Title

Metabolomic Profile of the Healthy Eating Index-2015 in the Multiethnic Study of Atherosclerosis

Journal

Journal of Nutrition, 153(8)

ISSN

0022-3166

Authors

Gadgil, Meghana D
Wood, Alexis C
Karaman, Ibrahim
et al.

Publication Date

2023-08-01

DOI

10.1016/j.tjnut.2023.05.030

Metabolomic Profile of the Healthy Eating Index-2015 in the Multi-Ethnic Study of Atherosclerosis

Running title: HEI-2015 and representative metabolites in MESA

Meghana D. Gadgil[a], Alexis C. Wood[b], Ibrahim Karaman[c], Goncalo Graça[d], Ioanna Tzoulaki[c], Victor W. Zhong[e], Philip Greenland[f], Alka M. Kanaya[a], David M. Herrington[g]

[a] Division of General Internal Medicine, Department of Medicine, University of California, San Francisco, CA; 1545 Divisadero Street, Suite 320, San Francisco, CA 94143-0320

[b] USDA/ARS Children's Nutrition Research Center, 1100 Bates Avenue, Houston, TX 77071

[c] National Phenome Centre, Imperial College London, IRDB Building 5th Floor, Hammersmith Hospital Campus, London, W12 0NN

[d] Section of Bioinformatics, Division of Systems Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College London, South Kensington Campus, Sir Alexander Fleming Building, London SW7 2AZ, U.K.

[e] School of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai, China.

[f] Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL

[g] Section on Cardiovascular Medicine, Department of Internal Medicine, Wake Forest School of Medicine; Medical Center Boulevard, Winston-Salem, NC 27157

Corresponding Author:
Meghana D. Gadgil
c/o University of California, San Francisco, Box 0320
1545 Divisadero Street
San Francisco, CA 94143-0320
Phone: (415) 353-7922
E-mail: meghana.gadgil@ucsf.edu

Word Count:
Abstract: 300
Text: 3481
Tables: 3

Conflict of Interest and Disclosures: The authors declare no conflicts of interest. ACW has received funding from Hass Avocado Board, The National Cattleman's Beef Association and Ionis Pharmaceuticals for work unrelated to the current analyses.

Abbr:
CVD: cardiovascular disease
HEI-2015: Health Eating Index 2015
MESA: Multi-Ethnic Study of Atherosclerosis
T2D: type 2 diabetes

Keywords: diet patterns; metabolomics; Healthy Eating Index

Funding Source

MDG is supported by K23DK119404 from the National Institute of Diabetes and Digestive and Kidney Diseases.

COMBI-BIO Research supported by EU COMBI-BIO project (FP7, 305422), and NIH/NHLBI (R01HL133932).

This research was supported by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168 and N01-HC-95169 from the National Heart, Lung, and Blood Institute, and by grants UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420 from the National Center for Advancing Translational Sciences (NCATS). The authors thank the other investigators, the staff, and the participants of the MESA study for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org.

ABSTRACT

INTRODUCTION

Poor diet quality is a risk factor for type 2 diabetes and cardiovascular disease. However, knowledge of metabolites marking adherence to Dietary Guidelines for Americans (2015 version; DGA-15) are limited. The goal was to determine a pattern of metabolites associated with the Healthy Eating Index-2015 (HEI-2015), which measures adherence to the DGA.

METHODS

The analysis examined 3557 adult men and women from the longitudinal cohort Multi-Ethnic Study of Atherosclerosis (MESA), without known cardiovascular disease and with complete dietary data. Fasting serum specimens, diet and demographic questionnaires were assessed at baseline. Untargeted ^{1}H NMR 1DNMR spectroscopy (600 MHz) was used to generate metabolomics and lipidomics. A metabolome-wide association study (MWAS) specified each spectral feature as outcomes, HEI-2015 score as predictor, adjusting for age, gender, race, and study site in linear regression analyses. Subsequently, hierarchical clustering defined discrete groups of correlated NMR features associated with named metabolites and linear regression analysis assessed for associations with HEI-2015 total and component scores.

RESULTS

The sample included 50% women with average age of 63 years, with 40% identifying as White, 23% Black, 24% Hispanic and 13% Chinese American. The average HEI-2015 score was 66. MWAS identified 179 spectral features significantly associated with HEI-2015 score. Cluster analysis identified seven clusters representing 4 metabolites; HEI-2015 score was significantly associated with all. HEI-2015 score was associated with proline betaine (ß 0.12 [0.02]; p=4.70 E-13) and was inversely related to proline (ß -0.13 [0.02]; p=4.45 E-14), 1,5 anhydrosorbitol (ß -

0.08 [0.02]; p=4.37 E-07) and unsaturated fatty acyl chains (ß 0.08 [0.02]; p=8.98 E-07). Intake of total fruit, whole grains and seafood and plant proteins was associated with proline betaine.

CONCLUSIONS

Diet quality was significantly associated with unsaturated fatty acyl chains, proline betaine, proline. Further analysis may clarify the link between diet quality, metabolites, and pathogenesis of cardiometabolic disease.

INTRODUCTION

Poor diet quality is independently associated with incidence of cardiovascular disease [1, 2], cancer[3] and type 2 diabetes (T2D). [4-6] The Healthy Eating Index 2015 (HEI-2015) is a measure of diet quality reflecting adherence to the Dietary Guidelines for Americans 2015-2020 (DGA 2015-2020.[7] The DGA 2015-2020 represents dietary guidance jointly published by the US Department of Agriculture and the US Department of Health and Human Services every five years, reflecting recommendations for ideal intake by the US Government. An important update to the HEI-2015 from earlier versions is a recommendation to limit intakes of both Added Sugars and Saturated Fats to <10% of energy.

The identification of small molecules, called metabolites, present in serum, urine or tissue, may help to shed light on the phenotypic links between habitual diet quality and disease. Diet quality is a complex, long-term exposure, likely affects multiple metabolic processes simultaneously, and habitual diet intake may produce a stable metabolic environment that is linked with risk for disease. Prior assessments of the HEI-2015 score and associated metabolites have been limited to targeted or commonly annotated metabolites, which may not capture the full metabolome representing consumption of a higher quality diet.[8] Previous work has also demonstrated that there may be stronger links between diet-associated circulating metabolites and disease than the original association between diet quality and disease outcomes.[9-11] A deeper assessment using NMR-based spectral features may allow for a more nuanced assessment of diet quality, which may support future assessment of diet quality and association with disease.

The objective of this investigation was to determine a pattern of metabolites associated with habitual diet quality as represented by the HEI-2015 and its components. This analysis

profiled serum untargeted NMR-based metabolomics to gain insight into metabolic features associated with high diet quality.

METHODS

Participants

We included 3557 adult men and women, determined through self-reported gender, from the Multi-Ethnic Study of Atherosclerosis (MESA) longitudinal cohort study without known cardiovascular disease at enrollment visit and with stored serum samples with available NMR-based COMBInatorial BIOmarkers for subclinical atherosclerosis[12] (COMBI-Bio) metabolomic profiling data available for analysis. MESA is a U.S.-based prospective cohort study of 6814 participants between the ages of 45 to 84 years recruited at six sites (Baltimore City and County, Maryland; Chicago, Illinois; Forsyth County, North Carolina; New York, New York; Los Angeles County, California; and St. Paul, Minnesota), designed to investigate the development and progression of subclinical atherosclerotic disease. Participants were enrolled between 2000-2002,[13] did not have cardiovascular disease at baseline and were purposively recruited from four race/ethnicity categories (Black, White, Chinese-American and Hispanic). Institutional review board approval was obtained at all participating centers, and all participants gave informed consent.

We included 3663 participants with available metabolomics data from the baseline examination. We further excluded 106 with implausible caloric intake (<600 kcal/day or >6000 kcal/day, in concordance with prior MESA publications[2, 4, 14]), or who were missing two-thirds or more of diet data. Of these, 3557 participants have available metabolomics measures from the baseline examination.

Data and Biospecimens

We assessed clinical and demographic data using questionnaires administered at baseline. Fasting biospecimens were collected at baseline and stored at -80C until analyzed. Participants were asked to fast for 12 h, avoid smoking on the morning of the exam, and avoid heavy exercise 12 h before the exam.

Metabolomic Profiling

Nuclear magnetic resonance measurements were carried out according to a previously published protocol using serum samples.[15] Briefly, a standard ^{1}H NMR one-dimensional (1D NMR) spectrum with water suppression was obtained for each sample, detecting signatures of all proton containing compounds, including sharp peaks from small molecule species and broad peaks from lipoproteins and proteins. Subsequent spectral processing was performed using the software TOPSPIN 3.1 (Bruker Biospin, Rheinstetten, Germany). The spectra were automatically phased and baseline corrected, and the chemical shifts were calibrated to the glucose signal at 5.233ppm. Spectral data were imported into MATLAB [Version 8.3 (R2014a) Mathworks Inc., Natick, MA, USA] for further processing, including peak alignment and normalization using PQN method.[16]

The spectral features were annotated using the following spectral information, chemical shift (ppm), the coupling constant (J in Hz), the peak multiplicity (singlet, doublet, and multiplet), and peak connectivity of the NMR signals from the 1D and 2D NMR spectra [2D JRES, cOrrelation SpectroscopY (COSY), tOtal Correlation SpectroscopY (TOCSY), Heteronuclear single quantum correlation spectroscopy (HSQC)] and statistical correlation methods [STOCSY (Statistical Total Correlation Spectroscopy) and STORM (Subset

Optimisation by Reference Matching)].[17] Annotations were also assessed using information from available in-house and publicly available spectral databases as well as with published data.

Diet assessment

Usual dietary intake over the past 12 months was assessed at baseline, from a self-administered 120-item food frequency questionnaire (FFQ) which evaluated diet intake over the past year. The MESA FFQ is a modified version of the Insulin Resistance Atherosclerosis Study (IRAS) FFQ, which was previously validated in non-Hispanic whites, those of Hispanic ethnicity and those who identify as Black.[18] The MESA FFQ was modified from that used in IRAS to include dietary intake common among Chinese-Americans. For each food item, participants indicated the average serving size and the frequency each food was eaten. Frequency ranged from "rare or never" to a maximum of "2+ times per day" for foods and "6+ times per day" for beverages. Average daily servings of forty-seven food groups were created using weighted recipes from the Nutrition Data System for Research (NDSR) and estimated per 100g of food and were used as the basis for creation of the diet score.

HEI-2015 score

The HEI-2015 was designed to align with the 2015-2020 Dietary Guidelines for Americans (DGAs).[7] The HEI-2015 contains 13 components, the sum of which totals to a maximum score of 100 points. As in HEI-2005 and HEI-2010, each of the components is scored on a density basis out of 1,000 calories, with the exception of fatty acids, which is a ratio of unsaturated to saturated fatty acids.

There are nine adequacy components: Total Fruits, Whole Fruits, Total Vegetables, Greens and Beans, Whole Grains, Dairy, Total Protein Foods, Seafood and Plant Proteins and Fatty Acids for which greater consumption is the goal. For four moderation components, we

assigned higher scores with minimization of intake for the following food groups: Refined Grains, Sodium, Added Sugars, Saturated Fats.

Statistical Analysis

Metabolome-wide Association Study (MWAS): The association of all 30,590 spectral features, which were mean-centered and scaled to unit variance, with HEI-2015 score was run using linear regression models specifying each spectral feature as the outcome in separate models, with standard deviation of HEI-2015 score as the predictor, and age (continuous), race (categorical), gender (binary), and data collection site (categorical) as covariates. A spectral decomposition based on the correlation matrix between all spectra suggested that the effective number of independent tests (ENT) was 22,857. Significance for associations between spectral features and HEI-2015 score was therefore set at Bonferroni-corrected significance level of $P<2.2*10^{-6}$ (.05/22857).

Elastic net regularized regression: To adjust for unreliable parameter estimates that may occur when using multiple regression models in the setting of multicollinearity, we performed an elastic net regularized regression model to evaluate metabolites that were significant in independent analyses. The elastic-net model allowed for a penalized logistic regression on all biomarkers simultaneously to identify the metabolites most highly associated with diet pattern score. Elastic net regularized regression models were run with HEI-2015 diet score as the predictor and spectral features showing a significant association with the HEI-2015 diet score in MWAS analysis as the outcomes. Optimal penalty parameters for the penalty value (mixing percentage; α) and the strength of the penalty (regularization penalty; λ) were ascertained via the package 'caret' in R using cross validation. Briefly, data in the full dataset were randomly assigned to one of two equal sized datasets. Parameter selection was conducted via resampling of

models with 100 values of λ chosen according to the caret algorithm. The final selected parameters were then applied to analyses on the whole dataset. Optimization was reached via feature-wise normalization change in successive coordinate descent iterations. Model performance was judged based on root mean square error of approximation (RMSEA), with α and λ parameters giving rise to the minimum mean cross-validated error used to generate new coefficients for the association of spectral features with HEI-2015 score.

Clustering analysis: Pearson correlations were run between all spectral features with non-zero coefficients in the elasticnet regularized regression models, to allow for identification of clusters or groups of spectral features. As groups of spectral features showed specific patterns of intercorrelations, all spectral features with non-zero coefficients from the regularized regression models were subject to hierarchical clustering analysis. Hierarchical clustering analysis was conducted using the package 'NbClust' in R. Euclidean distance was used to compute the dissimilarity matrix, with total within-cluster variance computed using Ward (1963) algorithm to minimize the total within-cluster variance. The optimal number of clusters was identified using the Duda-Hart stopping rule. For clusters with contributions of spectral features from more than one annotation, we assigned the metabolite with the most prominent signals. Methanol/proline was assigned as proline due to the presence of a coefficient of association of proline with the same beta coefficient as Proline/methanol and histidine/proline betaine was assigned as proline betaine based on the absence of non-overlapping signals from histidine within that spectral feature.

Final Associations between HEI-2015 diet score and metabolomics cluster scores As several spectral features may be representative of the same metabolite, to assist in interpretability and most accurately represent the presence of individual metabolites, sum scores for all the spectral

features within a cluster were created. Based on the annotations assigned to the spectral feature, the most likely metabolite or metabolites represented by each cluster score was assigned. Cluster scores were highly skewed, thus were winsorized and represented as 4 standard deviations (SDs) +/- the mean and transformed using a blom transformation. Associations were analyzed from linear regression models with HEI-2015 diet score standardized using z-score as the predictor, transformed cluster scores as the outcomes in separate models, and age, gender, race and site of data collection as fixed effects and were standardized. Significance was retained as a Bonferroni correction for the original number of ENTs in the MWAS (of $P<2.2*10^{-6}$). For all cluster scores significantly associated with HEI-2015, multivariable linear regression models were run with the cluster score as the outcomes in separate models, all thirteen components of HEI-2015 score as the predictors within the same model, and age, gender, race and site of data collection as fixed effects. Significance was set at a Bonferroni correction for 7 tests (.05/7 = P<.007).

RESULTS

The sample of participants self-identified as 50% women, and 13% of participants as Black, 23% of Hispanic ethnicity, 24% Chinese American and 40% non-Hispanic white, with a mean age of 63 years. (Table 1). Average HEI-2015 score was 66. HEI-2015 score was significantly associated with 179 1D-NMR-based spectral features determined through MWAS analysis. (Supplemental table 1 and Supplemental Figure 1).

The clustering analysis identified 7 main clusters of metabolomic spectral features each identified by a single metabolite or lipid (Table 2 and Supplemental Figure 2). Four out of seven clusters contained spectral features annotated to the amino acid proline. A higher HEI-2015 score, reflecting better diet quality, was associated with a lower abundance of proline ($p<0.007$,

corrected for 7 cluster comparisons). The strongest association was found between HEI-2015 score and proline betaine (0.12 [0.02]; p=4.70 E-13).

Intake of specific HEI-2015 components was differentially associated with the defined clusters of metabolomic spectral features. (Table 3) The HEI-2015 score component "Total Dairy" was associated with four clusters, representing 1,5-anhydrosorbitol and methanol/proline. Higher intake of dairy products was linked with lower abundance of both metabolites, mirroring the findings of total HEI-2015 score and these metabolites.

Intake of the HEI-2015 component "Total Fruits" had strong, positive associations with proline betaine (ß 0.18 [SE=0.02]; p=3.24E-12). Higher intake of Whole Grains (ß 0.05 [SE 0.01]; p=2.54E-03) and Seafood and Plant Protein (ß 0.08 [SE 0.018]; p=1.05E-03) was also associated with higher relative proline betaine abundance. Intake of refined grains was inversely associated with methanol/proline, most significantly in cluster 4, (ß -0.08 [0.02] p=7.07E-05). (Table 3)

DISCUSSION

In this investigation, diet quality as measured by the HEI-2015 score was associated with four metabolites in participants in the MESA cohort study. The strongest associations were between higher HEI-2015 score and the amino acid proline betaine, and an inverse association with the amino acid proline. Each cluster-associated metabolite was differentially associated with food groups. Greater intake of Total Fruits, Whole Grains and Seafood and Plant protein was associated with higher relative abundance of proline betaine. Intake of dairy products, total protein and refined grains was also negatively associated with abundance of proline.

Diet quality in the United States is low, with an average HEI-2015 score of 59/100 as surveyed by NHANES in 2015-2016.[19] Dietary intake representing high diet quality can vary,

representing broad food group categories rather than narrow associations with individual foods. Examinations of past HEI versions have found associations between a higher HEI score and a lower risk of cardiovascular disease and mortality.[20, 21] This finding supports copious observational evidence that diets of high quality, generally represented by high intake of fruit, vegetable, whole grain and plant-based protein and low intake of added sugars, salt, refined carbohydrates and red meat are associated with a lower incidence of chronic cardiometabolic disease.[22-25] The metabolic changes and mechanisms that may underlie these associations, however, less clear, and the goal was to clarify representative metabolites that may indicate high diet quality.

A higher HEI-2015 score, representing better diet quality, was associated with higher abundance of proline betaine. Proline betaine is also a biomarker of citrus consumption, [11] reflected in this analysis with the positive association between Total Fruit intake and this amino acid. In our prior work in the Mediators of Atherosclerosis in South Asians Living in America (MASALA) study, consumption of the Fruits, Vegetables, Nuts, Legumes diet pattern, a high-quality diet pattern, was similarly associated with proline betaine [26]. The DGA and most guidelines on diet intake emphasize fruit and vegetable intake as markers of high diet quality. As intake of fruits and vegetables likely occurs concurrently with other high quality foods, an increase in concentration of this metabolite may serve as a general indicator for improved consumption of a high-quality diet in the general population.

Previous epidemiologic studies have shown poor cardiometabolic risk [27] and insulin resistance [28] associated with lower concentrations of betaine in diverse populations. Proline betaine and its analogue, glycine betaine, were also associated with lower risk for T2D in the

Diabetes Prevention Program and other intervention and cohort studies.[29, 30] Deficiency of betaine was additionally linked with increased severity of non-alcoholic fatty liver disease (NAFLD).[31]

Betaine is derived from the amino acid glycine, and acts as a methyl donor to allow the conversion of homocysteine to methionine. [32] Betaine is also a precursor of TMAO, a possible marker of cardiometabolic risk[28, 33], and is likely processed by fecal microbiota into this compound. In the current analysis, whole grain intake was also associated with proline betaine levels. In an investigation in mice, consumption of rye bran increased the diversity of gut microbiota and provided a source of glycine betaine, which was metabolized into other betaine compounds which remained at high levels in the rye bran-fed group [34]. The presence of diverse microbiota from an overall healthful diet may promote higher concentrations of betaine and its metabolites throughout the gut and plasma. Despite these positive observational findings and promising preclinical data from animal studies, direct supplementation of betaine in humans during a randomized, controlled trial showed only minor improvements in fasting glucose, and no changes in dynamic measurements of insulin sensitivity and intrahepatic triglycerides.[35] All together, this suggests that diet intake including whole grains and cereal fiber may support a healthful gut microbial environment allowing for increasing levels of betaine and its metabolites, associated with lower risk for cardiometabolic disease. A deeper exploration of the choline-betaine metabolic pathways after whole grain intake may yield insights into the pathogenesis of diabetes and NAFLD.

Total HEI-2015 score was inversely associated with the amino acid proline. Increased levels of proline have previously been associated with insulin resistance in South Asian and

Chinese men of low body mass index, suggesting that this metabolite may reflect metabolic differences underlying T2D independent of those caused by obesity.[36] This metabolite has also been inversely associated with HEI-2015 in a study of African-American and European populations,[8] in an analysis restricted to known metabolites. Proline has recently been implicated in the gut-brain axis and an indicator for the severity of depression. In a multi-cohort analysis, circulating proline had the strongest association of all metabolites with worsened depression scores [37]. Those with high proline consumption and high plasma proline levels had a preponderance of the gut microbiota species *Parabacteroides* and *Acidaminococcus.* Interestingly, these gut microbiota species were also associated with higher depression scores. As we found a lower diet quality was associated with higher circulating proline, the promotion of a healthful gut environment through improved diet quality may help explain links between HEI-2015 score and depression [38].

1,5 anhydrosorbitol (1,5 anhydroglucitol) is a marker of short-term glycemic control, is inversely related to glucose concentration, and is used as a validated marker of daily glucose changes. In our study, a higher HEI-2015 score was associated with lower 1,5 anhydrosorbitol levels. Higher intake of Total Dairy was similarly associated with lower circulating concentrations of this metabolite, replicating a finding in normoglycemic individuals in Japan [39]. It is readily absorbed from a variety of foods and is generally present in stable levels in the body as it is excreted almost without metabolism. This metabolite was also indicative of high saturated fat intake in in a controlled diet trial of high saturated fat compared with n-6 fatty acids [40] – higher diet quality in our study is defined by lower saturated fat intake likely leading to this finding. However, circulating levels of this metabolite have been shown to decrease with a lower intake of overall carbohydrates or lower glycemic index under controlled dietary intake

conditions.[41] Lower levels of this metabolite have also been linked with an increase in major adverse cardiovascular events [42], however there is a stronger relationship among people with diabetes [43]. At a population level, lower intake of saturated fat and higher intake of dairy products as components of a higher HEI-2015 score may be reflected as lower 1,5 anhydrosorbitol levels. In populations with diabetes, however, the effect of glycemic variability on this marker likely supercedes changes from diet intake due to competitive inhibition with glucose excretion in the renal tubules, and it is not likely to be a good indicator of diet quality in this population.

HEI-2015 score was positively associated with unsaturated fatty acyl chains (C=CHCH2HC=C). Fatty acyls are one of eight categories of lipids and include many different fats. The HEI-2015 component Fatty acids, which represents the ratio of unsaturated to saturated fatty acid intake, was associated with higher Cluster 1 (unsaturated fatty acyl) score. The intake of unsaturated fatty acids has been linked to improved health outcomes, including omega-3 fatty acids and cardiovascular disease.[44] The association of higher HEI-2015 overall score to greater ratio of unsaturated:saturated fatty acids was in line with expected healthy eating guidelines.

Strengths of this analysis include a longitudinal cohort design with robust habitual dietary data collection through a comprehensive food frequency questionnaire, characterization of diet in multiple racial and ethnic groups and comprehensive evaluation of untargeted NMR spectral features beyond known metabolites. Despite multiple strengths, we acknowledge that our analysis also has limitations. These findings were not externally validated, although our sample size and methodology allows for adequate internal validation. This is a cross-sectional analysis performed at one time point, and data collected from food frequency questionnaires are subject to recall bias. The food frequency questionnaire data collected information on habitual diet

intake over the past 12 months, but do not quantify this intake at the time point of blood sampling; biomarkers may be affected by more proximate diet intake. The MESA Study food frequency questionnaire was modified to include unique Chinese foods and culinary practices, but was not validated in this population. Untargeted metabolomics is a broad-based analysis for identifying all possible markers as a snapshot of metabolism, and this observational analysis cannot establish causal relationships between controlled diet intake and metabolites. Still, our characterization of metabolites associated with HEI-2015 remains the first to broadly examine NMR spectral features associated with this dietary quality score rather than restricting the analysis to known metabolites.

Conclusion

HEI-2015 score was associated with spectral features representing proline betaine, proline, 1,5 anhydrosorbitol and fatty acyl chains in the MESA cohort study. These metabolites may represent increased whole grain, fruit, dairy and lower saturated fat intake as indicators of overall high diet quality. Further investigation into controlled diet intake will help to clarify links between diet quality and onset of cardiometabolic disease and areas for preventive action.

Statement of Author Contributions

MDG and DH designed research; MDG and AW analyzed data, MDG wrote the paper. MDG had primary responsibility for final content. AW, IK, GG, IT, VWZ, PG, DH, AMK contributed to the interpretation of the results and revised the manuscript. All authors read and approved the final manuscript.

This paper has been reviewed and approved by the MESA Publications and Presentations Committee.

REFERENCES

1. Yang, Q., et al., *Added Sugar Intake and Cardiovascular Diseases Mortality Among US Adults.* JAMA internal medicine, 2014.
2. Nettleton, J.A., et al., *Dietary patterns and incident cardiovascular disease in the Multi-Ethnic Study of Atherosclerosis.* The American Journal of Clinical Nutrition, 2009. **90**(3): p. 647-654.
3. Clinton, S.K., E.L. Giovannucci, and S.D. Hursting, *The World Cancer Research Fund/American Institute for Cancer Research Third Expert Report on Diet, Nutrition, Physical Activity, and Cancer: Impact and Future Directions.* J Nutr, 2020. **150**(4): p. 663-671.
4. Nettleton, J.A., et al., *Dietary patterns and risk of incident type 2 diabetes in the Multi-Ethnic Study of Atherosclerosis (MESA).* Diabetes care, 2008. **31**(9): p. 1777-1782.
5. InterAct, C., et al., *Mediterranean diet and type 2 diabetes risk in the European Prospective Investigation into Cancer and Nutrition (EPIC) study: the InterAct project.* Diabetes care, 2011. **34**(9): p. 1913-1918.
6. Chiuve, S.E., et al., *Alternative dietary indices both strongly predict risk of chronic disease.* The Journal of nutrition, 2012. **142**(6): p. 1009-1018.
7. Krebs-Smith, S.M., et al., *Update of the Healthy Eating Index: HEI-2015.* J Acad Nutr Diet, 2018. **118**(9): p. 1591-1602.
8. Kim, H., et al., *Serum Metabolites Associated with Healthy Diets in African Americans and European Americans.* J Nutr, 2021. **151**(1): p. 40-49.
9. Playdon, M.C., et al., *Identifying biomarkers of dietary patterns by using metabolomics.* The American Journal of Clinical Nutrition, 2016.
10. Esko, T., et al., *Metabolomic profiles as reliable biomarkers of dietary composition.* The American Journal of Clinical Nutrition, 2017.
11. Guertin, K.A., et al., *Metabolomics in nutritional epidemiology: identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations.* The American Journal of Clinical Nutrition, 2014. **100**(1): p. 208-217.
12. Karaman, I., et al., *Workflow for Integrated Processing of Multicohort Untargeted 1H NMR Metabolomics Data in Large-Scale Metabolic Epidemiology.* Journal of proteome research, 2016. **15**(12): p. 4188-4194.
13. Bild, D.E., et al., *Multi-Ethnic Study of Atherosclerosis: objectives and design.* American Journal of Epidemiology, 2002. **156**(9): p. 871-881.
14. Nettleton, J.A., et al., *Dietary patterns, food groups, and telomere length in the Multi-Ethnic Study of Atherosclerosis (MESA).* The American Journal of Clinical Nutrition, 2008. **88**(5): p. 1405-1412.
15. al., D.A.C.J.B.S.H.e., *Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale*

metabolic phenotyping. Analytical Chemistry, 2014. **86**(19): p. 9887-9894.
16. Frank Dieterle, A.R., Gotz Schlotterbeck, and Hans Senn, *Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex*
Biological Mixtures. Application in 1H NMR Metabonomics. Anal Chem, 2006(78): p. 4281-4290.
17. Posma, J.M., et al., *Subset optimization by reference matching (STORM): an optimized statistical approach for recovery of metabolic biomarker structural information from 1H NMR spectra of biofluids.* Analytical Chemistry, 2012. **84**(24): p. 10694-10701.
18. Mayer-Davis, E.J., et al., *Dietary intake in the diabetes prevention program cohort: baseline and 1-year post randomization.* Annals of Epidemiology, 2004. **14**(10): p. 763-772.
19. Prevention, C.f.D.C.a., *National Health and Nutrition Examination Survey Data, 2015-2016*. 2019, U.S. Department of Health and Human Services: Hyattsville, MD.
20. Shan, Z., et al., *Association Between Healthy Eating Patterns and Risk of Cardiovascular Disease.* JAMA Intern Med, 2020. **180**(8): p. 1090-1100.
21. Xu, Z., et al., *Diet quality, change in diet quality and risk of incident CVD and diabetes.* Public Health Nutr, 2020. **23**(2): p. 329-338.
22. Onvani, S., et al., *Adherence to the Healthy Eating Index and Alternative Healthy Eating Index dietary patterns and mortality from all causes, cardiovascular disease and cancer: a meta-analysis of observational studies.* J Hum Nutr Diet, 2017. **30**(2): p. 216-226.
23. Harmon, B.E., et al., *Associations of key diet-quality indexes with mortality in the Multiethnic Cohort: the Dietary Patterns Methods Project.* The American Journal of Clinical Nutrition, 2015. **101**(3): p. 587-597.
24. Hirahatake, K.M., et al., *Cumulative average dietary pattern scores in young adulthood and risk of incident type 2 diabetes: the CARDIA study.* Diabetologia, 2019.
25. Shikany, J.M., et al., *Associations between food groups, dietary patterns, and cardiorespiratory fitness in the Coronary Artery Risk Development in Young Adults study.* The American Journal of Clinical Nutrition, 2013. **98**(6): p. 1402-1409.
26. Gadgil, M.D., et al., *Dietary patterns are associated with metabolic risk factors in South asians living in the United States.* The Journal of nutrition, 2015. **145**(6): p. 1211-1217.
27. Konstantinova, S.V., et al., *Divergent associations of plasma choline and betaine with components of metabolic syndrome in middle age and elderly men and women.* J Nutr, 2008. **138**(5): p. 914-20.
28. Fu, B.C., et al., *Associations of plasma trimethylamine N-oxide, choline, carnitine, and betaine with inflammatory and cardiometabolic risk biomarkers and the fecal microbiome in the Multiethnic Cohort*

Adiposity Phenotype Study. Am J Clin Nutr, 2020. **111**(6): p. 1226-1234.
29. Walford, G.A., et al., *Metabolite Profiles of Diabetes Incidence and Intervention Response in the Diabetes Prevention Program.* Diabetes, 2016. **65**(5): p. 1424-33.
30. Sun, Y., et al., *Metabolomics Signatures in Type 2 Diabetes: A Systematic Review and Integrative Analysis.* J Clin Endocrinol Metab, 2020. **105**(4).
31. Morgan, E.K.K.M.G.N.B.C.S.M.A.C.T.B.S.W.F.a.T.R., *Betaine improves nonalcoholic fatty liver and associated insulin resistance: a potential mechanism for hepatoprotection by betaine.* Am J Physiol Gastrointest Liver Physiol, 2010. **299**(5): p. G1068–G1077.
32. Craig, S.A., *Betaine in Human Nutrition.* Am J Clin Nutr, 2004. **80**: p. 539-49.
33. Krueger, E.S., T.S. Lloyd, and J.S. Tessem, *The Accumulation and Molecular Effects of Trimethylamine N-Oxide on Metabolic Tissues: It's Not All Bad.* Nutrients, 2021. **13**(8).
34. Koistinen, V.M., et al., *Contribution of gut microbiota to metabolism of dietary glycine betaine in mice and in vitro colonic fermentation.* Microbiome, 2019. **7**(1): p. 103.
35. Grizales AM, P.M., Lin AP, et al., *Metabolic Effects of Betaine: A Randomized Clinical Trial of Betaine Supplementation in Prediabetes.* J Clin Endocrinol Metab. **103**(8): p. 3038-3049.
36. Tai, E.S., et al., *Insulin resistance is associated with a metabolic profile of altered protein metabolism in Chinese and Asian-Indian men.* Diabetologia, 2010. **53**(4): p. 757-67.
37. Mayneris-Perxachs, J., et al., *Microbiota alterations in proline metabolism impact depression.* Cell Metab, 2022. **34**(5): p. 681-701 e10.
38. Wang, K., et al., *Higher HEI-2015 Score Is Associated with Reduced Risk of Depression: Result from NHANES 2005-2016.* Nutrients, 2021. **13**(2).
39. Koga, M., et al., *Habitual intake of dairy products influences serum 1,5-anhydroglucitol levels independently of plasma glucose.* Diabetes Res Clin Pract, 2010. **90**(1): p. 122-5.
40. Pigsborg, K., et al., *Effects of changing from a diet with saturated fat to a diet with n-6 polyunsaturated fat on the serum metabolome in relation to cardiovascular disease risk factors.* Eur J Nutr, 2022. **61**(4): p. 2079-2089.
41. Juraschek, S.P., et al., *Effects of dietary carbohydrate on 1,5-anhydroglucitol in a population without diabetes: results from the OmniCarb trial.* Diabet Med, 2017. **34**(10): p. 1407-1413.
42. Ikeda, N., et al., *Impact of serum 1,5-anhydro-d-glucitol level on prediction of major adverse cardiac and cerebrovascular events in non-diabetic patients without coronary artery disease.* Atherosclerosis, 2016. **253**: p. 1-6.

43. Selvin, E., et al., *Association of 1,5-Anhydroglucitol With Cardiovascular Disease and Mortality.* Diabetes, 2016. **65**(1): p. 201-8.
44. Abdelhamid, A.S., et al., *Omega-3 fatty acids for the primary and secondary prevention of cardiovascular disease.* Cochrane Database Syst Rev, 2020. **3**: p. CD003177.

Acknowledgements

MDG is supported by 1K23DK119404 from the National Institute of Diabetes and Digestive and Kidney Diseases.

COMBI-BIO Research supported by EU COMBI-BIO project (FP7, 305422), and NIH/NHLBI (R01HL133932).

This research was supported by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168 and N01-HC-95169 from the National Heart, Lung, and Blood Institute, and by grants UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420 from the National Center for Advancing Translational Sciences (NCATS). The authors thank the other investigators, the staff, and the participants of the MESA study for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org.

This paper has been reviewed and approved by the MESA Publications and Presentations Committee.

Data described in the manuscript, code book, and analytic code will be made available upon request pending approval of the MESA cohort study.

Table 1: Baseline characteristics of MESA cohort participants by Healthy Eating Index 2015 (HEI-2015) Quartile (N=3557)

	Mean (SD)	Quartile 1	Quartile 2	Quartile 3	Quartile 4
N	3557	847	880	907	923
Women (%)	1787 (50)	334	403	465	585
Age, years	63 (10)	60 (10)	62 (10)	63 (10)	65 (10)
Race N (%)					
White	1428 (40)	337 (40)	318 (36)	368 (41)	405 (44)
Black	830 (23)	229 (27)	208 (24)	178 (20)	215 (23)
Hispanic	838 (24)	148 (17)	221 (25)	260 (29)	209 (23)
Chinese-American	461 (13)	133 (16)	133 (15)	101 (11)	94 (10)
Healthy Eating Index-2015 Score	66 (8)	56 (4)	64 (2)	69 (1)	76 (4)
BMI (kg/m^2)	28 (5)	28 (6)	29 (6)	28 (5)	28 (5)
Diabetes n (%)	470 (13)	99 (14)	114 (13)	136 (15)	121 (13)
Hypertension n (%)	1608 (45)	365 (43)	387 (44)	407 (45)	449 (49)

Table 2: Associations of Healthy Eating Index-2015 Diet Score with Representative Metabolites and Lipids

Cluster	Spectral features	Metabolite association	Beta[a,b]	SE	P
1	2.765603, 2.769304, 2.769641, 2.770313, 2.77065	C=CHCH2HC=C (fatty acyl chains)	0.08	0.02	8.98 E-07
2	3.100354, 3.10069, 3.101027, 3.101363	Proline betaine/histidine	0.12	0.02	4.70 E-13
3	3.268907	1,5-anhydrosorbitol	-0.08	0.02	4.37 E-07
4	3.3261, 3.326437	Proline	-0.09	0.02	5.46 E-08
5	3.342249, 3.347968	Methanol/proline	-0.10	0.02	4.06 E-10
6	3.34494, 3.345277	Methanol/proline	-0.12	0.02	1.63 E-12
7	3.34595, 3.346286	Methanol/proline	-0.13	0.02	4.45 E-14

[a]Standardized estimates

[b]Adjusted for age, gender, race, and study site

585 **Table 3: Associations of HEI2015 component scores with metabolomic cluster scores**
586

Component	Cluster	Most likely Annotation	Beta[a,b]	SE	P
Total fruit	2	Proline betaine/histidine	0.18	0.02	3.25E-12
Whole grains	2	Proline betaine/histidine	0.05	0.02	2.54E-03
Total dairy	3	1,5-anhydrosorbitol	-0.06	0.02	1.29E-03
	5	Methanol/proline	-0.12	0.01	7.31E-10
	6	Methanol/proline	-0.11	0.01	2.86E-09
	7	Methanol/proline	-0.13	0.01	1.28E-11
Total protein	5	Methanol/proline	-0.10	0.03	1.74E-04
	6	Methanol/proline	-0.11	0.03	3.65E-05
	7	Methanol/proline	-0.11	0.03	2.42E-05
Seafood and plant protein	2	Proline betaine/histidine	0.08	0.02	1.05E-03
Fatty acid	1	C=CHCH2HC=C (fatty acyl chains)	0.02	0.01	3.83E-03
Refined grains	5	Methanol/proline	-0.07	0.02	7.07E-05
	6	Methanol/proline	-0.06	0.02	3.12E-04
	7	Methanol/proline	-0.06	0.02	2.52E-04

587
588 [a]Standardized estimates
589 [b]Adjusted for age, gender, race, and study site