

A regulatory insertion-deletion polymorphism in FADS gene cluster determines PUFA synthesis and influences lipid profiles

Peiqin Li^{1*}, Jing Zhao^{1*}, Kumar S. D. Kothapalli^{2,3}, Xiang Li¹, Hui Li¹, Yuxuan Han¹, Shengquan Mi⁴, Wenhua Zhao⁵, Qizhai Li⁶, Hong Zhang⁷, Yiqing Song⁸, J. Thomas Brenna^{2,3}, Ying Gao^{1**}

¹Key Laboratory of Nutrition and Metabolism, Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, University of Chinese Academy of Sciences, Chinese Academy of Sciences

²Division of Nutritional Sciences, Cornell University, Savage Hall, Ithaca, NY 14853 USA;

³Dell Pediatric Research Institute, Dell Medical School, University of Texas Austin, 1400 Barbara Jordan Blvd, Austin, Texas 78723 USA;

⁴College of Arts and Science of Beijing Union University, Beijing 100191;

⁵National Institute for Nutrition and Health, Chinese Center for Disease Control and Prevention, Beijing, China;

⁶Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China;

⁷Institute of Biostatistics, School of Life Sciences, Fudan University, Shanghai 200438, China;

⁸Department of Epidemiology, Richard M. Fairbanks School of Public Health, Indiana University, Indianapolis, IN 46202;

* These authors had equal contributions.

Last names underlined for Pubmed indexing

** **Address correspondence to:** Ying Gao, 320 Yueyang Road, New Life Science Building, Room A1926, Shanghai, 200031, China. yinggao@sibs.ac.cn; +86 2154920985

This is the author's manuscript of the article published in final edited form as:

Li, P., Zhao, J., Kothapalli, K. S. D., Li, X., Li, H., Han, Y., ... Gao, Y. (2018). A regulatory insertion-deletion polymorphism in the FADS gene cluster influences PUFA and lipid profiles among Chinese adults: a population-based study. *The American Journal of Clinical Nutrition*, 107(6), 867–875. <https://doi.org/10.1093/ajcn/nqy063>

Sources of Funding: This research was supported by the National Key Research and Development Plan (2016YFD0400200 YG) and the 100 Talented Plan of Chinese Academy of Sciences (YG). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Short title: rs66698963 determines PUFA and affects lipids

Abbreviations used: TC, total cholesterol; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; TG, triglycerides; D5D, delta-5 desaturase; D6D, delta-6 desaturase; LCPUFA, long chain polyunsaturated fatty acids

Clinical Trial Registry number: ChiCTR-EOC-17012759

Abstract

Background: Arachidonic acid (AA) is the major polyunsaturated fatty acid (PUFA) substrate for potent eicosanoid signaling to modulate inflammation and thrombosis, and controlled in part by tissue abundance. Fatty acid desaturase 1 (FADS1) catalyzes synthesis of omega-6 AA and omega-3 eicosapentaenoic acid (EPA). The rs66698963 polymorphism, a 22-bp insertion-deletion (Indel) 137-bp downstream of a sterol regulatory element in *FADS2* intron 1, mediates expression of FADS1 in vitro, as well exerting positive selection in several human populations. The associations between polymorphism rs66698963 and plasma PUFA, as well as disease phenotypes are unclear.

Objective: This study aimed to evaluate the relationship between rs66698963 genotypes and plasma PUFA levels, and blood lipid profiles.

Design: Plasma fatty acids were measured from a single sample obtained at baseline in 1,504 healthy Chinese adults aged between 35–59 years using gas chromatography. Blood lipids were measured at baseline and a second time after 18 months follow up. The rs66698963 genotype was determined using agarose gel electrophoresis. Linear regression and logistic regression analyses were performed to assess the association between genotype and plasma PUFA and blood lipids.

Results: A shift from precursors linoleic acid and alpha-linolenic acid to product AA and EPA, respectively, was observed consistent with FADS1 activity increasing in the order of genotypes D/D to I/D to I/I. For I/I compared with D/D carriers, plasma levels of omega-6 AA and the ratio of AA to omega-3 EPA plus docosahexaenoic acid (DHA) were 57% and 32% higher respectively. Carriers of the deletion (D) allele of rs66698963 trended to higher triglycerides

(TG) (β (SE)=0.018(0.009), P=0.05) and lower high-density lipoprotein cholesterol (HDL-C) (β (SE)= -0.008(0.004), P=0.02) than carriers of the insertion (I) allele.

Conclusions: The rs66698963 genotype controls AA and AA to EPA+DHA levels reflecting basal risk for inflammatory and related chronic disease phenotypes, and is correlated with risk of dyslipidemia.

Keywords: dyslipidemia, gene polymorphism, fatty acid desaturases, insertion–deletion, long chain polyunsaturated fatty acids

Introduction

Arachidonic acid (20:4n-6, AA) is the major long chain polyunsaturated fatty acid (LCPUFA) precursor for more than 100 signaling molecules modulating inflammation (1), thrombosis (2), and blood pressure (3), amid a host of other specialized functions. Eicosanoid production is controlled at the level of substrate concentration, where phospholipase catalyzed membranes release free AA to be rapidly oxidized in the first steps of eicosanoid synthesis (5). Nutritional regulation of eicosanoid synthesis has long been tied to intake of the omega-3 LCPUFA eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) which, as available, populate membrane phospholipids and are also released by phospholipases. EPA and DHA compete for eicosanoid synthetic enzymes with AA, effectively lowering the production of AA-derived 2-series eicosanoids and yielding EPA-derived 3-series eicosanoids as well as the more recently described docosanoids, some of which signal for the resolution of inflammation (6).

The delta-5 desaturase (D5D) and delta-6 desaturase (D6D), encoded by *FADS1* and *FADS2* genes, respectively, are the key enzymes in the biosynthesis of long chain polyunsaturated fatty acids (LCPUFA), including AA, EPA, and DHA (1). Earlier studies have consistently found numerous SNPs in *FADS1* and *FADS2* genes, such as rs174537, rs174550, rs174553, rs174570, rs174575, and rs1535 to be associated with blood PUFA levels and lipid profiles (2-11). Moreover, omega-3 LCPUFA modulate peroxisome proliferator activated receptors (PPARs) which promote reverse cholesterol transport (12). Fish oil, rich in EPA and DHA, potently reduces blood triglycerides (TG), as well as total cholesterol (TC) to a minor degree, and increases high-density lipoprotein cholesterol (HDL-C) (13). Dyslipidemia is one of the independent risk factors for cardiovascular diseases (14).

In China, the prevalence of high TC, high TG, and low HDL-C has all increased by about 10%, from 2002 to 2012 (15).

Insertion/deletion(s) (Indels) represent the second most frequent human polymorphisms after SNPs (16). The rs66698963 polymorphism is a 22-bp insertion-deletion variant located 137 bp downstream from a sterol regulatory element in *FADS2* intron 1 identified in lymphoblasts from the International Hapmap project (17). Our previous data showed that the insertion (I) allele was associated with higher *FADS1* expression and higher erythrocyte AA levels in a small study (18). Examination of the 1000 Genomes data showed that this locus exerts positive selection related to ancestral diet: populations with low intake of animal foods, where AA is found, have predominantly the insertion-insertion (I/I) genotype that supports higher expression of *FADS1*, the AA synthase (18). Taken together these in vitro, in vivo, and in silico data are strongly suggesting that the rs66698963 is a causal locus modulating circulating AA levels via endogenous synthesis. They also strong suggest that circulating omega-3 PUFA should follow a similar pattern, but no relationship was found for the omega-3 PUFA so far (18). Here we aimed to use a prospective population-based study to further establish the relationship between rs66698963 genotype and PUFA levels, and explore the association with blood lipid profiles.

Subjects and methods

Study population

Samples were collected from a cohort referred to as *Study on Major Chronic Disease Risk Assessment System and Related Technology Developing and Application* (19). The study was conducted from April 2010 to December 2012, involving more than 7,000 subjects from four

project sites including two rural and two urban areas: Shunyi of Beijing (rural), Tongxiang of Zhejiang (rural), Haidian of Beijing (urban), and Jiaxing of Zhejiang (urban). Eligible participants had to live in the corresponding area for more than one year, had no plan to move within one year, and had no serious diseases of heart, lung, liver, and kidney. Among the four sites, only Shunyi was included in the current study. A total of 2,008 subjects aged between 35 and 59 years were recruited in Shunyi. All participants were interviewed in-person at baseline to collect demographic data, disease history, personal medical conditions, cigarette smoking, drinking, health education and related information. Fasting venous blood was obtained for measurements of plasma fatty acids, and biochemical tests including TC, HDL-C, TG, glucose, and others. Low-density lipoprotein cholesterol (LDL-C) was calculated according to the Friedewald formula: $LDL-C = TC - (HDL-C) - TG/2.2$. For the individuals with $TG \geq 4.5$ mmol/L, LDL-C was defined as missing (20). Non-high-density lipoprotein cholesterol (nHDL-C) was calculated by formula: $nHDL-C = TC - (HDL-C)$ (21). Plasma was separated at 1500g/min centrifugation for 15min. Whole blood and plasma were stored at $-80^{\circ}C$ until used. After about 18 months, 1622 (81%) of the subjects were successfully interviewed again. The information on health, cigarette smoking, drinking, and blood lipid profiles including TC, HDL-C, and TG were obtained again. For the current study, individuals selected: (1) had no history of cardiovascular diseases, malignant neoplasm, or other serious disease; (2) had valid data of both plasma PUFA levels and *FADS2* rs66698963 status. The final count was $n=1,504$ (75%) participants at baseline which were used in all analyses of genotype-plasma fatty acid status and genotype-blood lipid analysis. Among them, follow-up data was available from 1,206 (80%) subjects. (Supplemental Figure 1). Follow-up analyses for genotype blood lipids only

are presented in Supplemental materials only.

The study protocol was approved by the Institutional Review Board of Chinese Center for Disease Control and Prevention. All participants provided written informed consent. This study is registered at Chinese Clinical Trial Registry (Clinical Trial Registry number: ChiCTR-EOC-17012759).

Genotyping rs66698963

Indel rs66698963 is located 137-bp downstream from a sterol regulatory element (SRE) (25). Genomic DNA was extracted from whole blood and used to amplify a 200 base pair fragment flanking the SRE of *FADS2* intron 1 (GenBank Accession# NT_167190.1) by PCR. The following primer pairs were used: forward primer: 5'CAAAGCTGGTCACCACTGC3', reverse primer: 5'TCAGGGAAAAGAGACGTGCT3'. The PCR was performed in a 10ul reaction, consisting of 35 ng genomic DNA, 2.5 pmol of each primer, and 5ul PCR mix (Lifefeng Biotechnology Co. Ltd, Shanghai, China). Thermal Cycling conditions were: initial denaturation at 94 °C for 10 min followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 59 °C for 30 s and extension at 72 °C for 30 s, with a final extension at 72 °C for 10 min. The PCR products were run on 3% agarose gels until the bands were well separated. After staining with ethidium bromide (EB), DNA bands were visualized under UV. The concordance rate was greater than 98.7% based on 11% re-run samples (n = 158). Genotypes of samples were assigned as insertion/insertion (I/I), insertion/deletion (I/D), and deletion/deletion (D/D). Agarose gel image of rs66698963 polymorphism and genotype frequency in our study are shown in Supplemental Figure 2.

Plasma total fatty acid measurement

Gas chromatography/flame ion detector (GC/FID) was used to measure plasma total fatty acids profiles from the fasting blood samples as previously described(22). Briefly, 100ul plasma and 10 ug 1,2-dihexarachidoyl-sn-glycero-3-phosphocholine as internal standard was extracted with dichloromethane/methanol solution(22), and then converted to fatty acid methyl esters(FAMES) with methanol and sulfuric acid(23). After methyl esterification, FAMES were extracted with n-hexane and re-dissolved in isooctane. Analysis was performed by GC/FID (Agilent 6890 GC) equipped with a Supelco SP-2560 capillary column (100m x 0.25mm inside diameter x 0.20um thickness). The plasma level of each fatty acid was expressed as weight percentage of total fatty acids. Samples were organized in 20 batches. Two quality control samples from a standard pool were inserted in each batch. The coefficient of variation (CVs) of QC ranged from 3.69% to 16.58%.

Definition of hyperlipidemia and obesity

According to the Chinese Guidelines on Prevention and Treatment of Dyslipidemia in Adults (2007)(24), hyperlipidemia was defined as total cholesterol ($TC \geq 5.18$ mmol/L), and/or high-density lipoprotein cholesterol ($HDL-C < 1.04$ mmol/L), and/or triglycerides ($TG \geq 1.70$ mmol/L), and/or with history of hyperlipidemia diseases in the past one year. Body mass index (BMI) was defined base on criteria in China: Underweight with $BMI < 18.5$ kg/m², normal weight with $18.5 \leq BMI < 24$ kg/m², overweight with $24 \leq BMI < 30$ kg/m², and obese with $BMI \geq 30$ kg/m². Central obesity (WC) was defined as $WC \geq 90$ cm for men and ≥ 80 cm for women based on the International Diabetes Federation recommendations for Asians(25)

Statistical analysis

One degree of freedom Pearson Chi-squared test was employed to check whether the

Hardy Weinberg equilibrium (HWE) law holds in the total sample. The Kruskal-Wallis test on continuous variables and the Chi-squared test on categorical variables were used to examine the differences in basic characteristics between three genotype groups. Additive genetic model was used in the analysis. Genotypes I/I, I/D, and D/D (I for I-allele and D for D-allele) were coded as 0, 1, and 2, respectively, and were taken as a continuous variable in regression models. P-trend values for lipid profiles and LCPUFA from I/I to I/D to D/D were obtained from linear regression models. To control false positive error, p-trend were corrected by Bonferroni with $\alpha=0.0125$ (4 lipid profiles exploratory comparisons) and $\alpha=0.003$ (16 LCPUFA exploratory comparisons).

Multiple linear regression models were applied to examine the associations of rs66698963 with plasma lipids at baseline. Prior to the multiple linear regressions, TC, TG, HDL-C, LDL-C, and nHDL-C were \log_{10} transformed. Multivariable-adjusted odds ratios (ORs) and 95% confidence interval (CI) were obtained from logistic regression.

Hypertriglyceridemia, hypercholesterolemia, or low-HDL-C was considered as case of dyslipidemia in our study. As the highest frequency genotype, I/I genotype was used as reference. To evaluate the association between rs66698963 with plasma lipids in the 18-month follow-up, logistic regression was performed in the follow-up data of subjects with normal blood lipids at baseline. In the above linear regression and logistic models, several adjustment models were used: 1) model adjusted for age and sex (basic model); 2) model adjusted for sex, age, BMI, smoking, diabetes, hypertension, and farm working (full model); 3) model adjusted for sex, age, BMI, smoking, diabetes, hypertension, farm working, and AA/(EPA+DHA) (exploration model). Confounders in analysis were selected by prior known

and potential risk factors for dyslipidemia. We also used Kruskal-Wallis test and chi-square test to evaluate the differences of the variants between dyslipidemia and non- dyslipidemia in this study. Those variants whose $p < 0.05$ were considered as confounders and were included in the models. Farm working was one of variables that $p < 0.05$ and was suspected to be associated with lipid profiles, so farm working should be considered as a confounder and controlled in our analysis.

To explore the effect modification by plasma PUFA levels and some other factors on the association between rs66698963 and blood lipids, stratified analyses were conducted by sex(male, female), age(<median, \geq median), obesity(yes, no), hypertension(yes, no), current smoking (yes or no, defined by at least once a week over one year), current alcohol drinking (yes or no, defined by at least once a week over one year), farm working (yes or no, defined by doing farm work over one year), physical activity (yes or no, defined by at least once a week over one year), and various plasma PUFA(low, high). It was noted that plasma PUFA were classified into low and high groups based on the median levels in healthy subjects without dyslipidemia at baseline. Interactions were assessed by entering the cross product of genotype and interest modifying factor in the multiple regression models. P-interaction values were calculated by Wald test in linear regression models, and log-likelihood ratio test in logistic models.

All analyses were conducted with the SAS software (version 9.3). All P-values are two-sided. A P-value less than 0.05 was considered as statistically significant.

Results

Baseline characteristics

Participant characteristics (n=1,504) at baseline are shown in **Table 1**. In our sampling (n=1,504), D/D genotype was 11.84% (n = 178), I/D 42.35% (n = 637), and I/I 45.81% (n = 689); the observed allele frequency was D=33% and I=67%, and the genotypes are in HW equilibrium ($P = 0.10$). No obvious differences were found among genotypes for age, education, diabetes, current alcohol drinking, and physical activity. The D/D group had significantly lower HDL-C, and marginally significant ($p = 0.08$) higher TG than the I/I group.

Plasma PUFA levels among the three genotype at baseline

The D/D participants had significantly higher plasma 18:2n-6, 20:2n-6, 20:3n-6, 18:3n-3, and lower 18:3n-6, 20:4n-6, 22:4n-6, 22:5n-6, 20:5n-3, 22:5n-3, 22:6n-3, and AA/EPA+DHA, than those for I/I genotype (all $P_{\text{trend}} < 0.0001$) (**Figure 1**). The rs66698963 D/D was negatively associated with the ratios of 18:3n-6/18:2n-6 and 20:4n-6/20:3n-6, which are often used as a proxy for the activity of D6D and D5D, respectively (**Figure 1**). Compared to D/D genotype, omega-6 AA levels and the AA/EPA + DHA ratio were 57% and 32% increased respectively for individuals of I/I genotype.

Association between genotype and lipid profiles at baseline

At baseline, rs66698963 D/D was significantly associated with higher TG ($\beta(\text{SE})=0.018(0.009)$, $P=0.05$) and lower HDL-C ($\beta(\text{SE})= -0.008(0.004)$, $P=0.02$) in full model (**Table 2**, model 2). To examine whether the balance of AA and EPA+DHA may mediate the effect of rs66698963 on lipid profiles, we further adjusted the models for the

AA/EPA+DHA ratio and the associations were not significant (**Table 2**, model 3). Compared to I/I genotype, odds ratios (ORs) of D/D genotype were 1.10 (95%CI: 0.73, 1.67) and 1.47(95%CI: 1.03, 2.08) for hypertriglyceridemia and low-HDL-C dyslipidemia respectively in full model (**Table 3**).

Stratified analyses

We investigated the combined effects of rs66698963 and some potential modifiers of dyslipidemia on the lipid profiles. Significant interactions were observed for rs66698963 genotypes with hypertension and physical activity (**Supplemental Table 1**). The association between D/D genotype and low-HDL-C dyslipidemia tended to be stronger in subjects without physical activity (OR=1.79; 95%CI: 1.24, 2.6; $P_{\text{interaction}}=0.01$), or in subjects without hypertension (OR=2.09; 95%CI: 1.31, 3.36; $P_{\text{interaction}}=0.002$) (**Figure 2, Supplemental Table 2**). The interactions between rs66698963 and plasma PUFA were also analyzed. The significant associations between rs66698963 D/D and higher TG were restricted to individuals with lower plasma 18:2n-6, DHA, and EPA+DHA ($P_{\text{interaction}}=0.04$, 0.02, and 0.004, respectively) (**Figure 3, Supplemental Table 3**). However, the significant association between D/D genotype and lower HDL-C was only observed when 20:3n-6 and 20:5n-3/18:3n-3 levels were higher ($P_{\text{interaction}}=0.008$ and 0.008 respectively), or when EPA+DHA level was lower ($P_{\text{interaction}}=0.004$) (**Figure 3, Supplemental Table 3**). Lower TC was associated with rs66698963 D/D genotype only when the 20:5n-3/18:3n-3 ratio was higher ($P=0.0003$, $P_{\text{interaction}}=0.0005$) (**Supplemental Table 3**). No interaction was found between rs66698963 and plant-derived PUFA (18:2n-6 or 18:3n-3) on downstream PUFA such as AA,

EPA, and DHA (**Supplemental Table 4**).

Associations between genotype and lipid profiles in 18-month follow-up

For the individuals with normal HDL-C (n=1099) at baseline, we observed an association between rs66698963 D/D genotype and decreased HDL-C level at 18-month follow up (**Supplemental Table 5**). However, the significant association was limited to the individuals without physical activity (OR =2.18; 95%CI: 1.24, 3.82) (**Supplemental Table 6**), and individuals without hypertension (OR=2.77; 95% CI: 1.33, 5.78) (**Supplemental Table 6**).

Discussion

The D/D genotype was associated with lowest AA, DHA, EPA, and AA/EPA+DHA ratio; D/D genotype was also associated with higher TG and lower HDL-C levels. EPA+DHA levels modified the associations of rs66698963 genotype with plasma TG and HDL-C. Additionally, 18:2n-6 or DHA alone modified the association of genotype with plasma TG; while physical activity and hypertension modified the association of genotype with HDL-C.

The rs66698963 Indel polymorphism was strongly associated with plasma omega-6 PUFA. The relationship followed expectations from a precursor-product relationship with respect to FADS1 (Figure 1B, 1E). For omega-6 PUFA, the precursors (18:2 and 20:3) accumulate in the progression from I/I to I/D to D/D; and the products (20:4, 22:4, and 22:5) increase from D/D to I/D to I/I (Figure 1B). Higher 20:4n6/20:3n6 ratio suggests greater D5D desaturase activity for I/I compared with D/D genotype, consistent with our previous studies (17, 18).

For the first time we also observed a significant precursor-product relationship in omega-3 PUFA synthesis, following the pathway 18:3→20:5→22:5→22:6 (Figure 1B, 1C). The magnitudes of the percentage decreases in 20:4n-6 and 20:5n-3 were similar from I/I to D/D consistent with the hypothesis that *FADS1* expression and activity is regulating PUFA concentrations along both pathways similarly, as expected from biochemical competition.

Importantly, the ratio of AA to EPA+DHA increased by 32% from D/D to I/I (Figure 1D) due to a greater increase in AA than EPA and DHA. AA is precursor for mostly inflammatory, clotting signaling eicosanoids. EPA and DHA inhibit synthesis of those compounds and are precursors for resolving and other signaling molecules that oppose the

action of AA. Circulating omega-3 EPA and DHA are robustly associated with reduction of cardiovascular diseases (26). Sources of omega-3 LCPUFA, primarily seafood or marine oil supplements, are recommended for primary prevention of cardiovascular diseases, and for reduction of plasma triacylglycerol (13). Our data suggest that individuals with the D/D genotype may be especially vulnerable to dietary omega-3 insufficiency compared to individuals of I/I genotype. Importantly, controversy over the effects of omega-3 and cardiovascular disease(27-29) may be related to this genetic factor, considering that I/I individuals will have a greater mean basal EPA and DHA level, and therefore be less well differentiated from placebo groups than a supplemented group.

The rs66698963 genotype frequency in our study population was similar to that in India, but different from a small sampling of an urban US population (18). Differences in ethnic, geographical, and dietary intake related evolution between Chinese and Americans are likely to explain the different frequency. Rural inland regions including those in China and India tend to rely more on plants and have low intakes of PUFA rich foods (30) compared to diets of the ancestors of Americans. Populations subsisting for many generations on plant-based diets depend on greater endogenous LCPUFA biosynthesis, favoring the I/I genotype that is adaptive in evolution (18). Our data strongly support the hypothesis that those of I/I genotype consuming a conventional diet would particularly benefit from balanced dietary PUFA.

The Japanese HapMap population has two linkage disequilibrium (LD) blocks in the FADS gene cluster whereas the Chinese (and Europeans) populations have only 1 LD block (17, 31). Minor allele homozygotes (D/D) present in about one quarter of the Japanese population had significantly lower expression of *FADS1* than the I/I homozygotes (17), and

AA was lower in individuals of D/D genotype (18). Previous characterizations of LD were based on SNPs while Indels and other variations in this region have been mis-called or omitted. LD estimates for Indels must be approached with caution unless measured by definitive methods such as in the present work.

The rs66698963 Indel is 137-bp downstream of an SRE, approximately the length of a nucleosome (146-bp) and very likely modifying regulatory protein binding at the SRE. *FADS2* SNPs associated with PUFA and blood lipid profiles (rs174570, rs174575, rs1535) may be tagging the function of rs66698963 Indel polymorphism (32). For example, the rs174570 T allele, 5,400-bp upstream of rs66698963, was associated with reduced LCPUFA in the Greenlandic Inuit population (33). The rs174570 derived “T” and ancestral “C” allele tag the deletion and insertion alleles of the 22-bp rs66698963 Indel, respectively. Further, rs174570 “T” allele frequency is 0.99 in Greenlandic Inuit and 0.34 in CHB (33), the Indel rs66698963 D allele frequency (0.33) in the present study is similar to that in CHB. The rs174570 ancestral “C” allele exhibited strong LD with the insertion allele of rs66698963, which is adaptive and showed positive selection signals in South Asians ($nSL = -3.77$, $P = 0.00048$) and East Asians ($iHS = -2.41$, $P = 0.010$; $nSL = -2.41$, $P = 0.012$) (18). Combined with previous work, our results provide solid evidence indicating that *FADS2* rs66698963, is a causal variant influencing PUFA and through them, blood lipid levels. Future studies focused on the association of classical SNP with Indel rs66698963 in specific populations would enable high throughput SNP analyses to serve as proxies for the Indel, which is currently measured by Sanger sequencing.

The D/D genotype had 4.8% lower HDL-C and 7.5% higher TG than I/I genotype, in the

direction of increased dyslipidemia risk. EPA+DHA supplementation is a well-accepted treatment for hypertriglyceridemia. Our data also show concordance of increased EPA+DHA and lower TG. GWAS and genetic studies show that polymorphisms related to reductions in circulating EPA and DHA or increases in AA are associated with dyslipidemia (34). Several potential mechanisms link rs66698963 and blood lipids: (1) Lower omega-3 LCPUFA due to the D/D genotype may result in decreased activation of PPARs, which subsequently decreases the expression HDL-C component genes(21), including apo -lipoproteins (Apo) A-I, A-II, C-III, and lipoprotein lipase (4, 35), and results in low-HDL-C; (2) omega-3 PUFA regulate the activity of nuclear receptors, reducing TG synthesis and decreasing very-low-density lipoprotein (VLDL) secretion (36). Moreover, our data showed significant positive associations of AA with TG, and negative association of AA with HDL-C (data not shown). AA/EPA+DHA may mediate the association between rs66698963 and lipid profiles. This hypothesis is supported by the abolishment of rs66698963-lipid association after adjustment for AA/EPA+DHA.

FADS gene variants may interact with dietary omega-3 or omega-6 on lipid traits (20, 21). The association between rs66698963 D/D genotype and higher TG was stronger when 18:2n-6, DHA, and EPA+DHA levels were low. A striking association was also observed between rs66698963 D/D genotype and lower HDL-C when EPA+DHA level was low. The results are consistent with *FADS1*-rs174550 genotype association with lower HDL-C when 18:2n-6 and 18:3n-3 were low (4). Higher EPA+DHA intake lowers TG by down-regulating lipogenic enzymes, enhancing intra hepatic degradation of Apo-lipoprotein B (Apo-B) and reducing activity of lipolytic enzymes (37). The effect of rs66698963 polymorphism on

circulating lipid profiles is likely to depend on dietary PUFA intake.

This is the first study to examine the association between the Indel rs66698963 polymorphism and disease biomarkers. The main strength of our study is that all outcomes are measured rather than estimated indirectly. Several limitations should be considered. Anti-hyperlipidemia medication usage was not available, though according to Chinese National Nutrition and Health Survey in 2010-2013, the awareness rates (8.3%), treatment rates (7.0%), and detection rates (26.2%) are very low in rural areas of China(15).

In conclusion, the present study shows that the rs66698963 Indel was strongly associated with circulating LCPUFA levels. The D/D genotype was associated with higher TG and lower HDL-C, modifiable by plasma levels of 18:2n-6, DHA, EPA+DHA, and physical activity. Together with previous evidence for positive selection for this specific Indel polymorphism, these findings suggest that this locus is pivotal for the relationship between diet and endogenous LCPUFA levels.

Acknowledgments: We express our sincere appreciation to all of the study participants and research staff involved in the field work.

Author Contributions: Ying Gao and J. T. Brenna conceived the study. Ying Gao designed the study. Peiqin Li performed the experiment and obtained data. Jing Zhao and Peiqin Li analyzed the data. Peiqin Li, Jing Zhao, and Ying Gao wrote the manuscript. All authors were involved in the interpretation of the results. All authors read, gave comments, and approved the final version of the manuscript. All authors had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Conflict of Interest Statement: The authors have no conflicts of interest to declare.

References

1. Park HG, Park WJ, Kothapalli KS, Brenna JT. The fatty acid desaturase 2 (FADS2) gene product catalyzes Delta4 desaturation to yield n-3 docosahexaenoic acid and n-6 docosapentaenoic acid in human cells. *FASEB J* 2015;29(9):3911-9. doi: 10.1096/fj.15-271783.
2. Tanaka T, Shen J, Abecasis GR, Kisialiou A, Ordovas JM, Guralnik JM, Singleton A, Bandinelli S, Cherubini A, Arnett D, et al. Genome-wide association study of plasma polyunsaturated fatty acids in the InCHIANTI Study. *PLoS Genet* 2009;5(1):e1000338. doi: 10.1371/journal.pgen.1000338.
3. Sone Y, Kido T, Ainuki T, Sonoda M, Ichi I, Kodama S, Sone H, Kondo K, Morita Y, Egawa S, et al. Genetic variants of the fatty acid desaturase gene cluster are associated with plasma LDL cholesterol levels in Japanese males. *J Nutr Sci Vitaminol (Tokyo)* 2013;59(4):325-35.
4. Zhu J, Sun Q, Zong G, Si Y, Liu C, Qi Q, Ye X, Sun L, Sheng H, Li H, et al. Interaction between a common variant in FADS1 and erythrocyte polyunsaturated fatty acids on lipid profile in Chinese Hans. *J Lipid Res* 2013;54(5):1477-83. doi: 10.1194/jlr.P027516.
5. Kwak JH, Paik JK, Kim OY, Jang Y, Lee SH, Ordovas JM, Lee JH. FADS gene polymorphisms in Koreans: association with omega6 polyunsaturated fatty acids in serum phospholipids, lipid peroxides, and coronary artery disease. *Atherosclerosis* 2011;214(1):94-100. doi: 10.1016/j.atherosclerosis.2010.10.004.
6. Lemaitre RN, Tanaka T, Tang W, Manichaikul A, Foy M, Kabagambe EK, Nettleton JA, King IB, Weng LC, Bhattacharya S, 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):e1002193. doi: 10.1371/journal.pgen.1002193.
7. Aslibekyan S, Jensen MK, Campos H, Linkletter CD, Loucks EB, Ordovas JM, Deka R, Rimm EB, Baylin A. Fatty Acid desaturase gene variants, cardiovascular risk factors, and myocardial infarction in the costa rica study. *Front Genet* 2012;3:72. doi: 10.3389/fgene.2012.00072.
8. Lemaitre RN, Sitlani C, Song X, King IB, McKnight B, Spiegelman D, Sacks FM, Djousse L, Rimm EB, Siscovick DS, et al. Circulating and dietary alpha-linolenic acid and incidence of congestive heart failure in older adults: the Cardiovascular Health Study. *Am J Clin Nutr* 2012;96(2):269-74. doi: 10.3945/ajcn.112.037960.
9. Li SW, Wang J, Yang Y, Liu ZJ, Cheng L, Liu HY, Ma P, Luo W, Liu SM. Polymorphisms in FADS1 and FADS2 alter plasma fatty acids and desaturase levels in type 2 diabetic patients with coronary artery disease. *J Transl Med* 2016;14:79. doi: 10.1186/s12967-016-0834-8.
10. Takkenen MJ, de Mello VD, Schwab US, Kuusisto J, Vaitinen M, Agren JJ, Laakso M, Pihlajamaki J, Uusitupa MI. Gene-diet interaction of a common FADS1 variant with marine polyunsaturated fatty acids for fatty acid composition in plasma and erythrocytes among men. *Mol Nutr Food Res* 2016;60(2):381-9. doi: 10.1002/mnfr.201500594.
11. Wu Y, Zeng L, Chen X, Xu Y, Ye L, Qin L, Chen L, Xie L. Association of the FADS gene cluster with coronary artery disease and plasma lipid concentrations in the northern Chinese Han population. *Prostaglandins Leukot Essent Fatty Acids* 2017;117:11-6. doi: 10.1016/j.plefa.2017.01.014.
12. Dossi CG, Tapia GS, Espinosa A, Videla LA, D'Espessailles A. Reversal of high-fat diet-induced hepatic steatosis by n-3 LCPUFA: role of PPAR-alpha and SREBP-1c. *J Nutr Biochem* 2014;25(9):977-84. doi: 10.1016/j.jnutbio.2014.04.011.
13. Zhu W, Dong C, Du H, Zhang H, Chen J, Hu X, Hu F. Effects of fish oil on serum lipid profile in dialysis patients: a systematic review and meta-analysis of randomized controlled trials. *Lipids Health*

- Dis 2014;13:127. doi: 10.1186/1476-511X-13-127.
14. Nordestgaard BG, Varbo A. Triglycerides and cardiovascular disease. *Lancet* 2014;384(9943):626-35. doi: 10.1016/S0140-6736(14)61177-6.
 15. Chang J, Wang Y. Chinese National Nutrition and Health Survey in 2010-2013. Peking University Medical Press 2016;Beijing.
 16. Ratan A, Olson TL, Loughran TP, Jr., Miller W. Identification of indels in next-generation sequencing data. *BMC Bioinformatics* 2015;16:42. doi: 10.1186/s12859-015-0483-6.
 17. Reardon HT, Zhang J, Kothapalli KS, Kim AJ, Park WJ, Brenna JT. Insertion-deletions in a FADS2 intron 1 conserved regulatory locus control expression of fatty acid desaturases 1 and 2 and modulate response to simvastatin. *Prostaglandins Leukot Essent Fatty Acids* 2012;87(1):25-33. doi: 10.1016/j.plefa.2012.04.011.
 18. Kothapalli KS, Ye K, Gadgil MS, Carlson SE, O'Brien KO, Zhang JY, Park HG, Ojukwu K, Zou J, Hyon SS, 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. doi: 10.1093/molbev/msw049.
 19. Chen XR, Zhang J, Ding GQ, Dong Z, Zhang XW, Li JH, Chen B, Yan LX, Mi SQ, Zhao WH. Negative Association of Domestic Activity and Active Commuting with Metabolic Syndrome in a Chinese Population Aged 35-64 Years. *Biomed Environ Sci* 2015;28(7):486-94. doi: 10.3967/bes2015.070.
 20. Hellstrand S, Sonestedt E, Ericson U, Gullberg B, Wirfalt E, Hedblad B, Orho-Melander M. Intake levels of dietary long-chain PUFAs modify the association between genetic variation in FADS and LDL-C. *J Lipid Res* 2012;53(6):1183-9. doi: 10.1194/jlr.P023721.
 21. Lu Y, Feskens EJ, Dolle ME, Imholz S, Verschuren WM, Muller M, Boer JM. Dietary n-3 and n-6 polyunsaturated fatty acid intake interacts with FADS1 genetic variation to affect total and HDL-cholesterol concentrations in the Doetinchem Cohort Study. *Am J Clin Nutr* 2010;92(1):258-65. doi: 10.3945/ajcn.2009.29130.
 22. Han X, Rozen S, Boyle SH, Hellegers C, Cheng H, Burke JR, Welsh-Bohmer KA, Doraiswamy PM, Kaddurah-Daouk R. Metabolomics in early Alzheimer's disease: identification of altered plasma sphingolipidome using shotgun lipidomics. *PLoS One* 2011;6(7):e21643. doi: 10.1371/journal.pone.0021643.
 23. Sun Q, Ma J, Campos H, Hankinson SE, Hu FB. Comparison between plasma and erythrocyte fatty acid content as biomarkers of fatty acid intake in US women. *Am J Clin Nutr* 2007;86(1):74-81.
 24. Joint committee for developing Chinese guidelines on prevention and treatment of dyslipidemia in adults. Chinese Guidelines on Prevention and Treatment of Dyslipidemia in Adults (2007). *Chin J Cardiol* 2007;35(5).
 25. Alberti KG, Zimmet P, Shaw J, Group IDFETFC. The metabolic syndrome--a new worldwide definition. *Lancet* 2005;366(9491):1059-62. doi: 10.1016/S0140-6736(05)67402-8.
 26. Siscovick DS, Barringer TA, Fretts AM, Wu JH, Lichtenstein AH, Costello RB, Kris-Etherton PM, Jacobson TA, Engler MB, Alger HM, et al. Omega-3 Polyunsaturated Fatty Acid (Fish Oil) Supplementation and the Prevention of Clinical Cardiovascular Disease: A Science Advisory From the American Heart Association. *Circulation* 2017;135(15):e867-e84. doi: 10.1161/CIR.0000000000000482.
 27. Galli C, Brenna JT. Omega-3 fatty acid supplementation and cardiovascular disease events. *JAMA* 2013;309(1):28-9. doi: 10.1001/jama.2012.116648.

28. Rizos EC, Ntzani EE, Bika E, Kostapanos MS, Elisaf MS. Association between omega-3 fatty acid supplementation and risk of major cardiovascular disease events: a systematic review and meta-analysis. *JAMA* 2012;308(10):1024-33. doi: 10.1001/2012.jama.11374.
29. Chowdhury R, Warnakula S, Kunutsor S, Crowe F, Ward HA, Johnson L, Franco OH, Butterworth AS, Forouhi NG, Thompson SG, et al. Association of dietary, circulating, and supplement fatty acids with coronary risk: a systematic review and meta-analysis. *Ann Intern Med* 2014;160(6):398-406. doi: 10.7326/M13-1788.
30. Nath R, Luan Y, Yang W, Yang C, Chen W, Li Q, Cui X. Changes in Arable Land Demand for Food in India and China: A Potential Threat to Food Security. *Sustainability* 2015;7(5):5371-97. doi: 10.3390/su7055371.
31. Tian Y, Zhang W, Zhao S, Sun Y, Bian Y, Chen T, Du Y, Zhang J, Wang Z, Huang T, et al. FADS1-FADS2 gene cluster confers risk to polycystic ovary syndrome. *Sci Rep* 2016;6:21195. doi: 10.1038/srep21195.
32. Lu JT, Wang Y, Gibbs RA, Yu F. Characterizing linkage disequilibrium and evaluating imputation power of human genomic insertion-deletion polymorphisms. *Genome Biol* 2012;13(2):R15. doi: 10.1186/gb-2012-13-2-r15.
33. Fumagalli M, Moltke I, Grarup N, Racimo F, Bjerregaard P, Jorgensen ME, Korneliussen TS, Gerbault P, Skotte L, Linneberg A, et al. Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science* 2015;349(6254):1343-7. doi: 10.1126/science.aab2319.
34. Davidson MH. Omega-3 fatty acids: new insights into the pharmacology and biology of docosahexaenoic acid, docosapentaenoic acid, and eicosapentaenoic acid. *Curr Opin Lipidol* 2013;24(6):467-74. doi: 10.1097/MOL.000000000000019.
35. Standl M, Lattka E, Stach B, Koletzko S, Bauer CP, von Berg A, Berdel D, Kramer U, Schaaf B, Roder S, et al. FADS1 FADS2 gene cluster, PUFA intake and blood lipids in children: results from the GINIplus and LISAplus studies. *PLoS One* 2012;7(5):e37780. doi: 10.1371/journal.pone.0037780.
36. Harris WS. Fish oils and plasma lipid and lipoprotein metabolism in humans: a critical review. *J Lipid Res* 1989;30(6):785-807.
37. Chaudhary R, Saadin K, Bliden KP, Harris WS, Dinh B, Sharma T, Tantry US, Gurbel PA. Risk factors associated with plasma omega-3 fatty acid levels in patients with suspected coronary artery disease. *Prostaglandins Leukot Essent Fatty Acids* 2016;113:40-5. doi: 10.1016/j.plefa.2016.08.009.

Table1. Population characteristics according to three genotypes of the Indel polymorphism at baseline

Characteristics ¹	I/I(n=689)	I/D(n=637)	D/D(n=178)	P ²	P ³ for trend
Sex				0.002	
male	228(33.09)	187(29.36)	77(43.26)		
female	461(66.91))	450(70.64)	101(56.74)		
Age(years)	49(45-54)	49(44-54)	48(43-53)	0.36	
BMI(kg/m²)	26.0(23.5-28.6)	25.86(23.7-28.5)	25.79(23.3-28.5)	0.99	
WC(cm)	85.2(77.8-93.2)	84.5(78.1-91.7)	86.0(78.2-94.1)	0.58	
Education				0.09	
illiteracy	25(3.63)	11(1.73)	2(1.12)		
Primary School	67(9.72)	81(12.72)	26(14.61)		
Middle School	451(65.46)	393(61.70)	115(64.61)		
High School	139(20.17)	142(22.29)	32(17.98)		
College	7(1.02)	10(1.57)	3(1.69)		
Current smoking	153(22.21)	129(20.25)	41(23.03)	0.59	
Current alcohol drinking	167(24.24)	156(24.49)	58(32.58)	0.06	
Obesity	95(13.79)	90(14.13)	28(15.73)	0.75	
Diabetes	12(1.74)	10(1.57)	6(3.37)	0.28	
Hypertension	301(43.69)	314(49.29)	77(43.26)	0.09	
Farm working	351(50.94)	305(47.72)	91(51.12)	0.46	
Physical activity	91(13.21)	81(12.72)	21(11.80)	0.88	
Lipid profiles					
Total Cholesterol, mmol/l	5.18(4.66-5.84)	5.18(4.58-5.72)	5.04(4.37-5.76)	0.14	0.60
Triglyceride, mmol/l	1.07(0.75-1.58)	1.17(0.77-1.70)	1.15(0.75-1.63)	0.08	0.25
HDL Cholesterol, mmol/l	1.45(1.26-1.70)	1.41(1.22-1.67)	1.38(1.17-1.61)	0.01*	0.02
LDL cholesterol (mmol/l)	1.35(1.02-1.71)	1.37(1.04-1.72)	1.29(0.99-1.75)	0.84	1

Abbreviation: D/D, homozygous for the deletion allele; I/D: heterozygous for the deletion allele; I/I, homozygous for the insertion allele; BMI, Body Mass index; WC, Waist Circumference

¹Continuous variables were expressed as median (interquartile) and categorical variables were expressed as frequency (percentage among I/I, I/D and D/D groups).

²Overall p values were calculated from Kruskal-Wallis test for continuous variables and χ^2 -test for categorical variables. * D/D vs I/I, P<0.05

³P for trend values were obtained from linear regression using the additive model (increasing copy of FADS2-rs66698963 D-allele), and were adjusted by Bonferroni correction for lipid profiles.

Table2. Associations of FADS2 rs66698963 with blood lipids at baseline¹

Blood lipids	β(SE)	<i>P</i> value
TC		
Model 1	-0.004(0.003)	0.13
Model 2	-0.004(0.003)	0.11
Model 3	-0.004(0.003)	0.16
TG		
Model 1	0.020(0.009)	0.04
Model 2	0.018(0.009)	0.05
Model 3	-0.015(0.009)	0.11
HDL-C		
Model 1	-0.009(0.004)	0.02
Model 2	-0.008(0.004)	0.02
Model 3	0.0005(0.004)	0.9
LDL-C		
Model 1	-0.001(0.007)	0.83
Model 2	-0.002(0.006)	0.73
Model 3	-0.005(0.007)	0.41
nHDL-C		
Model 1	-0.002(0.004)	0.56
Model 2	-0.003(0.004)	0.42
Model 3	-0.005(0.004)	0.19

LDL-C/HDL-C		
Model 1	0.007(0009)	0.45
Model 2	0.006(0.008)	0.50
Model 3	-0.005(0.009)	0.54
HDL-C/TG		
Model 1	-0.30(0.012)	0.016
Model 2	-0.026(0.011)	0.02
Model 3	0.016(0.012)	0.18

¹TC, TG, HDL-C, LDL-C, nHDL-C, LDL-C/HDL-C, and HDL-C/TG were log₁₀-transformed before analyses. β (SE) were obtained from multiple linear regression in additive model and represented changes in blood lipids for increasing number of D-alleles.

Model 1: basic model adjusting for age and sex

Model 2: full model adjusting for sex, age, BMI, smoking, diabetes, hypertension, and farm working.

Model 3: model adjusting for sex, age, BMI, smoking, diabetes, hypertension, farm working, and

AA/ EPA+DHA

Table3 The association of FADS2 rs66698963 and risk of dyslipidemia at baseline.

	genotype	N	OR(95%CI) ¹	OR(95%CI) ¹
		(case/total)	(basic model ²)	(full model ³)
TC	I/I	348/689	1	1
	I/D	322/637	1.006 (0.81,1.25)	0.98 (0.79,1.23)
	D/D	81/178	0.87 (0.62,1.21)	0.87 (0.62,1.22)
TG	I/I	144/689	1	1
	I/D	162/637	1.33 (1.03,1.72)	1.33 (1.02,1.75)
	D/D	41/178	1.11 (0.74,1.65)	1.10 (0.73,1.67)
HDL-C	I/I	291/689	1	1
	I/D	316/637	1.38 (1.11,1.72)	1.41 (1.12,1.77)
	D/D	95/178	1.48 (1.06,2.06)	1.47 (1.03,2.08)

Case outcomes were defined as hypercholesterolemia ($TC \geq 5.18$ mmol/l), and/or hypertriglyceridemia ($TG \geq 1.70$ mmol/l), and/or low high-density lipoprotein cholesterol (HDL-C < 1.04 mmol/l).

¹The odds ratio (OR) of hyper-TC, hyper-TG, and lower -HDL-C were obtained from logistic regression.

²Basic model: adjusted for age and sex;

³Full model: adjusted for sex, age, BMI, smoking, diabetes, hypertension, and farm working.

Figures with Figure Legends

Figure 1: The association between rs66698963 and plasma fatty acids at baseline.

Percent changes relative to the D/D were calculated for each fatty acids. (A) omega-6 PUFA (B) LCPUFA pathway (C) omega-3 PUFA (D) the ratio of AA to EPA+DHA, and (E) the ratio of product to precursor which reflected D6D, D5D, and aggregate desaturase activity in omega-6 PUFA pathway and omega-3PUFA pathway. Kruskal-Wallis was used to test the difference PUFA level among three genotype groups, * D/D vs I/I, $P < 0.05$; **D/D or I/D vs I/I, $P < 0.05$. P trend values were expressed as the linear association of plasma PUFA level and genotype from I/I to I/D to D/D and were obtained from simple linear regression. P trend values were adjusted by Bonferroni correction. All p trend are significant ($p < 0.001$).

Figure 2: Association of *FADS2*-rs66698963 and risk of dyslipidemia by potential modifiers at baseline.

Stratification analysis was conducted in (A-C) physical activity and (D-F) hypertension. The odds ratios (ORs) of hypercholesterolemia, hypertriglyceridemia, and lower-HDL-C were obtained from logistic regression. ORs were adjusted for sex, age, BMI, smoking, diabetes, hypertension, and farm working. * represents $P < 0.05$ in D/D compared with the I/I genotype (referent category)

Figure 3: Interactions of *FADS2*- rs66698963 and plasma PUFA on blood lipids at baseline.

Plasma fatty acids were classified into high vs low levels based on the median levels in subjects without dyslipidemia. Linear regression was conducted in (A-C) 18:2n6, (D-F) DHA, and (G-I) EPA plus DHA. P values for additive model per each group were indicated as P-low and P-high. P for interaction values were adjusted for sex, age, BMI, smoking, diabetes, hypertension, and farm working using an additive model. Data were expressed as geometric mean.

Fig1

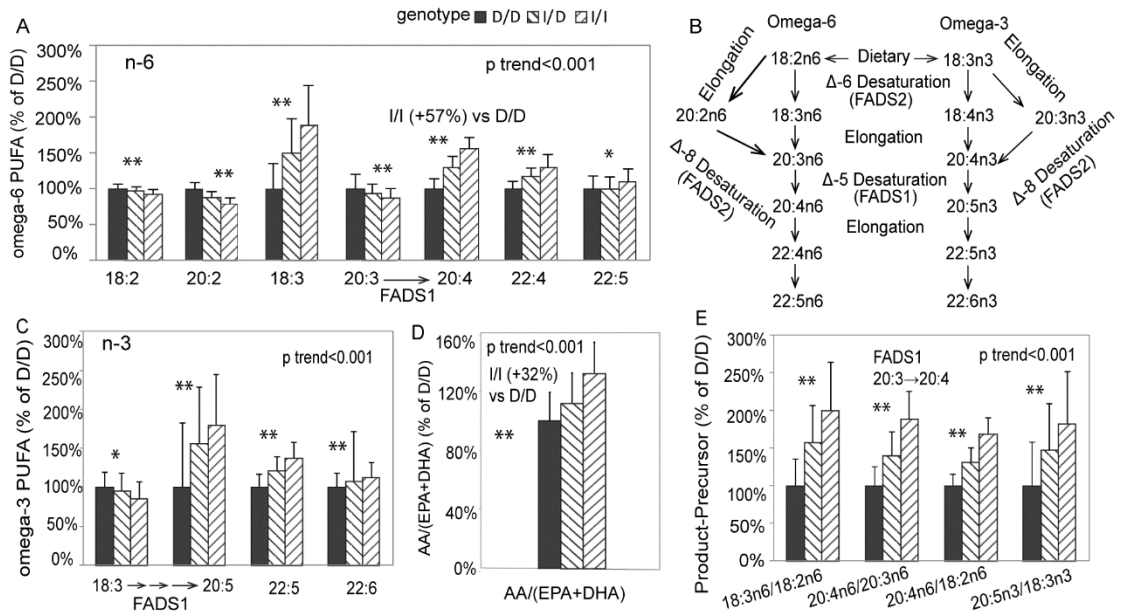


Fig2

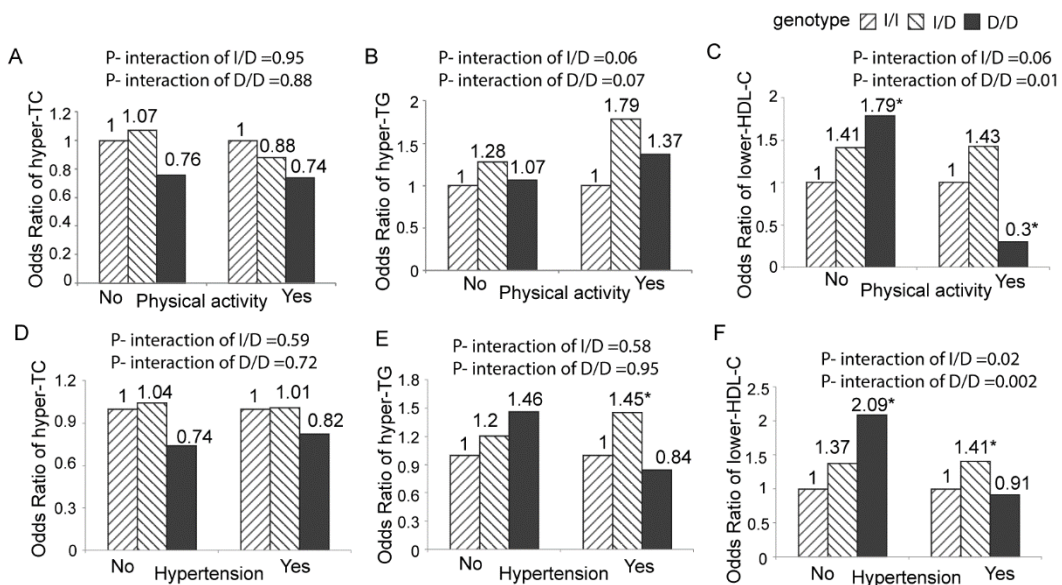


Fig3

