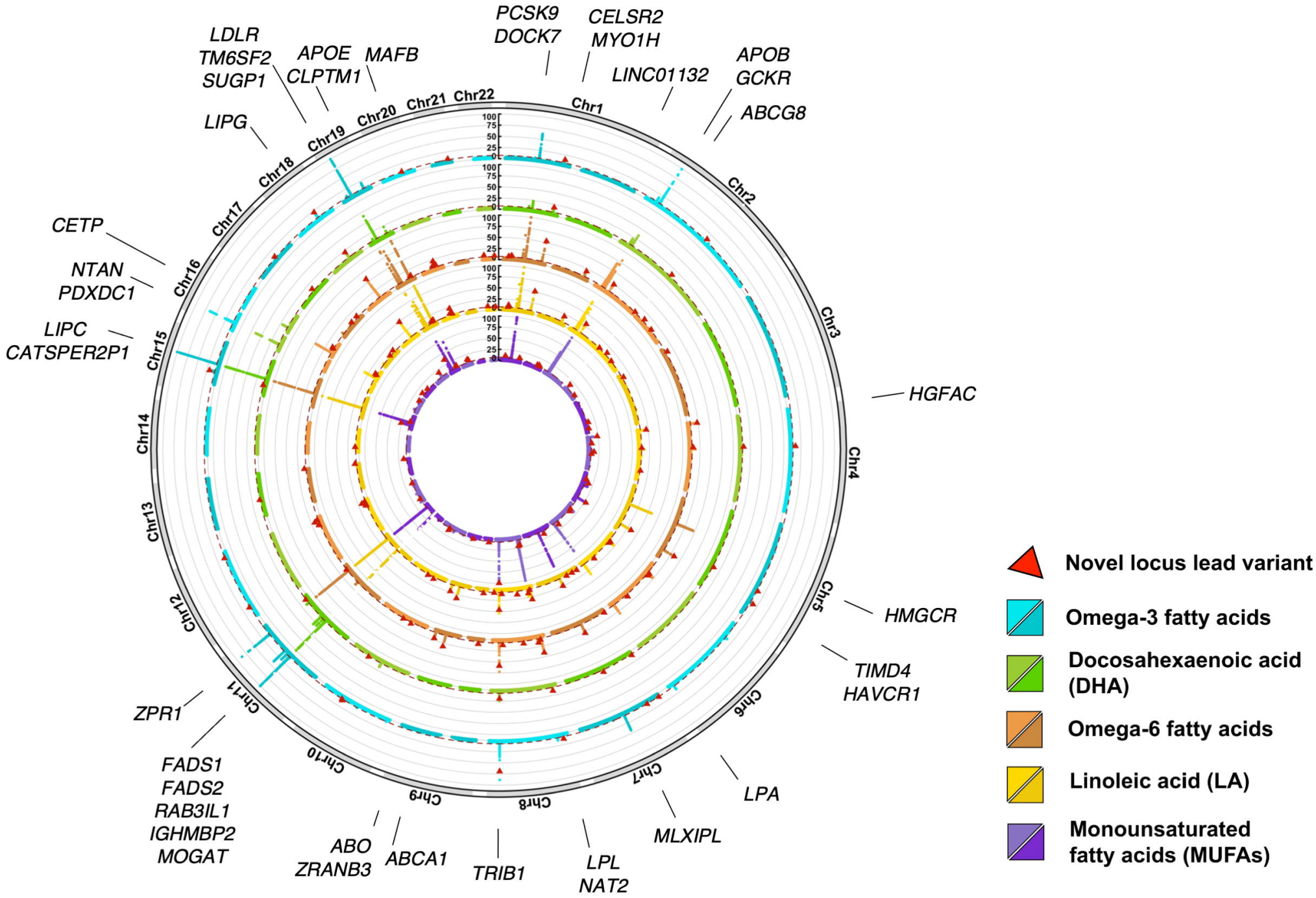
S18. Meta-analysis significant loci. FUMA genomic risk loci from meta-analysis of UKB-EUR, FinMetSeq, and Kettunen et al. studies. Corresponding summary statistics from each study are also shown.

S19. Gene set enrichment. GWAS Catalog gene set enrichment of PUFA-variant associations for five meta-analyzed PUFA traits. Genes in query gene sets mapped from significant meta-analysis associations by position, GTEx liver eQTLs, and HiC liver chromatin data using GENE2FUNC implemented by FUMA.

S20. Gene set enrichment by novelty. GWAS Catalog gene set enrichment, stratified by novelty of PUFA-variant associations for five meta-analyzed PUFA traits. Genes in query gene sets mapped from significant meta-analysis associations by position, GTEx liver eQTLs, and HiC liver chromatin data using GENE2FUNC implemented by FUMA.

S21. GWAS Catalog Accessions. GWAS Catalog accession codes for the full summary statistics generated in this study.


Discovery
UKB-EUR;
14 PUFA traits
P < 1.68*10-8
UKB-EUR
GWAS (BOLT-LMM)
(n=101,729)
146,308 sig. associations (612 loci)
35,869 sig. variants across traits (115 loci)
Genetic correlation + heritability
(LDSC; BOLT-REML)
Independent associated variants
(GCTA-COJO)
Characterize genomic loci
(FUMA)
TWAS
(S-PrediXcan + S-MultiXcan)
12 additional sig. novel
gene loci
Replication
External EUR cohorts;
UKB multi-ancestry
P < 0.05
FinMetSeq
(n=8,751)
Kettunnen et al.
(n=3,644 to 13,544)
19,929 replicated associations
(8,543 unique variants)
UK Biobank multi-ancestry
GWAS (MLMA)
(AFR n=1,564;
CSA n = 2,203;
EAS n = 633)
27,353 replicated
associations
(7,847 unique variants)
Meta-analysis
All EUR cohorts;
5 PUFA traits
P < 2.44*10-8
UKB-EUR + FinMetSeq + Kettunnen et al.
Meta-analysis (METAL)
(n= up to 123,924)
45,371 sig. associations (254 loci)
20,730 sig. variants across traits (109 loci)
22,334 sig. + replicated associations
102 sig. + novel + replicated loci
51 significant + novel + replicated loci
across traits
Heritability
(LDSC)
Independent associated variants
(GCTA-COJO)
Characterize genomic loci
Functional gene mapping
(FUMA)
Tissue expression analysis
Gene set analysis
(MAGMA)


PCSK9
DOCK7
CELSR2
MYO1H
LINC01132
APOB
GCKR
ABCG8
LDLR
TM6SF2
SUGP1
APOE
CLPTM1
MAFB
LIPG
CETP
NTAN
PDXDC1
LIPC
CATSPER2P1
HGFAC
HMGCR
TIMD4
HAVCR1
LPA
MLXIPL
LPL
NAT2
TRIB1
ABCA1
ABO
ZRANB3
FADS1
FADS2
RAB3IL1
IGHMBP2
MOGAT
ZPR1
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
Chr14
Chr15
Chr16
Chr17
Chr18
Chr19
Chr20
Chr21
Chr22
100
75
50
25
0
Novel locus lead variant
Omega-3 fatty acids
Docosahexaenoic acid (DHA)
Omega-6 fatty acids
Linoleic acid (LA)
Monounsaturated fatty acids (MUFAs)