Genes & Nutrition

RESEARCH

Open Access

Decomposed interaction testing improves detection of genetic modifiers of the relationship of dietary omega-3 fatty acid intake and its plasma biomarkers with hsCRP in the UK Biobank

Kenneth E. Westerman[1,2,3*], Chirag J. Patel[4], James B. Meigs[2,3,5], Daniel I. Chasman[6,7,8,9] and Alisa K. Manning[1,2,3]

Abstract

Discovery and translation of gene-environment interactions (GxEs) influencing clinical outcomes is limited by low statistical power and poor mechanistic understanding. Molecular omics data may help address these limitations, but their incorporation into GxE testing requires principled analytic approaches. We focused on genetic modification of the established mechanistic link between dietary long-chain omega-3 fatty acid (dN3FA) intake, plasma N3FA (pN3FA), and chronic inflammation as measured by high sensitivity CRP (hsCRP). We considered an approach that decomposes the overall genetic effect modification into components upstream and downstream of a molecular mediator to increase the potential to discover gene-N3FA interactions. Simulations demonstrated improved power of the upstream and downstream tests compared to the standard approach when the molecular mediator for many biologically plausible scenarios. The approach was applied in the UK Biobank ($N = 188{,}700$) with regression models that used measures of dN3FA (based on fish and fish oil intake), pN3FA (% of total fatty acids measured by nuclear magnetic resonance), and hsCRP. Mediation analysis showed that pN3FA fully mediated the dN3FA-hsCRP main effect relationship. Next, we separately tested modification of the dN3FA-hsCRP ("standard"), dN3FA-pN3FA ("upstream"), and pN3FA-hsCRP ("downstream") associations. The known *FADS1-3* locus variant rs174535 reached $p = 1.6 \times 10^{-12}$ in the upstream discovery analysis, with no signal in the downstream analysis ($p = 0.94$). It would not have been prioritized based on a naïve analysis with dN3FA exposure and hsCRP outcome ($p = 0.097$), indicating the value of the decomposition approach. Gene-level enrichment testing of the genome-wide results further prioritized two genes from the downstream analysis, *CBLL1* and *MICA*, with links to immune cell counts and function. In summary, a molecular mediator-focused interaction testing approach enhanced statistical power to identify GxEs while homing in on relevant sub-components of the dN3FA-hsCRP pathway.

Keywords Gene-diet interaction, Nutrigenetics, PUFA, Omega-3, Inflammation, Mediation

*Correspondence:
Kenneth E. Westerman
kewesterman@mgb.org
Full list of author information is available at the end of the article

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Background

Gene-environment interactions (GxEs), in which the relationship between an exposure and outcome is modified by a genetic variant, continue to hold promise for the development of more precise clinical and behavioral interventions for the prevention of cardiometabolic disease. Compared to intervention trials, observational datasets allow for the exploration of GxEs in much larger sample sizes and with a broader array of exposures and thus a greater opportunity to uncover new biology. However, these analyses are still limited in statistical power [9].

Molecular omics data are being increasingly collected in observational cohort studies, with promise to add mechanistic insight and dynamic longitudinal measures to genotype-only analyses. This potential applies directly to some of the challenges present in GxE studies. Molecular data can increase power for detection by acting as objective proxies for exposures (e.g., lifestyle or environment) that otherwise require noisy estimates based on self-report [20]. They also represent intermediates on biological pathways linking genotypes and exposures to outcomes of interest, allowing for mechanistically informative analyses [18]. However, such molecular data require associated analytical approaches to coherently incorporate them into GxE testing.

A particularly straightforward example of such a mediated pathway is the relationship between dietary long-chain omega-3 fatty acid (dN3FA) intake, plasma N3FA (pN3FA), and chronic inflammation [6]. dN3FA intake, coming primarily from oily fish and fish oil supplements, is a primary determinant of pN3FA, which acts as a physiological N3FA status indicator and is sometimes used as a dietary intake proxy. pN3FA is consistently associated with lower hsCRP, and long-chain N3FA supplementation has been shown to decrease the chronic inflammatory biomarker high-sensitivity C-reactive protein (hsCRP), though this finding is less consistent. Genetic variation, especially at the fatty acid desaturase (*FADS*) locus, may play a role in this variable response to dN3FA [23].

We reasoned that a decomposed interaction testing approach, separating an exposure-outcome pathway into components upstream and downstream of a molecular mediator, would increase the potential to discover associated gene-N3FA interactions impacting hsCRP. We first use simulations to illustrate the expected gain in statistical power to detect interactions using this approach. We proceed to explore genetic modification of the dN3FA-hsCRP relationship after its decomposition into dN3FA-pN3FA (upstream) and pN3FA-hsCRP (downstream) sub-pathways, uncovering interactions at variants and genes that would not have been found using a standard GxE testing approach.

Methods

Simulations

Simulations were performed to understand how statistical power for interaction detection changes when varying a small set of underlying parameters governing the strength of the interactions, mediating pathways, and measurement error. First, genotypes (G) were generated from a binomial distribution with a minor allele frequency of 0.25 and exposures (E) were generated from a standard normal distribution. Next, a mediator (M) was generated, incorporating signal from a G*E product term (for upstream simulations) or from an E main effect (for downstream simulations). From this, a "measured" mediator ($M_{measured}$) was generated, tracking the true M but with added noise according to a specified intraclass correlation coefficient (ICC_M). Finally, an outcome (Y) was generated, incorporating signal from a main effect of M (for upstream simulations) or from a G*M product term (for downstream simulations). For each scenario, interaction terms were tested for significance in two of three regressions: $Y \sim G + E + GE$ (standard), $M_{measured} \sim G + E + GE$ (upstream only), and $Y \sim G + M + GM_{measured}$ (downstream only). Finally, power for each scenario was calculated as the fraction of tests passing the chosen significance threshold out of the total number of repeated simulations.

For simplicity, in the primary set of simulations, the sample size was fixed at $N = 1{,}000$, the significance threshold was set to 0.05, and the number of repeated simulations per scenario was fixed at 500. For upstream simulations, three parameters were varied: the proportion of variance in M explained by the G×E interaction, the proportion of variance in Y explained by M, and the measurement error in M (ICC_M). For downstream simulations, three parameters were varied: the proportion of variance in M explained by E, the proportion of variance in Y explained by the G×M interaction, and the measurement error in M (ICC_M). We note that, despite the small sample size used here compared to biobank datasets, the relative changes in power based on simulated interaction strength, degree of pathway mediation, and measurement error should be consistent.

UK Biobank population and genotype data

The primary analysis was conducted under a Not Human Subjects Research determination for UKB data analysis (NHSR-4298 at the Broad Institute of MIT and Harvard) and UK Biobank application 27,892. UKB is a large prospective cohort with both deep phenotyping and molecular data, including genome-wide genotyping, on over 500,000 individuals ages 40–69 living throughout the UK between 2006–2010 [25]. Genotyping, imputation, and initial quality control on the genetic dataset have

been described previously [5]. We used a multi-ancestry sample of individuals that had not withdrawn consent by the time of analysis. Additionally, we subset to a group of unrelated samples (by including only those that were used for genetic principal components analysis during central genetic data preprocessing) and removed participants who were pregnant or had diabetes, coronary heart disease, liver cirrhosis, or cancer. Genetic variants with minor allele frequency (MAF) > 1% in the full analysis population were included in genome-wide and follow-up studies.

Phenotypic data

The primary outcome trait, hsCRP, was originally measured in plasma using an immunoturbidimetric assay (Beckman Coulter AU5800). Values were log-transformed and outliers (more than 5 standard deviations from the mean of the log-transformed values) were removed prior to analysis.

The primary N3FA variable for analysis was intended to reflect only long-chain, marine sources, thus excluding plant-based sources of dietary omega-3 such as alpha-linolenic acid (ALA; see Discussion). N3FA intake data came from multiple self-reported sources. All participants completed a 30-item food frequency questionnaire (FFQ) at the baseline assessment center visit, which has been validated for reliably ranking participants according to intake of major food groups but is not sufficient for calculating specific nutrient intake estimates [4]. Estimates for typical intake of both oily fish and non-oily fish (servings/day; fields 1329 and 1339) were retrieved from this FFQ. Fish oil supplementation was recorded based on reported use as a medication, with variables derived from both touchscreen questionnaire (UKB field 6179) and verbal interview (field 20,003).

Additional covariates collected for analysis included genetically-determined sex, age, age^2, a sex-by-age product term, income (5 categories), educational attainment (6 categories), smoking (categorical: never, past, or current), alcohol intake (categorical: weekly frequency estimates), and additional diet variables from the FFQ (cooked vegetables, raw vegetables, fresh fruit, processed meat), and a categorical diet variable (bread type; wholemeal or wholegrain bread versus other types). For variables coded as categorical, ambiguous categories such as "do not know" or "prefer not to answer" were left as non-missing to allow them to constitute an independent category for adjustment. Additional covariates used for sensitivity analysis included body mass index (kg/m^2), total physical activity (continuous; MET-min/wk), and self-reported medication usage (separate binary indicators for cholesterol medication, blood pressure medication, and insulin, based on UKB fields 6153 for females and 6177 for males).

Plasma N3FA measurements

Various measurements of plasma N3FA status were available from the Nightingale platform (N = 199,059), These data were preprocessed using the *ukbnmr* R package [22], which includes imputation of zero values, log-transformation, adjustment for key batch variables such as shipment plate and time between sample preparation and measurement, and transformation back into absolute concentrations. Ultimately, relevant available N3FA species included: direct N3FA concentrations, docosahexanoic acid (DHA), and both of these quantities as fractions of total fatty acids and total polyunsaturated fatty acids. Non-DHA pN3FA was calculated by subtracting DHA% from total N3FA%. Eicosapentanoic acid was not measured directly but should constitute approximately two thirds of this non-DHA quantity [1].

An overall estimate of dietary N3FA intake ("dN3FA") was calculated as a weighted sum of four key sources: oily and non-oily fish intake based on FFQ, touchscreen-reported fish oil intake, and verbal interview-reported fish oil intake. To determine the relative contribution of each source, total N3FA% was regressed on the four dietary sources in the full population. The dN3FA variable was then calculated as a linear combination of these components, weighted by their corresponding multivariable regression effect estimates. Missing values for fish oil variables were imputed as "no intake" when taking this weighted sum.

Main effect analysis and mediation testing

All statistical analyses were performed using R version 4.2.2 [21] unless otherwise noted. Linear regression models were used to understand the associations between various self-reported dietary N3FA sources and hsCRP using standard model-based standard errors (in contrast to the robust standard errors used in genome-wide interaction testing) and the covariates described above. Supp. Fig. S2 shows the directed acyclic graph and preliminary main effect regression results used to guide the choice of covariates. Initial models included additional adjustment for assessment center variable (one indicator variable per center), but this adjustment was removed from downstream analysis due to the minimal impact on regression estimates.

Mediation analysis examined the degree to which the plasma fatty acids mentioned above mediated the relationship between dN3FA and hsCRP. Mediation tests were performed using the *mediation* package for R [10, 27], using robust standard errors and 100 Monte Carlo

draws. Exposures and outcomes were scaled to mean zero and standard deviation one to create comparable mediation effect estimates across exposures. The output of each analysis included estimates of the total effect, the average causal mediation effect (ACME), and the direct effect. We note that, given the continuous exposure and outcome variables explored here, this mediation framework reduces conceptually to the linear structural equation modeling approach originally described by Baron and Kenny [3]. Self-reported raw vegetable intake was used as a negative control for the mediation analysis, given the similarity of its likely confounding structure based on the examination of its Pearson correlations with selected outcome and socioeconomic variables (see Results).

Gene-environment interaction modeling

Genome-wide interaction studies (GWIS) were performed using GEM v1.5.2 [29] with robust standard errors. The primary interaction model was as follows:

$$y \sim g + E + g * E + covariates$$

where y represents the outcome of interest, g represents the imputed genotype dosage, and E represents the exposure of interest. Covariates were the same as used for main effect models, with the addition of 10 genetic principal components as calculated centrally by the UKB team. GWIS were performed for three pathways: "standard" (E = dN3FA, Y = hsCRP), "upstream" (E = dN3FA, Y = pN3FA), and "downstream" (E = pN3FA, Y = hsCRP). Significance was assessed based on a standard genome-wide threshold of $p < 5 \times 10^{-8}$. GWIS results were pruned using PLINK 2.0 [7], using an LD reference panel consisting of a random 20,000-participant subset of the UKB and with parameters as follows: index variant p-value threshold = 5×10^{-8}, LD r^2 threshold = 0.2, and clumping radius = 5,000 kb. Sensitivity models at top loci included adjustment for: (1) exposure-by-gPC interaction terms, (2) genetic interaction terms for all covariates, and (3) body mass index (BMI) as a measure of adiposity.

GWIS results were subject to enrichment analysis to prioritize genes with enrichment of interaction signal in the surrounding genetic region. Interaction p-values from the GWIS were used as input to the MAGMA program [8], using the same LD reference panel as used for pruning and gene regions defined from 2 kb upstream to 1 kb downstream of the gene limits based on the NCBI database (GRCh37). Sensitivity analyses as described above were performed using the most significant variant from each gene-based finding.

Results

Simulations revealed the extent to which the decomposed approach is advantageous over the standard GxE testing approach as a function of the magnitude of the genetic interaction and the relationships between the exposure (E), mediator (M), and outcome (Y). As shown in Fig. 1, upstream interaction tests were more powerful when the true genetic interaction was with E (rather than M). Intuitively, the power of the upstream test matched that of the standard test as the M-Y relationship became stronger (i.e., in the limit that M fully determines Y). However, at more plausible values, the difference in power was substantial: when the GxE explained 0.5% of the variance in M, which then explained only 10% of the variance in Y, the upstream test was more powerful by a factor of 1.79 (38% versus 21% power). Even more stark patterns were observed for the downstream tests as a function of the strength of the E-M relationship. Power for the downstream test was equal when E fully determined M, but was greater by a factor of 4.72 (43% versus 9% power) when we simulated a GxM explaining 0.5% of the variance in Y and an E explaining 10% of the variance in M. We note that the simulations presented in Fig. 1 assume perfect measurement of the mediator; power of the decomposed approach decreases as this measurement error increases (see Supp. Fig. S1). Nonetheless, these results support the greater power of the decomposed approach for GxE discovery in most scenarios in which molecular mediators are known and measured.

A summary of the UKB population, which was primarily of European ancestry but included individuals from six ancestry groups based on assignments from the Pan-UKBB project [11] can be found in Supp. Table S1. Preliminary regressions confirmed the expected negative association between dietary N3FA sources (fish and fish oil) and hsCRP, with the association partially attenuated by adjustment for confounders (Supp. Fig. S2). Of the available pN3FA measures, total N3FA, as a percentage of total plasma fatty acids, showed the strongest correlations with dietary N3FA sources. Based on this, an overall estimate of dietary N3FA intake ("dN3FA") was calculated as a weighted sum of four key sources (oily and non-oily fish, touchscreen-reported fish oil intake, and verbal interview-reported fish oil intake), with oily fish contributing most substantially to the derived dN3FA estimate (Supp. Fig. S3, Supp. Table S2).

This derived dN3FA metric was positively associated with pN3FA, while both dN3FA and pN3FA were negatively associated with hsCRP (Fig. 2a-c). Mediation analysis showed that the dN3FA-hsCRP relationship was mediated by pN3FA (mediated effect: −0.037, total effect: −0.023; Fig. 2d, Supp. Table S3). This result, with a mediated effect greater than the total estimated effect, is a case

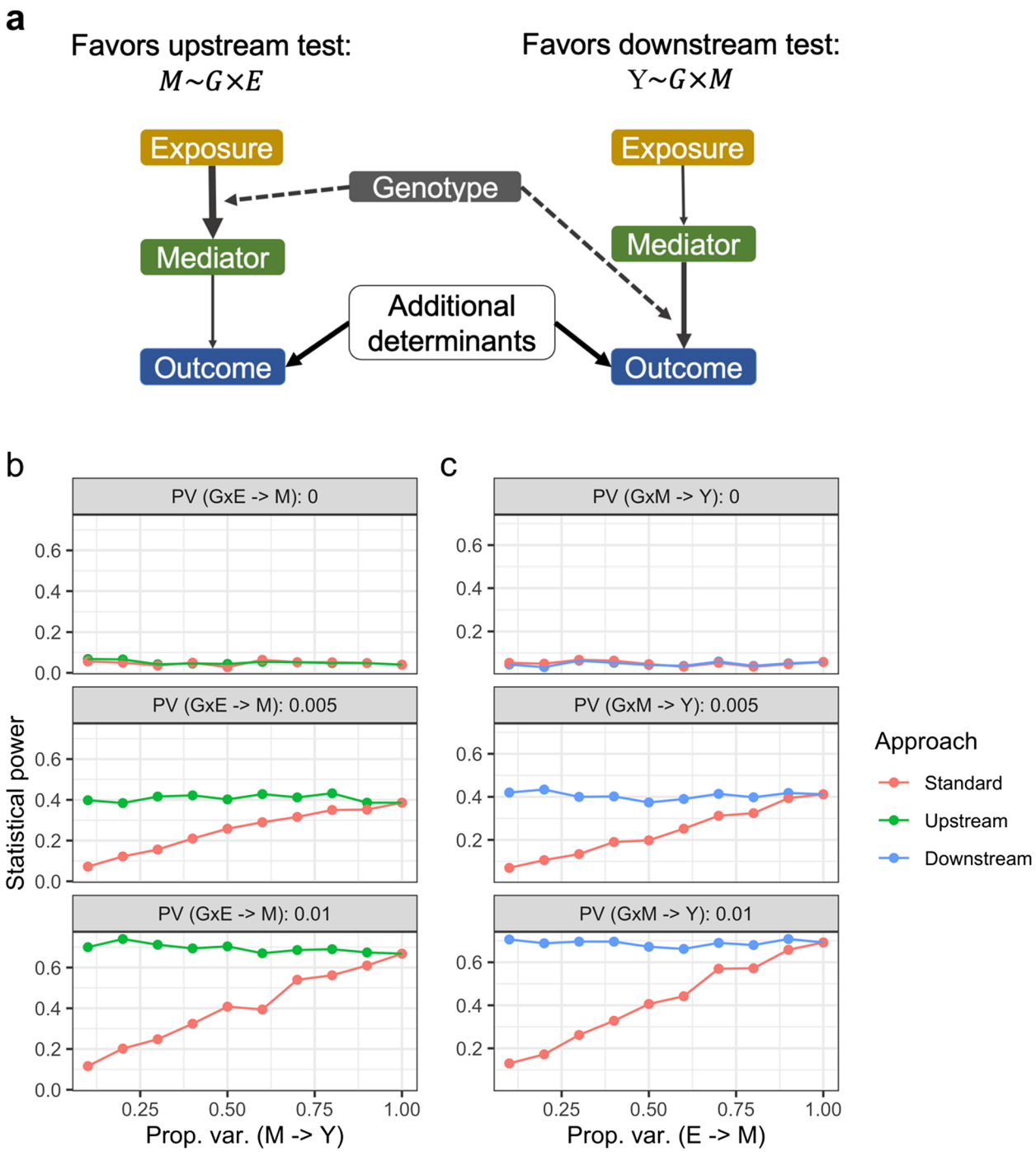

Fig. 1 Power simulations and conceptual basis for the decomposed GxE testing approach. **a** Conceptual diagram indicates the relationships between the exposure (E), mediator (M), and continuous outcome (Y), with line thickness denoting proportion of variance explained. Dotted lines correspond to interaction effects. **b**,**c** Power plots display simulation results. *X*-axes correspond to the strength (quantified by proportion of variance explained) of the M-Y relationship (upstream, left) or E-M relationship (downstream, right). Faceted panels correspond to the strength of the simulated genetic interaction with E (upstream) or M (downstream). No measurement error is included in the simulations summarized here (see Supp. Fig. S1)

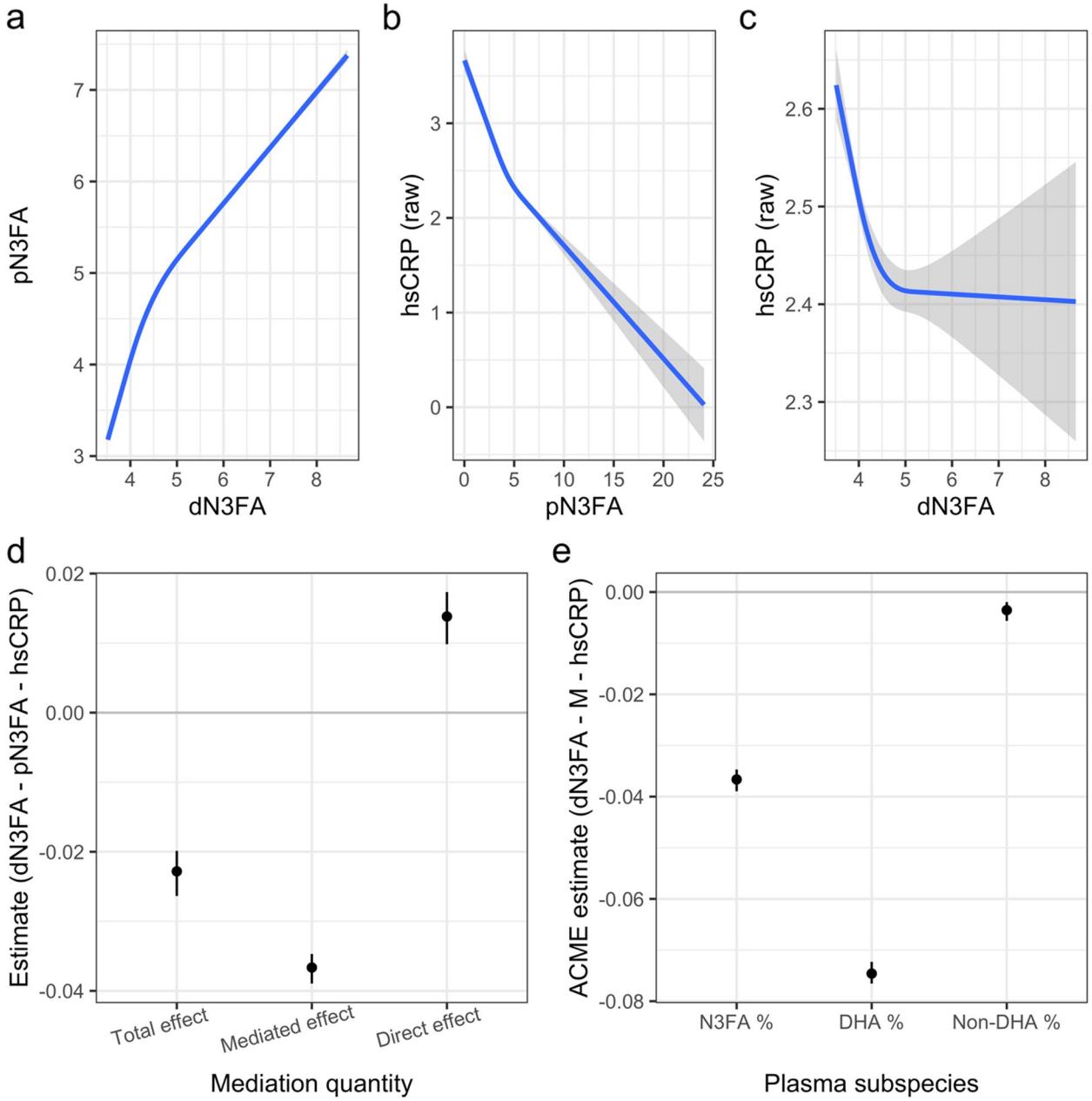

Fig. 2 Relationships and mediation between dietary N3FA, plasma N3FA, and hsCRP. **a**-**c** Curves show the relationship between dN3FA and pN3FA (**a**), pN3FA and non-transformed hsCRP (**b**), and dN3FA and non-transformed hsCRP (**c**). Curves were estimated using a restricted cubic spline with 3 knots. *X*-axes correspond to estimated dN3FA (units of % total blood fatty acids due to the derivation of this dietary intake proxy; see Methods) or pN3FA (same units). *Y*-axes correspond to pN3FA or hsCRP (mg/L). **d** Estimates and 95% CIs for the total, mediated (indirect; ACME), and direct effects from mediation analysis are shown for the pathway in which pN3FA mediates the dN3FA-hsCRP relationship. **e** ACME (mediated effect) estimates are shown using either total % N3FA (i.e., total pN3FA) or subsets of this quantity (DHA and non-DHA; see Methods) as the mediating quantity

of inconsistent mediation that can indicate residual negative confounding in the opposite direction of the mediated effect [17]. Thus, we explored several additional mediation analyses to confirm the result. As shown in Supp. Fig. S4, the total and mediated effects were larger for oily compared to non-oily fish, as expected due to the higher N3FA content. Additionally, raw vegetable intake, which is related similarly correlated with other confounding variables and outcomes (Supp. Fig. S4a), served as a negative control with a comparable total effect on hsCRP,

but with minimal estimated mediation by pN3FA. When comparing mediation estimates across pN3FA subspecies (DHA vs. non-DHA), most of the mediated effect was attributable to DHA (Fig. 2e).

Next, we performed a series of three GWIS, corresponding to the upstream, downstream, and standard pathways. One locus (on chromosome 11 containing the *FADS* genes) reached $p < 5 \times 10^{-8}$ in the upstream analysis, while there were none for the downstream or standard pathways (Fig. 3; Supp. Fig. S5; index variants reaching $p < 5 \times 10^{-6}$ listed in Supp. Table S4). Given the strong signal and known biology for the *FADS* locus, we explored the set of interactions at this locus more in-depth. The lead variant for the upstream analysis, rs174535, reached $p_{int} = 1.6 \times 10^{-12}$. Importantly, it showed no interaction signal for the downstream pathway ($p_{int} = 0.94$) and would not have been prioritized based on the standard pathway analysis ($p_{int} = 0.097$; Fig. 4a). This variant-specific difference in findings was supported by an additional power simulation applying relevant parameters: using a sample size of 200,000, GxE proportion of variance explained of $0.025^2 = 0.000625$, M-Y proportion of variance explained of 0.009, and MAF of 0.25, the upstream pathway had a power of approximately 1 at genome-wide significance, versus 0 for the standard pathway. Exploring the signal at this variant further, stratified plots indicated a stronger association between dN3FA and pN3FA for carriers of the pN3FA-decreasing allele (Fig. 4b). Sensitivity analyses for the interaction with rs174535 indicated robustness to additional covariates, including gPC-exposure and genotype-covariate interaction terms and adjustment for adiposity, physical activity, or use of selected medications (see Methods; Supp. Table S5), and did not reveal any direct association with dietary intake behavior (from regression of dN3FA on rs174535; $p = 0.80$).

We then performed gene-level enrichment analysis using the MAGMA program to further increase power by pooling signal across variants within each gene region (Fig. 5). Using a false discovery rate threshold of $q < 0.05$ (based on the Benjamini–Hochberg method calculated separately for each of the three pathways), this analysis showed the same signals for the upstream and standard

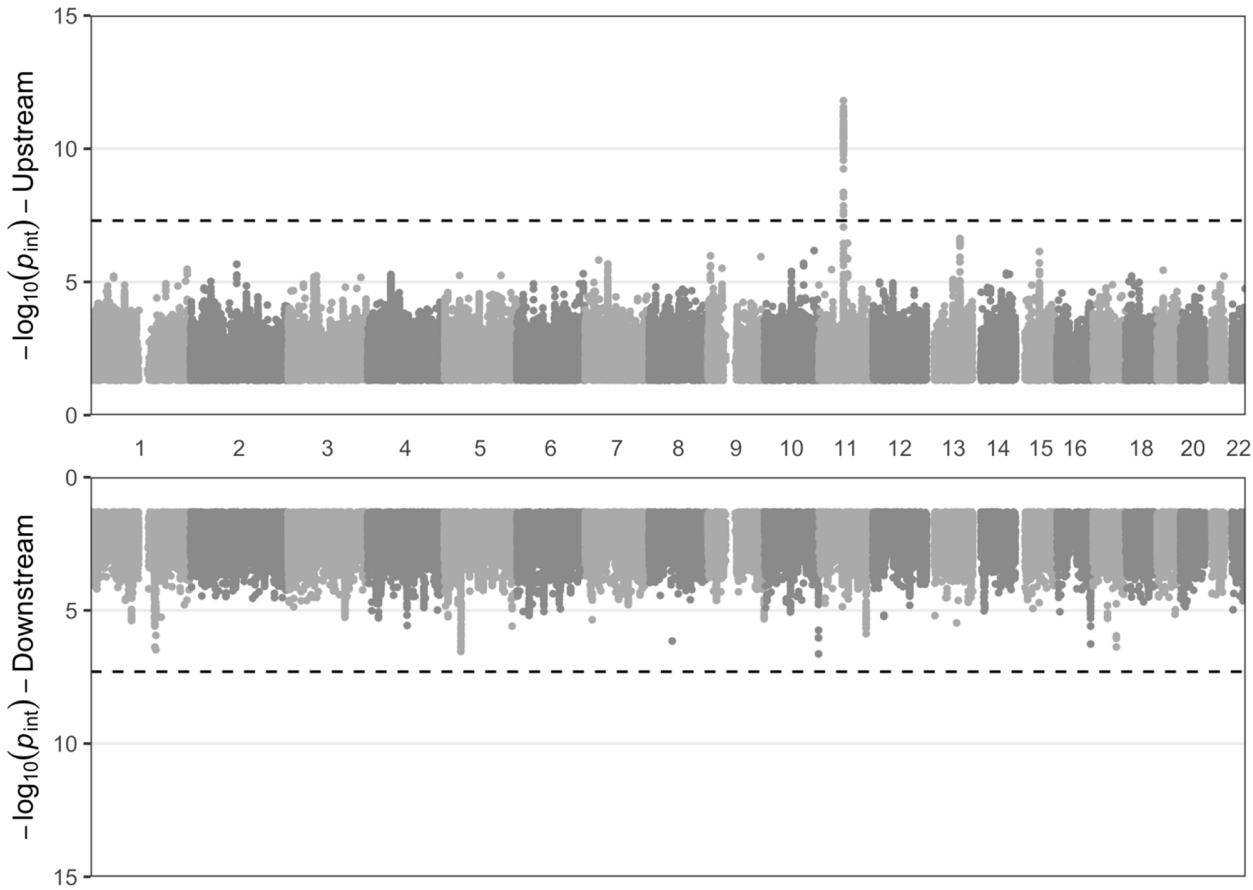

Fig. 3 Chicago plot displays variant interaction *p*-values as a function of chromosomal position for the upstream (top) and downstream (bottom) pathways. *Y*-axes show -log(*p*) for interaction tests (based on robust standard errors), while the *x*-axis indicates chromosomal position. Dotted lines indicate the genome-wide significance threshold of 5×10^{-8}. See Supp. Fig. S5 for results from the standard pathway

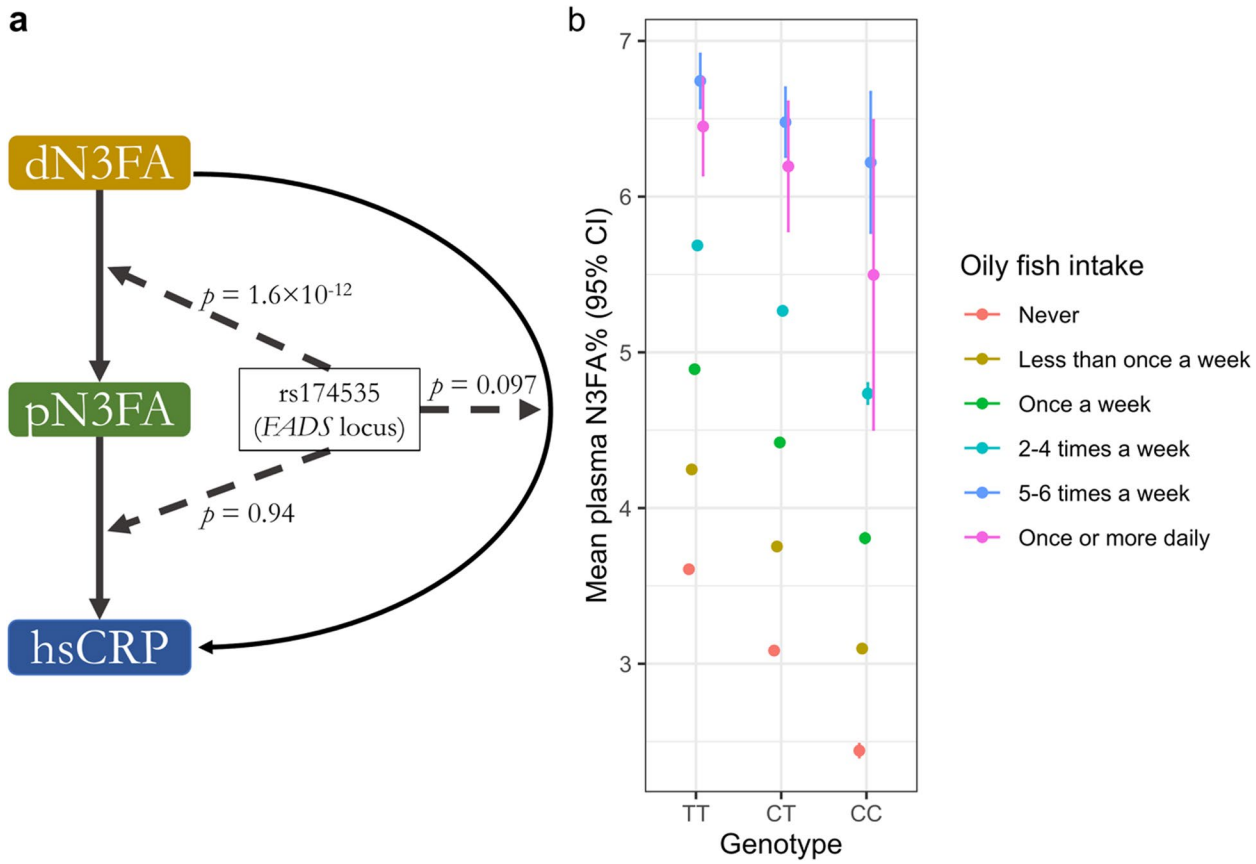

Fig. 4 Investigation of the *FADS* locus interactions. **a** Summary of the decomposed interaction testing for rs174535, including the standard, upstream, and downstream test *p*-values for interaction. **b** Depiction of the interaction at rs174535 using mean pN3FA levels stratified by genotype (*x*-axis) and categories of oily fish intake (colors)

analyses (a series of genes within the larger *FADS* locus, and none, respectively). The downstream analysis revealed two additional genes: *CBLL1* ($p_{gene}=4.9\times10^{-7}$) and *MICA* ($p_{gene}=3.9\times10^{-6}$). *CBLL1* codes for a gene involved in immune-related pathways and has genetic associations with circulating immune cell counts. *MICA* also has very strong evidence for an effect on immune cell counts and codes for a gene with a key role in adaptive immune function. Thus, while having only modest genetic main effects on hsCRP itself in association studies, both genes have biologically plausible mechanisms for involvement in the modulation of inflammation by immune cells in response to changes in pN3FA status. These genes did not show any signal in gene-level enrichment of results from the standard pathway (both $p>0.05$).

Discussion

We explored the use of a molecular mediator-focused interaction testing approach to improve the statistical power and biological precision of GxE discovery analysis, using N3FA and hsCRP in a mechanistically informed case study. We showed that variants in the key *FADS* locus modify the dN3FA-pN3FA relationship, with an upstream interaction appearing that would not have been discovered independently of the molecular mediator. Furthermore, we described additional genes appearing in enrichment analysis of the downstream pathway analysis.

Analytical approaches combining omics and mediation analysis are increasingly used [18], with continual methods development improving statistical models for multiple models and incomplete data [13, 31]. There has been some discussion in the statistical literature of frameworks combining mediation and interaction, with distinctions made between mediation upstream or downstream of the interaction itself [14, 16]. One example from genetic epidemiology leveraged this concept to explain how sex-associated blood cell proportion differences might explain sex heterogeneity in expression-quantitative trait loci [12]. The approach described here differs in that its goal is to increase GxE discovery by leveraging

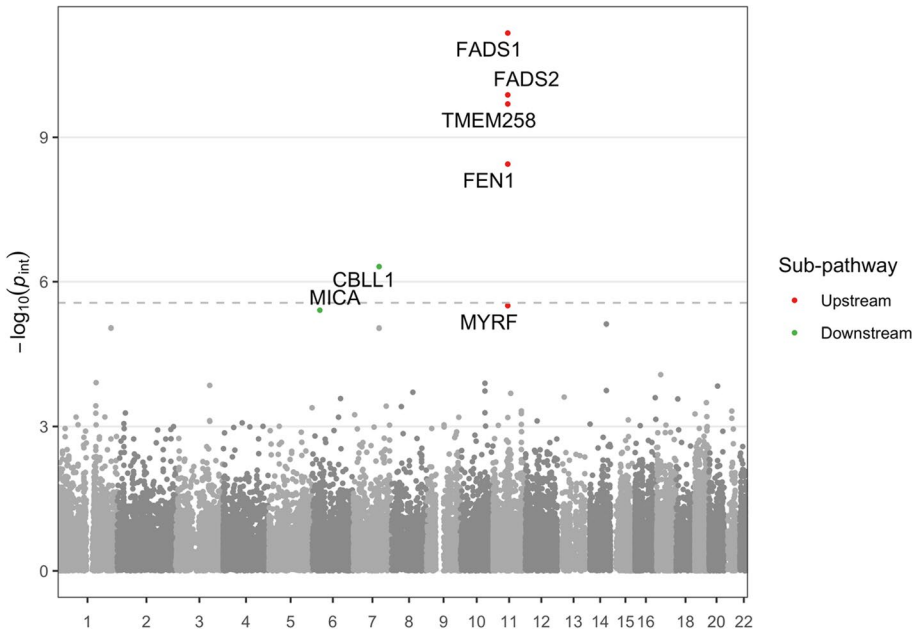

Fig. 5 Manhattan plot displays gene-level interaction *p*-values as a function of chromosomal position for the upstream (red) and downstream (green) pathways. The *y*-axis shows -log(*p*) for gene-level enrichment tests of interaction effects from MAGMA, while the *x*-axis indicates the chromosomal position of the 5′ gene boundary. The dotted line denotes a Bonferroni significance threshold accounting for 18,249 genes tested. Colored points correspond to genes reaching FDR $q < 0.05$

knowledge about molecular mediators, rather than to explain previously observed interactions.

The strongest and only variant-specific finding was part of the upstream pathway analysis. Conceptually, this relates to the relationship between dietary intake and plasma N3FA status, which is affected by both the efficiency of absorption of N3FAs and the rate of their endogenous production from essential fatty acid precursors. Notably, this interaction signal would not have been uncovered in the standard pathway analysis (genetic modification of the dN3FA-hsCRP relationship). This lack of signal using the standard approach can be explained by the complex and multifactorial set of inputs determining hsCRP, with pN3FA explaining only a small portion of the variability in hsCRP in our dataset (0.9%). This discrepancy is supported by the post hoc power comparison indicating a power of 1 for the upstream pathway versus 0 for the downstream pathway. However, this does not necessarily indicate a lack of relevance of this interaction for chronic inflammation, due to both the imperfection of hsCRP as a marker of chronic inflammation and potential error in hsCRP measurement. Furthermore, by revealing modifiers of pN3FA status, this type of upstream analysis has implications for not only inflammation, but also any N3FA-related risk factor or disease state. Our finding also suggests caution in the use of pN3FA as a biomarker of dietary intake, since the dN3FA-pN3FA relationship that underlies this dietary proxy may be biased according to genetic variation.

The *FADS* locus has one of the strongest known main effects on pN3FA [15]. The biological function of these genes, especially *FADS1* and *FADS2*, is related to the production of long-chain N3FAs from alpha-linolenic acid. Interactions of this locus with dN3FA intake have been explored, with inconsistent findings [23]. Based on our stratified models (such as in Fig. 4b), variants at this locus related to lower mean pN3FA also associate with a modestly stronger association between dN3FA and pN3FA, consistent with an overall endogenous feedback mechanism that enables greater absorption and incorporation of N3FA given lower baseline status. This is additionally consistent with the metabolic role of the *FADS* genes in long-chain N3FA production. Specifically, variability in the abundance or function of these genes products plays a direct role in their production from ALA, and thus their circulating concentrations, but has a minimal role in metabolizing the diet-derived fatty acids themselves. Thus, a mechanism based on endogenous feedback loops (in which higher baseline status decreases production or absorption) is most likely to be playing a role in the observed interaction.

Smith and colleagues did not find evidence of a similar interaction impacting pN3FA for two variants in the *FADS1* locus in a meta-analysis of cohorts from the

CHARGE consortium [24]. This study differed in the collection of cohorts and overall sample size ($N = 11{,}668$) as well as the combination of circulating fatty acid measurements from plasma (as analyzed here) and erythrocyte membranes, which they note as a source of potential heterogeneity.

Beyond this finding in the upstream analysis, we also report multiple genes passing a false discovery rate threshold in gene-level enrichment analysis of the downstream analysis linking pN3FA to hsCRP. Both *CBLL1* and *MICA* have biologically plausible explanations for impact on hsCRP via immune cell counts and function, establishing hypotheses for more in-depth exploration that would not have been prioritized using a standard GxE testing strategy without molecular mediators (all $p > 0.05$ in the standard pathway analysis). We also note that this downstream analysis has conceptual links to GxE studies using molecular quantities as exposures or "contexts" [19, 32].

In interpreting these results, it should be noted that the decomposition approach, including this N3FA application, depends on the quality and comprehensiveness of the measurement of mediating molecular species. For example, this study only considers N3FA as measured in plasma due to data availability, despite potential heterogeneity in genetic interaction results compared to other measurement compartments such as erythrocyte membranes [24]. Additionally, despite the statistical significance of the uncovered *FADS* locus interaction, the practical relevance is quite small, with the interaction explaining only about 0.06% of the variance in pN3FA. We studied only long-chain N3FA because of their greater prevalence in plasma and more sparse and clear dietary sources, meaning that these interactions may not apply to plant-based sources of N3FA providing ALA. However, there is some evidence of an effect of ALA intake on hsCRP and other inflammatory biomarkers, making this a compelling area for future study [2]. Our use of only common variants in this study may have limited the space of possible variant effects,more clinically relevant effects may be discoverable by analyzing rare variation [30] or using polygenic score-based approaches [26, 28].

Future work could expand this analysis in multiple directions. First, the biological case study presented here explores a relatively narrow biological question involving the pathway from dN3FA to pN3FA to hsCRP; future analyses focused more directly on N3FA biology could include more fine-grained measurements of specific dietary and plasma N3FA subspecies as well as a broader range of outcome metrics. Second, this strategy could be expanded to include multiple mediators, as has been described for omics-related main effect mediation analysis [31] but not extended to the realm of GxEs to our knowledge. Finally, while we used only unrelated individuals in this study, mixed models accounting for familial or other sample relatedness (e.g., longitudinal data) can be straightforwardly incorporated into mediation testing frameworks [31].

In summary, our interaction decomposition approach leverages molecular mediators to improve the power to discover GxEs while improving the biological interpretability of the results. Applying this strategy using dietary and plasma N3FAs and hsCRP, we showed signal at the known *FADS* locus that was not present for the standard analysis, while additionally reporting multiple genes enriched in signal for the downstream analysis. We anticipate that this framework can guide more effective future GxE studies that leverage molecular quantities to increase discovery and mechanistic understanding.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12263-025-00765-w.

Supplementary Material 1.

Supplementary Material 2.

Authors' contributions

KEW and AKM designed the analysis plan. KEW conducted the analysis. CJP, JBM, DIC, and AKM provided critical feedback on the analysis and manuscript. All authors reviewed the final manuscript.

Funding

KEW was supported by K01DK133637. JBM was supported by UM1DK078616 and R01HL151855. AKM was supported by R01HL145025.

Data availability

Code supporting the simulations and analyses described here can be found at https://github.com/kwesterman/ukb-n3fa-decomp. The UK Biobank data can be obtained through application at https://www.ukbiobank.ac.uk/.

Declarations

Competing interests

The authors declare no competing interests.

Author details

[1]Clinical and Translational Epidemiology Unit, Mongan Institute, Massachusetts General Hospital, Boston, MA, USA. [2]Programs in Metabolism and Medical & Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA. [3]Department of Medicine, Harvard Medical School, Boston, MA, USA. [4]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. [5]Division of General Internal Medicine, Massachusetts General Hospital, Boston, MA, USA. [6]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA. [7]Division of Genetics, Brigham and Women's Hospital, Boston, MA 02115, USA. [8]Medical and Population Genetics Program, Broad Institute, Cambridge, MA, USA. [9]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

Received: 10 September 2024 Accepted: 19 February 2025
Published online: 05 March 2025

References

1. Abdelmagid SA, Clarke SE, Nielsen DE, et al. Comprehensive profiling of plasma fatty acid concentrations in young healthy canadian adults. PLoS One. 2015;10(2):e0116195. https://doi.org/10.1371/journal.pone.0116195.
2. Baker EJ, Miles EA, Burdge GC, et al. Metabolism and functional effects of plant-derived omega-3 fatty acids in humans. Prog Lipid Res. 2016;64:30–56.
3. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. J Pers Soc Psychol. 1986;51(6):1173–82. https://doi.org/10.1037/0022-3514.51.6.1173.
4. Bradbury KE, Young HJ, Guo W, Key TJ. Dietary assessment in UK Biobank: an evaluation of the performance of the touchscreen dietary questionnaire. J Nutr Sci. 2018;7:e6. https://doi.org/10.1017/jns.2017.66.
5. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562:203–9. https://doi.org/10.1038/s41586-018-0579-z.
6. Calder PC. Marine omega-3 fatty acids and inflammatory processes: Effects, mechanisms and clinical relevance. Biochim Biophys Acta Mol Cell Biol Lipids. 2015;1851(4):469–84.
7. Chang CC, Chow CC, Tellier LCAM, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7. https://doi.org/10.1186/s13742-015-0047-8.
8. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. PLoS Comput Biol. 2015;11(4):e1004219. https://doi.org/10.1371/journal.pcbi.1004219.
9. Gauderman WJ, Mukherjee B, Aschard H, et al. Update on the State of the Science for Analytical Methods for Gene-Environment Interactions. Am J Epidemiol. 2017;186:762–70. https://doi.org/10.1093/aje/kwx228.
10. Imai K, Keele L, Tingley D. A General Approach to Causal Mediation Analysis. Psychol Methods. 2010;15(4):309–34. https://doi.org/10.1037/a0020761.
11. Karczewski KJ, Gupta R, Kanai M, et al. Pan-UK Biobank GWAS improves discovery, analysis of genetic architecture, and resolution into ancestry-enriched effects. medRxiv [Preprint]. 2024. Available from: https://doi.org/10.1101/2024.03.13.24303864.
12. Kasela S, Aguet F, Kim-Hellmuth S, et al. Interaction molecular QTL mapping discovers cellular and environmental modifiers of genetic regulatory effects. Am J Human Genet. 2024;111:133–49. https://doi.org/10.1016/j.ajhg.2023.11.013.
13. Kidd J, Raulerson CK, Mohlke KL, Lin DY. Mediation analysis of multiple mediators with incomplete omics data. Genet Epidemiol. 2023;47(1):61–77. https://doi.org/10.1002/gepi.22504.
14. Kwan JLY, Chan W. Variable system: An alternative approach for the analysis of mediated moderation. Psychol Methods. 2018;23(2):262–77. https://doi.org/10.1037/met0000160.
15. Lemaitre RN, Tanaka T, Tang W, 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. https://doi.org/10.1371/journal.pgen.1002193.
16. Little TD, Card NA, Bovaird JA, et al. Structural equation modeling of mediation and moderation with contextual factors. In: Modeling Contextual Effects in Longitudinal Studies. Chapter 9. Mahwah (NJ): Lawrence Erlbaum Associates Publishers; 2007.
17. MacKinnon DP, Krull JL, Lockwood CM. Equivalence of the mediation, confounding and suppression effect. Prev Sci. 2000;1(4):173–81. https://doi.org/10.1023/A:1026595011371.
18. Mohammadi-Shemirani P, Sood T, Paré G. From 'Omics to Multi-omics Technologies: the Discovery of Novel Causal Mediators. Curr Atheroscler Rep. 2023;25(2):55–65.
19. Moore R, Casale FP, Jan Bonder M, et al. A linear mixed-model approach to study multivariate gene–environment interactions. Nat Genet. 2019;51:180–6. https://doi.org/10.1038/s41588-018-0271-0.
20. Patel CJ, Kerr J, Thomas DC, et al. Opportunities and challenges for environmental exposure assessment in population-based studies. Cancer Epidemiol Biomarkers Prev. 2017;26(9):1370–80.
21. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. 2022.
22. Ritchie SC, Surendran P, Karthikeyan S, et al. Quality control and removal of technical variation of NMR metabolic biomarker data in ~120,000 UK Biobank participants. Sci Data. 2023;10(1):64. https://doi.org/10.1038/s41597-023-01949-y.
23. Schulze MB, Minihane AM, Saleh RNM, Risérus U. Intake and metabolism of omega-3 and omega-6 polyunsaturated fatty acids: nutritional implications for cardiometabolic diseases. Lancet Diabetes Endocrinol. 2020;8(11):915–30.
24. Smith CE, Follis JL, Nettleton JA, et al. Dietary fatty acids modulate associations between genetic variants and circulating fatty acids in plasma and erythrocyte membranes: Meta-analysis of nine studies in the CHARGE consortium. Mol Nutr Food Res. 2015;59(7):1373–83. https://doi.org/10.1002/mnfr.201400734.
25. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLoS Med. 2015;12:e1001779. https://doi.org/10.1371/journal.pmed.1001779.
26. Sun Y, McDonald T, Baur A, et al. Fish oil supplementation modifies the associations between genetically predicted and observed concentrations of blood lipids: a cross-sectional gene-diet interaction study in UK Biobank. Am J Clin Nutr. 2024;120:540–9. https://doi.org/10.1016/j.ajcnut.2024.07.009.
27. Tingley D, Yamamoto T, Hirose K, et al. Mediation: R package for causal mediation analysis. J Stat Softw. 2014;59:1–38. https://doi.org/10.18637/jss.v059.i05.
28. Westerman K, Liu Q, Liu S, et al. A gene-diet interaction-based score predicts response to dietary fat in the Women's Health Initiative. Am J Clin Nutr. 2020;111:893–902. https://doi.org/10.1093/ajcn/nqaa037.
29. Westerman KE, Pham DT, Hong L, et al. GEM: scalable and flexible gene–environment interaction analysis in millions of samples. Bioinformatics. 2021;37(20):3514–20. https://doi.org/10.1093/bioinformatics/btab223.
30. Westerman KE, Walker ME, Gaynor SM, et al. Investigating Gene-Diet Interactions Impacting the Association Between Macronutrient Intake and Glycemic Traits. Diabetes. 2023;72(5):653–65. https://doi.org/10.2337/db22-0851.
31. Yang T, Niu J, Chen H, Wei P. Estimation of total mediation effect for high-dimensional omics mediators. BMC Bioinformatics. 2021;22(1):414. https://doi.org/10.1186/s12859-021-04322-1.
32. Zhernakova DV, Deelen P, Vermaat M, et al. Identification of context-dependent expression quantitative trait loci in whole blood. Nat Genet. 2017;49(1):139–45. https://doi.org/10.1038/ng.3737.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.