S3 Text. Cross-sectional analyses in EPIC InterAct and United States NHANES

Metabolic factors

Regression analyses were performed to examine relationships of the fatty acid pattern score with metabolic risk factors and with dietary factors. The results are presented in Table 3 and Figure 3. In EPIC-InterAct, estimates from each country were pooled by meta-analysis adjusting for each of potential confounders as shown in the footnote of Table 2. In the National Health and Nutrition Examination Survey (NHANES), weighting individuals by sampling probability, the model included selected potential confounders: age, gender, race (White, African American, Asian, Mexican Hispanic, other Hispanic), education (<high school, high school, >high school), annual family income (<$15,000, $15,000-$34,999, $35,000 or more), smoking status (never, former, current), alcohol (spline terms), prevalent diseases (positive for either coronary heart disease, hyperlipidaemia, or hypertension), family history of diabetes, and family history of myocardial infarction. Body-mass index (BMI) was additionally adjusted for in the analyses for metabolic risk factors, with exception of analysis of BMI. Risk for non-alcoholic fatty liver disease was evaluated by modified Poisson regression [1]. Analyses using different cutpoints for non-alcoholic fatty liver disease showed similar results.

Genetic factors

In the analysis to assess the association between genetic risk factors and the fatty acid pattern score, genetic variables were considered as causal determinants of the fatty acid pattern score. Therefore, genetic risk scores were treated as independent variables and the fatty acid pattern score was treated as a dependent variable. For interpretability, these variables were scaled so that a regression coefficient can be interpreted as % difference in the fatty acid pattern score per the interdecile range (10th and 90th percentile range). Regression results were obtained by country and pooled by random-effects meta-analysis. The measure of heterogeneity, $I^2$, was 0%, indicating no heterogeneity in the impact of the selected genetic effects on the fatty acid pattern. A few considerations were taken in regards to confounding. First, the analyses were performed with adjustment for potential confounders.
as selected above and also for genotyping sources (metabochip or not). Second, as genetic variability was determined at the time of conception, any modifiable factors at cohort baseline could not be confounders. Thus, analyses were repeated with adjustment only for centres, age, sex, and genotyping sources (not shown because results indicated no meaningful impact). Third, in this case-cohort design [2], cases were oversampled, but there was little concern of reverse causality in this genetic analysis. Thus, analyses were performed, stratified by the cohort status (sub-cohort or not) and pooled by random-effects meta-analysis.

**Dietary factors**

We performed dietary analysis in each of EPIC-InterAct and NHANES to identify dietary exposure associated with the main fatty acid pattern identified in EPIC-InterAct. In EPIC-InterAct, we examined adults included in the sub-cohort. We evaluated adults with information both on fatty acid profiles and dietary consumption, retaining 15,566 adults in EPIC-InterAct (n excluded=353 from the sub-cohort) and 1,500 adults in NHANES (n excluded=66). In EPIC-InterAct, dietary variables were derived from responses to food-frequency questionnaires and dietary records in a standardized manner [3–5]. In NHANES, dietary information from two 24-hour dietary recalls was used [6,7]. In each study population, dietary factors were modelled as independent variables in linear regression analysis adjusting for potential confounders. In EPIC-InterAct, country-specific estimates were first obtained and pooled by random-effects meta-analysis.

In nutrient-based analysis, saturated fatty acids, monounsaturated fatty acids, polyunsaturated fatty acids, carbohydrates, and protein were evaluated. We considered that dietary fatty acids influence circulating fatty acids directly and carbohydrates and proteins from a diet are converted to fatty acids through enzymatic pathways including citric acid cycle and de novo lipogenesis [8]. Fibre intake was examined because of its association with type 2 diabetes and its effects on metabolism of lipids [9,10]. For specific dietary fat, carbohydrates, and protein, we performed isocaloric substitution analysis [11]. We fitted a multivariable-adjusted linear regression model to estimate hypothetical effects of replacing 5% of total calories from carbohydrates intakes with 5% of total calories from each type of
dietary fats and protein [11]. The model simultaneously included fibre intake and adjusted for socio-demographic variables, lifestyle factors, prevalent clinical characteristics, total energy intake, and consumption of alcoholic beverages which was examined in food-based analysis.

In food-based analysis, we performed analysis first in EPIC-InterAct and then selected a small subset of food variables that showed trends toward positive or negative associations with the fatty acid pattern score. Then, in NHANES, we examined the associations of the small subset of foods or beverages with the fatty acid pattern score. Variables evaluated in the EPIC-InterAct were selected on the basis of associations with fatty acids or type 2 diabetes or fatty acid profiles, including potatoes, rice or pasta, cereals, fruits, vegetables, nuts, legumes, vegetable oils, margarine, cheese, yoghurt, meats, processed meats, fish, soft drinks, coffee, tea, and alcoholic beverages [10,12–14]. We first evaluated each of these dietary factors without co-adjustment for each other (S4 Fig). To focus a fewer number of these dietary factors, we performed a variable selection procedure (backward selection) in EPIC-InterAct, using results from meta-analysis combining country-specific estimates for the food variables. In the backward selection, p<0.2 was used to keep variables showing significant associations or non-significant trends of associations; the p-value cutpoint was considered to be informative for descriptive purpose and important for adjustment for dietary confounding [15]. In NHANES, 4,573 food variables were compiled with 8-digit food code numbers assigned. Using the food code numbers and associated information [6], we created food variables that corresponded to a small subset of food variables identified in EPIC-InterAct. Including the food variables simultaneously in a multivariable-adjusted linear regression model, we examined associations of consumption of the foods with the fatty acid pattern score in NHANES.

For interpretability, the fatty acid pattern score was scaled to one standard deviation (SD) so that regression coefficient meant a difference of the fatty acid pattern score by percentages of one SD according to 1 unit difference in dietary variables. Each of dietary variables was also scaled so that 1 unit became practically informative (e.g. 1 serving per day).
References


6. USDA Food and Nutrient Database for Dietary Studies, 2.0 [Internet]. Beltsville, MD: Agricultural Research Service, Food Surveys Research Group; 2006.

7. Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS), National Health and Nutrition Examination Survey Data, 2003-2004. [Internet].


