

# **A gene-diet interaction-based score predicts response to dietary fat in the Women's Health Initiative**

Kenneth Westerman<sup>1</sup>, Qing Liu<sup>2</sup>, Simin Liu<sup>2</sup>, Laurence D. Parnell<sup>1</sup>, Paola Sebastiani<sup>3</sup>, Paul Jacques<sup>1</sup>, Dawn L. DeMeo<sup>4</sup>, José M. Ordovás<sup>1,5,6\*</sup>

<sup>1</sup>*JM-USDA Human Nutrition Research Center on Aging, Boston, MA*, <sup>2</sup>*Department of Epidemiology, Brown University School of Public Health, Providence, RI*, <sup>3</sup>*Department of Biostatistics, Boston University School of Public Health, Boston, MA*, <sup>4</sup>*Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA*, <sup>5</sup>*IMDEA Alimentación, CEI, UAM, Madrid, Spain*, <sup>6</sup>*Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid, Spain*

\*Corresponding author: José M. Ordovás

711 Washington St.

Boston, MA 02111

(617) 556-3103

[jose.ordovas@tufts.edu](mailto:jose.ordovas@tufts.edu)

## **Abstract**

While diet response prediction for cardiometabolic risk factors (CRFs) has been demonstrated using single SNPs and main-effect genetic risk scores, little investigation has gone into the development of genome-wide diet response scores. We sought to leverage the multi-study setup of the Women's Health Initiative cohort to generate and test genetic scores for the response of six CRFs (body mass index, systolic blood pressure, LDL-cholesterol, HDL-cholesterol, triglycerides, and fasting glucose) to dietary fat. A genome-wide interaction study was undertaken for each CRF in women ( $n \sim 10000$ ) not participating in the Dietary Modification (DM) trial, which focused on the reduction of dietary fat. Genetic scores based on these analyses were developed using a pruning-and-thresholding approach and tested for the prediction of one-year CRF changes as well as long-term chronic disease development in DM trial participants ( $n \sim 5000$ ). One of these genetic scores, for LDL-cholesterol (LDL-C), predicted changes in the associated CRF. This 1760-variant score explained 3.4% of the variance in one-year LDL-C changes in the intervention arm, but was unassociated with changes in the control arm. In contrast, a main-effect genetic risk score for LDL-C was not useful for predicting dietary fat response. Further investigation of this score with respect to downstream disease outcomes revealed suggestive differential associations across DM trial arms, especially with respect to coronary heart disease and stroke subtypes. These results lay the foundation for the combination of many genome-wide gene-diet interactions for diet response prediction while highlighting the need for further research and larger samples in order to achieve robust biomarkers for use in personalized nutrition.

## Introduction

Nutrigenetics approaches, in which genetic information is used to predict response to dietary inputs, are central to the emerging promise of personalized nutrition for cardiometabolic risk reduction. Inter-individual differences in food preferences, metabolism, detoxification, excretion, etc. affect our responses to diet, in a similar manner to the well-studied field of pharmacogenomics (1). Ideally, genotype-based nutrigenetic investigations would be conducted in large-scale dietary interventions. Two notable examples are the PREDIMED and POUNDS LOST trial, with significant findings including the interaction of a *TCF7L2* variant with a Mediterranean diet pattern for glycemic traits (2) and the interaction of a *PCSK9* variant with dietary carbohydrate for insulin resistance (3). However, such intervention-based studies are able to examine only a single dietary change (whether food, nutrient, or pattern) at a time, and are often limited to lower sample sizes (4).

To allow for more flexibility and greater sample sizes, gene-diet interactions (GDIs) are more commonly investigated in observational datasets. There is a rich literature of GDI discovery in the cardiometabolic realm. Typically, these focus on cardiometabolic risk factors in relation to biology-based candidate genes/variants (5,6), but some have looked at clinical outcomes (e.g. MI (7)). Other approaches use main-effect genetic risk scores, such as that for obesity interacting with sugar-sweetened beverage intake to influence anthropometric traits (8,9).

Characterization of individuals based on single or small groups of single nucleotide polymorphisms (SNPs) likely neglects important signal elsewhere in the genome, especially when dealing with highly polygenic cardiometabolic traits. Thus, for effective personalized nutrition approaches to be realized, it is necessary to integrate signals across the genome. A few investigations have explored GDIs genome-wide, such as for dairy and BMI (10) and for various dietary components and colorectal cancer (11). However, genome-wide interaction studies (GWIS) can be problematic due to the lower statistical power inherent in gene-environment interaction analyses (12). Furthermore, the potential for confounding and reverse causation (i.e. cardiometabolic risk impacting dietary behavior) in statistical interactions from observational data means that GDIs may not always predict modification of the risk factor in question after a dietary intervention.

In order to provide proof-of-concept for the use of GDIs in developing comprehensive diet response genetic scores, we sought to develop a genome-wide, GDI-based dietary fat response score for each of a series of cardiometabolic risk factors (CRFs). We performed genome-wide interaction studies for six CRFs, prioritizing sites with documented nominal main effects in prior genome-wide association studies, and used these intermediate results to derive fat response scores (FRS) for each CRF. We tested the performance of these scores in the fat reduction-focused Women's Health Initiative Dietary Modification trial, finding that an FRS for LDL-cholesterol (LDL-C) predicts 1-year LDL-C changes selectively in the intervention arm. Furthermore, we found associations of the LDL-C FRS with incident coronary heart disease and stroke subtypes specifically in the fat-reduction arm of the trial over approximately 22 years of follow-up.

## Methods

### *Women's Health Initiative Dataset*

The Women's Health Initiative study consists of a series of substudies: three clinical trials (related to cancer, cardiovascular disease, and osteoporosis) and an observational study (13). Over 160,000 participants were enrolled between 1993-1998, with the ability to enroll in up to 3 of the clinical trials simultaneously. For the purposes of this analysis, participants were categorized based only on whether or not they were enrolled in the dietary modification (DM) trial, which randomized almost 50,000 women to a low-fat diet or a control diet with no recommended dietary changes, with primary outcomes being incidence of breast and colorectal cancers and heart disease [(14). Study of these participants conformed to the ethical guidelines outlined in the Declaration of Helsinki, and this research was approved by the Tufts Health Sciences IRB (protocol 12592).

Participants were comprehensively screened at baseline, including physical measurements, blood sample collection, and questionnaire administration, while only a subset of participants provided blood samples or returned questionnaires during later visits. The food frequency questionnaire (FFQ) was designed specifically for the WHI study, emphasizing specific foods and preparation methods to maximize its sensitivity to changes in fat intake (15).

Phenotype data were accessed from dbGaP (accession: phs000746.v2.p3). Values shown in Table 1 only pertain to women whose genotypes were measured in one of a series of follow-up studies. For gene-diet interaction analyses, SBP, LDL-C, and GLU were adjusted for medication use: LDL-C and GLU values were divided by 0.75 for those on lipid-lowering and anti-diabetic medication, respectively, and SBP values were increased by 15 mmHg for those on anti-hypertensive medication. This type of adjustment for medication use has precedent in gene-environment interaction analyses (16). Cardiovascular risk factors (CRFs) were winsorized at 5 standard deviations from the mean and those other than LDL-C (BMI, SBP, HDL-C, TG, and GLU) were log-transformed prior to analysis. Longitudinal risk factor changes were calculated in DM trial participants as the difference between baseline and year 1. Adjudicated time-to-event data for chronic disease outcomes (coronary heart disease, myocardial infarction, ischemic stroke, hemorrhagic stroke, and non-CVD death) were collected, while diabetes incidence was defined as the self-report of any of: diabetes pills, insulin treatment, or general treatment for diabetes. Follow-up data was available for approximately 22 years following enrollment. Phenotype data processing was performed using R version 3.4.3 (17) and Python version 3.6.0.

### *Genotype data and preprocessing*

Imputed genotype data were retrieved from dbGaP (accession: phs000746.v2.p3) as a harmonized set of imputation outputs from a series of genotyping studies involving WHI participants. Prior to imputation, study-specific quality control steps had been undertaken on directly-genotyped SNPs, with filters based on sample and call rate, Hardy-Weinberg equilibrium, and minor allele frequency. Phasing had been performed for autosomes using BEAGLE, followed by imputation using minimac (MACH for the SHARE study subset). After download from dbGaP, variants were converted from dose format using dose2plink (<http://genepi.qimr.edu.au/staff/sarahMe/dose2plink>), filtered for imputation R-squared > 0.3 and minor allele frequency (MAF) > 0.001, and annotated with rsIDs, loci, and allelic information

using the 1000 Genomes Phase 3 download from dbSNP (download date: April 13, 2018). Only variants passing the imputation quality threshold in all genotyping sub-studies were included in the final dosage dataset. For score development and calculation, imputed variant dosages were converted to hard-calls and set to missing if the estimated dosage was not within 0.1 of an integer allele count. Post-imputation genotype data processing was performed using PLINK 2.0, while clumping and score calculation were performed using PLINK 1.9 (18).

### *Genome-wide interaction studies*

A genome-wide interaction study was performed for each of the six cardiometabolic risk factors. The genome-wide scan used an additive genotype model, adjusted for fixed effects including dietary fat (binary: % of kcals above or below the median), total kcals per day, age, five ancestry principal components, and genotyping sub-study. Genotyping was performed in a series of ancillary studies in WHI including Hip Fracture, GARNET, WHIMS+, GECCO (initial or CytoSNP), and AS264/MOPMAP. (Many participants were also genotyped as a part of the SHARe effort, but those women were of African American and Hispanic ancestry and thus were not included in the GWIS portion of this study.) Five of six sub-studies were adjusted for as binary variables to avoid collinearity. The primary estimand of interest was the interaction term between dietary fat and minor allele count at the SNP of interest. Interaction analyses were carried out using PLINK 2.0 (18). Variants of interest were annotated to genes using Annovar (19).

Gene-environment interaction power calculations for single SNPs were performed using the Quanto tool (20). The following assumptions were made: additive model; variance explained by genotype alone = 0.5%; and binary environment with 50% prevalence and explaining 10% of variance. (Note: there is no effect of minor allele frequency in this case given that variances explained are directly specified.)

### *Genetic responder score construction and evaluation*

To prioritize variants for inclusion in genetic responder scores, nominal ( $p < 0.05$ ) variants for main-effect on each risk factor were retrieved from large-scale meta-analyses: GIANT for BMI (21); International Consortium for Blood Pressure for SBP (22); Global Lipid Genetics Consortium (GLGC) for LDL-C, HDL-C, and TG (23); and MAGIC for fasting glucose (24). Each FRS was constructed using summary statistics for the diet-SNP interaction terms from the associated GWIS. SNPs were filtered for nominal main-effect relationships using the above meta-analyses, and interaction summary statistics were used as input to a pruning-and-thresholding (P&T) procedure (using the “-clump” function in PLINK 1.9), with a seed threshold of  $p=0.05$  and an LD threshold of  $r^2=0.5$ . The linkage disequilibrium references for the procedure was calculated from genotypes of the white DM trial participants. Genetic fat response scores (FRS) for each individual were then calculated as the weighted sum of allelic dosages for variants selected by the P&T procedure, with weights corresponding to the GWIS interaction term estimates. This type of diet interaction-based genetic score development has been described previously, for example in a Korean cohort with respect to body fat changes (25). A genetic risk score for LDL-C was created using the GLGC LDL-C meta-analysis summary statistics and the same P&T method and parameters as was used for the interaction analyses, resulting in a 26467-SNP score.

FRS were used to test for discrimination of changes in CRFs over the first year of the DM trial. Risk factor changes were assessed using linear models in participants in the intervention arm, with and without adjustment for baseline CRF levels. As a sensitivity analysis, p-values were calculated in separate models for interaction of the genetic score with 1) trial arm (control vs. dietary modification), and 2) observed fat reduction (negative vs. positive 3-year change in FFQ-estimated dietary fat). GRS were further tested for prediction of chronic disease development during follow-up across DM trial strata. Time-to-event for each of coronary heart disease, myocardial infarction, ischemic stroke, hemorrhagic stroke, and non-CVD death were used to fit age-adjusted Cox proportional hazards models, including a random effect term for genotyping sub-study (*cluster()* term in the *coxph* function call). Estimated log-hazard ratios were extracted from regressions conducted in the following strata: 1) DM trial intervention arm, 2) DM trial control arm, 3) DM trial intervention arm filtered for participants with 1-year fat reduction based on FFQ, and 4) DM trial control arm filtered for participants with 1-year fat increase based on FFQ.

## Results

### *Dietary fat responder score development*

A series of genome-wide interaction studies (GWIS) were undertaken in cross-sectional data from the Women's Health Initiative. These GWIS incorporated only women who did not participate in the dietary modification (DM) trial, using imputed genotypes along with baseline self-reported dietary intakes (from food frequency questionnaires) and fasting blood biomarkers. Baseline characteristics of these women, along with those participating in the DM trial, are shown in Table 1.

Preliminary power calculations were undertaken, based on parameter assumptions including a modest SNP main effect (0.5%) under an additive model and a binary environment with 50% prevalence explaining 10% of the outcome phenotypic variance. The results showed that, at the sample sizes available for European ancestry non-DM participants (7000-10000 individuals for each cardiometabolic risk factor (CRF)), this analysis was powered to detect only moderately large interaction effects (GxE variance explained greater than approximately 0.5%) at genome-wide significance (Supp. Table S1).

Dietary fat response scores were generated for each CRF using results from the corresponding GWIS analysis. Linear regression models were fit for log-transformed baseline CRF values (other than LDL-C), incorporating gene-dietary fat interactions (dietary fat represented as a binary % of total calories above or below the median) while adjusting for age, total calories, five genetic principal components, and genotyping sub-study. Q-Q plots of the GWIS results showed that genomic inflation was fairly well-controlled (Supp. Figure S1). For each CRF, the associated summary statistics (corresponding to the fat-genotype interaction term estimates) were filtered to include only those with nominal main-effect associations in large-scale published GWAS. This filter was informed by the power analysis above and chosen as a compromise between discovery and statistical power (alternative results using either a more stringent threshold or no filtering are shown in Supp. Table S2). A pruning-and-thresholding method was used to generate six FRS from these individual sets of summary statistics along with genotypes from the WHI DM participants as a linkage disequilibrium (LD) reference. Using parameters of seed p-value=0.05

and LD  $r^2 < 0.5$ , six sets of score weights were generated, with relevant SNP set sizes ranging from 1536 (SBP) to 6042 (BMI). Scores were then calculated as the weighted sum of allele counts across SNPs, normalized by the number of non-missing SNPs per individual.

#### *Dietary fat responder score assessment*

As the scores were developed to predict a positive interaction with dietary fat intake, the expected direction of the FRS effect on risk factors in the present fat-reduction trial would be negative. Of the fat response scores examined, only the LDL-C fat response score (LDL-FRS) was predictive at  $p < 0.05$  of the associated CRF change in DM trial participants in the fat-reduction arm (passing a Bonferroni correction for the 6 CRFs tested in baseline-adjusted sensitivity models). For this score, the standardized effect size was -0.18 (corresponding to a 5.1 mg/dL greater decrease in LDL-C per score standard deviation;  $p=0.026$ ). We note that the sample size of European-ancestry DM trial participants with follow-up measurements was much smaller for biochemical variables ( $n \sim 150$ ) compared to BMI and SBP ( $n \sim 2000$ ). Using the score developed in European-ancestry individuals, score performance was then tested in a combined-ancestry group including Black and Hispanic individuals, which almost doubled the sample size (Supp. Table S3). While some traits showed strong relationships (e.g. SBP), the signs of many were in a counter-intuitive direction, including a flip in sign for the previously-strong LDL-C relationship, suggesting that these results reflect primarily differences in ancestry rather than the intended biological differences. This observation was reinforced by the lack of association of the score with CRF changes in either Blacks or Hispanics alone.

Based on its observed association in European-ancestry participants, the LDL-FRS was investigated further. Linear models showed that the LDL-FRS accounted for 3.4% of the variance in 1-year LDL-C changes in the DM intervention arm. In baseline-adjusted models, this figure rose slightly to 3.9, based on the  $R^2$  change from a baseline-only model. Additional models confirmed an interaction between DM trial arm and the LDL-FRS ( $p=0.011$ ), supporting the specificity of this score for the fat-reduction arm. The LDL-FRS also showed specificity for LDL-C in that it did not predict changes in any other CRF (Supp. Table S4). The 1760 component SNPs were annotated using Annovar (19), revealing a predominance of intergenic and intronic variants and a set of genes with high numbers of independent SNPs contributing to the score (Figure 1a-d). Top genes by number of contributing SNPs included *CSMD1*, *PTPRD*, and *RGS12*.

Differences in mean LDL-C changes during the DM trial across genetic score strata are shown in Figure 1e,f. As suggested by the regression results, those in the control arm trended towards less strong LDL-C reductions in higher LDL-FRS strata, while those in the fat-reduction arm showed the opposite trend. Furthermore, isolation of individuals at the highest extreme of the score (top 10%) revealed an LDL-C reduction of almost double that of the rest of the DM intervention group (-39.0 versus -20.0 mg/dL). For comparison to the FRS, a main-effect genetic risk score (GRS) for LDL-C was developed using summary statistics from the Global Lipid Genetics Consortium meta-analysis (23) and an identical pruning-and-thresholding procedure to that used for the GDI-based scores. As expected, this score was strongly predictive of baseline LDL-C concentrations ( $p=4.53e-27$ ). However, unlike the GDI-based score, the GRS did not predict LDL-C changes in the DM intervention group ( $p=0.15$ ; stratum-specific mean changes in Figure 1f).

### *LDL-FRS association with chronic disease outcomes*

Next, the LDL-FRS was tested for relationships with incident disease outcomes over approximately 22 years of follow-up (Figure 2). Cox proportional hazards models were used, adjusting for age at baseline and including a random effect for genotyping sub-study. This frailty/random effects model has been recommended for optimizing power in multi-center time-to-event models (26). In addition to intervention versus control arm, another set of “per protocol-like” strata was produced by additionally filtering for FFQ-based self-reported fat reduction (in the intervention group) or fat increase (in the control group). CHD qualitatively showed the expected interaction, i.e. a stronger inverse association between LDL-FRS and disease risk in the fat reduction group). Ischemic stroke showed a similar pattern, with a risk reduction only in the fat reduction group ( $p = 0.020$ ). In contrast, hemorrhagic stroke, while having a low number of events (44 in total), showed a positive association only in the fat reduction group ( $p = 0.011$ ). Results for diabetes qualitatively mirrored those for CHD and ischemic stroke, while those for non-CVD death did not vary across groups. These cross-arm differences were generally strengthened when comparing the per protocol-like strata, with a much stronger effect for CHD in the confirmed fat reduction stratum ( $p = 4.07e-5$ ). In DM trial arm interaction models (score x arm), only ischemic stroke reached nominal statistical significance ( $p < 0.05$ ).

## **Discussion**

Diet response scores have shown some success in predicting the response of cardiovascular risk factors (CRFs) to nutritional interventions, but they are often based solely on main effects or single GDI SNPs. Here, we explored the potential for gene diet interaction (GDI)-based diet responder score development, leveraging the multi-trial setup of the Women’s Health Initiative. We developed what to our knowledge is the first example of a diet response score based on a hypothesis-free genome scan for each of six risk factors, and showed preliminary evidence for the viability of a LDL-C fat response score. The set of SNPs used for each score were limited to those showing nominal main effects in large-scale GWAS as a compromise between discovery and utilization of prior information, which was supported by the weaker results in sensitivity models incorporating either stronger (suggestive main-effect) or weaker (all SNPs) variant filters (Supp. Table S2).

Though FRS for six CRFs were developed and tested, only that for LDL-C showed nominal significance in predicting 1-year changes in the corresponding CRF. Multiple factors could explain this lack of predictive performance in general. First, analysis in observational datasets carries with it the potential for confounding of the observed relationships. Second, FFQs are imprecise instruments for measuring dietary intake, and though the FFQs used in the present study were optimized for detection of dietary fat, there was potential for substantial misclassification of this environmental exposure. Third, power calculations shown here and elsewhere suggest that a cohort of this size may not be powered to detect many very small gene-environment interactions such as are sought in genome-wide approaches like the one used here.

*CSMD1*, *PTPRD*, and *RGS12* stood out as genes containing the highest number of SNPs in the LDL-FRS (11, 9, and 9, respectively, after LD-pruning for  $r^2 < 0.5$ ). *CSMD1* variants are notably associated with LDL-C response to statin treatment (27) as well as SBP response to a high-salt diet [Newton-Cheh2009]. *CSMD1* has also shown epigenetic associations with LDL-C (28) as

well as response to modification of dietary fat composition (29). *PTPRD* variants modulate the response of T2D patients to pioglitazone therapy (30) and show suggestive associations with eating behaviors (caloric intake at dinner) (31). *RGS12* has been linked to LDL-C in GWAS (32). Altogether, these genes have literature evidence for relationships to dietary intake, response to cardiometabolic therapies, and LDL-C, but have not until now been shown to directly modify the LDL-C response to dietary fat proportions. We note that there is a bias towards identifying LDL-C-related variants in the LDL-FRS, as only nominally-associated main-effect SNPs were used as input to the score development algorithm.

A reasonable body of literature exists establishing GDIs for both dietary fat on CRFs (33,34) and general dietary exposures on LDL-C (35). Multiple studies have looked specifically at genetic variants modulating the LDL-C response to dietary fat. For example, a caloric restriction intervention in type 2 diabetics was more effective in reducing LDL-C in ApoE4 carriers (-15.6% versus -0.7%) (36). Two studies using the POUNDS LOST trial found relationships of specific polymorphisms with 2-year changes in response to a dietary fat intervention. Carriers of a risk allele at the APOA5 variant rs964184 show a decrease of 7.5 mg/dL in LDL-C in low-fat but not high-fat interventions (37), and carriers of the minor allele at CETP rs3764261 show an 8.9 mg/dL greater decrease in LDL-C on a low-fat diet (38). Our observed effect size of a 5.4 mg/dL decrease in LDL-C is of a similar magnitude to these findings, and emerged despite the multi-factorial nature of the WHI DM trial (which incorporated additional nutrition recommendations). The observed variance explained of 3.4% for the LDL-FRS means that the score, while contributing meaningfully to the prediction, does not capture most of the interindividual variability in LDL-C response to the WHI DM trial intervention. Based on prior observations of an inflection point in the impact of various genetic risk scores near the 90th percentile (39), we additionally evaluated the impact of LDL-FRS in the top 10%, finding almost double the LDL-C reduction in DM intervention participants with values at this extreme.

There has been interest in the past in using main-effect genetic risk scores (GRS) as genetic variables in order to improve statistical power to detect gene-environment interactions (40). Such interactions may be viewed from a lens in which genetic risk corresponds to a predisposition that is only triggered in certain environments (e.g. dietary behaviors). Here, we observed only a minor association of a main-effect GRS for LDL-C with greater LDL-C reductions in the DM trial ( $p=4.5e-27$ ). This trend runs counter to a prior observation of greater lifestyle intervention effectiveness for LDL-C reduction in those with low genetic risk of hyperlipidemia (41). This discrepancy may be due to differences between the DM trial and the personalized diet and lifestyle changes recommended in the intervention in question. Regardless, the meaningful increase in predictive power of the GDI-based FRS compared to the main-effect GRS for LDL-C indicates the value in using interaction-based scores rather than simple genetic predispositions for the development of personalized dietary recommendations.

A diet response score such as that developed here is most useful if its value extends beyond just risk factor changes and predicts downstream changes in chronic disease and mortality risk. Suggestive interactions for CHD, ischemic stroke, and diabetes were apparent across strata (Figure 2a,b,d), corresponding to a decreased risk in fat-reduction participants (whose LDL-C would be expected to drop more prominently according to the score). In contrast, hemorrhagic stroke showed the opposite trend, with a positive score-disease relationship only in the fat reduction group. This result is in line with existing evidence for the detrimental effects of low LDL-C on hemorrhagic stroke risk (42). Non-CVD death showed no major associations, which



could be expected due to the dominance of this category by cancer outcomes and the equivocal associations of cancer with lipids (43). We note that all disease outcome relationships assessed here are subject to the major caveat that dietary evolution and decreased adherence likely developed over time in many subjects, diluting the utility of the randomization and 1-year changes used for stratification in Figure 2.

The present study had the advantage of developing a GDI-based dietary fat score in almost 10,000 women and testing in a dietary intervention trial using independent individuals from the same population. However, nominal main-effect SNPs were prioritized to improve statistical power given this moderate sample size, an approach which may fail to identify interactions with effect directions opposite that of the main effect. Smaller fractions of alternate ancestries in this population also made development of ancestry-specific response scores unrealistic. Additionally, the DM trial intervention in which the scores were tested may not exactly match the intervention relevant to the purely fat reduction-focused score developed here; it included additional non-fat-related dietary recommendations that may have effected the interactions examined here, and did not ultimately achieve its intended 20% fat reduction. Finally, this study only examined women, despite the fact that CRF profiles and their genetic trait architectures are known to vary across sexes (44).

In summary, we present a method for the development of diet response scores based on genome-wide gene-diet interaction study summary statistics. While the resulting dietary fat response scores are not all informative, a score focused on LDL-C is predictive of 1-year LDL-C changes during a fat reduction trial. Its performance is risk factor-specific, and is superior to that of a main-effect genetic risk score for LDL-C. Furthermore, it displays suggestive relationships with chronic disease outcomes, following known coronary heart disease biology and adding a new perspective to known discrepancies between associations of LDL-C with cardiovascular disease and stroke subtypes. These findings support the utility of gene-diet interactions for personalized nutrition while highlighting the need for increased sample sizes and improved diet measures for the discovery of robust genetic predictors of diet response.

## References

1. Ma Q, Lu AYH. Pharmacogenetics, Pharmacogenomics, and Individualized Medicine. *Pharmacological Reviews*. 2011;63:437–59.
2. Corella D, Carrasco P, Sorli JV, Estruch R, Rico-Sanz J, Martinez-Gonzalez MA, Salas-Salvado J, Covas MI, Coltell O, Aros F, et al. Mediterranean Diet Reduces the Adverse Effect of the TCF7L2-rs7903146 Polymorphism on Cardiovascular Risk Factors and Stroke Incidence: A randomized controlled trial in a high-cardiovascular-risk population. *Diabetes Care*. 2013;36:3803–11.
3. Huang T, Huang J, Qi Q, Li Y, Bray GA, Rood J, Sacks FM, Qi L. PCSK7 Genotype Modifies Effect of a Weight-Loss Diet on 2-Year Changes of Insulin Resistance: The POUNDS LOST Trial. *Diabetes Care*. 2015;38:439–44.
4. Ordovas JM, Ferguson LR, Tai ES, Mathers JC. Personalised nutrition and health. *BMJ*. 2018;bmj.k2173.
5. Corella D. APOA2, Dietary Fat, and Body Mass Index. *Archives of Internal Medicine*. 2009;169:1897.
6. Cuda C, Badawi A, Karmali M, El-Sohemy A. Polymorphisms in Toll-like receptor 4 are associated with factors of the metabolic syndrome and modify the association between dietary saturated fat and fasting high-density lipoprotein cholesterol. *Metabolism*. 2011;60:1131–5.
7. Cornelis MC, El-Sohemy A, Kabagambe EK, Campos H. Coffee, CYP1A2 Genotype, and Risk of Myocardial Infarction. *JAMA*. 2006;295:1135.
8. Qi Q, Chu AY, Kang JH, Jensen MK, Curhan GC, Pasquale LR, Ridker PM, Hunter DJ, Willett WC, Rimm EB, et al. Sugar-Sweetened Beverages and Genetic Risk of Obesity. *New England Journal of Medicine*. 2012;367:1387–96.
9. Olsen NJ, Ängquist L, Larsen SC, Linneberg A, Skaaby T, Husemoen LLN, Toft U, Tjønneland A, Halkjær J, Hansen T, et al. Interactions between genetic variants associated with adiposity traits and soft drinks in relation to longitudinal changes in body weight and waist circumference. *The American Journal of Clinical Nutrition*. 2016;104:816–26.
10. Smith CE, Follis JL, Dashti HS, Tanaka T, Graff M, Fretts AM, Kilpeläinen TO, Wojczynski MK, Richardson K, Nalls MA, et al. Genome-Wide Interactions with Dairy Intake for Body Mass Index in Adults of European Descent. *Molecular Nutrition & Food Research*. 2018;62:1700347.
11. Figueiredo JC, Hsu L, Hutter CM, Lin Y, Campbell PT, Baron JA, Berndt SI, Jiao S, Casey G, Fortini B, et al. Genome-Wide Diet-Gene Interaction Analyses for Risk of Colorectal Cancer. Amos CI, editor. *PLoS Genetics*. 2014;10:e1004228.

12. Dempfle A, Scherag A, Hein R, Beckmann L, Chang-Claude J, Schäfer H. Gene–environment interactions for complex traits: definitions, methodological requirements and challenges. *European Journal of Human Genetics*. 2008;16:1164–72.
13. Anderson GL, Cummings SR, Freedman LS, Furberg C, Henderson MM, Johnson SR, Kuller LH, Manson JE, Oberman A, Prentice RL, et al. Design of the Women’s Health Initiative Clinical Trial and Observational Study. *Controlled Clinical Trials*. 1998;19:61–109.
14. Ritenbaugh C, Patterson RE, Chlebowski RT, Caan B, Fels-Tinker L, Howard B, Ockene J. The women’s health initiative dietary modification trial: overview and baseline characteristics of participants. *Annals of Epidemiology*. 2003;13:S87–97.
15. Patterson RE, Kristal AR, Tinker LF, Carter RA, Bolton MP, Agurs-Collins T. Measurement Characteristics of the Women’s Health Initiative Food Frequency Questionnaire. *Annals of Epidemiology*. 1999;9:178–87.
16. Rao DC, Sung YJ, Winkler TW, Schwander K, Borecki I, Cupples LA, Gauderman WJ, Rice K, Munroe PB, Psaty BM. Multiancestry Study of Gene–Lifestyle Interactions for Cardiovascular Traits in 610 475 Individuals From 124 Cohorts. *Circulation: Cardiovascular Genetics*. 2017;10.
17. R Core Team. R: A language and environment for statistical computing. [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2017. Available from: <https://www.r-project.org/>
18. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015;4:7.
19. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*. 2010;38:e164–4.
20. Gauderman WJ. Sample Size Requirements for Association Studies of Gene-Gene Interaction. *American Journal of Epidemiology*. 2002;155:478–84.
21. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, Frayling TM, Hirschhorn J, Yang J, Visscher PM. Meta-analysis of genome-wide association studies for height and body mass index in 700000 individuals of European ancestry. *Human Molecular Genetics*. 2018;27:3641–9.
22. Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, Smith AV, Tobin MD, Verwoert GC, Hwang SJ, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature*. 2011;478:103–9.
23. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, Mora S, et al. Discovery and refinement of loci associated with lipid levels. *Nature Genetics*. 2013;45:1274–83.
24. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, Wheeler E, Glazer NL, Bouatia-Naji N, Gloyn AL, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nature Genetics*. 2010;42:105–16.

25. Cha S, Kang J, Lee J-H, Kim J, Kim H, Yang Y, Park W-Y, Kim J. Impact of Genetic Variants on the Individual Potential for Body Fat Loss. *Nutrients*. 2018;10:266.
26. Munda M, Legrand C. Adjusting for centre heterogeneity in multicentre clinical trials with a time-to-event outcome. *Pharmaceutical Statistics*. 2014;13:145–52.
27. Thompson JF, Hyde CL, Wood LS, Paciga SA, Hinds DA, Cox DR, Hovingh GK, Kastelein JJ. Comprehensive Whole-Genome and Candidate Gene Analysis for Response to Statin Therapy in the Treating to New Targets (TNT) Cohort. *Circulation: Cardiovascular Genetics*. 2009;2:173–81.
28. Bell JT, Tsai P-C, Yang T-P, Pidsley R, Nisbet J, Glass D, Mangino M, Zhai G, Zhang F, Valdes A, et al. Epigenome-Wide Scans Identify Differentially Methylated Regions for Age and Age-Related Phenotypes in a Healthy Ageing Population. Li J, editor. *PLoS Genetics*. 2012;8:e1002629.
29. Perfilyev A, Dahlman I, Gillberg L, Rosqvist F, Iggman D, Volkov P, Nilsson E, Risérus U, Ling C. Impact of polyunsaturated and saturated fat overfeeding on the DNA-methylation pattern in human adipose tissue: a randomized controlled trial. *The American Journal of Clinical Nutrition*. American Society for Nutrition; 2017;105:991–1000.
30. Pei Q, Huang Q, Yang G-p, Zhao Y-c, Yin J-y, Song M, Zheng Y, Mo Z-h, Zhou H-h, Liu Z-q. PPAR- $\gamma$ 2 and PTPRD gene polymorphisms influence type 2 diabetes patients' response to pioglitazone in China. *Acta Pharmacologica Sinica*. 2013;34:255–61.
31. Comuzzie AG, Cole SA, Laston SL, Voruganti VS, Haack K, Gibbs RA, Butte NF. Novel Genetic Loci Identified for the Pathophysiology of Childhood Obesity in the Hispanic Population. Crawford DC, editor. *PLoS ONE*. 2012;7:e51954.
32. Spracklen CN, Chen P, Kim YJ, Wang X, Cai H, Li S, Long J, Wu Y, Wang YX, Takeuchi F, et al. Association analyses of East Asian individuals and trans-ancestry analyses with European individuals reveal new loci associated with cholesterol and triglyceride levels. *Human Molecular Genetics*. 2017;26:1770–84.
33. Badawi A, Cuda, Garcia-Bailo, Karmali, El-Sohemy. A common polymorphism near the interleukin-6 gene modifies the association between dietary fat intake and insulin sensitivity. *Journal of Inflammation Research*. 2012;1.
34. Zheng Y, Huang T, Zhang X, Rood J, Bray GA, Sacks FM, Qi L. Dietary Fat Modifies the Effects of FTO Genotype on Changes in Insulin Sensitivity. *The Journal of Nutrition*. 2015;145:977–82.
35. Ordovás JM, Robertson R, Cléirigh EN. Gene–gene and gene–environment interactions defining lipid-related traits. *Current Opinion in Lipidology*. 2011;22:129–36.
36. Saito M, Eto M, Nitta H, Kanda Y, Shigeto M, Nakayama K, Tawaramoto K, Kawasaki F, Kamei S, Kohara K, et al. Effect of Apolipoprotein E4 Allele on Plasma LDL Cholesterol Response to Diet Therapy in Type 2 Diabetic Patients. *Diabetes Care*. 2004;27:1276–80.

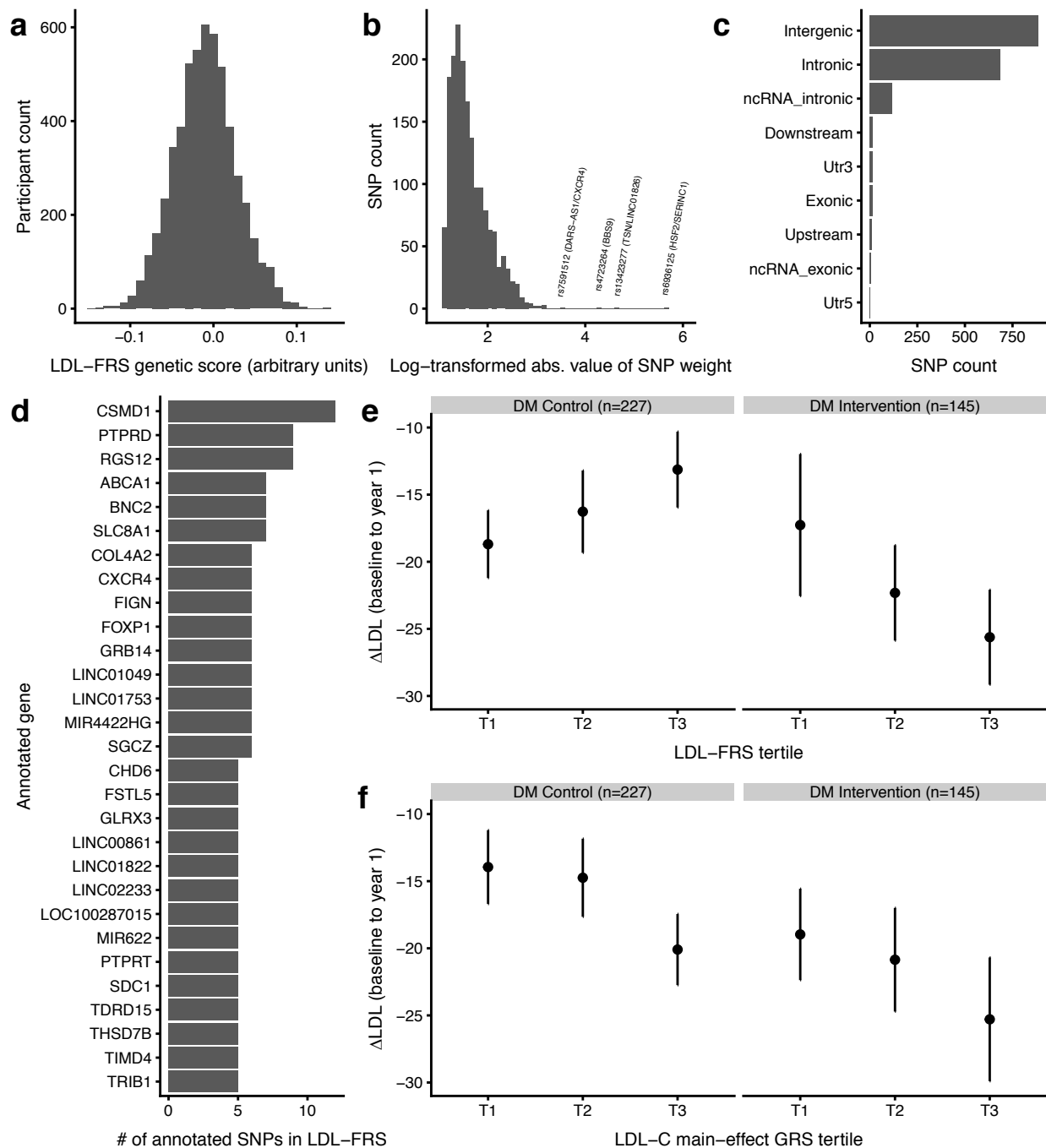
37. Zhang X, Qi Q, Bray GA, Hu FB, Sacks FM, Qi L. APOA5 genotype modulates 2-y changes in lipid profile in response to weight-loss diet intervention: the Pounds Lost Trial. *The American Journal of Clinical Nutrition*. 2012;96:917–22.
38. Xu M, Ng SS, Bray GA, Ryan DH, Sacks FM, Ning G, Qi L. Dietary Fat Intake Modifies the Effect of a Common Variant in the LIPC Gene on Changes in Serum Lipid Concentrations during a Long-Term Weight-Loss Intervention Trial. *The Journal of Nutrition*. 2015;145:1289–94.
39. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P, Lander ES, Lubitz SA, Ellinor PT, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics*. 2018;50:1219–24.
40. Aschard H. A perspective on interaction effects in genetic association studies. *Genetic Epidemiology*. 2016;40:678–88.
41. Zubair N, Conomos MP, Hood L, Omenn GS, Price ND, Spring BJ, Magis AT, Lovejoy JC. Genetic Predisposition Impacts Clinical Changes in a Lifestyle Coaching Program. *Scientific Reports*. 2019;9:6805.
42. Sun L, Clarke R, Bennett D, Guo Y, Walters RG, Hill M, Parish S, Millwood IY, Bian Z, Chen Y, et al. Causal associations of blood lipids with risk of ischemic stroke and intracerebral hemorrhage in Chinese adults. *Nature Medicine*. 2019;25:569–74.
43. Koene RJ, Prizment AE, Blaes A, Konety SH. Shared Risk Factors in Cardiovascular Disease and Cancer. *Circulation*. 2016;133:1104–14.
44. Knopp RH, Paramsothy P, Retzlaff BM, Fish B, Walden C, Dowdy A, Tsunehara C, Aikawa K, Cheung MC. Sex differences in lipoprotein metabolism and dietary response: Basis in hormonal differences and implications for cardiovascular disease. *Current Cardiology Reports*. 2006;8:452–9.

**Table 1:** Baseline characteristics of European-ancestry participants in the Women's Health Initiative Study (n = 17304 in total)

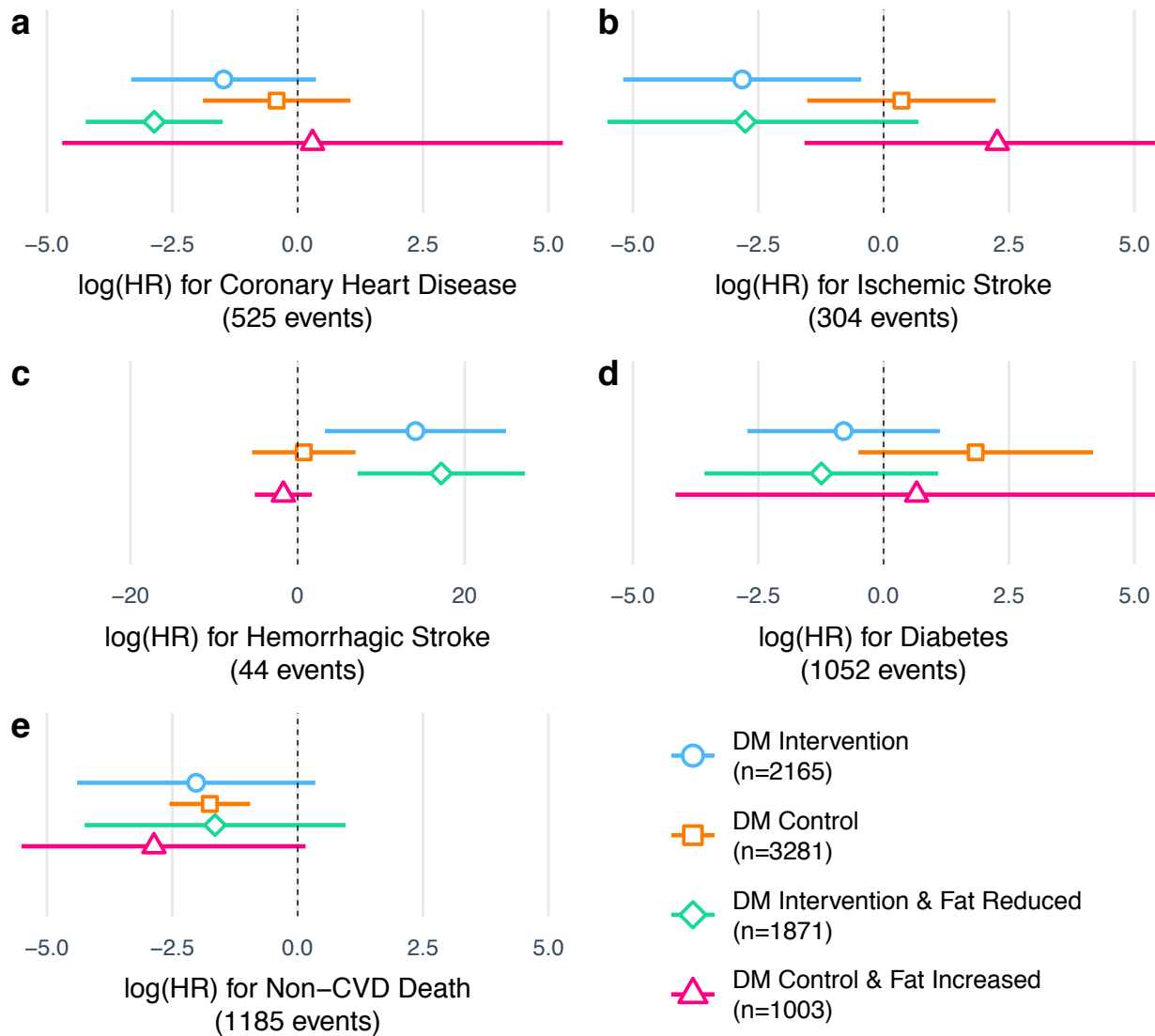
|   | DM trial          | Non-DM trial     |
|---|-------------------|------------------|
| Sample size                               | 6173              | 11131            |
| Age                                       | 66 (61-70)        | 68 (64-72)       |
| Current smoking                           | 403 (7%)          | 994 (9%)         |
| Lipid-lowering medication                 | 655 (11%)         | 1454 (13%)       |
| Hypertension medication                   | 2334 (38%)        | 4023 (36%)       |
| Diabetes medication                       | 300 (5%)          | 560 (5%)         |
| Body mass index (BMI; kg/m <sup>2</sup> ) | 28.7 (25.2-32.8)  | 26.9 (23.7-30.9) |
| Systolic blood pressure (SBP; mm Hg)      | 128.5 (117.5-140) | 144 (132-156)    |
| LDL cholesterol (LDL-C; mg/dL)            | 150.8 (127.5-175) | 161 (134-191)    |
| HDL cholesterol (HDL-C; mg/dL)            | 50 (42.5-58)      | 51 (44-60)       |
| Triglycerides (TG; mg/dL)                 | 137 (100-194)     | 128 (92-179.6)   |
| Fasting glucose (FG; mg/dL)               | 97 (90-107)       | 97 (90-113.3)    |

**Table 2:** Responder score effects on 1-year CRF changes in DM trial participants

| Risk factor | # SNPs in risk score | N    | Std. effect size | P-value | Std. effect size | P-value |
|-------------|----------------------|------|------------------|---------|------------------|---------|
| BMI         | 6042                 | 1988 | 0.03             | 0.221   | 0.02             | 0.253   |
| SBP         | 1536                 | 2004 | 0.03             | 0.196   | 0.04             | 0.061   |
| LDL-C       | 1760                 | 145  | -0.18            | 0.026   | -0.19            | 0.007   |
| HDL-C       | 1731                 | 150  | -0.06            | 0.471   | -0.06            | 0.483   |
| TG          | 1774                 | 150  | -0.14            | 0.066   | -0.14            | 0.064   |
| FG          | 1924                 | 281  | 0.02             | 0.791   | 0.03             | 0.570   |



**Figure 1: LDL-FRS characterization.** a) Distribution of LDL-FRS in WHI DM trial participants. b) Distribution of SNP weights constituting the LDL-FRS (shown as the natural log-transformed absolute values of the true weights). c) SNP counts in different loci types for LDL-FRS constituent SNPs. d) Genes are summarized by the number of annotated SNPs in the LDL-FRS (genes with at least 5 component SNPs are shown). e-f) 1-year changes in LDL-C in DM trial participants as a function of genetic scores. Mean changes in LDL-C (y-axis) are shown as a function of either LDL-FRS (e) or LDL-GRS (f) tertile (x-axis). Error bars represent standard errors. LDL-FRS: LDL-C fat response genetic score, LDL-GRS: LDL-C main-effect genetic score.



**Figure 2:** LDL-FRS prediction of chronic disease development. X-axis shows log-hazard ratio estimates for the LDL-FRS from Cox proportional hazards regression for a) coronary heart disease, b) ischemic stroke c) hemorrhagic stroke, d) diabetes, and e) non-CVD death. Separate estimates are shown for DM trial intervention arm, control arm, and the same strata filtered for FFQ-reported fat reduction or increase, respectively. Cox models are adjusted for age at baseline and include a random effect for WHI genotyping sub-study. Error bars represent 95% confidence interval estimates for the regression coefficients.