

Optimized metabotype definition based on a limited number of standard clinical parameters in the population-based KORA study

Chetana Dahal, Nina Wawro, Christa Meisinger, Taylor A. Breuninger, Barbara Thorand, Wolfgang Rathmann, Wolfgang Koenig, Hans Hauner, Annette Peters, Jakob Linseisen

Angaben zur Veröffentlichung / Publication details:



Dahal, Chetana, Nina Wawro, Christa Meisinger, Taylor A. Breuninger, Barbara Thorand, Wolfgang Rathmann, Wolfgang Koenig, Hans Hauner, Annette Peters, and Jakob Linseisen. 2022. "Optimized metabotype definition based on a limited number of standard clinical parameters in the population-based KORA study." *Life* 12 (10): 1460.
<https://doi.org/10.3390/life12101460>.

Nutzungsbedingungen / Terms of use:

CC BY 4.0

Article

Optimized Metabotype Definition Based on a Limited Number of Standard Clinical Parameters in the Population-Based KORA Study

Chetana Dahal ^{1,2} , Nina Wawro ^{1,2} , Christa Meisinger ^{1,2}, Taylor A. Breuninger ^{1,2} , Barbara Thorand ^{3,4} , Wolfgang Rathmann ^{4,5}, Wolfgang Koenig ^{6,7,8}, Hans Hauner ^{9,10} , Annette Peters ^{3,4,†}  and Jakob Linseisen ^{1,2,11,*} 

- ¹ Independent Research Group Clinical Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health (GmbH), Ingolstädter Landstr. 1, 85764 Neuherberg, Germany
- ² Epidemiology, Faculty of Medicine, University Hospital Augsburg, University of Augsburg, Stenglinstraße 2, 86156 Augsburg, Germany
- ³ Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health (GmbH), Ingolstädter Landstr. 1, 85764 Neuherberg, Germany
- ⁴ German Center for Diabetes Research, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany
- ⁵ Institute for Biometrics and Epidemiology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Auf'm Hennekamp 65, 40225 Düsseldorf, Germany
- ⁶ German Centre for Cardiovascular Research, Partner Site Munich Heart Alliance, Pettenkoferstr. 8a & 9, 80336 Munich, Germany
- ⁷ Deutsches Herzzentrum München, Technische Universität München, Lazarettstr. 36, 80636 Munich, Germany
- ⁸ Institute of Epidemiology and Medical Biometry, University of Ulm, Helmholtzstr. 22, 89081 Ulm, Germany
- ⁹ Else Kröner-Fresenius-Center for Nutritional Medicine, TUM School of Life Sciences, Technical University of Munich, 85354 Freising, Germany
- ¹⁰ Institute of Nutritional Medicine, School of Medicine, Technical University of Munich, Georg-Brauchle-Ring 62, 80992 Munich, Germany
- ¹¹ Institute for Medical Information Processing, Biometry, and Epidemiology, Ludwig-Maximilians-Universität München, Marchioninistr. 15, 81377 Munich, Germany
- * Correspondence: jakob.linseisen@med.uni-augsburg.de; Tel.: +49-821-598-6471
- † These authors contributed equally to this work.



Citation: Dahal, C.; Wawro, N.; Meisinger, C.; Breuninger, T.A.; Thorand, B.; Rathmann, W.; Koenig, W.; Hauner, H.; Peters, A.; Linseisen, J. Optimized Metabotype Definition Based on a Limited Number of Standard Clinical Parameters in the Population-Based KORA Study. *Life* **2022**, *12*, 1460. <https://doi.org/10.3390/life12101460>

Academic Editor: Tao Huang

Received: 10 August 2022

Accepted: 16 September 2022

Published: 20 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The aim of metabotyping is to categorize individuals into metabolically similar groups. Earlier studies that explored metabotyping used numerous parameters, which made it less transferable to apply. Therefore, this study aimed to identify metabotypes based on a set of standard laboratory parameters that are regularly determined in clinical practice. K-means cluster analysis was used to group 3001 adults from the KORA F4 cohort into three clusters. We identified the clustering parameters through variable importance methods, without including any specific disease endpoint. Several unique combinations of selected parameters were used to create different metabotype models. Metabotype models were then described and evaluated, based on various metabolic parameters and on the incidence of cardiometabolic diseases. As a result, two optimal models were identified: a model composed of five parameters, which were fasting glucose, HDLc, non-HDLc, uric acid, and BMI (the metabolic disease model) for clustering; and a model that included four parameters, which were fasting glucose, HDLc, non-HDLc, and triglycerides (the cardiovascular disease model). These identified metabotypes are based on a few common parameters that are measured in everyday clinical practice. These metabotypes are cost-effective, and can be easily applied on a large scale in order to identify specific risk groups that can benefit most from measures to prevent cardiometabolic diseases, such as dietary recommendations and lifestyle interventions.

Keywords: metabotype; cluster analysis; parameter selection; clinical marker; metabolic diseases; cardiovascular diseases

1. Introduction

Metabotyping describes the process of forming subgroups based on similarities in subjects' metabolic or phenotypic characteristics. These subgroups are termed as metabotypes or metabolic phenotypes [1–4]. All individuals within a subgroup show a high metabolic similarity, while the different subgroups are all distinct from each other. This allows for the identification and description of specific subgroups according to their cardiometabolic disease risk [5–8]. Evidence suggests that dietary recommendations that are provided at personalized and metabotype levels tend to be more effective than providing general dietary advice [1,9–11]. Thus, metabotyping is a promising approach for the development of personalized preventive measures, such as dietary recommendations and lifestyle interventions [1,2,6,12].

Several studies have been performed to define metabotypes [2,3,13]. However, due to the use of different methods and inconsistent definitions, studies have shown large heterogeneities in the types and numbers of parameters used to identify metabotypes or metabolic phenotypes [3]. Some studies have even used a large number of metabolic variables from different metabolic pathways, leading to comprehensive metabotyping [14–18]. Similarly, in our previous study by Riedl and co-authors [15], we identified comprehensive metabotypes in the German population-based KORA study using a range of biochemical and anthropometric parameters.

However, many metabolic parameters are not routinely measured in primary care, making it difficult to implement a comprehensive metabotype concept in general research settings. Therefore, in order to identify a metabotype concept that can be broadly applicable in daily practice, a set of routinely measured clinical parameters, so-called “standard laboratory parameters”, should be explored. The importance of having an easily applicable metabotype definition for use on a large scale to identify subjects with a specific cardiometabolic disease risk, has also been highlighted in a recent perspective paper by Palmnäs et al. [12]. Currently, only a few studies have investigated the use of a reduced set of available parameters to identify metabotypes [5,8,19,20]. In this study, we aimed to develop a statistically guided selection of variables for metabotypes, without disregarding the availability and clinical relevance of parameters.

Therefore, this study aimed to optimize the metabotype definition by reducing the clinical parameters to a few that were economical and routinely measured. For this purpose, we used machine learning-based variable importance methods to assess the suitability of parameters for identifying different metabotypes. In order to evaluate the results, we described the metabotype clusters using various metabolic parameters, as well as the incidences of various cardiometabolic diseases.

2. Materials and Methods

2.1. Study Population

All data for this study were obtained from the population-based KORA F4 (2006–2008) and KORA FF4 (2013/2014) studies. The data set represents the first and second follow-up examinations of the KORA S4 study, conducted between 1999 and 2001 ($n = 4261$ participants aged 25–74 years) in the region of Augsburg in southern Germany [21]. In total, 3080 individuals took part in the KORA F4 study, and 2279 individuals participated in the 7-year follow-up examination (KORA FF4). Among these, 2161 individuals participated in both the KORA F4 and KORA FF4 studies. In both studies, participants were invited to the study center where bio-samples were collected, and trained study nurses performed standardized physical examinations as well as computer-assisted face-to-face interviews. Likewise, all participants answered self-administered questionnaires. An in-depth description of the primary study design [21] and of the KORA F4 [22,23] and KORA FF4 [24] studies was reported previously. Written informed consent was provided by all participants, and the studies were approved by the Ethics Committee of the Bavarian Medical Association, and were conducted in accordance with the Declaration of Helsinki.

2.2. Biochemical and Anthropometric Parameters

We identified the metabotypes based on fasting biochemical parameters, along with body mass index (BMI) data that were available from the KORA F4 study. BMI was used as a continuous measure in kg/m^2 . Parameters such as high-density lipoprotein cholesterol (HDLc), total cholesterol (TC), triglycerides (TG), glucose, insulin, uric acid, high-sensitive C-reactive protein (hs-CRP), gamma-glutamyltransferase (GGT), glutamate-pyruvate transaminase (GPT), glutamate-oxaloacetate transaminase (GOT), and alkaline phosphatase (AP), were measured in serum samples. Non-HDLc cholesterol was calculated by subtracting HDLc from TC. Leukocyte count and glycated hemoglobin (HbA1c) were measured from fresh venous whole EDTA blood samples. More technical details on the handling of blood samples and the derivation of biomarker measurements can be found elsewhere [15].

2.3. Socio-Demographic and Lifestyle Variables

Socio-demographic data included sex, age (in years), and education; the latter was categorized according to the German education system into < 10 years, 10–<12 years, and ≥ 12 years at school. Lifestyle data included physical activity (active: active for ≥ 2 h per week; inactive: active for <2 h per week), smoking status (smoker, ex-smoker, and never-smoker), and alcohol consumption (≥ 40 g/day, 20–<40 g/day, 0–<20 g/day, and 0 g/day). According to the WHO [25], BMI was categorized into underweight (BMI < $18.5 \text{ kg}/\text{m}^2$), normal weight (BMI 18.5 –< $25 \text{ kg}/\text{m}^2$), overweight (BMI 25 –< $30 \text{ kg}/\text{m}^2$), and obese (BMI $\geq 30 \text{ kg}/\text{m}^2$).

2.4. Health Status

In both the F4 and FF4 studies, cardiometabolic diseases were assessed during standardized face-to-face computer-assisted interviews and physical examinations. Metabolic diseases were defined as follows: hypertension by a blood pressure of $\geq 140/90$ mmHg in the resting state during physical examination or treatment with antihypertensive medication; type 2 diabetes mellitus was defined by self-reported diagnosis validated by the respective treating physician and by current intake of glucose-lowering medication. In addition, undiagnosed diabetes cases were identified through oral glucose tolerance test based on ADA criteria. Cases with a diagnosis of dyslipidemia or hyperuricemia/gout were defined by the current intake of lipid-lowering drugs or hyperuricemia/gout medication, respectively. Similarly to our previous studies [8,15], we analyzed all metabolic diseases individually, and as a combined outcome variable, “any metabolic diseases” (defined as suffering from at least one of the four metabolic diseases: hypertension, type 2 diabetes, dyslipidemia, and hyperuricemia/gout). Similarly, the diagnosis of myocardial infarction or stroke was defined on the basis of self-reports, and was further validated by means of medical records such as hospital or general practitioners’ records. Both myocardial infarction and stroke were analyzed individually and summarized as a dichotomous variable, “any cardiovascular diseases” (defined as suffering from at least one of two cardiovascular diseases: myocardial infarction and/or stroke).

Prevalent cases were cases identified in the KORA F4 study, and incident cases were defined as newly occurring cases after a follow-up of seven years in the KORA FF4 study (in those participants who were not yet diagnosed with the respective disease in the KORA F4 study).

2.5. Data Preprocessing

Among the 3080 participants in the F4 data, we excluded 61 participants, as 54 participants did not fast for at least 8 h before blood collection, and 7 participants had missing information regarding fasting glucose levels. Furthermore, we excluded 18 participants who had more than 10% missing data for the above-mentioned parameters; ultimately, there were 3001 study participants in total. Among these, 2120 participated in both the F4 study and the seven-year follow-up FF4 study (Figure 1). We imputed the remaining missing

variables of the biomarkers using the multivariate imputation by chained equations ‘mice’ package version 3.8.0 in R [26], which generated five complete data sets with ten iterations each. Subsequently, to avoid the biases of different scales and units, we z-standardized all biochemical and anthropometric parameters before using these imputed and standardized data for clustering purposes only.

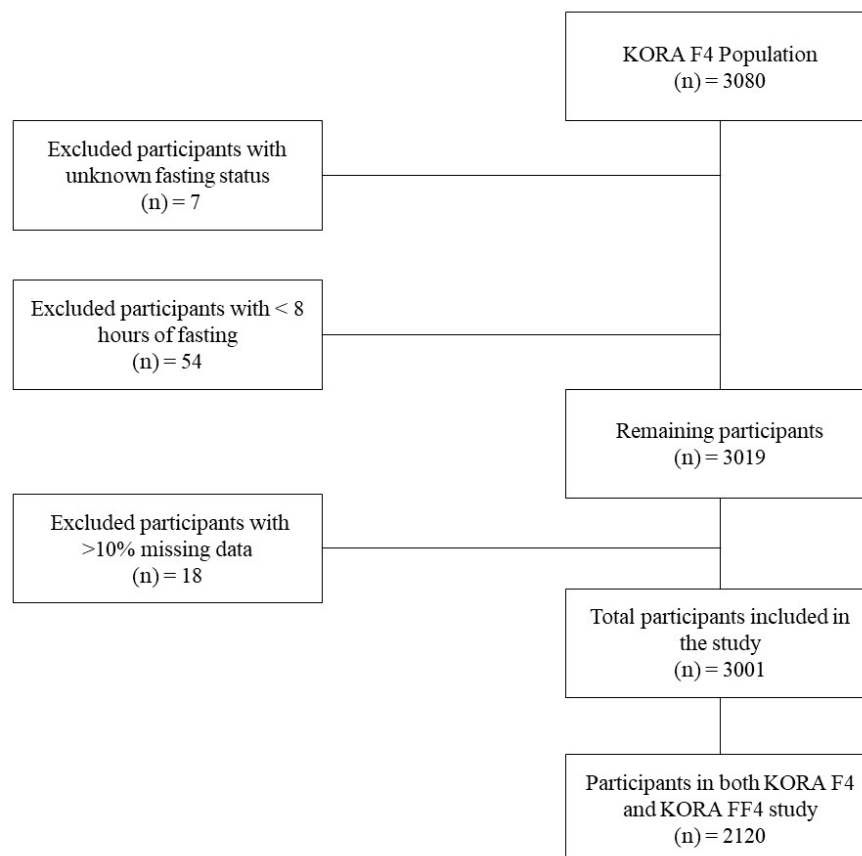


Figure 1. Study flow diagram.

2.6. Descriptive Statistics

We reported the baseline characteristics, including socio-demographic, lifestyle, and health status, of the KORA F4 study population in total and stratified these by sex. Median and interquartile range (IQR) were shown for continuous variables, and absolute frequency and percentage for categorical variables. In order to analyze the differences in the distributions between sexes as well as metabolotypes, we used the Kruskal–Wallis test for continuous variables, and Pearson’s chi-square test for categorical variables. Additionally, we also carried out respective post hoc tests with the Bonferroni correction to examine differences in metabolic parameters between metabolotype clusters. As there was missing information for some participants, the maximum number of data available was used, leading to different sample sizes; the exact numbers are provided in the footnotes of the tables.

2.7. Parameter Selection

In our previously published studies, we developed the metabolotyping concept using a comprehensive set of 32 [15] and 16 [27] widely available clinical parameters. However, in this study, we optimized the metabolotype concept by reducing the metabolotyping parameters to a few relevant standard clinical parameters. As an initial step, we reduced the 16 parameters to 14 parameters by replacing three-biochemical parameters, TC, low-density lipoprotein (LDL) cholesterol, and the TC/HDL ratio, with non-HDLc cholesterol, as recent findings showed non-HDLc to have high prognostic value [28–31].

In order to investigate the contribution of individual variables to metabotyping, variable importance analysis was performed using a machine learning-based method, beginning with the 14 parameters. We applied a commonly used feature selection method for biomarker discovery [32–34] called permutation variable importance (PVI) [35]. It was implemented using the R function “PIMP” (algorithm for the permutation variable importance measure) [36]. In this method, variable importance is calculated with the help of the variable importance measure of the random forest (RF) algorithm [37]. Initially, an RF model is trained on the original data set. Then, variable importance of each variable is calculated using a decrease in Gini impurity, which is the likelihood of falsely identifying the occurrence of a random variable (for details see [37]). Then, the outcome variable is randomly permuted a fixed number of times. For each permutation of the outcome variable, the variable importance for each predictor variable is calculated, which is referred to as “null importance”. Then, a probability distribution (selected using Kolmogorov–Smirnov tests) is fitted to the null importances. The fitted distribution is then used to derive p -values of true importance from the null importances. In this study, we obtained the variable importance by creating 500 trees with all 14 parameters included, and derived the p -value of the predictors by randomly permuting the outcome variables 100 times, as described in detail by Altman et al. [35].

As a sensitivity analysis and to validate the results of the PVI method, we applied two other methods to examine variable importance. First, we used the cross-validated permutation variable importance measure (CVPVI), which is an average of all k -fold cross-validated permutation importances [38]. In this method, data sets are divided into k equal folds; for each fold, an RF model is trained. The prediction error from each tree in the RF model is calculated. The same is repeated after permuting each predictor variable. The difference between the two prediction errors is calculated and averaged over all trees. Finally, the differences from all k folds are averaged, and the final relevance of predictor variables is assessed. We implemented this function in R using the “CVPVI” function from the “Vita” package [39]. In our implementation, we carried out 10-fold cross-validation, and created 1000 trees in each fold. The predictor variables were permuted 100 times. For the second method, we performed gradient-boosted feature selection [40]. Gradient boosting is a boosted tree-based supervised learning algorithm [41]. In this method, variable importance is calculated using the fractional contribution each variable provides to the model, based on the total gain of the variable’s splits. These importance scores are then averaged across all decision trees within the model [41,42]. For this task, we used the R package “xgboost” [43].

Based on the top 50% of contributing variables in all methods, as well as their availability in general primary care, we selected a subset of 7 out of the 14 variables.

2.8. Metabotyping

Similarly to previous studies [5,15,19,27], we identified metabolically homogenous subgroups (metabotypes) by performing a clustering method called k -means clustering algorithm. Likewise, in order to identify the appropriate number of clusters for the k -means algorithm, we used a statistical function in R called “NbClust” which provides 30 indices for determining the optimal number of clusters [44]. We used the clustering algorithm in all five imputed data sets via the R package “mclust” (multiple imputation in cluster analysis), version 1.2.5 [45].

We used the seven identified parameters (namely TG, BMI, uric acid, fasting glucose, insulin, HDLc, and Non-HDLc) to derive metabotypes for KORA F4 participants. We created several random unique combinations of parameters and computed cluster analyses in each combination, which resulted in different metabotype models. The models with at least 5% or 150 participants in the smallest cluster were regarded as acceptable metabotype models, and were included in this study. In the majority of the models, a three-cluster solution was the best option, followed by two clusters; all other options were ranked low. Similar results were obtained in our previous studies, where three clusters were identified

as an appropriate number of clusters [8,15,46]. Therefore, to make the models consistent and comparable, we derived a three-cluster solution in all models.

In each model, we termed cluster 1 as the cluster with the metabolically most favorable clustering biochemical parameters (“healthy metabotype”), in contrast with cluster 3 which was characterized by the metabolically least favorable clustering parameters (“unfavorable metabotype”). Cluster 2 was termed as the intermediate cluster, where the clustering biochemical parameters were in between those of clusters 1 and 3 (“intermediate metabotype”).

The incidence of cardiometabolic disease in KORA FF4 data was used to identify the most appropriate metabotype models that were identified in KORA F4 participants. However, the diseases were not included in the metabotyping process. We ranked the models based on the highest incidence of all metabolic diseases and cardiovascular diseases in cluster 3. The model with the highest rank for “any metabolic disease” was regarded as the best model for metabolic disease, and the model with the highest rank for “any cardiovascular disease” was regarded as the best model for cardiovascular disease. These models were further evaluated on the basis of various metabolic parameters.

We performed all statistical analyses for this study using the statistical software R version 3.6.2 (R Development Core Team, 2010, <http://www.r-project.org>, (accessed on 17 February 2020) and RStudio Version 1.1.423, which is an integrated development environment (IDE) for R. All tests were two-tailed, and we considered a *p*-value < 0.05 to be statistically significant.

3. Results

Table 1 describes the baseline characteristics of the study population, including both demographic parameters and data on the prevalence of diseases identified in the KORA F4 study, in total and stratified by gender. Among the total study population, 52% were female and 48% were male, with a median age of 56 years (IQR = 22 years). The median BMI was 27 kg/m² (IQR = 5.9 kg/m²), and almost 55% of the total study population was physically active. The prevalence of “any metabolic disease” and “any cardiovascular disease” was 43.6% and 4.7%, respectively. Compared to men, women had a lower median age and BMI, were more often never-smokers, consumed less alcohol, and showed a lower prevalence of diseases. We observed similar differences between groups in the follow-up study population (KORA FF4) as well (Table S1).

Table 1. Baseline characteristics of the KORA F4 study population.

	Total n = 3001	Men n = 1450	Women n = 1551	<i>p</i> -Value
Socio-demographic characteristics				
Age (years)				
Median (IQR)	56.0 (22.0)	57.0 (23.0)	55.0 (22.0)	0.03
Education				
<10 years	261 (8.7%)	58 (4.0%)	203 (13.1%)	<0.001
10–<12 years	1499 (50.0%)	661 (45.6%)	838 (54.0%)	
≥12 years	1236 (41.2%)	728 (50.2%)	508 (32.8%)	
Missing	5 (0.2%)	3 (0.2%)	2 (0.1%)	
BMI (kg/m ²)				
Median (IQR)	27.0 (5.9)	27.3 (5.1)	26.3 (7.1)	<0.001
Normal weight (18.5–<25)	941 (31.4%)	345(23.8%)	596 (38.4%)	
Overweight (25–<30)	1253 (41.8%)	726(50.1%)	527 (34.0%)	
Obese (≥30)	793 (26.4%)	375(25.9%)	418 (27.0%)	
Missing	14 (0.5%)	4 (0.3%)	10 (0.6%)	

Table 1. Cont.

	Total n = 3001	Men n = 1450	Women n = 1551	p-Value
Physical Activity				0.409
Active	1641 (54.7%)	780 (53.8%)	861 (55.5%)	
Inactive	1356 (45.2%)	666 (45.9%)	690 (44.5%)	
Missing	4 (0.1%)	4 (0.3%)	0 (0.0%)	
Smoking				<0.001
Smoker	524 (17.5%)	281 (19.4%)	243 (15.7%)	
Ex-Smoker	1218 (40.6%)	715 (49.3%)	503 (32.4%)	
Never-Smoker	1254 (41.8%)	450 (31.0%)	804 (51.8%)	
Missing	5 (0.2%)	4 (0.3%)	1 (0.1%)	
Alcohol consumption				<0.001
≥40 g/day	336 (11.2%)	290 (20.0%)	46 (3.0%)	
20–<40 g/day	540 (18.0%)	351 (24.2%)	189 (12.2%)	
0–<20 g/day	1221 (40.7%)	508 (35.0%)	713 (46.0%)	
0 g/day	900 (30.0%)	297 (20.5%)	603 (38.9%)	
Missing	4 (0.1%)	4 (0.3%)	0 (0.0%)	
Prevalence of disease n (%)				
Type 2 diabetes mellitus	242 (8.1%)	144 (9.9%)	98 (6.3%)	<0.001
Hypertension	1150 (38.3%)	639 (44.1%)	511 (32.9%)	<0.001
Missing	7 (0.2%)	4 (0.3%)	3 (0.2%)	
Hyperuricemia	113 (3.8%)	90 (6.2%)	23 (1.5%)	<0.001
Missing	2 (0.1%)	2 (0.1%)	0 (0.0%)	
Dyslipidemia	386 (12.9%)	219 (15.1%)	167 (10.8%)	<0.001
Missing	3 (0.1%)	3 (0.2%)	0 (0.0%)	
Any metabolic disease	1309 (43.6%)	730 (50.3%)	579 (37.3%)	<0.001
Missing	8 (0.3%)	4 (0.3%)	3 (0.2%)	
Stroke	71 (2.4%)	46 (3.2%)	25 (1.6%)	0.007
Myocardial infarction	79 (2.6%)	62 (4.3%)	17 (1.1%)	<0.001
Any cardiovascular disease	142 (4.7%)	100 (6.9%)	42 (2.7%)	<0.001

Median (IQR) for continuous variables and n (column %) for categorical variables. *p*-values are from the Kruskal–Wallis test for continuous variables, and from Pearson’s chi-squared test for categorical variables. Prevalence: due to missing information, there were reduced data sets for hypertension *n* = 2994, hyperuricemia *n* = 2999, dyslipidemia *n* = 2998, and “any metabolic disease” *n* = 2993. KORA, Cooperative Health Research in the Region of Augsburg.

Figure 2 presents the variable importance of the grouping parameters included in the 14-parameter model. All three methods showed similar results. The seven most important parameters out of fourteen were TG, uric acid, BMI, HDLc, glucose, insulin, and non-HDLc. According to the PVI method, all seven selected parameters had a significant effect. TG was identified as the most important variable in all methods, whereas uric acid was identified as the second most important parameter in two (PVI and CVPVI) methods. Next, BMI, HDLc, and glucose were identified as the third, or either fourth, or fifth most important parameters, respectively, which was followed by insulin. Non-HDLc was also identified as either the sixth or the eighth most important variable. Except for insulin, all identified parameters were labeled a priori as standard laboratory parameters. Therefore, insulin was not included any further in our models. The variable importance carried out on the comprehensive set of 29 parameters also showed similar results (Figure S1). Following parameter selection, we explored unique combinations of the seven selected parameters that resulted in 18 different metabotype models. All models were described on the basis of the cumulative incidence of diseases in the most unfavorable cluster (cluster 3). Compared to the model with 14 parameters, all 18 models based on the selected parameters resulted in comparatively higher incidences of both metabolic and cardiovascular disease

in participants of cluster 3 (Table S2). Among all models, cluster 3 of model 7 showed the highest incidence (62%) of “any metabolic disease”, whereas cluster 3 of model 17 revealed the highest incidence of “any cardiovascular disease” (9.1%) (Table S2). As a result, these two models were selected for further exploration.

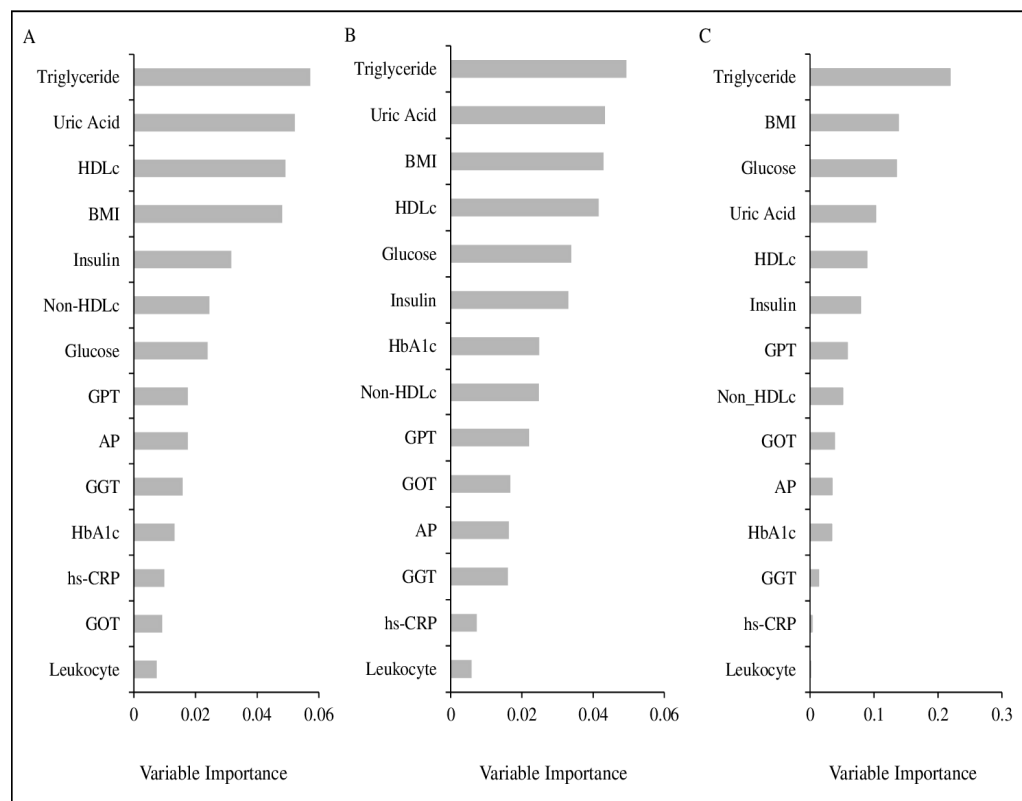


Figure 2. Variable importance of the clustering parameters based on the permutation variable importance (A), cross-validated permutation variable importance (B), and gradient-boosted feature selection (C).

Table 2 shows the distribution of socio-demographic variables across all three clusters for the best models, models 7 and 17. In model 7, about 40% ($n = 1189$) of participants were assigned to cluster 1, 48% ($n = 1140$) to cluster 2, and 12% ($n = 372$) to cluster 3. Meanwhile, in model 17, 42% ($n = 1253$) were assigned to cluster 1, 49% ($n = 1467$) to cluster 2, and 9% ($n = 281$) to cluster 3. In both models, a high proportion of men (60% to 70%) were in cluster 3, whereas a high proportion of women (~70%) were in cluster 1. In both models, cluster 3 had the highest median age (65 years and 64 years, respectively) and median BMI (33.2 and 30.5 kg/m², respectively). Similarly, almost 60% of the participants in cluster 3 were physically inactive. Moreover, participants in cluster 3 were more often heavy drinkers (more than 40 g/day). Compared to cluster 3, cluster 2 and cluster 1 included a higher number of never-smokers and participants with higher education levels.

Table 3 displays the distributions of clinical parameters across the different clusters in the two selected metabotype models. All five clustering parameters in model 7 were significantly different across clusters. Moreover, in both models, other biochemical parameters that were not included in the metabotyping process, such as GGT, GOT, GPT, HbA1c, hs-CRP, AP, insulin, and leukocyte count, also showed significant differences across clusters, with the most unfavorable values in cluster 3.

Table 2. Characteristics of the study population across the three clusters in model 7 and model 17, KORA F4 study.

	Total	Cluster 1	Metabotype Cluster 2	Cluster 3	p-Value
Model 7	n = 3001	n = 1189	n = 1440	n = 372	
Age Median (IQR)	56.0 (22.0)	51.0 (20.0)	57.0 (22.0)	65.0 (15.0)	<0.001
Sex					
Male	1450 (48.3%)	282 (23.7%)	942 (65.4%)	226 (60.8%)	<0.001
Female	1551 (51.7%)	907 (76.3%)	498 (34.6%)	146 (39.2%)	
Education					
<10 years	261 (8.7%)	86 (7.2%)	124 (8.6%)	51 (13.7%)	<0.001
10–<12 years	1499 (50.0%)	584 (49.2%)	721 (50.1%)	194 (52.2%)	
≥12 years	1236 (41.3%)	516 (43.5%)	593 (41.2%)	127 (34.1%)	
BMI					
Median (IQR)	27.0 (5.9)	24.2 (3.9)	28.2 (4.6)	33.2 (6.5)	<0.001
Normal weight (18.5–<25)	941 (31.5%)	723 (61.2%)	205 (14.3%)	13 (3.5%)	<0.001
Overweight (25–<30)	1253 (41.9%)	387 (32.8%)	793 (55.2%)	73 (19.8%)	
Obese (≥30)	793 (26.5%)	71 (6.0%)	439 (30.5%)	283 (76.7%)	
Physical Activity					
Active	1641 (54.8%)	719 (60.5%)	767 (53.4%)	155 (41.7%)	<0.001
Inactive	1356 (45.2%)	469 (39.5%)	670 (46.6%)	217 (58.3%)	
Smoking					
Smoker	524 (17.5%)	210 (17.7%)	282 (19.6%)	32 (8.6%)	<0.001
Ex-smoker	1218 (40.7%)	419 (35.3%)	594 (41.3%)	205 (55.1%)	
Never smoker	1254 (41.9%)	558 (47.0%)	561 (39.0%)	135 (36.3%)	
Alcohol consumption					
≥40 g/day	336 (11.2%)	97 (8.2%)	181 (12.6%)	58 (15.6%)	<0.001
20–<40 g/day	540 (18.0%)	211 (17.8%)	273 (19.0%)	56 (15.1%)	
0–<20 g/day	1221 (40.7%)	534 (44.9%)	561 (39.0%)	126 (33.9%)	
0 g/day	900 (30.0%)	346 (29.1%)	422 (29.4%)	132 (35.5%)	
Model 17	n = 3001	n = 1253	n = 1476	n = 281	
Age Median (IQR)	56.0 (22.0)	52.0 (22.0)	57.0 (21.0)	64.0 (16.0)	<0.001
Sex					
Male	1450 (48.3%)	384 (30.6%)	868 (59.2%)	198 (70.5%)	<0.001
Female	1551 (51.7%)	869 (69.4%)	599 (40.8%)	83 (29.5%)	
Education					
<10 years	261 (8.7%)	99 (7.9%)	132 (9.0%)	30 (10.7%)	0.024
10–<12 years	1499 (50.0%)	620 (49.6%)	719 (49.1%)	160 (56.9%)	
≥12 years	1236 (41.3%)	532 (42.5%)	613 (41.9%)	91 (32.4%)	
BMI					
Median (IQR)	27.0 (5.9)	25.0 (5.3)	28.0 (5.2)	30.5 (5.9)	<0.001
Normal weight (18.5–<25)	941 (31.5%)	632 (50.6%)	287 (19.7%)	22 (7.9%)	<0.001
Overweight (25–<30)	1253 (41.9%)	437 (35.0%)	716 (49.0%)	100 (36.0%)	
Obese (≥30)	793 (26.5%)	180 (14.4%)	457 (31.3%)	156 (56.1%)	
Physical Activity					
Active	1641 (54.8%)	763 (61.0%)	763 (52.1%)	115 (40.9%)	<0.001
Inactive	1356 (45.2%)	488 (39.0%)	702 (47.9%)	166 (59.1%)	

Table 2. Cont.

	Total	Cluster 1	Metabotype Cluster 2	Cluster 3	p-Value
Smoking					
Smoker	524 (17.5%)	195 (15.6%)	277 (18.9%)	52 (18.5%)	<0.001
Ex-smoker	1218 (40.7%)	476 (38.1%)	604 (41.2%)	138 (49.1%)	
Never smoker	1254 (41.9%)	579 (46.3%)	584 (39.9%)	91 (32.4%)	
Alcohol consumption					
≥40 g/day	336 (11.2%)	128 (10.2%)	155 (10.6%)	53 (18.9%)	<0.001
20–<40 g/day	540 (18.0%)	212 (16.9%)	282 (19.2%)	46 (16.4%)	
0–<20 g/day	1221 (40.7%)	568 (45.4%)	560 (38.2%)	93 (33.1%)	
0 g/day	900 (30.0%)	343 (27.4%)	468 (31.9%)	89 (31.7%)	

Median (IQR) for continuous variables and n (column %) for categorical variables, NA excluded. *p*-values are from the Kruskal–Wallis test for continuous variables and from Pearson’s chi-squared test for categorical variables. Model 7 included 5 parameters (glucose, BMI, uric acid, HDLc, and non-HDLc), and Model 17 included 4 parameters (glucose, triglyceride, HDLc, and non-HDLc), and BMI body mass index. Due to missing information, there were reduced data sets for education *n* = 2996, BMI *n* = 2987, physical activity *n* = 2997, smoking *n* = 2996, and alcohol consumption *n* = 2997. The highest values across the clusters are marked in bold. KORA, Cooperative Health Research in the Region of Augsburg. KORA, Cooperative Health Research in the Region of Augsburg.

Table 3. Comparison of the clustering parameters and other metabolic parameters across the three clusters of two selected clustering models (7 and 17), KORA F4 study.

	Total	Cluster 1	Metabotype Cluster 2	Cluster 3	p-Value
Model 7	n = 3001	n = 1189	n = 1440	n = 372	
Parameters used for metabotyping					
BMI [kg/m ²]	26.90 (5.93)	24.19 (4.31) ^a	28.24 (4.85) ^b	33.17 (6.47) ^c	<0.001
Uric acid [μmol/L]	299.41 (114.12)	243.52 (79.29) ^a	334.12 (93.88) ^b	375.29 (113.76) ^c	<0.001
Glucose [mg/dL]	94.00 (14.00)	89.00 (10.80) ^a	96.00 (12.40) ^b	122.00 (28.40) ^c	<0.001
HDLc [mmol/L]	1.39 (0.52)	1.70 (0.45) ^a	1.26 (0.37) ^b	1.18 (0.39) ^c	<0.001
Non-HDLc [mmol/L]	4.05 (1.32)	3.59 (1.21) ^a	4.49 (1.28) ^b	4.01 (1.25) ^c	<0.001
Other Parameters					
TG [mmol/L]	1.19 (0.90)	0.82 (0.56) ^a	1.46 (0.87) ^b	1.76 (1.22) ^c	<0.001
AP [μmol/L]	1.10 (0.42)	1.00 (0.43) ^a	1.14 (0.39) ^b	1.21 (0.43) ^c	<0.001
GPT [μkat/L]	0.35 (0.23)	0.28 (0.16) ^a	0.40 (0.24) ^b	0.48 (0.32) ^c	<0.001
GOT [μkat/L]	0.41 (0.14)	0.38 (0.12) ^a	0.43 (0.14) ^b	0.45 (0.19) ^c	<0.001
GGT [μkat/L]	0.43 (0.4)	0.31 (0.26) ^a	0.50 (0.42) ^b	0.63 (0.53) ^c	<0.001
HbA1c [%]	5.50 (0.5)	5.30 (0.42) ^a	5.50 (0.5) ^b	6.10 (0.98) ^c	<0.001
hs-CRP [mg/L]	1.18 (2.03)	0.76 (1.38) ^a	1.38 (2.08) ^b	2.49 (3.27) ^c	<0.001
Leukocytes (n/L)	5.70 (2)	5.30 (1.84) ^a	5.90 (1.93) ^b	6.30 (2) ^c	<0.001
Insulin [μU/mL]	8.80 (6.70)	6.60 (4.14) ^a	10.00 (6.46) ^b	18.00 (11.74) ^c	<0.001
Model 17	n = 3001	n = 1253	n = 1467	n = 281	
Parameters used for metabotyping					
TG [mmol/L]	1.19 (0.90)	1.18 (0.90) ^a	1.47 (0.78) ^b	2.71 (1.74) ^c	<0.001
Glucose [mg/dL]	94.00 (14.00)	90.00 (11.60) ^a	96.00 (13.20) ^b	124.00 (38.00) ^c	<0.001
HDLc [mmol/L]	1.39 (0.52)	1.70 (0.45) ^a	1.24 (0.37) ^b	1.08 (0.35) ^c	<0.001
Non-HDLc [mmol/L]	4.05 (1.32)	3.48 (1.05) ^a	4.49 (1.19) ^b	4.49 (1.27) ^b	<0.001
Other Parameters					
BMI [kg/m ²]	26.99 (5.93)	24.95 (5.42) ^a	27.95 (5.35) ^b	30.54 (5.93) ^c	<0.001
Uric acid [μmol/L]	299.41 (114.12)	260.5 (96.70) ^a	321.17 (107.05) ^b	378.23 (117.8) ^c	<0.001
AP [μmol/L]	1.10 (0.42)	1.01 (0.42) ^a	1.15 (0.41) ^b	1.20 (0.44) ^b	<0.001
GPT [μkat/L]	0.35 (0.23)	0.30 (0.18) ^a	0.39 (0.24) ^b	0.48 (0.31) ^c	<0.001

Table 3. Cont.

	Total	Cluster 1	Metabotype Cluster 2	Cluster 3	p-Value
GOT [μ kat/L]	0.41 (0.14)	0.39 (0.13) ^a	0.42 (0.15) ^b	0.45 (0.19) ^c	<0.001
GGT [μ kat/L]	0.43 (0.40)	0.34 (0.28) ^a	0.48 (0.42) ^b	0.71 (0.59) ^c	<0.001
HbA1c [%]	5.50 (0.50)	5.40 (0.50) ^a	5.50 (0.50) ^b	6.20 (1.22) ^c	<0.001
hs-CRP [mg/L]	1.18 (2.03)	0.86 (1.55) ^a	1.39 (2.23) ^b	1.97 (2.57)	<0.001
Leukocytes (n/L)	5.70 (2.00)	5.40 (1.84) ^a	5.90 (2.00) ^b	6.40 (2.16)	<0.001
Insulin [μ U/mL]	8.80 (6.70)	6.90 (4.62) ^a	10.00 (6.54) ^b	17.00 (10.94)	<0.001

Median and interquartile range (IQR) were calculated through the mean of median and IQR in all five imputed data sets. *p*-values are from the Kruskal–Wallis test. Different superscript letters represent a significant difference between clusters obtained from the Kruskal–Wallis post hoc test with Bonferroni correction. The highest median values across the three clusters are marked in bold. BMI: body mass index; HDLc: high-density lipoprotein; TG: triglyceride; GGT: gamma-glutamyltransferase; GOT: glutamate-oxaloacetate transaminase; GPT: glutamate-pyruvate transaminase; HbA1c: glycated hemoglobin, hs-CRP: high-sensitive C-reactive protein, AP; alkaline phosphatase. KORA, Cooperative Health Research in the Region of Augsburg.

Table 4 presents the prevalence and incidence of metabolic and cardiovascular diseases of study participants in the three clusters of models 7 and 17, respectively. Regarding the incidence of individual metabolic and cardiovascular diseases in both models, we obtained significant differences across clusters, except for hypertension. Compared to cluster 1 and cluster 2, participants in cluster 3 of both models (models 7 and 17) showed the highest incidence of all metabolic and cardiovascular diseases. This holds for “any metabolic disease” (model 7 (13% vs. 23.7% vs. 62%), model 17 (13.5% vs. 25.3% vs. 58.4%)) and “any cardiovascular disease” (model 7 (1.4% vs. 3.4% vs. 7.4%), model 17 (1.5% vs. 3.3% vs. 9.1%)).

Table 4. Prevalence and incidence of diseases across the three clusters of two selected clustering models (7 and 17), KORA F4 and FF4 studies.

	Total	Cluster 1	Metabotype Cluster 2	Cluster 3	p-Value
Model 7					
Prevalence of disease in KORA F4; n (%)	n = 3001	n = 1189	n = 1440	n = 372	
Type 2 diabetes	242 (8.06%)	15 (1.26%)	52 (3.6%)	175 (47.0%)	<0.001
Hypertension	1150 (38.4%)	249 (21.0%)	616 (42.9%)	285 (76.6%)	<0.001
Hyperuricemia	113 (3.8%)	15 (1.3%)	58 (4.0%)	40 (10.8%)	<0.001
Dyslipidemia	386 (12.9%)	101 (8.5%)	173 (12.0%)	112 (30.2%)	<0.001
Any metabolic diseases	1309 (43.7%)	299 (25.2%)	684 (47.6%)	326 (87.9%)	<0.001
Stroke	71 (2.4%)	19 (1.6%)	32 (2.2%)	20 (5.4%)	<0.001
Myocardial infraction	79 (2.6%)	12 (1.0%)	38 (2.6%)	29 (7.8%)	<0.001
Any cardiovascular disease	142 (4.7%)	27 (2.3%)	68 (4.7%)	47 (12.6%)	<0.001
Incidence of disease in KORA FF4; n (%)					
Type 2 diabetes	94 (4.7%)	13 (1.5%)	43 (4.4%)	38 (30.2%)	<0.001
Hypertension	230 (10.9%)	86 (9.6%)	114 (11.4%)	30 (13.6%)	0.187
Hyperuricemia	45 (2.1%)	0 (0.0%)	24 (2.4%)	21 (9.5%)	<0.001
Dyslipidemia	157 (7.4%)	27 (3.0%)	92 (9.2%)	38 (17.2%)	<0.001
Any metabolic diseases	442 (21.9%)	114 (13.0%)	233 (23.7%)	95 (62.0%)	<0.001
Stroke	35 (1.7%)	10 (1.1%)	15 (1.5%)	10 (4.6%)	0.001
Myocardial infraction	27 (1.3%)	2 (0.2%)	20 (2.0%)	5 (2.4%)	<0.001
Any cardiovascular disease	60 (2.9%)	12 (1.4%)	33 (3.4%)	15 (7.4%)	<0.001

Table 4. Cont.

	Total	Cluster 1	Metabotype Cluster 2	Cluster 3	p-Value
Model 17					
Prevalence of disease in KORA F4; n (%)	n = 3001	n = 1235	n = 1467	n = 281	
Type 2 diabetes	242 (8.1%)	39 (3.1%)	67 (4.6%)	136 (48.4%)	<0.001
Hypertension	1150 (38.4%)	332 (26.6%)	620 (42.4%)	198 (70.5%)	<0.001
Hyperuricemia	113 (3.8%)	22 (1.8%)	53 (3.6%)	38 (13.6%)	<0.001
Dyslipidemia	386 (12.9%)	140 (11.2%)	168 (11.5%)	78 (27.9%)	<0.001
Any metabolic diseases	1309 (43.7%)	385 (31.0%)	694 (47.4%)	230 (82.1%)	<0.001
Stroke	71 (2.4%)	26 (2.1%)	35 (2.4%)	10 (3.6%)	0.334
Myocardial infraction	79 (2.6%)	24 (1.9%)	37 (2.5%)	18 (6.4%)	<0.001
Any cardiovascular disease	142 (4.7%)	46 (3.6%)	70 (4.8%)	26 (9.2%)	<0.001
Incidence of disease in KORA FF4; n (%)	n = 2120	n = 916	n = 1040	n = 164	
Type 2 diabetes	94 (4.7%)	15 (1.7%)	54 (5.4%)	25 (26.9%)	<0.001
Hypertension	230 (10.9%)	85 (9.3%)	125 (12.0%)	20 (12.3%)	0.127
Hyperuricemia	45 (2.1%)	9 (0.1%)	29 (2.8%)	7 (4.3%)	0.002
Dyslipidemia	157 (7.4%)	30 (3.3%)	92 (8.8%)	35 (21.5%)	<0.001
Any metabolic diseases	442 (21.9%)	120 (13.5%)	256 (25.3%)	66 (58.4%)	<0.001
Stroke	35 (1.7%)	11 (1.2%)	14 (1.4%)	10 (6.2%)	<0.001
Myocardial infraction	27 (1.3%)	2 (0.2%)	21 (2.0%)	4 (2.6%)	<0.001
Any cardiovascular disease	60 (2.9%)	13 (1.5%)	33 (3.3%)	14 (9.1%)	<0.001

n (column%), NA excluded. *p*-values from Pearson's chi-squared test (Fisher's exact test for low frequencies). Model 7 included 5 parameters (glucose, BMI, uric acid, HDLc, and non-HDLc), and Model 17 included 4 parameters (glucose, triglyceride, HDLc, and non-HDLc). Prevalence: due to missing information, there were reduced data sets for hypertension *n* = 2994, hyperuricemia *n* = 2999, dyslipidemia *n* = 2998, and "any metabolic disease" *n* = 2993. Incidence: due to missing information, there were reduced data sets for type 2 diabetes *n* = 1988, hypertension *n* = 2115, hyperuricemia *n* = 2117, dyslipidemia *n* = 2117, "any metabolic disease" *n* = 2017, stroke *n* = 2091, myocardial infraction *n* = 2076, and "any cardiovascular disease" *n* = 2055. The highest prevalence and incidence of diseases across three different clusters are marked in bold. KORA, Cooperative Health Research in the Region of Augsburg.

4. Discussion

We further improved the metabotyping concept in the population-based KORA F4/FF4 study by using a few routinely measured clinical parameters (namely TG, BMI, uric acid, fasting glucose, HDLc, and non-HDLc) that were identified through the PVI method. By computing a k-means cluster analysis, we identified three-cluster solutions and described them on the basis of metabolic parameters and disease occurrence. We selected two models as the most appropriate solutions.

To the best of our knowledge, this is the first study to assess the importance of parameters for metabotypes, using the PVI method to select them. We further reinforced the results from PVI using two additional variable importance methods (i) CVPVI and (ii) gradient-boosted feature selection. The similar results from these two additional methods validated the results from the PVI method. Furthermore, we created multiple unique subsets of the selected parameters to create 18 different metabotype models, in contrast to using just the initially selected parameters. In accordance with our earlier studies [15,27] and also with other similar papers [5,16,47–51], we used the unsupervised method of k-means clustering for metabotyping, resulting in metabotype models that were independent of disease. Two models out of eighteen were chosen, based on the incidence of disease in the 7-year follow-up KORA FF4 study, and were further analyzed.

Metabotype model 7 was described as the metabotype model with the highest incidence of metabolic diseases in the unfavorable cluster 3, and was based on five parameters (glucose, BMI, uric acid, HDLc, and non-HDLc); meanwhile, cluster 3 of model 17 showed the highest incidence of cardiovascular diseases, and included four parameters (TG, glucose, HDLc, and non-HDLc). Both models 7 and 17 were evaluated using an additional set of biochemical parameters that were not included in identifying the metabotype groups. The concentrations of the biochemical parameters across the three clusters of both models were

consistently and significantly different, showing the unique metabolic characteristics of each cluster. This validated our identification of distinct metabolotype subgroups (clusters).

The differences regarding the prevalence and incidence of diseases between clusters in both models were statistically significant, except for the incidence of hypertension. Although the incidence of hypertension was higher in cluster 3, it did not reach statistical significance in either model 7 or 17 ($p = 0.18$ and $p = 0.12$, respectively). However, there was a significant difference in the prevalence of hypertension across all three clusters in both models. This may have been due to the high prevalence of hypertension (more than 70%) among participants in cluster 3. Another explanation could also be that there may have been participation bias in the KORA FF4 study, as those who did not participate in the follow-up study were less healthy [52]. Regarding the socio-demographic characteristics, participants in cluster 3 were more likely to have received less than 10 years of education, had a higher median age, a higher BMI, were more physically inactive, and were more often heavy drinkers compared to participants in cluster 2 and cluster 1. Additionally, cluster 3 included the lowest number of current smokers compared to other clusters, but the highest number of ex-smokers and non-drinkers, which likely reflects behavioral changes in response to worsening health status. Thus, these clear differences in risk factors and occurrence of diseases across clusters show that the identified metabolotypes represent specific characteristics where the clusters can be meaningfully classified into healthy, intermediate, and unfavorable clusters. Thus, the identified metabolotypes can be used as a tool to stratify populations according to their metabolic features. However, the intention of the present research was not to create a risk prediction model; rather, it aimed to define metabolically homogenous subgroups in the population.

We previously used the metabolotyping concept to investigate associations between diet and type-2 diabetes (T2D), and identified different associations by metabolotype subgroups [27]. T2D risk increased in the healthy subgroup with a higher intake of total meat and processed meat, while in the unhealthy subgroup T2D risk was positively associated with consumption of sugar-sweetened beverages, and inversely associated with fruit intake. We also found significant associations between dietary patterns and T2D in the total population; however, when stratified by the metabolotype subgroups, a significant association was only seen in the unhealthy metabolotype subgroup [46]. Likewise, in studies by Fiamoncini et al. [53] and O'Sullivan et al. [48], the effect of dietary intervention was only evident after dividing the population into metabolic phenotypes. Similarly, in our recent publication, we successfully applied the metabolotypes that were developed in this manuscript in a different study population where participants in different metabolotype subgroups showed significantly differential reactions to the oral glucose tolerance test (OGTT) [54]. These studies clearly demonstrate that metabolotyping can be used to identify a metabolically similar high-risk subgroup that can benefit from targeted dietary advice and lifestyle intervention.

O'donovan et al. [5] and Hillesheim et al. [19] used decision tree methods in their studies to develop targeted dietary advice for specific metabolic subgroups. When the dietary advice from the decision tree was compared to the individual-based approach that was delivered by a dietitian, they found that the advice matched in more than 80% of the study population. Similar results were seen in the Food4Me study [55], where decision trees were used to provide personalized dietary advice to adults in seven European countries. Providing dietary advice at the individual level is the epitome of personalized nutrition; however, this approach is costly, and involves extensive data collection [19]. On the other hand, the metabolotype approach provides a simpler and more feasible approach [12]. These results show that using metabolotypes can be a promising tool in the field of personalized nutrition. Furthermore, this approach can also help clinicians provide dietary recommendations quickly [5,19] by overcoming the usual barriers such as lack of time, heavy workloads, and inadequate training [56,57].

Several studies have implemented both small and large ($n > 50$) sets of anthropometric and biochemical parameters, in order to identify metabolically similar

groups [3,5,16,17,19,49,58–61]. Bouwman et al. [14] even included omics data and identified two distinct subgroups to study visualization and identification of health space. As omics data may provide a more comprehensive outlook, we also investigated the inclusion of omics as well as genetics data for metabotyping in one of our previous studies [15]. However, the model did not perform better than the model with the extensive set of 32 biochemical and anthropometric parameters (details not published). A few other studies have also identified metabolotypes as a useful measure to identify subjects with a high risk of cardiometabolic disease [48,50,54,62–64]. However, these studies did not include a wide range of metabolic and cardiovascular diseases. Furthermore, previous studies were often based on parameters that are not easily accessible in daily practice. For example, in a study by Urpi-Sarda et al. [63], four different subgroups were identified on the basis of a urinary metabolomics fingerprint associated with type-2 diabetes.

The two metabolotype models identified in this study are based on a small number of routinely used clinical parameters. We found a distinct difference across the clusters in regard to metabolic parameters. Similarly, the distribution of disease was also different between the subgroups; in addition, the unfavorable cluster 3 showed the highest percentage for both prevalence and incidence of diseases in both models. These consistent results show that our clustering model successfully identified valid metabolotypes, which have the advantage of being simpler yet no less valid than previously identified metabolotype models [8,15,27]. Moreover, the successful application of the metabolotypes identified in this study in a different study population further validates our metabolotypes concept [54]. These findings illustrate that identified metabolotypes can be easily applied at a population level, as well as in general research settings and primary care, to detect metabolically similar subgroups. Furthermore, they may also help to develop new, targeted, and precise dietary recommendations based on their different metabolic features.

The present study was conducted in a large population-based cohort, which makes the finding of our study generalizable to the adult German population. The use of a few readily available parameters to define valid metabolotypes makes the findings of this study simple, cost-effective, and instantly applicable on a large scale, if replicated in other cohorts. We were also able to consider subjects with previously unknown or undiagnosed type 2 diabetes by using an oral glucose tolerance test at baseline and follow-up, which could have otherwise remained undetected. Another strength of this study is that the incidence data on self-reported diseases such as myocardial infarction, stroke, and type 2 diabetes were validated by means of medical records. However, there may have been an underestimation of the prevalence and incidence of dyslipidemia and hyperuricemia, as they were defined only on the bases of reported intake of lipid-lowering drugs and hyperuricemia/gout medication. Additionally, we lost many participants in the follow-up KORA FF4 study, which may have biased our results [52]. The highest dropout rate of participants in cluster 3 may also have underestimated the disease incidence in this cluster.

In conclusion, we successfully identified two valid and practical metabolotype solutions based on a minimal number of routinely measured clinical parameters that are generally available in research settings and primary care. Thus, the identified metabolotypes can easily be applied to the general population for the purposes of identifying individuals who could benefit from receiving additional preventive measures targeted to metabolic derangements, such as dietary recommendations and lifestyle modifications. The replication of identified metabolotypes in a different cohort could further aid in the development of a simple and consistent metabolotype definition.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/life12101460/s1>, Table S1: Characteristics of the KORA FF4 study population; Table S2: Incidence of diseases in the unfavorable metabolotype clusters identified in different clustering models based on combinations of seven parameters, versus the 14-parameter clustering model, KORA FF4 study; Figure S1: Variable importance of the 29 clustering parameters using the permutation variable importance (A), cross-validated permutation variable Importance (B), and gradient-boosted feature selection method (C).

Author Contributions: Conceptualization, J.L.; methodology, C.D.; formal analysis, C.D. and N.W.; data curation, B.T., W.R., W.K. and A.P.; writing—original draft preparation, C.D.; writing—review and editing, N.W., C.M., T.A.B., B.T., W.R., W.K. and H.H.; visualization, C.D.; supervision, J.L.; project administration, J.L.; funding acquisition, J.L. and H.H. All authors have read and agreed to the published version of the manuscript.

Funding: The Cooperative Health Research in the Region of Augsburg (KORA) studies are initiated and financed by the Helmholtz Zentrum München—German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF), and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians- Universität, as part of LMUinnovativ. CD and NW were supported by a grant from the German Ministry for Education and Research (BMBF) FK 01EA1807E and the enable Competence Cluster of Nutrition Research. This manuscript is cataloged by the enable Steering Committee as enable #064. The funders had no role in the study design, data collection, and analysis; in the decision to publish, or in the preparation of the manuscript.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of the Bavarian Medical Association.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The authors confirm that, for approved reasons, access restrictions apply to the data underlying the findings, and thus the information cannot be made freely available in the manuscript, in the supplemental files, or at a public repository. The data are subject to national data protection laws and restrictions that were imposed by the ethics committee of the Bavarian Medical Association (“Bayerische Landesärztekammer”), in order to ensure the data privacy of the study participants, since they did not explicitly consent to the data being made publicly available. Data can be applied through an individual project agreement with KORA that allows researchers to access the data in the same way used by this study’s authors to access the data. Applications for access to the data sets can be made via the KORA-passt platform (<https://helmholtz-muenchen.managed-otrs.com/external/>, accessed on 10 August 2022).

Acknowledgments: We thank all of the participants of the KORA F4 and KORA FF4 studies for their contributions.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

BMI	body mass index
HDL	high-density lipoprotein
TG	triglyceride
GGT	gamma-glutamyltransferase
GOT	glutamate-oxaloacetate transaminase
GPT	glutamate-pyruvate transaminase
HbA1c	glycated hemoglobin
hs-CRP	high-sensitive C-reactive protein
AP	alkaline phosphatase
KORA	Cooperative Health Research in the Region of Augsburg

References

1. Holmes, E.; Wilson, I.D.; Nicholson, J.K. Metabolic Phenotyping in Health and Disease. *Cell* **2008**, *134*, 714–717. [[CrossRef](#)] [[PubMed](#)]
2. O’donovan, C.B.; Walsh, M.C.; Gibney, M.J.; Gibney, E.R.; Brennan, L. Can metabotyping help deliver the promise of personalised nutrition? *Proc. Nutr. Soc.* **2015**, 106–114. [[CrossRef](#)] [[PubMed](#)]
3. Riedl, A.; Gieger, C.; Hauner, H.; Daniel, H.; Linseisen, J. Metabotyping and its application in targeted nutrition: An overview. *Br. J. Nutr.* **2017**, *117*, 1631–1644. [[CrossRef](#)] [[PubMed](#)]
4. Hillesheim, E.; Brennan, L. Metabotyping and its role in nutrition research. *Nutr. Res. Rev.* **2019**, *33*, 1–10. [[CrossRef](#)] [[PubMed](#)]
5. O’Donovan, C.B.; Walsh, M.C.; Nugent, A.P.; McNulty, B.; Walton, J.; Flynn, A.; Gibney, M.J.; Gibney, E.R.; Brennan, L. Use of metabotyping for the delivery of personalised nutrition. *Mol. Nutr. Food Res.* **2015**, *59*, 377–385. [[CrossRef](#)]
6. Brennan, L. Use of metabotyping for optimal nutrition. *Curr. Opin. Biotechnol.* **2017**, *44*, 35–38. [[CrossRef](#)]

7. De Roos, B. Personalised nutrition: Ready for practice? In Proceedings of the Conference on 'Future Food and Health' Symposium II: Diet–Gene Interactions; Implications for Future Diets and Health, Aberdeen, UK, 26–27 March 2012. [\[CrossRef\]](#)
8. Riedl, A.; Hillesheim, E.; Wawro, N.; Meisinger, C.; Peters, A.; Roden, M.; Kronenberg, F.; Herder, C.; Rathmann, W.; Völzke, H.; et al. Evaluation of the Metabotype Concept Identified in an Irish Population in the German KORA Cohort Study. *Mol. Nutr. Food Res.* **2020**, *64*, 1900918. [\[CrossRef\]](#)
9. Zeevi, D.; Korem, T.; Zmora, N.; Israeli, D.; Rothschild, D.; Weinberger, A.; Ben-Yacov, O.; Lador, D.; Avnit-Sagi, T.; Lotan-Pompan, M.; et al. Personalized Nutrition by Prediction of Glycemic Responses. *Cell* **2015**, *163*, 1079–1094. [\[CrossRef\]](#)
10. Celis-Morales, C.; Livingstone, K.M.; Marsaux, C.F.; Macready, A.L.; Fallaize, R.; O'donovan, C.B.; Woolhead, C.; Forster, H.; Walsh, M.C.; Navas-Carretero, S.; et al. Effect of personalized nutrition on health-related behaviour change: Evidence from the Food4Me European randomized controlled trial. *Int. J. Epidemiol.* **2017**, *46*, 578–588. [\[CrossRef\]](#)
11. Livingstone, K.M.; Celis-Morales, C.; Navas-Carretero, S.; San-Cristobal, R.; Forster, H.; Woolhead, C.; O'Donovan, C.B.; Moschonis, G.; Manios, Y.; Traczyk, I.; et al. Personalized Nutrition Advice Reduces Intake of Discretionary Foods and Beverages: Findings From the Food4Me Randomized Controlled Trial. *Curr. Dev. Nutr.* **2021**, *5*, 152. [\[CrossRef\]](#)
12. Palmnäs, M.; Brunius, C.; Shi, L.; Rostgaard-Hansen, A.; Torres, N.E.; González-Domínguez, R.; Zamora-Ros, R.; Ye, Y.L.; Halkjær, J.; Tjønneland, A.; et al. Perspective: Metabotyping—A Potential Personalized Nutrition Strategy for Precision Prevention of Cardiometabolic Disease. *Adv. Nutr.* **2019**, *10*, S308–S319. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Brennan, L. Symposium on “The challenge of translating nutrition research into public health nutrition” Session 2: Personalised nutrition Metabolomic applications in nutritional research. *Proc. Nutr. Soc.* **2020**, *67*, 404–408. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Bouwman, J.; Vogels, J.T.; Wopereis, S.; Rubingh, C.M.; Bijlsma, S.; van Ommen, B. Visualization and identification of health space, based on personalized molecular phenotype and treatment response to relevant underlying biological processes. *BMC Med. Genomics.* **2012**, *5*, 1. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Riedl, A.; Wawro, N.; Gieger, C.; Meisinger, C.; Peters, A.; Roden, M.; Kronenberg, F.; Herder, C.; Rathmann, W.; Völzke, H.; et al. Identification of Comprehensive Metabotypes Associated with Cardiometabolic Diseases in the Population-Based KORA Study. *Mol. Nutr. Food Res.* **2018**, *62*, 1–9. [\[CrossRef\]](#)
16. Vázquez-Fresno, R.; Llorach, R.; Perera, A.; Mandal, R.; Tinahones, F.J.; Wishart, D.S.; Andrés-Lacueva, C. Clinical phenotype clustering in cardiovascular risk patients for the identification of responsive metabotypes after red wine polyphenol intake. *J. Nutr. Biochem.* **2016**, *28*, 114–120. [\[CrossRef\]](#)
17. Moazzami, A.A.; Shrestha, A.; Morrison, D.A.; Poutanen, K.; Mykkänen, H. Metabolomics Reveals Differences in Postprandial Responses to Breads and Fasting Metabolic Characteristics Associated with Postprandial Insulin Demand in Postmenopausal Women. *J. Nutr.* **2014**, *144*, 807–814. [\[CrossRef\]](#)
18. Mäkinen, V.; Soininen, P.; Forsblom, C.; Parkkonen, M.; Ingman, P.; Kaski, K.; Groop, P.-H.; Ala-Korpela, M. ¹H NMR metabolomics approach to the disease continuum of diabetic complications and premature death. *Mol. Syst. Biol.* **2008**, *4*, 167. [\[CrossRef\]](#)
19. Hillesheim, E.; Ryan, M.F.; Gibney, E.; Roche, H.M.; Brennan, L. Optimisation of a metabotype approach to deliver targeted dietary advice. *Nutr. Metab.* **2020**, *17*, 1–12. [\[CrossRef\]](#)
20. Zubair, N.; Kuzawa, C.W.; McDade, T.W.; Adair, L.S. Cluster analysis reveals important determinants of cardiometabolic risk patterns in Filipino women. *Asia Pac. J. Clin. Nutr.* **2012**, *21*, 271–281. [\[CrossRef\]](#)
21. Holle, R.; Happich, M.; Löwel, H.; Wichmann, H.E. KORA-A research platform for population based health research. *Gesundheitswesen* **2005**, *67*, 19–25. [\[CrossRef\]](#)
22. Herder, C.; Bongaerts, B.W.C.; Rathmann, W.; Heier, M.; Kowall, B.; Koenig, W.; Thorand, B.; Roden, M.; Meisinger, C.; Ziegler, D. Association of subclinical inflammation with polyneuropathy in the older population: KORA F4 study. *Diabetes Care* **2013**, *36*, 3663–3670. [\[CrossRef\]](#)
23. Stöckl, D.; Meisinger, C.; Peters, A.; Thorand, B.; Huth, C.; Heier, M.; Rathmann, W.; Kowall, B.; Stöckl, H.; Döring, A. Age at menarche and its association with the metabolic syndrome and its components: Results from the Kora F4 study. *PLoS ONE* **2011**, *6*, 26076. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Kowall, B.; Rathmann, W.; Stang, A.; Bongaerts, B.; Kuss, O.; Herder, C.; Roden, M.; Quante, A.; Holle, R.; Huth, C.; et al. Perceived risk of diabetes seriously underestimates actual diabetes risk: The KORA FF4 study. *PLoS ONE* **2017**, *12*, e0171152. [\[CrossRef\]](#) [\[PubMed\]](#)
25. WHO/Europe | Nutrition-Body Mass Index-BMI. Available online: <http://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi> (accessed on 3 April 2020).
26. van Buuren, S.; Groothuis-Oudshoorn, K. Mice: Multivariate Imputation by Chained Equations in R. *J. Stat. Softw.* **2011**, *45*, 1–67. [\[CrossRef\]](#)
27. Riedl, A.; Wawro, N.; Gieger, C.; Meisinger, C.; Peters, A.; Rathmann, W.; Koenig, W.; Strauch, K.; Quante, A.S.; Thorand, B.; et al. Modifying effect of metabotype on diet–diabetes associations. *Eur. J. Nutr.* **2019**, *59*, 1357–1369. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Tanabe, N.; Iso, H.; Okada, K.; Nakamura, Y.; Harada, A.; Ohashi, Y.; Ando, T.; Ueshima, H. Serum total and non-high-density lipoprotein cholesterol and the risk prediction of cardiovascular events: -The JALS-ecc-. *Circ. J.* **2010**, *74*, 1347–1356. [\[CrossRef\]](#)
29. Cui, Y.; Blumenthal, R.S.; Flaws, J.A.; Whiteman, M.K.; Langenberg, P.; Bachorik, P.S.; Bush, T.L. Non-high-density lipoprotein cholesterol level as a predictor of cardiovascular disease mortality. *Arch. Intern. Med.* **2001**, *161*, 1413–1419. [\[CrossRef\]](#)

30. Angoorani, P.; Khademian, M.; Ejtahed, H.S.; Heshmat, R.; Motlagh, M.E.; Vafaeenia, M.; Shafiee, G.; Mahdivi-Gorabi, A.; Qorbani, M.; Kelishadi, R. Are non-high-density lipoprotein fractions associated with pediatric metabolic syndrome? The CASPIAN-V study. *Lipids Health Dis.* **2018**, *17*. [CrossRef]
31. Brunner, F.J.; Waldeyer, C.; Ojeda, F.; Salomaa, V.; Kee, F.; Sans, S.; Thorand, B.; Giampaoli, S.; Brambilla, P.; Tunstall-Pedoe, H.; et al. Application of non-HDL cholesterol for population-based cardiovascular risk stratification: Results from the Multinational Cardiovascular Risk Consortium. *Lancet* **2019**, *394*, 2173–2183. [CrossRef]
32. Putin, E.; Mamoshina, P.; Aliper, A.; Korzinkin, M.; Moskalev, A.; Kolosov, A.; Ostrovskiy, A.; Cantor, C.; Vijg, J.; Zhavoronkov, A. Deep biomarkers of human aging: Application of deep neural networks to biomarker development. *Aging* **2016**, *8*, 1021. [CrossRef]
33. Huynh-Thu, V.A.; Saeys, Y.; Wehenkel, L.; Geurts, P. Statistical interpretation of machine learning-based feature importance scores for biomarker discovery. *Bioinformatics* **2012**, *28*, 1766–1774. [CrossRef] [PubMed]
34. Prosperi, M.C.; Marinho, S.; Simpson, A.; Custovic, A.; Buchan, I.E. Predicting phenotypes of asthma and eczema with machine learning. *BMC Med. Genomics*. **2014**, *7*. [CrossRef] [PubMed]
35. Altmann, A.; Tološi, L.; Sander, O.; Lengauer, T. Permutation importance: A corrected feature importance measure. *Bioinformatics* **2010**, *26*, 1340–1347. [CrossRef] [PubMed]
36. PIMP Function | R Documentation. Available online: <https://www.rdocumentation.org/packages/vita/versions/1.0.0/topics/PIMP> (accessed on 9 March 2020).
37. Leo Breiman, A.C. Random Forests. Signal and Image Processing for Remote Sensing. 2003. Available online: <http://www.stat.berkeley.edu/breiman/RandomForests/cchome.htm> (accessed on 3 September 2020).
38. Janitz, S.; Celik, E.; Boulesteix, A.L. A computationally fast variable importance test for random forests for high-dimensional data. *Adv. Data Anal. Classif.* **2018**, *12*, 885–915. [CrossRef]
39. CVPVI Function | R Documentation. Available online: <https://www.rdocumentation.org/packages/vita/versions/1.0.0/topics/CVPVI> (accessed on 17 February 2021).
40. Xu, Z.; Huang, G.; Weinberger, K.Q.; Zheng, A.X. Gradient boosted feature selection. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014; pp. 522–531. [CrossRef]
41. Hastie, T.; Tibshirani, R.; Friedman, J. Boosting and Additive Trees. In *The Elements of Statistical Learning*, 2nd ed.; Springer: New York, NY, USA, 2017; pp. 337–384. Available online: <https://web.stanford.edu/~hastie/ElemStatLearn/> (accessed on 18 February 2021).
42. Feature Importance and Feature Selection with XGBoost in Python. Available online: <https://machinelearningmastery.com/feature-importance-and-feature-selection-with-xgboost-in-python/> (accessed on 18 February 2021).
43. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [CrossRef]
44. NbClust Function - RDocumentation. Available online: <https://www.rdocumentation.org/packages/NbClust/versions/3.0.1/topics/NbClust> (accessed on 5 September 2022).
45. Barrera-Gómez, J.; Maintainer, X.B. Mclust: Multiple Imputation in Cluster Analysis. Available online: <https://cran.r-project.org/web/packages/mclust/mclust.pdf> (accessed on 9 March 2020).
46. Wawro, N.; Pestoni, G.; Riedl, A.; Breuninger, T.A.; Peters, A.; Rathmann, W.; Koenig, W.; Huth, C.; Meisinger, C.; Rohrmann, S.; et al. Association of dietary patterns and type-2 diabetes mellitus in metabolically homogeneous subgroups in the KORA FF4 study. *Nutrients* **2020**, *12*, 1684. [CrossRef]
47. van Bochove, K.; van Schalkwijk, D.B.; Parnell, L.D.; Lai, C.-Q.; Ordovás, J.M.; de Graaf, A.A.; van Ommen, B.; Arnett, D.K. Clustering by Plasma Lipoprotein Profile Reveals Two Distinct Subgroups with Positive Lipid Response to Fenofibrate Therapy. *PLoS ONE* **2012**, *7*, e38072. [CrossRef]
48. O'Sullivan, A.; Gibney, M.J.; Connor, A.O.; Mion, B.; Kaluskar, S.; Cashman, K.D.; Flynn, A.; Shanahan, F.; Brennan, L. Biochemical and metabolomic phenotyping in the identification of a vitamin D responsive metabolotype for markers of the metabolic syndrome. *Mol. Nutr. Food Res.* **2011**, *55*, 679–690. [CrossRef]
49. Kamrath, C.; Hartmann, M.F.; Pons-Kühnemann, J.; Wudy, S.A. Urinary GC–MS steroid metabolotyping in treated children with congenital adrenal hyperplasia. *Metabolism* **2020**, *112*, 154354. [CrossRef]
50. Li, K.; Brennan, L.; McNulty, B.A.; Bloomfield, J.F.; Duff, D.J.; Devlin, N.F.C.; Gibney, M.J.; Flynn, A.; Walton, J.; Nugent, A.P. Plasma fatty acid patterns reflect dietary habits and metabolic health: A cross-sectional study. *Mol. Nutr. Food Res.* **2016**, *60*, 2043–2052. [CrossRef]
51. Prendiville, O.; Walton, J.; Flynn, A.; Nugent, A.P.; McNulty, B.A.; Brennan, L. Classifying Individuals Into a Dietary Pattern Based on Metabolomic Data. *Mol. Nutr. Food Res.* **2021**, *65*, 2001183. [CrossRef]
52. Sujana, C.; Seissler, J.; Jordan, J.; Rathmann, W.; Koenig, W. Associations of cardiac stress biomarkers with incident type 2 diabetes and changes in glucose metabolism: KORA F4/FF4 study. *Cardiovasc. Diabetol.* **2020**, *19*, 1–12. [CrossRef] [PubMed]
53. Fiamoncini, J.; Rundle, M.; Gibbons, H.; Thomas, E.L.; Geillinger-Kästle, K.; Bunzel, D.; Trezzi, J.-P.; Kiselova-Kaneva, Y.; Wopereis, S.; Wahrheit, J.; et al. Plasma metabolome analysis identifies distinct human metabolotypes in the postprandial state with different susceptibility to weight loss-mediated metabolic improvements. *FASEB J.* **2018**, *32*, 5447–5458. [CrossRef] [PubMed]

54. Dahal, C.; Wawro, N.; Meisinger, C.; Brandl, B.; Skurk, T.; Volkert, D.; Hauner, H.; Linseisen, J. Evaluation of the metabolotype concept after intervention with oral glucose tolerance test and dietary fiber-enriched food: An enable study. *Nutr. Metab. Cardiovasc. Dis.* **2022**. [[CrossRef](#)] [[PubMed](#)]
55. O'donovan, C.B.; Walsh, M.C.; Woolhead, C.; Forster, H.; Celis-Morales, C.; Fallaize, R.; Macready, A.L.; Marsaux, C.F.M.; Navas-Carretero, S.; San-Cristobal, S.R.; et al. Metabotyping for the development of tailored dietary advice solutions in a European population: The Food4Me study. *Br. J. Nutr.* **2017**, *118*. [[CrossRef](#)] [[PubMed](#)]
56. Rogers, H.L.; Fernández, S.N.; Hernando, S.P.; Sanchez, A.; Martos, C.; Moreno, M.; Gonzalo, G. "My Patients Asked Me if I Owned a Fruit Stand in Town or Something." Barriers and Facilitators of Personalized Dietary Advice Implemented in a Primary Care Setting. *J. Pers. Med.* **2021**, *11*, 747. [[CrossRef](#)]
57. Brotons, C.; Björkelund, C.; Bulc, M.; Ciurana, R.; Godycki-Cwirko, M.; Jurgova, E.; Kloppe, P.; Lionis, C.; Mierzecki, A.; Piñeiro, R.; et al. Prevention and health promotion in clinical practice: The views of general practitioners in Europe. *Prev. Med.* **2005**, *40*, 595–601. [[CrossRef](#)]
58. Morris, C.; O'Grada, C.; Ryan, M.; Roche, H.M.; Gibney, M.J.; Gibney, E.R.; Brennan, L. Identification of Differential Responses to an Oral Glucose Tolerance Test in Healthy Adults. Federici M, editor. *PLoS ONE* **2013**, *8*, e72890. [[CrossRef](#)]
59. Frazier-Wood, A.C.; Glasser, S.; Garvey, W.T.; Kabagambe, E.K.; Borecki, I.B.; Tiwari, H.K.; Tsai, M.Y.; Hopkins, P.N.; Ordovas, J.M.; Arnett, D.K. A clustering analysis of lipoprotein diameters in the metabolic syndrome. *Lipids Health Dis.* **2011**, *10*. [[CrossRef](#)]
60. Chua, E.C.P.; Shui, G.; Lee, I.T.G.; Lau, P.; Tan, L.C.; Yeo, S.C.; Lam, B.D.; Bulchand, S.; Summers, S.A.; Puvanendran, K.; et al. Extensive diversity in circadian regulation of plasma lipids and evidence for different circadian metabolic phenotypes in humans. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 14468–14473. [[CrossRef](#)]
61. Wilcox, M.A.; Wyszynski, D.F.; Panhuysen, C.I.; Ma, Q.; Yip, A.; Farrell, J.; Farrer, L.A. Empirically derived phenotypic subgroups—qualitative and quantitative trait analyses. *BMC Genet.* **2003**, *4* (Suppl. 1). [[CrossRef](#)]
62. Muniandy, M.; Velagapudi, V.; Hakkarainen, A.; Lundbom, J.; Lundbom, N.; Rissanen, A.; Kaprio, J.; Pietiläinen, K.H.; Ollikainen, M. Plasma metabolites reveal distinct profiles associating with different metabolic risk factors in monozygotic twin pairs. *Int. J. Obes.* **2019**, *43*, 487–502. [[CrossRef](#)] [[PubMed](#)]
63. Urpi-Sarda, M.; Almanza-Aguilera, E.; Llorach, R.; Vázquez-Fresno, R.; Estruch, R.; Corella, D.; Sorli, J.V.; Carmona, F.; Sanchez-Pla, A.; Salas-Salvadó, J.; et al. Non-targeted metabolomic biomarkers and metabolotypes of type 2 diabetes: A cross-sectional study of PREDIMED trial participants. *Diabetes Metab.* **2019**, *45*, 167–174. [[CrossRef](#)] [[PubMed](#)]
64. Tzeng, C.R.; Chang, Y.C.I.; Chang, Y.C.; Wang, C.W.; Chen, C.H.; Hsu, M.I. Cluster analysis of cardiovascular and metabolic risk factors in women of reproductive age. *Fertil Steril.* **2014**, *101*, 1404–1414. [[CrossRef](#)] [[PubMed](#)]