Metabolic Profiling of High Egg Consumption and the Associated Lower Risk of Type 2 Diabetes in Middle-Aged Finnish Men

Noerman, S

Wiley

Tieteiliset aikakauslehtiartikkelit
© WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim
All rights reserved
http://dx.doi.org/10.1002/mnfr.201800605

https://erepo.uef.fi/handle/123456789/7356
Downloaded from University of Eastern Finland's eRepository
Metabolic profiling of high egg consumption and the associated lower risk of type 2 diabetes in middle-aged Finnish men

Stefania Noerman¹, Olli Kärkkäinen¹, Anton Mattsson³, Jussi Paananen³, Marko Lehtonen², Tarja Nurmi¹, Tomi-Pekka Tuomainen¹, Sari Voutilainen¹, Kati Hanhineva¹,4,* and Jyrki K Virtanen¹,*

¹Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio
²School of Pharmacy, University of Eastern Finland, Kuopio
³Bioinformatics Center, University of Eastern Finland, Kuopio
⁴LC-MS Metabolomics Center, Biocenter Kuopio, Kuopio
⁵Department of Mathematics and System Analysis, Aalto University, Helsinki
*Both authors contributed equally to the study

Correspondence: Jyrki K. Virtanen, University of Eastern Finland, Institute of Public Health and Clinical Nutrition, P.O. Box 1627, 70211 Kuopio, Finland. Phone: +358-30-355 2957, Fax: +358-17-162 936, E-mail: jyrki.virtanen@uef.fi.

Abbreviations

BCAA, branched-chain amino acids; FDR, false-discovery rate; KIHD, Kuopio Ischaemic Heart Disease Risk Factor Study; LC-PUFA, long-chain polyunsaturated fatty acid; (Lyso)PC, (lyso)phosphatidylcholine; OCFA, odd-chain fatty acid; OR, odds ratio; T2D, type 2 diabetes; TMAO, trimethylamine N-oxide
**Keywords:** eggs, LC-MS, metabolomics, serum, type 2 diabetes

**Abstract**

*Scope:* Higher egg intake was previously associated with a lower risk of developing type 2 diabetes (T2D) in the prospective, population-based Kuopio Ischaemic Heart Disease Risk Factor Study (KIHD) in eastern Finland. We explored potential compounds that could explain this association using non-targeted LC-MS-based metabolic profiling.

*Methods and results:* We analyzed 239 baseline serum samples from the KIHD in 4 groups: subjects with higher (mean intake 1 egg/d) or lower (mean intake 2 eggs/wk) egg intake who developed T2D (cases) or remained healthy (controls) during the mean follow-up of 19.3 years. We observed different serum profiles of subjects who had either higher or lower egg intakes, and of those who developed type 2 diabetes or remained healthy. The higher baseline tyrosine level predicted higher odds of T2D (OR 1.94; 95% CI 1.45, 2.60; \( P < 0.001; \) FDR 0.023) along with an unknown hexose-containing compound (OR 2.13; 95% CI 1.57, 2.88; \( P < 0.001; \) FDR 0.005). Certain predominant metabolites in T2D cases were correlated positively with ones in lower-egg-intake group and negatively with ones in higher-egg-intake group.

*Conclusion:* Our current findings may underline some potential metabolites that could explain how egg intake was associated with a lower risk of T2D.

1. **Introduction**

High egg intake has traditionally been discouraged because the cholesterol content may, in addition to the effects on serum lipid profile, impair glucose metabolism and promote inflammation, thereby may potentially increasing the risk of developing type 2 diabetes (T2D).\(^{[1-3]}\) Interestingly, clinical trials have shown beneficial rather than harmful effects of high egg intake on many risk factors of cardiovascular diseases and T2D. For example,
increase in egg intake for a few months has lowered plasma insulin and insulin resistance,[4] inflammatory markers, and improved the serum lipid profile.[5] Some recent studies showed a null association[6] between high egg intake and the T2D risk, albeit the conclusions differ across study populations.[7] An association with a higher T2D risk has been observed mainly in the USA, possibly confounded by other lifestyle and dietary practices associated with a higher egg intake.[7-9] These include higher BMI, smoking, lower physical activity, and higher consumption of red meat, which all are potential risk factors for T2D.[6-10]

In middle-aged and older men in the prospective, population-based Kuopio Ischaemic Heart Disease Risk Factor Study (KIHD) in eastern Finland, egg intake was not systematically linked with either better or worse lifestyle or dietary habits that could explain the observed associations.[11] Even so, we found that higher egg intake was associated with a lower risk of incident T2D in that study population.[11] This finding suggested that the observed inverse association may be due to the modulation of endogenous metabolism mediated or moderated by compounds present in eggs. Noteworthy, eggs are an especially rich source of several bioactive compounds, such as carotenoids and choline, which have been shown to have beneficial effects on, for example, insulin resistance,[12] inflammation,[13] and lipid oxidation and metabolism.[5]

Here we present the application of non-targeted metabolic profiling to identify potential metabolites that would differ based on the amount of egg intake and would be associated with risk of developing T2D. Using a nested case-control study setting, we aim to explore the metabolites that could at least partly explain the association between egg intake and the lower T2D risk in the participants of KIHD study. Furthermore, we investigate the correlations between metabolic profiles linked with each trait to highlight the metabolites that deserve attention in future studies directed to elucidate the causal mechanisms underlying the possible health benefits of eggs, especially in decreasing the risk of developing T2D.
2. Materials and Methods

2.1. Study population

The study population consisted of a total of 2682 male participants of KIHD (83% of those eligible) aged 42, 48, 54 or 60 years old at baseline examinations in 1984-1989. The baseline characteristics of the whole study population have been described. The KIHD protocol got approval from the Research Ethics Committee of the University of Kuopio and was performed in compliance with Declaration of Helsinki. All subjects gave written informed consent for participation. All data were handled anonymously.

2.1.1. Dietary assessment

Food consumption at baseline was assessed with a 4-d food record, guided by the book containing a list of 126 foods and drinks most commonly consumed in Finland when the study started in the 1980s, each with a corresponding picture and portion size. Total egg intake also included eggs in mixed dishes and recipes. A nutritionist gave the instructions for filling the dietary records and reviewed the completed records. Nutrient intakes were estimated using the NUTRICA® 2.5 software (Social Insurance Institution, Turku, Finland) which mainly based on the nutrient composition values of Finnish foods. Because there is no information on choline and phosphatidylcholine (PC) values in this Finnish composition table, we used the nutrients’ values from the USDA database.

2.1.2. Diagnosis of type 2 diabetes

T2D diagnosis at baseline was defined based on self-reported physician diagnosis and/or fasting plasma glucose ≥7.0 mmol/L. The diagnosis of incident T2D during follow-up was based on self-administered questionnaires, fasting and 2-h oral glucose tolerance test, blood glucose measurement at the re-examination rounds 4, 11, and 20 years after the baseline, and by record linkage to the registry of hospital discharge and reimbursement on...
diabetes medication expenses. During the mean follow-up of 19.3 years, 432 men developed T2D as previously described.\cite{11}

### 2.1.3. Collection of blood samples and other measurements

At the baseline examination, subjects were requested to abstain from alcohol for 3 d and from smoking and eating for 12 hours before the blood sample collection. The venous blood samples were collected between 0800 and 1000. The recording of habitual physical activity,\cite{18} smoking and alcohol consumption in the past 12 months,\cite{19} and the analytical procedures of the blood lipid profile,\cite{19} plasma glucose, serum insulin and C-reactive protein have been described.\cite{11} The serum samples were stored at -80 °C.

### 2.1.4. Selection of samples

Men with a diagnosis of T2D or impaired fasting glucose at baseline, or men with a diagnosis of coronary heart disease or cancer at baseline and before the T2D diagnosis during follow-up were excluded (Supporting Information Figure S1). To limit the heterogeneity of the subjects included in the current study, men without baseline data on dietary intakes, with an energy intake of <1700 kcal/day, or with a body mass index (BMI) of <20 or >30 kg/m², were also excluded. Among the remaining men, we randomly selected 264 participants who met the following criteria and sorted them into four groups: subjects with higher or lower egg intake who developed (cases) or remained free (controls) of T2D during the mean follow-up of 19.3 years. Among the selected 264 men, the serum samples were not available for 25 men, leaving 239 men for the metabolomics analysis, with following composition: 1) 61 controls with higher egg intake, 2) 60 controls with lower egg intake, 3) 60 T2D cases with higher egg intake, and 4) 58 T2D cases with lower egg intake.

### 2.1.5. Dosage information

Because we used data from an observational, population-based study among free-living subjects, we did not pre-specified the amounts of egg consumption. The definitions of higher
and lower egg intakes were based on the range of *ad libitum* egg intakes in the study population, estimated using 4-day food recording.

2.2. Non-targeted metabolic profiling analysis

2.2.1. Sample preparation

The serum samples were randomized, recoded, and properly thawed on ice. Then, a 100 µl of each sample was added to each well in the Captiva ND filter plate (Agilent Technologies) containing 400 µl of acetonitrile (LC-MS grade) and mixed by pipette. The protein-free filtrate was collected on a 96-well polypropylene plate after centrifugation (700 rcf, 4 °C, 5 minutes). A pooled sample was prepared for quality control and injected after every 12 analytical samples.

2.2.2. LC-MS platform and data acquisition

The utilized platform in this study was liquid chromatography quadrupole time-of-flight mass spectrometry (LC-QTOF-MS). Liquid chromatography was run on 1290 Infinity Binary UPLC system (Agilent Technologies, Santa Clara, CA, USA) with both reverse phase (RP, Zorbax Eclipse XDB-C18, particle size 1.8 µm, dimensions of 2.1x100 mm, Agilent Technologies, USA) and hydrophilic interaction chromatography (HILIC, Acquity UPLC® BEH Amide 1.7 µm, 2.1x100 mm, Waters, Ireland) columns. Mass spectrometry used a 6540 UHD Accurate-Mass Q-TOF MS (Agilent Technologies, Santa Clara, CA, USA) with both positive and negative electrospray ionization (ESI) modes in the previously described conditions. The data-dependent tandem mass spectrometry (MS/MS) analysis used three different collision-induced dissociation voltages (10, 20, and 40 V) in subsequent runs. Mass calibration was done continuously (Supporting Information Table S1). Data were acquired using Agilent MassHunter Acquisition B.07.00 software (Agilent Technologies). The differential metabolites that were not included in the data-dependent
MS/MS analysis underwent targeted LC-MS/MS analysis on one of the samples that contained the most abundant metabolites of interest (Supporting Method and Supporting Information Table S1).

2.2.3. Data pre-processing

Initial peak picking was performed using “Find by Molecular Feature” algorithm in Agilent MassHunter Qualitative Analysis B.07.00 (Agilent Technologies). Centroid spectra peaks higher than 400 counts were restricted to ion species [M-H]{-} and [M+Cl]{-} in negative and [M+H]{+} and [M+Na]{+} in positive modes. Data files (.cef) were aligned and transformed using the Z-baselining method in Mass Profiler Professional software (MPP version 2.2, Agilent Technologies). After the initial alignment, the data were combined into a reference file (.cef) against which the original raw data was reanalyzed. For this recursive analysis, the tolerance for compound mass was ±15 ppm, retention time ±0.2 min, and symmetric expansion value ±10 ppm for chromatograms. The output was then re-exported to MPP for peak alignment and data cleanup. Subsequent filtering in MPP included only entities that were present in at least 50% of samples in at least one study group. As the results, we got 1190 and 555 metabolic features from HILIC, and 2405 and 1227 features from RP, in positive and negative mode, respectively. The combined data matrix comprised 5376 signals from each of 239 subjects which then underwent statistical analysis.

2.2.4. Identification of differential metabolites

Differential metabolites after statistical analysis were filtered based on several inclusion criteria: $P < 0.05$, found in $\geq 80\%$ of samples in at least one group, mass $< 1000$ Da, and retention time $\geq 0.5$ min. Fold change was used to assess the magnitude of differences across compared groups, calculated by dividing the mean peak area of each signal from the cases by that of the controls or from the high by that of the low-egg-intake groups. The metabolites of interest were reported based on the guidelines from Sumner et al.\textsuperscript{[22]} The compounds in level
were matched against mass, retention time, and MS/MS spectra of fragmented ions from the in-house library built by running commercial standards using the same instrument and experimental condition. Level II includes compounds with matching mass, ion charge, and spectra of fragmented ions from METLIN [23] or HMDB [24] version 3.6 with a mass tolerance of 10 ppm. Lipophilic compounds were matched against the built-in MS-DIAL library version 2.58 [25] or LIPID MAPS. [26] Identification of phospholipids [27] in positive and negative modes, [28] acylcarnitines [29,30] and γ-glutamylated-branched-chain amino acids (BCAA) [31] have been previously described. Level III includes compounds belong to a particular group based on their mass, retention time or a specific fragmentation pattern but we could not identify due to lack of reference spectra or other necessary information. Level IV includes unknown compounds.

2.3. Statistical analysis

We checked the baseline differences using the Chi-Square test for categorical variables and analysis of variance (ANOVA) for continuous variables (IBM SPSS Statistics 21, IBM Corporation). Metabolite values were normalized and transformed to a normal distribution using rank-inverse normalization. Associations between egg consumption and serum metabolites were analyzed using linear regression, where egg consumption was modelled as the independent variable and serum metabolites were modelled as the dependent variable. Hence, the positive slope indicated that egg intake predicted the higher levels of serum metabolite and negative slope indicated otherwise. Association between serum metabolites and the incidence of T2D was analyzed using logistic regression and presented as odds ratio (OR). OR<1 indicated that the serum metabolite predicted a lower risk of T2D, whilst OR>1 indicated a higher risk. P-values for all the metabolomics analyses were adjusted using Benjamini-Hochberg false-discovery rate (FDR). P<0.05 was considered statistically significant.
significant. The normalization, regression, and correlation of metabolite signals were analyzed using R 3.2.2 (R Foundation for Statistical Computing). Analysis of covariance (ANCOVA, IBM SPSS Statistics 21, IBM Corporation) was used to adjust the levels of (lyso)PC that contains odd-chain fatty acid (OCFA) for intakes of total energy, dairy, and fiber.

3. Results

3.1. Baseline characteristics of subjects

Assuming one medium-sized egg weighs approximately 50 g, subjects in the high-egg-intake groups had a mean intake of 1 egg/d, while those in the low-egg-intake groups consumed on average 2 eggs/wk (Table 1). Table 1 shows that the subjects in all groups did not significantly differ in most parameters, except for higher BMI and fasting serum insulin in the T2D cases. Compared to the subjects who ate less eggs, those who ate more eggs had lower intakes of carbohydrates and fiber, and higher intakes of total energy, total fat, saturated fatty acids (SFA), total dietary and egg-derived cholesterol, choline, and PC (Supporting Information Table S2).

3.2. Metabolites reflecting egg intake

Using linear regression analysis, we found that 107 signals differed in abundances ($P<0.05$) between the high and low-egg-intake groups. When the FDR correction was applied to the whole data matrix consisting of 5376 entries for each of the 239 subjects, none of the differences remained statistically significant. Removal of redundant ions resulted in 54 differential signals, which mostly could not be identified based on spectral matching in the databases (Supporting Information Table S3). Fifteen of them with the level of identification I and II are shown in Figure 1. Egg intake was positively associated with
indolelactic acid, lysoPC(21:2), PC(18:0/22:2), and inversely associated with long-chain acylcarnitines with unsaturated acyl chains (acylcarnitine 16:1, 18:1, and 18:2), lysophosphatidylethanolamines (lysoPE) (18:0 and 18:2), monoglyceride 18:1, lysoPC(16:1), PC(16:0/20:5), inosine, and two γ-glutamylated-branched-chain amino acids (BCAA), namely γ-glutamyl-valine and γ-glutamyl-isoleucine/leucine. There were no statistical differences between the egg intake groups in fasting serum choline (slope 0.11; 95% CI -0.14, 0.37; \( P = 0.388 \), Supporting Information Figure S2A) or its microbial-derived liver metabolite, trimethylamine N-oxide (TMAO, slope 0.08; 95% CI -0.18, 0.33; \( P = 0.545 \), Supporting Information Figure S2B).

### 3.3. Potential predictors of T2D risk

The T2D cases and controls showed different baseline metabolite profiles, as apparent in different abundances of 388 signals based on the logistic regression analysis. Removal of redundant ions resulted in 155 differential signals (Supporting Information Table S4). Two of them remained significantly differential after FDR: tyrosine (OR 1.94; 95% CI 1.45, 2.60; \( P<0.001 \); FDR 0.023) and an unknown hexose-containing compound with mass 368.113 Da and retention time 5.15 mins in HILIC negative mode (OR 2.13; 95% CI 1.57, 2.88; \( P<0.001 \); FDR 0.005). Of those 155 differential signals, 39 metabolites which could be identified with identification level I or II were sorted based on the OR and plotted in Figure 2. The metabolic phenotypes of the cases featured higher abundances of five PC species with medium-chain or SFA (such as PC(14:0/16:0), (14:0/16:1), (14:0/18:2), (16:0/22:2), (16:1/18:1)), and other metabolites such as creatinine, uridine, piperine, caffeine, and its derivative, 1,3-dimethylurate. Serum tyrosine was positively associated with BMI and fasting insulin, along with uridine, creatinine, and some unknown metabolites (Supporting Information Figure S3).
Strikingly, all identified lysoPC species were higher in controls than in cases. The control-specific metabolic phenotypes included choline, phosphocholine, glycerophosphocholine, two plasmalogens (lysoPC(O-18:0) and lysoPC(O-18:1)), and (lyso)PCs that contain OCFAs (lysoPC(17:0), (17:1), (19:1), PC(16:0/23:5), (18:0/23:5)) as well as long-chain polyunsaturated fatty acids (LC-PUFA, such as lysoPC(20:4), (20:5), (22:5), (22:6), PC(16:0/24:4)). Free fatty acids, notably PUFA (such as linoleic acid and DHA) and hydroxylated fatty acids (such as 20-OH-DHA, 16-OH-hexadecanoic acid, and lactic acid) were also more abundant in the controls than in the cases. Noteworthy, these metabolites were also negatively correlated with BMI and fasting insulin (Supporting Information Figure S4).

After further adjustment for dairy and fiber intakes that may have an impact on circulating OCFAs levels,\textsuperscript{[32,33]} fiber intake did not seem to be a confounder, whilst the consumption of dairy products was a significant predictor only for (lyso)PCs that contain 23:5 fatty-acid chain (Supporting Information Table S5), although the total dairy consumption did not differ across the study groups (Supporting Information Table S2). Furthermore, we did not see any differences between the T2D cases and controls, or between the subjects with high and low egg intake in their serum acylcarnitines, BCAA and aromatic amino acids (Supporting Information Table S6).

### 3.4. Correlation analyses between potential biomarkers of egg intake and of T2D incidence

The positive predictors of T2D risk, such as uridine, creatinine, PE (25:2/14:2), unknown PE, unknown peptide, and other unknown lipophilic metabolites had a negative correlation with lysoPC(16:0) and (21:2), unknown (lyso)PC, that were found to be more predominant in the high-egg-intake groups (Figure 3). Strikingly, these same positive predictors of T2D risk...
had positive correlation with serum metabolites of subjects who had lower consumption of eggs, such as lysoPE (18:2), inosine, γ-glutamylated-BCAA, monoglyceride 16:1, and acylcarnitines (16:1, 18:1, 18:2) (Figure 4). Serum metabolites indicative for controls’ phenotypes tended to have a more positive correlation with ones in the high-egg-intake group (Supporting Information Figure S5). Conversely, the controls-associated metabolites had both positive and negative correlations with the predominant metabolites in the lower-egg-intake group (Supporting Information Figure S6). The potentially involved metabolic pathways constructed based on these findings is depicted in the Figure 5.

4. Discussion

In this study, we observed differences in the metabolic profiles indicative for both the egg intake, as well as the T2D incidence, with some indication of metabolic overlap that could provide some explanation for the previously observed inverse association between egg intake and risk of T2D in this study population.\textsuperscript{[11]} The correlation analyses suggested that certain cases-predominant metabolites had a negative correlation with metabolites indicative for higher egg intake. The same cases-predominant metabolites also showed a positive correlation with metabolites indicative for lower egg intake. This consistency may evoke further studies to identify these metabolites, since many remained unknown in this study, and to examine if they could be involved in the biological mechanism to explain the inverse association between egg intake and T2D risk. Similarly, the controls-predominant metabolites tended to correlate positively with those of the higher-egg-intake group. The controls-predominant metabolites, however, had both positive and negative correlations with the metabolic profiles of the subjects with lower egg intake, suggesting a role of factors other than low intake of eggs in shaping the metabolic profile of the lower-egg-intake group.
4.1. Metabolic consequences of egg consumption

Regardless of the significantly different intakes of choline between the egg consumption groups, we confirmed the previous finding that dietary choline did not affect plasma choline levels. This most likely is due to the endogenous maintenance of choline homeostasis in the fasting circulation. Consistent with the choline levels, we did not find a difference in the levels of TMAO, either. This could be expected, because increased TMAO levels were observed in experimental studies only with higher egg intake (≥2 eggs/d) and there is significant interindividual variation in responses to increased egg intake.

The potential biomarkers, especially ones indicative for egg intake, mostly consisted of unidentified metabolites, suggesting that egg intake may generate unique compounds that are currently unamenable using regular LC-MS/MS identification route. Moreover, the different phenotypes between subjects with higher and lower egg consumption involved endogenous metabolites that may represent the influence of egg intake on metabolism rather than the original compounds in eggs per se. The identification of potential biomarkers of egg intake hence warrants further exploration.

4.2. Proposed mechanisms linking egg intake and T2D incidence

The distinct profiles between the T2D cases and controls, despite the long mean follow-up of 19.3 y was interesting. We confirmed the previous findings from the Finnish Diabetes Prevention Study, which showed that tyrosine reflected higher risk of T2D whereas PUFA- and OCFA-containing (lyso)PCs indicated the opposite. Other aromatic amino acids or BCAA, however, did not differ between T2D cases and controls, though we observed higher levels of γ-glutamylated-BCAAs in the lower-egg-intake groups which positively correlated with the metabolic profiles of the T2D cases. This observation might suggest an impaired glutathione-dependent transport of BCAAs that have been previously reported in.
diabetic subjects.\textsuperscript{[39]} Furthermore, the positive correlation between T2D-cases-specific metabolites and medium-chain acylcarnitines in the low-egg-intake groups might imply impaired lipid transport and mitochondrial β-oxidation.\textsuperscript{[40]} Specifically, acylcarnitine 18:1 has been previously linked with impaired glucose tolerance.\textsuperscript{[41]} Therefore, a speculation that higher egg intake might suppress acylcarnitines and γ-glutamylated-BCAAs as early indicators T2D risk warrants further studies.

Another interesting observation is the positive correlation between (lyso)PCs, choline, and PUFA-containing lipid pools in controls with serum metabolites in the higher-egg-intake group and between PCs in T2D cases with serum metabolites in the lower-egg-intake groups. This finding may highlight the possible involvement of eggs in maintaining PC homeostasis. For example, the activity of lecithin:cholesterol acyltransferase (LCAT) which transfers a fatty acid from PC to yield a lysoPC \textsuperscript{[42]} has been shown to increase after a high egg intake (1-3 eggs/d).\textsuperscript{[43]} PC-LC-PUFAs, especially PC-DHA are mainly derived by the activity of phosphatidylethanolamine N-methyltransferase (PEMT),\textsuperscript{[44]} whereas PCs that contain saturated or medium-chain fatty acids are mostly produced by CDP-choline pathway.\textsuperscript{[45]} Production of PC-DHA via PEMT was reduced after a low-choline diet.\textsuperscript{[44]} This observation hence may create possibility for a preference of PEMT-based PC production and lower choline utilization in the CDP-choline pathway, thereby explaining the observed higher levels of choline in controls. Further investigations are necessary to confirm these pathways.

The correlations between lysoPC(21:2) in higher-egg-intake groups positively with some OCFA-containing lipid pools in controls and negatively with T2D-cases-specific metabolites may highlight the importance of the OCFA-containing lipids metabolism as a potential mediator linking egg intake with the lower T2D risk in this study population.\textsuperscript{[11]} The remained associations in the current study after adjustment analyses for fiber and dairy intakes confirmed the previous finding in this cohort\textsuperscript{[11]} that the adjustment for fiber and
dairy intakes did not attenuate the inverse association between egg intake and T2D risk. Several studies have previously shown how the profiles of various lipids, including OCFAs, have been associated with the composition of gut microbiota\cite{46} regardless of age, sex, BMI, or genetic factors.\cite{47} These led us to hypothesize that the higher abundances of OCFAs as the potential biomarkers of lower T2D risk might imply the potential role of gut microbiota which deserves further exploration.

4.3. **Strengths and limitations**

The strengths of the KIHD include the detailed information on egg intake that also incorporated eggs used in mixed dishes and recipes. KIHD also accommodated thorough investigations of other factors besides egg intake, allowing adjustment for potential confounders. The other potential confounders such as genetic and environmental factors were controlled by selecting Finnish middle-aged and older men as the study population. Correspondingly, this selection also limits the generalizability of the study outcomes to women or other populations.

Secondly, the application of non-targeted metabolic profiling platform in this cohort creates possibility of a novel insight of plausible mechanisms underlying the previously observed association. The causal relationship, cannot be inferred from this study due to its cross-sectional nature. Hence, we also could not prevent the inclusion of other metabolites that might not be completely relevant with the analyses. For example, higher level of caffeine in the T2D cases might be due to higher consumption of coffee, especially in the T2D cases with higher egg intake (Supporting Information Table S2), or other reasons that may explain the high variation of caffeine metabolisms\cite{48} and does not necessarily link to BMI or fasting insulin as the traditional biomarker of T2D (Supporting Information Figure S3).
Moreover, the nominally significant findings call for a cautious interpretation. One potential reason for the nominally significant associations could be the exploratory property of the non-targeted metabolomics approach employed in this study. Second, although we excluded men with T2D diagnosis at baseline, this diagnosis was based only on self-reported T2D diagnosis or on fasting plasma glucose level at the study visit, and not on post-glucose challenge values. Using these criteria, we might have failed to exclude all undetected T2D cases at baseline. However, even during the first two years of follow-up in the whole cohort, only 3 men were diagnosed with T2D,[11] suggesting that this is not a major concern in our study. Finally, the estimation of egg consumption by using a single 4-d food recording may not accurately represent typical egg intakes. This could potentially introduce random misclassification, which would then attenuate the true associations.

Technically, the storage of samples for about 30 years may raise a question about the stability of the samples, as previously investigated.[49] Since we stored all samples with the same storage condition and always compared the arbitrary values of metabolites between one group and another, the deviation of metabolites content in the samples, if any, will not change the final findings.

4.4. Concluding remarks

The non-targeted metabolic profiling approach we employed in this study revealed differences in the serum profile of the study participants who remained healthy and of those who developed T2D during the mean 19.3-year follow-up and who had either higher or lower egg intakes. Certain cases-predominant metabolites which correlated negatively with metabolic profile in the higher-egg-intake group and positively with profile in the lower-egg-intake group tend to favor the inverse association between egg intake and T2D. Based on those metabolites, we present a novel insight of plausible mechanisms that may contribute to
protective role of egg consumption against T2D risk. Further trials that include comprehensive analyses about the relevant metabolic pathways and the possible involvement of gut microbiota are essential to verify the causal mechanism underlying the possible health benefits of eggs, especially in decreasing the risk of developing T2D.

References


Acknowledgments

SN received financial support from Indonesian Endowment Fund for Education (LPDP), Erasmus+ Mobility, Juho Vainio Foundation, Saara Kuusisto and Salme Pennanen Grant for Cardiovascular Research, Kuopio University Hospital (grant number VTR 510RA17), and North Savo fund of Finnish Cultural Foundation (grant number 65171618).

OK received financial support from the Finnish Foundation for Alcohol Studies. KH was funded by a grant from the Academy of Finland (grant numbers 277986, 283454, 305396).

We also want to thank Biocenter Finland for supporting our core LC-MS laboratory facility. The funders had no contribution in study design, data collection, data analysis, manuscript preparation, or decision to publish.

Authors contributions

JKV, TPT, SV, TN, KH designed research; ML performed laboratory analysis for metabolomics; SN, JKV, KH conducted research; SN, AM, JP, JKV performed statistical analyses; OK, KH verified the identification of metabolites; SN drafted the manuscript; OK, AM, JP, ML, TN, TPT, SV, KH, JKV critically revised the manuscript for important intellectual content; JKV and KH had primary responsibility for the final content. The contribution of laboratory technician Miia Reponen is acknowledged. All authors read and approved the final manuscript. All authors declared no conflict of interest.
**Figure captions**

**Figure 1.** Serum metabolites associated ($P < 0.05$) with egg intake, estimated with linear regression. *indicates metabolite with fold change more than 20%. Fold change was calculated as ratio of mean peak area of metabolite in higher and lower-egg-intake groups. Error bars indicate 95% confidence interval; Slope indicates the estimates of linear regression between egg intake (higher/lower) as the independent variable and the peak area of a metabolite as the dependent variable. The values of the slope, fold change, 95% confidence interval, and $P$ are listed in the Supporting Information Table S3. (lyso)PC: (lyso)phosphatidylcholine; lysoPE: lysophosphatidylethanolamine.
Figure 2. Serum metabolites associated (P < 0.05) with the incidence of type 2 diabetes (T2D) after the mean follow up of 19.3 years, estimated with logistic regression. *indicates metabolite with fold change more than 20%. Fold change was calculated as the ratio of the mean peak area of each metabolite in T2D cases and in controls. Odds ratio indicates the estimates of logistic regression between the peak area of a serum metabolite as the independent variable and the incidence of T2D (yes/no) as the dependent variables; values <1 were associated with lower risk of T2D whereas values >1 were associated with higher T2D risk. Error bars indicate 95% confidence interval. The values of the slope, fold change, 95% confidence interval, and P are listed in Supporting Information Table S4. (lyso)PC: (lyso)phosphatidylcholine; DHA: docosahexaenoic acid (C22:6).
Figure 3. Association between potential biomarkers of egg intake and potential predictors of incident type 2 diabetes (T2D), estimated with Pearson’s correlation. Red color indicates positive correlation and blue color indicates the negative one. PE: phosphatidylethanolamine; (lyso)PC: (lyso)phosphatidylcholine.
**Figure 4.** Association between serum metabolites that were more abundant in lower-egg-intake group and potential predictors of incident type 2 diabetes (T2D), estimated with Pearson’s correlation. Red color indicates positive correlation and blue color indicates the negative one. lysoPE: lysophosphatidylethanolamine. (lyso)PC: (lyso)phosphatidylcholine; MG: monoglyceride. Glu: glutamate. Val: valine. Ile/Leu: isoleucine/leucine.
Figure 5. Proposed biological mechanisms that may partially explain the inverse association between egg intake and risk of type 2 diabetes (T2D) in the Kuopio Ischaemic Heart Disease Risk Factor Study Study population. γ-Glu-BCAA: gamma-glutamylated branch-chain amino acids; MCFA: medium-chain fatty acids; LC-PUFA: long-chain polyunsaturated fatty acids; OCFA: odd-chain fatty acids; SFA: saturated fatty acids; lyso(PC): (lyso)phosphatidylcholine; PEMT: phosphatidylethanolamine N-methyltransferase; LCAT: lecithin:cholesterol acyltransferase; PO4-choline: phosphocholine. The solid lines represent established relationship that have been previously known, dashed lines indicate observations in this current study, and dotted lines indicate proposed mechanisms that are suggested by the observations.
Table 1. Baseline characteristics of the subset of 239 participants from the Kuopio Ischaemic Heart Disease Risk Factor Study, divided into 4 groups based on their egg intake and incidence of type 2 diabetes

<table>
<thead>
<tr>
<th>Study groups</th>
<th>Controls + Higher Intake (n=61)</th>
<th>Controls + Lower Intake (n=60)</th>
<th>Cases + Higher Intake (n=60)</th>
<th>Cases + Lower Intake (n=58)</th>
<th>P-difference Higher vs. lower-egg-intake groups</th>
<th>Case vs. control groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Egg intake*, g d⁻¹</td>
<td>54 ± 6</td>
<td>15 ± 3</td>
<td>55 ± 18</td>
<td>15 ± 6</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age, y</td>
<td>51.8 ± 4.9</td>
<td>52.0 ± 5.9</td>
<td>53.9 ± 4.4</td>
<td>51.8 ± 4.6</td>
<td>0.068</td>
<td></td>
</tr>
<tr>
<td>Body mass index, kg m⁻²</td>
<td>26.1 ± 2.0</td>
<td>25.6 ± 2.2</td>
<td>26.8 ± 2.1</td>
<td>26.7 ± 2.0</td>
<td>0.008</td>
<td>0.243</td>
</tr>
<tr>
<td>Waist-to-hip ratio</td>
<td>0.94 ± 0.04</td>
<td>0.94 ± 0.07</td>
<td>0.97 ± 0.13</td>
<td>0.94 ± 0.04</td>
<td>0.259</td>
<td></td>
</tr>
<tr>
<td>Leisure-time physical activity, kcal d⁻¹</td>
<td>134 ± 163</td>
<td>145 ± 157</td>
<td>122 ± 117</td>
<td>112 ± 124</td>
<td>0.611</td>
<td></td>
</tr>
<tr>
<td>Current smoker, %</td>
<td>0.00</td>
<td>3.33</td>
<td>0.00</td>
<td>1.72</td>
<td>0.292</td>
<td>0.078</td>
</tr>
<tr>
<td>Alcohol consumption*, g week⁻¹</td>
<td>26.4 ± 27.3</td>
<td>18.8 ± 21.0</td>
<td>23.8 ± 28.4</td>
<td>23.1 ± 24.5</td>
<td>0.365</td>
<td></td>
</tr>
<tr>
<td>Fasting serum insulin*, mU L⁻¹</td>
<td>10.8 ± 4.6</td>
<td>9.3 ± 3.4</td>
<td>11.7 ± 4.6</td>
<td>11.2 ± 3.7</td>
<td>0.005</td>
<td>0.056</td>
</tr>
<tr>
<td>Blood glucose, mmol L⁻¹</td>
<td>4.6 ± 0.4</td>
<td>4.4 ± 0.4</td>
<td>4.8 ± 0.5</td>
<td>4.7 ± 0.5</td>
<td>0.068</td>
<td></td>
</tr>
<tr>
<td>Serum total cholesterol, mmol L⁻¹</td>
<td>5.56 ± 0.79</td>
<td>5.96 ± 1.11</td>
<td>5.90 ± 1.10</td>
<td>5.61 ± 1.11</td>
<td>0.084</td>
<td></td>
</tr>
<tr>
<td>Serum VLDL cholesterol*, mmol L⁻¹</td>
<td>0.51 ± 0.40</td>
<td>0.59 ± 0.51</td>
<td>0.55 ± 0.37</td>
<td>0.46 ± 0.23</td>
<td>0.233</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Group 1</td>
<td>Group 2</td>
<td>Group 3</td>
<td>Group 4</td>
<td>p-value</td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>Serum LDL cholesterol, mmol L⁻¹</td>
<td>3.74 ± 0.76</td>
<td>4.08 ± 1.09</td>
<td>4.04 ± 1.11</td>
<td>3.81 ± 1.05</td>
<td>0.158</td>
<td></td>
</tr>
<tr>
<td>Serum HDL cholesterol, mmol L⁻¹</td>
<td>1.32 ± 0.28</td>
<td>1.29 ± 0.28</td>
<td>1.31 ± 0.27</td>
<td>1.34 ± 0.28</td>
<td>0.799</td>
<td></td>
</tr>
<tr>
<td>Serum triglycerides, mmol L⁻¹</td>
<td>1.09 ± 0.62</td>
<td>1.31 ± 1.04</td>
<td>1.16 ± 0.58</td>
<td>1.17 ± 0.56</td>
<td>0.383</td>
<td></td>
</tr>
<tr>
<td>Serum C-reactive protein, mg L⁻¹</td>
<td>1.85 ± 5.06</td>
<td>1.41 ± 2.23</td>
<td>1.95 ± 2.48</td>
<td>1.80 ± 3.80</td>
<td>0.855</td>
<td></td>
</tr>
</tbody>
</table>

a) Values are means ± SDs. Differences between the groups were tested with Chi-square for the categorical variable and ANOVA for continuous variables (one-way ANOVA for equal variances and Welch’s ANOVA for unequal variances, indicated with *).
Eating eggs was previously connected with a lower risk of developing type 2 diabetes in Finnish men. We analyzed blood samples taken from 239 men to check if there are any compounds that could explain the lower diabetes risk. The blood of men who ate more eggs was less likely to contain compounds that were found in the men who developed type 2 diabetes at the end of the study. These findings warrant further research to explain how egg intake could lower the risk of type 2 diabetes.