Metabotyping and its application in targeted nutrition: an overview

Q1 Anna Riedl[1,2]*, Christian Gieger[1,2,3], Hans Hauner[4,5,6,7], Hannelore Daniel[7] and Jakob Linseisen[1,2,5,8]

[1]*Helmholtz Zentrum München, German Research Center for Environmental Health (GmbH), Institute of Epidemiology II, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany*

[2]*German Center for Diabetes Research (DZD e.V.), Ingolstädter Landstr. 1, 85764 Neuherberg, Germany*

[3]*Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health (GmbH), Ingolstädter Landstr. 1, 85764 Neuherberg, Germany*

[4]*Else Kröner-Fresenius Centre for Nutritional Medicine, Technical University Munich, Gregor-Mendel-Str. 2, 85354 Freising-Weihenstephan, Germany*

[5]*ZIEL – Institute for Food and Health, Technical University of Munich, Weihenstephaner Berg 1, 85354 Freising, Germany*

[6]*Klinikum rechts der Isar, Institute of Nutritional Medicine, Technical University of Munich, Uptown München Campus D, Georg-Brauchle-Ring 60/62, 80992 Munich, Germany*

[7]*Technical University of Munich, Gregor-Mendel-Str. 2, 85354 Freising-Weihenstephan, Germany*

Q2 [8]*LMU München, UNIKA-T, Neusässer Str. 47, 86156 Augsburg, Germany*

Abstract

Metabolic diversity leads to differences in nutrient requirements and responses to diet and medication between individuals. Using the concept of metabotyping – that is, grouping metabolically similar individuals – tailored and more efficient recommendations may be achieved. The aim of this study was to review the current literature on metabotyping and to explore its potential for better targeted dietary intervention in subjects with and without metabolic diseases. A comprehensive literature search was performed in PubMed, Google and Google Scholar to find relevant articles on metabotyping in humans including healthy individuals, population-based samples and patients with chronic metabolic diseases. A total of thirty-four research articles on human studies were identified, which established more homogeneous subgroups of individuals using statistical methods for analysing metabolic data. Differences between studies were found with respect to the samples/ populations studied, the clustering variables used, the statistical methods applied and the metabotypes defined. According to the number and type of the selected clustering variables, the definitions of metabotypes differed substantially; they ranged between general fasting metabotypes, more specific fasting parameter subgroups like plasma lipoprotein or fatty acid clusters and response groups to defined meal challenges or dietary interventions. This demonstrates that the term 'metabotype' has a subjective usage, calling for a formalised definition. In conclusion, this literature review shows that metabotyping can help identify subgroups of individuals responding differently to defined nutritional interventions. Targeted recommendations may be given at such metabotype group levels. Future studies should develop and validate definitions of generally valid metabotypes by exploiting the increasingly available metabolomics data sets.

Key words: Metabotypes: Metabotyping: Metabolic phenotypes: Targeted nutrition: *enable* Cluster

The human metabolome is influenced by genetic, transcriptional and post-transcriptional factors as well as by the gut microbiome and environmental factors like diet and other lifestyle determinants[1,2]. It is well known that individuals show large differences in their nutrient requirements and responses to diet and medication according to their metabolic characteristics[2–5]. Specific dietary recommendations or drug treatments for disease states should thus be tailored to optimise the benefit to the individual. Equally important, specific treatments should not be provided to individuals with only a minor response or a lack of positive response to the intervention. The concept of personalisation is supposed to be more effective with respect to individual benefit:risk ratio and health-care costs than currently used general dietary recommendations and standard treatments for chronic disease[3–8].

Such efforts have led to the concept of metabotyping or metabolic phenotyping, which describes the categorisation of individuals based on their metabolic or phenotypic characteristics into more homogeneous subgroups, the so-called metabotypes or metabolic phenotypes. This concept implies that individuals within a subgroup show a high metabolic similarity and those in different subgroups show a high dissimilarity. Metabotyping could, thus, allow the identification of subpopulations or specific patient groups responding differently to a defined dietary or medical intervention, promising better nutritional and medical treatment at the metabotype group level[6,9–13].

* **Corresponding author:** A. Riedl, fax +49 89 3187 2951, email anna.riedl@helmholtz-muenchen.de

The metabotyping approach has been used widely in healthy animals[14,15] as well as in rodent models of disease for testing drug effects[16,17]. On this basis, it was possible to separate strain-specific metabolic phenotypes or strain subtypes based on the plasma, urine or faecal metabolic profiles, thereby finding diagnostic and prognostic biomarker differences between groups[14–26]. Strain subtypes could be established by sex[19,23–25], age[22], diet[20,26] or diurnal time of sample collection[18,21,25].

Further, several human studies have been conducted to define specific metabotypes, but these studies used a variety of methods and inconsistent definitions, indicating that the term 'metabotype' is often used with quite a different meaning. In reviews on personalised nutrition, O'Donovan *et al.*[6] and Brennan[13] proposed the concept of metabotyping and provided examples of articles using the metabotyping approach.

The aim of this paper was to review the existing literature on metabotyping in human studies, to show its application in targeted nutrition and, thus, to provide recommendations for future studies in this field.

Methods

A comprehensive literature search was performed using PubMed, Google and Google Scholar up to May 2016. However, this is not a strictly systematic review as described, for example, by the Cochrane Collaboration[27] because of many open questions. The first search strategy addressed the definition of metabotypes in healthy individuals or population-based samples to find evidence for differences in metabolism and corresponding subgroups. The second search was conducted on the definition of metabotypes in patients with chronic diet-related metabolic diseases (obesity, metabolic syndrome, diabetes, dyslipidaemia, hyperlipidaemia, hyperuricemia, gout and hypertension) for diagnosing or establishing metabolically homogeneous patient subgroups.

Different combinations of the following keywords were used to search for studies that performed metabotyping in healthy subjects or in population-based samples: 'metabotype', 'metabolic phenotype', 'metabolomic phenotype', 'molecular phenotype', 'clinical phenotype', 'biochemical phenotype', 'metabolic profile', 'metabolomic profile', 'metabolic pattern', 'nutritional phenotype', 'nutritype', 'metabolome', 'metabolomics', 'metabolism' or 'metabolic response' and 'cluster', 'pattern', 'subgroup', 'subtype', 'cluster analysis' or 'principal component analysis'. In addition, an extended search was conducted on this topic including information on underlying causes for differences in metabolism between individuals, namely with regard to genetics, epigenetics, transcriptomics or the microbiome[5]. To this end, the search terms 'genetics', 'genotype', 'SNP', 'epigenetics', 'transcriptomics', 'gut microbiota' or 'enterotype' were added to the search strategy mentioned above.

The literature search concerning the definition of metabotypes in patients was restricted to frequent chronic metabolic diseases with a strong relation to diet. This selection was based on the worldwide growing prevalence of diet-related metabolic diseases such as obesity and type 2 diabetes, on the one hand, and on the fact that, besides tailored medical treatments, targeted dietary intervention could also have an important effect on diet-related diseases, on the other[28]. Thus, in addition to the keywords mentioned above concerning the definition of metabotypes in healthy subjects or population-based samples, the following search terms referring to common metabolic diseases were included in the search strategy: 'obesity', 'adiposity', 'metabolic syndrome', 'diabetes', 'dyslipidaemia', 'hyperlipidaemia', 'hyperuricemia', 'gout' or 'hypertension'. Again, extended searches with keywords addressing underlying causes of metabolic differences were performed.

Relevant articles were selected by first checking titles and abstracts and subsequently the full text of the search results in accordance with the inclusion criteria. Additional studies were identified through supplementary screening of the reference lists of all articles analysed.

The following inclusion and exclusion criteria were used in the literature search: original research articles in English language on human studies, which established homogeneous groups of individuals using statistical analyses based on metabolic data from the body fluids blood and urine. Studies using exclusively other information like genetic, epigenetic, transcriptomic, microbiome, anthropometric or lifestyle data for group establishment were excluded, except in combination with metabolic and/or metabolomics data. In addition, studies in which metabotyping was based only on the combination of simple cut-off points of metabolic variables instead of on statistical analyses, as in the definition of the metabolic syndrome, were not included in this review. In general, all types of study designs were accepted and there were no restrictions on sample size. However, the study populations were limited to healthy subjects or population-based samples in the first search and – for the definition of patient subgroups – to individuals affected by common chronic metabolic diseases in the second search. Extreme or rare chronic diet-related metabolic diseases were not included.

Results

In total, thirty-four articles met the inclusion criteria, of which twenty-five articles were related to the definition of metabotypes in healthy subjects or population-based samples, and nine articles were related to the definition of patient subgroups with common metabolic diseases revealed by metabotyping.

Definition of metabotypes in healthy subjects or population-based samples

Tables 1 and 2 summarise the key features of the twenty-five articles identified according to the definition of metabotypes in healthy subjects or population-based samples. Table 1 gives an overview of twenty articles defining metabotypes based on fasting data. Table 2 shows an additional five articles defining metabotypes on the basis of metabolic response data for different dietary interventions. Both tables present the respective study objectives, designs and samples, the variables for clustering and their preprocessing, the clustering methods used and their validation as well as the main findings. With the exception of four articles[36,41,42,45], the studies were published within the past decade. The studies were conducted mainly, but

Table 1. Definition of metabotypes based on metabolic data in the fasting state

References	Objective	Study design and study sample	Variables for clustering	Preprocessing of variables	Clustering method	Validation of cluster solutions	Main findings
Van Bochove et al.[29]	Plasma lipoprotein clusters	Genetics of Lipid-Lowering Drugs and Diet Network (GOLDN) study (n 775) in the USA	NMR plasma lipoprotein profiles of ten particles: three VLDL (large, medium and small), four LDL (IDL, large, medium small and very small) and three HDL (large, medium and small) particles	Normalisation by standard deviation	k-Means cluster analysis (squared Euclidean distance)	Well-differentiated lipoprotein profiles by discriminatory variables (t test) Stability of cluster results (500 replications of clustering to select the result with the lowest total sum of distances) Biologically meaningful groups (Particle Profiler model)	Three distinct subgroups with differences in lipid characteristics (low, medium and high degree of dyslipidaemia) and in Prevalence of cardiovascular risk factors Positive lipid response of two subgroups (medium and high degree of dyslipidaemia) to fenofibrate therapy; the resulting group is larger than groups based on standard cut-off points for TAG and HDL
O'Sullivan et al.[30]	Metabolic phenotypes	Intervention study (n 135 healthy subjects) of participants aged 18–63 years in Ireland	Thirteen blood ^{1}H NMR biochemical markers of the metabolic syndrome (leptin, resistin, adiponectin, IL-6, CRP, TNF-α, insulin, C-peptide, cholesterol, TAG, NEFA, glucose, HOMA) and 25 (OH)D concentrations	z-Standardisation	k-Means cluster analysis (Euclidean distance)	Well-differentiated metabotypes by discriminatory variables (ANOVA, GLM analysis, Bonferroni post hoc multiple comparison test, PLS-DA with R^2, Q^2 and variable importance in the projection value) Stability of cluster results (ten iterations, 5-fold cross-validation) Biologically meaningful groups	Five subgroups with distinct biochemical profiles One subgroup with lower serum 25(OH)D and higher levels of adipokines and resistin (cluster 5) responsive to vitamin-D supplementation concerning markers of the metabolic syndrome
O'Donovan et al.[31]	Metabolic phenotypes	National Adult Nutrition Survey (NANS) (n 896 adults) aged 18–90 years in Ireland	Four routinely measured and widely applicable serum markers of metabolic health (TAG, total cholesterol, direct HDL-cholesterol and glucose)	z-Standardisation Outlier exclusion	Two-step cluster analysis with k-means cluster analysis	Well-differentiated groups by discriminatory variables (GLM analysis, Bonferroni post hoc test) Stability of cluster results (two-step cluster analysis) Biologically meaningful groups	Three distinct subgroups Identification of a risk cluster with high fasting levels of TAG, total cholesterol and glucose Development and validation of a decision tree based on biochemical characteristics, anthropometry and BP for personalised dietary advice per cluster
Vázquez-Fresno et al.[32]	Clinical phenotypes	Prospective, randomised, cross-over and controlled study (n 57 cardiovascular risk patients aged ≥55 years) in Spain	Sixty-nine biochemical (blood, urinary ^{1}H NMR) and anthropometric parameters	No preprocessing	k-Means cluster analysis (Euclidean distance)	Well-differentiated groups by discriminatory variables (ANOVA/Kruskal–Wallis test, Tukey's post hoc multiple comparison test/Mann–Whitney test, OSC-PLS-DA) Internal coherence (Dunn analysis), external homogeneity (Figure of Merit analysis) Stability of cluster results (1000 different random initialisations of clustering, 100 iterations, 7-fold internal cross-validation) Biologically meaningful groups	Four distinct subgroups Identification of the two most discriminant clusters 3 and 4 Different responses to red wine polyphenols of the two subgroups (cluster 3 and 4)
Frazier-Wood et al.[33]	Plasma lipoprotein clusters	Genetics of Lipid-Lowering Drugs and Diet (GOLDN) study (n 1036 aged 48·8 (sd 16·2) years) in the USA	Plasma lipoprotein diameters (VLDL, LDL, HDL) by NMR spectroscopy	Standardisation	Latent class analysis	Well-differentiated groups by discriminatory variables (mixed effects models) Stability of cluster results (good internal reliability) Biologically meaningful groups	Eight distinct subgroups with different plasma lipoprotein diameters Association of the subgroups with the metabolic syndrome
Zubair et al.[34]	Cardiometabolic risk patterns	Cebu Longitudinal Health and Nutrition Survey (CLHNS) (n 1768 women aged 36–69 years) in the Philippines	Eight cardiometabolic biomarkers (TAG, HDL, LDL, CRP, systolic and diastolic BP, HOMA-IR and glucose)	z-Standardisation	k-means cluster analysis (Euclidean distance)	Well-differentiated groups by discriminatory variables (multinomial logistic regression) Stability of cluster results (1000 iterations, different cluster numbers) Biologically meaningful groups	Five distinct subgroups of cardiometabolic risk: 'healthy', 'high BP', 'low HDL', 'insulin resistant' and 'high CRP'

Table 1. *Continued*

References	Objective	Study design and study sample	Variables for clustering	Preprocessing of variables	Clustering method	Validation of cluster solutions	Main findings
Zubair et al.[35]	Cardiometabolic risk patterns	Cebu Longitudinal Health and Nutrition Survey (CLHNS) (n 1621 individuals aged 21 (SD 0·0) years) in the Philippines	Eight cardiometabolic biomarkers (TAG, HDL, LDL, CRP, systolic and diastolic BP, HOMA-IR and glucose)	z-Standardisation	k-Means cluster analysis	Well-differentiated groups by discriminatory variables (multinomial logistic regression) Stability of cluster results (iterations, different cluster numbers) Biologically meaningful groups	Five distinct sex-specific subgroups of cardiometabolic risk: 'healthy/high HDL', 'healthy/low BP', 'high BP', 'insulin resistant/high TAG' and 'high CRP' Prediction of clusters by diet, adiposity and environment
Wilcox et al.[36]	Metabolic phenotypes	Framingham Heart Study (FHS) cohort (n 2885) in the USA	CVD risk factors	Categorisation of variables Data reduction by multiple-correspondence analysis	Two-staged clustering: k-means cluster analysis and hierarchical cluster analysis	Well-differentiated groups by discriminatory variables (probability of cluster membership by binary logistic regression, genome-wide linkage analyses) Stability of cluster results (iterations, two cluster analyses) Biologically meaningful groups	Four distinct subgroups: one healthy group, two groups with mild to moderately elevated lipid levels, and one group with strongly elevated lipid levels Assessment of heritability of traits
Wilcox et al.[37]	Metabolic phenotypes	Framingham Heart Study (FHS) offspring cohort (n 2760) in the USA	CVD risk factors	Categorisation of variables Data reduction by multiple-correspondence analysis	Two-staged clustering: k-means cluster analysis and hierarchical cluster analysis	Well-differentiated groups by discriminatory variables (probability of cluster membership by binary logistic regression, genome-wide association analyses) Stability of cluster results (iterations, two cluster analyses) Biologically meaningful groups	Five distinct subgroups: One group dropped because of missing data, two healthy groups, one group with features of the metabolic syndrome and one group with features of the metabolic syndrome and obesity Genetic associations, but loss of significance after stratification/ adjustments
Tzeng et al.[38]	Metabolic phenotypes	Study (n 573 women of reproductive age) in Taiwan	Ten cardiovascular and metabolic risk factors (systolic and diastolic BP, waist size, fasting insulin, fasting glucose, 2-h glucose, cholesterol, TAG, HDL and LDL)	No preprocessing	Hierarchical cluster analysis (Ward's method and within-group linkage)	Well-differentiated groups by discriminatory variables (χ^2 test, Fisher's exact test, ANOVA, one-way ANOVA *post hoc* range (Dunnett's) test) Stability of cluster results (two cluster analyses) Biologically meaningful groups	Two distinct subgroups (low- and high-risk group) Association between endocrine disturbances and increased risk for metabolic diseases
Li et al.[39]	Plasma fatty acid patterns	Irish National Adult Nutrition Survey (NANS) (n 1052 aged 42·9 (SD 16.5) years) in Ireland	Twenty-six plasma fatty acids	Log-transformation of skewed data Exclusion of outliers Standardisation (Subtraction of minimum and division by range)	k-Means cluster analysis (squared Euclidean distance)	Well-differentiated groups by discriminatory variables (GLM, χ^2 test, ANOVA, Bonferroni correction) Stability of cluster results (validation analysis, scree plot examination, two-step cluster analysis) Biologically meaningful groups	Four subgroups with distinct fatty acid profile Relationship between plasma fatty acid patterns, dietary intake and biomarkers of metabolic health The subgroup (cluster 3) higher in very-long-chain SFA and lower in α-linolenic acid was associated with metabolic health
Bermúdez et al.[40]	Selection of metabolically healthy and sick individuals for waist circumference cut-off point selection	Maracaibo City Metabolic Syndrome Prevalence Study (MMSPS) (n 1902 aged 38·70 (SD 15·06) years) in Venezuela	Eleven metabolic variables (mean arterial pressure, TAG, cholesterol, HDL, HOMA2-IR, HOMA2-βcell, HOMA2-S, fasting glucose, non-HDL-C cholesterol, TAG/HDL-C index and hs-CRP)	Log-transformation of skewed data Classification according to BMI before the two-step cluster analysis	Two-step cluster analysis: hierarchical (centroid-based) and k-means cluster analysis (Euclidean distance)	Well-differentiated groups by discriminatory variables (t test, ANOVA, cohesion, separation, silhouette coefficient) Stability of cluster results (training and validation data set with Cohen's kappa coefficient) Biologically meaningful groups	Six subgroups with distinct cardiometabolic profiles Most predictive variables (HOMA2-IR, HOMA2-βcell and TAG) Selection of a cut-off point for waist circumference (91 cm for women and 98 cm for men)
Micciolo[41]	Metabolic phenotypes	Patients of one general practice in Castel D'Azzano (n 458 aged 21–60 years) in Italy	Seven metabolic variables (glucose, uric acid, TAG, cholesterol, LDL and HDL (both total and percentage)) and BP levels or nine anthropometric characteristics (six skinfolds and three circumferences)	Log-transformation of skewed data Standardisation (subtraction of mean and division by standard deviation)	k-Means cluster analysis (separately on anthropometric and metabolic variables for each sex)	Well-differentiated groups by discriminatory variables (hierarchical algorithm for number of clusters, one-way ANOVA, χ^2 statistics) Stability of cluster results (five iterations, cross-classification of cluster results using correspondence analysis, γ coefficient and correlation coefficient) Biologically meaningful groups	Seven distinct subgroups for men and women, respectively Solution with anthropometric variables more stable than solution with metabolic variables Significantly different metabolic patterns with anthropometric and metabolic variables Associations between anthropometric characteristics and metabolic profiles

Table 1. *Continued*

References	Objective	Study design and study sample	Variables for clustering	Preprocessing of variables	Clustering method	Validation of cluster solutions	Main findings
Baumgartner et al.[42]	Cardiovascular risk factor groups	Cross-sectional study (n 317 individuals aged 18–88 years) in the USA	Cardiovascular risk factors (BP, plasma lipids, lipoprotein cholesterols and serum glucose)	Log-transformation of skewed data Standardisation (subtraction of mean and division by standard deviation)	k-Means cluster analysis (Euclidean distance)	Well-differentiated groups by discriminatory variables (PCA for number of clusters, one-way ANOVA, χ^2 test, discriminant analysis) Biologically meaningful groups	Four distinct subgroups for men and women, respectively Significant association of cluster membership with indices of adiposity but not with adipose tissue distribution
Huang et al.[43]	Metabolic phenotypes	West Australian Cohort (Raine) Study (n 1094 adolescents aged 14 years) in Australia	The Metabolic syndrome components (TAG, BMI, HOMA, systolic BP)	Log-transformation of skewed data	Two-step cluster analysis separately by sex (log-likelihood distance)	Well-differentiated groups by discriminatory variables (one-way ANOVA) Biologically meaningful groups	Two distinct subgroups (high-risk and low-risk cluster of cardiovascular and metabolic disorders) Relationships between inflammatory markers and components of a metabolic syndrome cluster
Andreeva-Gateva et al.[44]	Metabolic phenotypes	Cross-sectional study (n 113 subjects aged 21–70 years with an increased risk for type 2 diabetes) in Bulgaria	Components of the metabolic syndrome: anthropomorphic measurements, lipid and carbohydrate parameters (during oral glucose-tolerance test), insulin, C-peptide, creatinine, CRP, liver tests, β-cell function assessment, insulin sensitivity and insulin resistance	z-Standardisation	Hierarchical cluster analysis (squared Euclidean distance, Ward's method)	Well-differentiated groups by discriminatory variables (test statistics) Stability of cluster results (PCA with Varimax-normalised rotation for latent factor identification) Biologically meaningful groups	Two distinct subgroups Association of clusters with different patterns and stages of cardiovascular risk → diversity of metabolic disorders in subjects with an increased risk for type 2 diabetes
Ventura et al.[45]	Risk profiles for the metabolic syndrome	Longitudinal study (non-clinical sample of n 154 adolescent girls aged 13 years) in the USA	Six metabolic syndrome factors (systolic and diastolic BP, HDL, TAG, waist circumference and blood glucose)	Standardisation	Mixture model (or latent profile analysis)	Well-differentiated groups by discriminatory variables (GLM, ANOVA, Fisher's least significant difference comparison, χ^2 test, Fisher's exact test) Stability of cluster results (AIC, BIC, multiple iterations, different cluster numbers) Biologically meaningful groups	Four distinct subgroups of risk profiles for the metabolic syndrome Differences in developmental, lifestyle and family history variables between the subgroups
Bucci et al.[46]	Cardiovascular risk phenotypes	Data sets from France of the Pole Cardiovasculaire Hopital Europeen Georges Pompidou (n 618) and from Uruguay (n 123)	Five clinical variables (age, systolic and diastolic BP, LDL and HDL)	No preprocessing	k-Means cluster analysis (Euclidean distance)	Validation using Framingham index Well-differentiated groups by discriminatory variables (t test) Stability of cluster results (iterations, silhouette index) Biologically meaningful groups	Two distinct subgroups in the data sets of France and Uruguay, respectively Association of clusters with cardiovascular risk patterns
Moazzami et al.[47]	Metabolic phenotypes	Randomised, controlled, cross-over meal study (n 19 postmenopausal women aged 61 (SD 4·8) years) in Finland	189 metabolites from LC-MS metabolomics analysis (twenty-one amino acids, seventeen biogenic amines, forty-seven acyl-carnitines, thirty-eight phosphatidylcholines, thirty-nine acyl-alkyl phosphatidylcholines, fourteen lysophosphatidylcholines, fifteen sphingomyelins and one hexose)	No preprocessing	Hierarchical cluster analysis, O-PLS and PCA	Well-differentiated groups by discriminatory variables (O-PLS-DA, GLM, ANOVA) Stability of cluster results (three cluster analyses, cross-validated ANOVA, constant over three different sampling days) Biologically meaningful groups	Two distinct subgroups Different postprandial metabolic responses to breads (refined wheat, whole-meal rye and refined rye breads) → identification of individuals with reduced insulin sensitivity Different metabolic responses after consumption of different breads
Qureshi et al.[48] (only abstract of a presentation available)	Metabolic phenotypes	Insulin Resistance Atherosclerosis Study (n 500 individuals) in the USA	Ninety-three serum metabolites from liquid chromatography-MS analysis	–	Hierarchical cluster analysis and PCA	Well-differentiated groups by discriminatory variables (test statistics) Stability of cluster results (different cluster numbers) Biologically meaningful groups	133 individuals developed incident hypertension Identification of a cluster (n 154) with high risk for incident hypertension Identification of metabolites associated with a high risk for incident hypertension

IDL, intermediate-density lipoprotein; hs-CRP, high-sensitivity C-reactive protein; GLM, general linear model; HOMA-IR, homoeostasis model assessment of insulin resistance; OSC-PLS-DA, orthogonal signal-correction partial least squares discriminant analysis; BP, blood pressure; HOMA2-S, homoeostasis model assessment of insulin sensitivity; PCA, principal component analysis; AIC, Akaike information criterion; BIC, Bayesian information criterion.

Table 2. Definition of metabotypes based on metabolic response data to interventions

References	Objective	Study design and study sample	Variables for clustering	Preprocessing of variables	Clustering method	Validation of cluster solutions	Main findings
Morris et al.[9]	Response groups to an oGTT	Metabolic Challenge (MECHE) study (n 116 healthy adults aged 18–60 years) in Ireland	Response curves of blood glucose to oGTT (blood glucose measured during the oGTT at 0, 10, 20, 30, 60, 90 and 120 min)	No preprocessing	Mixed-model clustering	Well-differentiated response groups by discriminatory variables (ANOVA, GLM, Bonferroni post hoc multiple comparison test) Stability of cluster results (oral lipid-tolerance test) Biologically meaningful groups	Four distinct subgroups with different responses to oGTT One subgroup (cluster 1) as 'at risk' phenotype having the highest BMI, TAG, hs-CRP, C-peptide, insulin and HOMA-IR score and lowest VO_{2max}
Krishnan et al.[49]	Response groups to meal challenges with different glycaemic indices	Cross-over study (n 24 healthy premenopausal women aged 20–50 years) in the USA	Blood glucose, insulin and leptin	Range-scaling	PCA	Well-differentiated response groups by discriminatory variables (ANOVA, Tukey's post hoc test, Bonferroni post hoc multiple comparison test) Biologically meaningful groups	Three distinct subgroups with different responses to meal challenges One subgroup with higher insulin resistance and another subgroup with higher leptin values
Wang et al.[50]	Response groups to dietary carotenoids in watermelon juice and tomato juice	Cross-over study (n 23 healthy subjects) in the USA	Temporal response of individual plasma carotenoids (β-carotene, lycopene, phytoene and phytofluene)	Normalisation to baseline values	k-Means cluster analysis	Well-differentiated response groups by discriminatory variables (t test) Biologically meaningful groups	Five distinct subgroups with different plasma responses to dietary carotenoids → Identification of strong and weak responders Response differences between individual carotenoids and between interventions Association of response with genetic variants of carotenoid-metabolising enzyme
Bouwman et al.[51]	Response groups to a 5-week dietary intervention with anti-inflammatory ingredients	Controlled cross-over study (n 33 men) in the Netherlands	145 metabolites, seventy-nine proteins and 10 812 transcripts	Selection of significantly changed plasma parameters due to the intervention Normalisation (subtraction of the mean and division by the distance between mean scores of intervention and placebo group)	Hierarchical cluster analysis (Euclidean distance, group average linkage)	Well-differentiated groups by discriminatory variables (PLS-DA, ANOVA) Stability of cluster results (double cross-validation of PLS-DA) Biologically meaningful groups	Two distinct subgroups of inter-individual responses to intervention → Difference in metabolic stress profile, inflammatory and oxidative response Effects of the nutritional intervention on oxidative stress, inflammation, and metabolism → Differentiation between treated and untreated individuals
Chua et al.[52]	Circadian metabolic phenotypes	Study (n 20 ethnic-Chinese male aged 21–28 years) in Singapore	Time course of 263 plasma lipids	Iterative feature selection Elimination of linear trends of time courses z-Standardisation	k-Means cluster analysis and hierarchical cluster analysis	Well-differentiated groups by discriminatory variables (ANOVA, Kruskal–Wallis test, Bayes method) Stability of cluster results (consensus clustering: 1000 iterations of k-means cluster analysis, two cluster methods) Biologically meaningful groups	Three distinct subgroups 13 % of lipids showed circadian variation Diversity in circadian regulation of plasma lipids, (glucose and insulin)

oGTT, oral glucose-tolerance test; GLM, general linear model; hs-CRP, high-sensitivity C-reactive protein; HOMA-IR, homoeostasis model assessment of insulin resistance; PCA, principal component analysis; PLS-DA, partial least squares discriminant analysis.

not exclusively, in Europe and the USA, either with population-based samples or random samples of healthy individuals. The sample size of the studies varied considerably from twenty to up to 3000 participants. Also, the age range of the study populations differed across the studies with a main focus on adults. Regarding sex, two studies investigated only men[51,52], five studies only women[34,38,45,47,49] and all other studies included both sexes.

For the identification of metabotypes, different numbers of clustering variables were used. Besides the use of full ^{1}H NMR spectra or metabolomics data in some studies[32,47,48,51,52], all other studies used selected metabolites for clustering similar components of the metabolic syndrome[43,45] or cardiovascular risk factors[36,37,42]. The type of the cluster variables differed between the studies using blood or urine metabolites, diverse metabolite classes or specifically selected individual metabolite subclasses like lipoproteins or fatty acids and those using fasting metabolites (Table 1) or metabolic responses to dietary interventions (Table 2). According to the number and type of the selected clustering variables, the definitions of metabotypes differed considerably; they ranged between general fasting metabotypes, more specific fasting parameter subgroups like plasma lipoprotein[29,33] or fatty acid clusters[39] and response groups to defined meal challenges or dietary interventions. However, in most studies, at least some standard clinical markers such as glucose, TAG and cholesterol were included. Besides metabolic data, the inclusion of additional phenotypic factors for the definition of metabotypes was implemented in some studies: for example, the consideration of anthropometric parameters like BMI or waist circumference[32,36–38,41,43–45] and blood pressure[34–38,40–43,45,46]. However, only the study by Bouwman et al.[51] also assessed some underlying causes for differences in metabolism between subpopulations in the clustering process using transcriptomics data.

Before grouping individuals into metabotypes, diverse preprocessing steps were applied in the studies analysed to the cluster variables such as outlier exclusion, log-transformation of skewed data, dimension reduction (e.g. by multiple-correspondence analysis) and standardisation (e.g. range-scaling or z-standardisation). Different unsupervised learning methods were used in the studies to define relatively homogeneous metabolic groups of individuals. These included k-means cluster analysis, hierarchical clustering and combinations of the two, principal component analysis (PCA), latent class analysis[33] and mixed-model clustering[9,45]. Then, supervised learning methods, such as partial least squares regression as well as statistical tests like the t test and ANOVA, were used to find discriminatory variables between the established groups. Clustering indices, cross-validation procedures, repetitions with different cluster seeds and cluster numbers as well as different clustering methods were applied to validate the clustering results. Biologically meaningful metabotypes, which were differentiated using discriminatory variables, also confirmed the clustering results. Using the clustering methods, different numbers of metabotypes were found, ranging between two and eight groups. Some studies identified subgroups of individuals with differential response to nutritional interventions; others only described differences between the subgroups, mainly in the fasting state.

The following two studies are examples for the establishment of metabotypes using metabolite profiles obtained in the fasting state and the subsequent investigation of differences in response to dietary interventions between the subgroups. O'Sullivan et al.[30] described metabotypes in an Irish intervention study with 135 healthy individuals aged 18–63 years. After z-standardisation, thirteen blood ^{1}H NMR biochemical markers of the metabolic syndrome and serum vitamin-D levels were used in a k-means cluster analysis. Five distinct biologically meaningful clusters were found. Among these, one group with lower serum vitamin-D levels and higher levels of adipokines showed a positive response to vitamin-D supplementation on parameters of the metabolic syndrome. The stability of the cluster result was verified using a 5-fold cross-validation method. Second, Vázquez-Fresno et al.[32] investigated fifty-seven subjects at a high cardiovascular risk aged ≥55 years in a randomised and controlled cross-over study. k-Means cluster analysis revealed four well-differentiated and biologically meaningful clusters using sixty-nine blood and urine ^{1}H NMR biochemical markers and anthropometric variables identifying red wine polyphenol-responsive metabotypes. In addition to cross-validation, cluster indices like Dunn analysis and Figure of Merit analysis were used.

An example for the definition of metabotypes based on metabolic response data to a dietary intervention is the Irish Metabolic Challenge (MECHE) study, which included 116 participants aged 18–60 years[9]. Mixed-model clustering of blood glucose curves revealed four distinct metabotypes with different responses to an oral glucose-tolerance test, of which one group was identified as a high-risk phenotype. The stability of the differentiated clusters was confirmed by another intervention, an oral lipid-tolerance test. Wang et al.[50] described metabotypes in a dietary intervention with carotenoid-rich beverages in a cross-over design based on twenty-three healthy subjects in the USA. In each carotenoid arm, the responses to all plasma carotenoids were analysed individually. k-Means cluster analysis revealed five distinct subgroups with different temporal responses. Subsequently, strong and weak responders to individual dietary carotenoids were identified. The different responses were induced by genetic variants of the carotenoid-metabolising enzyme β-carotene 15,15'-monooxygenase 1.

Definition of patient subgroups with metabolic diseases by metabotyping

Table 3 presents nine publications that were selected during the literature search on the definition of metabotypes in patients with chronic diet-related metabolic diseases for diagnosing or establishing metabolically homogeneous patient subgroups. All articles were published within the last 10 years and, again, a majority of the studies were performed in Europe and the USA with differences in study design, sample size (between fifty and 50 000 participants) and the age range of adults. Both sexes were considered in all studies. The articles describe the diagnosis and subgrouping of patients affected by diabetes, obesity, the metabolic syndrome or dyslipidaemia. Here, again, the definitions of patient subgroups varied according to the use of different numbers of metabolic clustering variables. In addition, the types

Table 3. Definition of patient subgroups with metabolic diseases by metabotyping

References	Objective	Study design and study sample	Variables for clustering	Preprocessing of variables	Clustering method	Validation of cluster solutions	Main findings
Zák et al.[53]	Diagnosis and identification of distinct phenotypes of the metabolic syndrome	Study (n 354 individuals (166 patients with the metabolic syndrome and 188 controls)) in the Czech Republic	Initially twenty-two but reduced to six plasma fatty acids in plasma phosphatidylcholine (dihomo-γ-linolenic, stearic, myristic, DHA, DPA and linoleic acids)	Examination of extreme values Power transformation for symmetry and constant variance Variable reduction by linear discriminant analysis with forward variable selection using Wilk's λ criterion	Hierarchical cluster analysis (Ward's method with Euclidean distance)	Well-differentiated individuals by discriminatory metabolites (t test, Wilcoxon's test, Benjamin–Hochberg correction, ANCOVA adjustments) Biologically meaningful groups	Diagnosis of the metabolic syndrome Two distinct subgroups of the metabolic syndrome with differences in concentrations of glucose, NEFA, HOMA-IR and conjugated dienes in LDL
Schader[54]	Subtypes of type 2 diabetes	GWAS (Framingham Heart Study (FRAM), MESA SHARe Study (MESA), Atherosclerosis Risk in Communities study (ARIC)) (13 459 study participants aged 30–84 years (832 cases during follow-up for clustering and 12 066 controls) in the USA	Ten metabolic and anthropometric characteristics before diagnosis of type 2 diabetes (sex, BMI, waist: hip ratio, TAG, HDL, glucose, insulin, cholesterol, systolic BP and diastolic BP)	Standardisation	k-Means cluster analysis (Euclidean distance)	Well-differentiated individuals by discriminatory metabolites (t test, Cox proportional hazards model) Stability of cluster results (Calinski method, twenty-five iterations) Biologically meaningful groups	Two distinct subtypes No statistical significant differences in genetic risk factors between the subtypes
Li et al.[55]	Subtypes of type 2 diabetes	Mount Sinai BioMe Biobank Program (n 11 210 individuals mean aged 55·5 years, of whom 2551 were patients with type 2 diabetes) in the USA	Seventy-three clinical data from high-dimensional electronic medical records	Feature selection (>50 % of patients with non-missing values)	Topological analysis (cosine distance)	Well-differentiated individuals by discriminatory metabolites (t test, ANOVA, χ^2 test) Stability of cluster results (random training and test sets, stability and robustness statistics) Biologically meaningful groups	Three distinct subtypes characterised by increased diabetic nephropathy and retinopathy in subtype 1, cancer malignancy and CVD in subtype 2 and CVD, neurological diseases, allergies and HIV infections in subtype 3 Association of subtypes with specific SNP
Amato et al.[56]	Subtypes of type 2 diabetes	Cross-sectional study (n 96 patients with type 2 diabetes aged 62·40 (SD 6·36) years (range = 51–75 years)) in Italy	Three fasting serum incretins (GLP-1, GIP and ghrelin)	Log-transformation of skewed data	Two-step cluster analysis (preclustering and hierarchical methods, log-likelihood distance)	Well-differentiated individuals by discriminatory metabolites (t test, χ^2 test, Fisher's exact test) Stability of cluster results (silhouette coefficient) Biologically meaningful groups	Two distinct subgroups with higher levels of glycated Hb, glucagon, fasting glucose and lower levels of C-peptide in subgroup 1
Frei et al.[57]	Subtypes of obesity	Study (n 50 patients aged 21–61 years) in Brazil	Blood parameters before and after the surgery (BMI, LDL, HDL, VLDL, Hb, platelets, leucocytes, TAG, glucose and bilirubin)	z-standardisation	Hierarchical cluster analysis (Euclidean distance)	Well-differentiated individuals by discriminatory metabolites (ANOVA, Bonferroni test) Stability of cluster results (Calinski–Harabasz, silhouette index, different cluster algorithms (complete linkage, average linkage, Ward's method)) Biologically meaningful groups	Two distinct subtypes with differences in indicators of the metabolic syndrome (glucose, LDL, VLDL and TAG) Identification of patterns that hinder recovery after the bariatric surgery

Table 3. *Continued*

References	Objective	Study design and study sample	Variables for clustering	Preprocessing of variables	Clustering method	Validation of cluster solutions	Main findings
Arguelles *et al.*[(58)]	Subtypes of the metabolic syndrome	Hispanic Community Health Study/Study of Latinos (HCHS/SOL) (*n* 15 825 Hispanics/ Latinos aged 18–74 years) in the USA	The Metabolic syndrome components (waist circumference, systolic and diastolic BP, HDL, TAG, glucose, medication use)	Log-transformation and multiplication with 100 skewed data	Latent class analysis separately by sex	Well-differentiated individuals by discriminatory metabolites (logistic regression) Stability of cluster results (different cluster numbers, AIC, BIC, ABIC, entropy and posterior probabilities) Biologically meaningful groups	Two distinct subgroups for men and women, respectively ('metabolic syndrome' cluster and 'non-metabolic syndrome' cluster) Association of subgroups with covariates and CVD No identification of additional subtypes of the metabolic syndrome
Kim *et al.*[(59)]	Subtypes of prediabetes	Large Cohort (*n* 52 139 adult Mayo Clinic patients) in the USA	Diagnoses (obesity, hyperlipidaemia, hypertension, renal failure, various cardiovascular conditions), vital signs (BP, pulse), laboratory results (glucose, lipids), use of medication (aspirin, medication for hypertension and hypercholesterolaemia)	Binary transformation of variables	Bisecting divisive hierarchical cluster analysis	Well-differentiated individuals by discriminatory metabolites Biologically meaningful groups	A subgroup with higher and another subgroup with lower risk for diabetes than the general population Identification of twelve highest-risk groups (out of twenty-six clusters) and their relevant risk factors Use of clustering as a diabetes index outperforming the Framingham risk score
Mäkinen *et al.*[(60)]	Subtypes of type 1 diabetes	Finnish Diabetic Nephropathy (FinnDiane) Study (*n* 613 patients with type 1 diabetes) in Finland	Blood serum ^{1}H NMR spectrum	Several preprocessing steps of ^{1}H NMR spectra Adjustment of intensity units to equal variance	Self-organising map (9×9 hexagonal sheet of map units, Gaussian neighbourhood function))	Well-differentiated individuals by discriminatory metabolites Stability of cluster results (non-NMR measurements of a number of metabolites) Biologically meaningful groups	Six subgroups Different diabetic complications, clinical and metabolic characteristics between subgroups
Botelho *et al.*[(61)]	Subgroups of dyslipidaemia	Patient data bank at the Dante Pazzanese Institute of Cardiology (*n* 57 individuals aged 30–80 years with dyslipidaemia controlled by statins) in Brazil	Four plasma biomarkers of oxidative stress (malondialdehyde, ferric reducing ability power, 2,2-diphenyl-1-picrylhydrazyl radical and oxidised-LDL)	Dimension reduction by PCA	Hierarchical cluster analysis (Ward's method, Euclidean distance)	Well-differentiated individuals by discriminatory metabolites (ANOVA, Tukey's *post hoc* test) Biologically meaningful groups	Five distinct subgroups No difference in dietary pattern between the subgroups

HOMA-IR, homoeostasis model assessment of insulin resistance; GWAS, genome-wide association study; BP, blood pressure; GLP-1, glucagon-like peptide-1; GIP, glucose-dependent insulinotropic polypeptide; AIC, Akaike information criterion; BIC, Bayesian information criterion; ABIC, sample size-adjusted BIC; PCA, principal component analysis.

of clustering variables differed, often depending on the particular disease investigated. For example, Mäkinen *et al.*[60] used a full blood serum ^{1}H NMR spectrum for the subgrouping of patients with type 1 diabetes. In contrast, Arguelles *et al.*[58] tried to identify subgroups of the metabolic syndrome using only components of this syndrome (waist circumference, systolic and diastolic blood pressure, HDL, TAG, fasting glucose and medication use) for the clustering procedure. Few studies used additional variables such as anthropometry[54,57,58] or medication use[58,59] along with the metabolic information in the clustering process. As a result, the studies identified different patient subgroups depending on the metabolic data assessed. After the application of various preprocessing steps to the cluster variables as described above, clustering methods like *k*-means cluster analysis, hierarchical clustering and combinations of the two, topological analysis[55], latent class analysis[58] and self-organising maps[60] were applied. Discriminatory variables between the resulting disease subgroups were again identified using test statistics. Moreover, biological meaning, clustering indices, cross-validation procedures, repetitions with different cluster seeds and cluster numbers as well as different clustering algorithms were applied to validate the clustering results. Different numbers of disease subgroups were formed, mainly two to four groups.

An example for the establishment of type 2 diabetes subgroups is the study by Schader[54] using three studies in the USA with a total of 832 patients with type 2 diabetes aged 30–84 years. Applying *k*-means cluster analysis with ten standardised metabolic and anthropometric characteristics assessed before the diagnosis of type 2 diabetes, two subgroups of the disease were found. Despite the stability of the clustering results, measured using the Calinski method and twenty-five repetitions of the clustering method, and strong differentiation of individuals based on discriminatory variables, no statistically significant difference was found between the genetic risk factors among the subgroups. In a smaller sample size of ninety-six patients with type 2 diabetes, Amato *et al.*[56] used three fasting incretins in a two-step cluster analysis to identify two subgroups of this disease.

Discussion

This review analysed the literature on metabotyping of individuals in metabolic and nutrition research. In total, thirty-four studies were included in this analysis covering a wide range of populations and using various clustering variables and statistical methods to identify different numbers of metabotypes. Consequently, it is difficult to draw meaningful conclusions regarding the establishment of metabotypes based on these rather heterogeneous studies using different approaches in metabotyping. However, this paper includes all available human studies using metabotyping in healthy subjects, population-based samples and patients with chronic metabolic diseases, and thereby represents the current state of knowledge.

Differences in study populations

We found a considerable variation in metabotypes across the countries in which the studies were performed, and this could be due to different genetic characteristics, environmental influences (like dietary and cultural behaviour), risk factors and disease rates[5,62–64]. This variation was seen to be particularly large between Western countries and East Asian countries, whereas metabotypes across different Western countries displayed substantial overlapping[62,64]. As most studies we review here were conducted in Western populations in Europe and the USA, the defined metabotypes seem to be transferable and comparable between these studies. However, there is a lack of data as to whether these metabotypes can be transferred to other ethnic populations.

Comparing metabotypes between different age ranges may be hampered by the physiological ageing process itself, which is characterised by marked changes in metabolism or metabolic flexibility[65]. However, it was shown in some studies that the plasma metabotypes (metabolite profiles) of individuals remain relatively stable over a few years[66,67] and only large differences in age seem to be relevant. As many metabolites differ between men and women – for example, steroid hormones or branched chain amino acids[62,68,69] – studies need to consider sex differences. This could be achieved by the exclusion of these sex-specific variables from the clustering process or by separate analyses for men and women.

Differences in variables used for clustering

The use of diverse types and numbers of clustering variables does not allow a reasonable comparison of the metabotypes identified in different studies. At present, the debate on the most important criteria and variables to be used for the definition of a biologically meaningful metabotype remains open. Equally important, the aim of metabotype definition has to be defined *a priori*. In 2000, Gavaghan *et al.*[15] defined a metabotype as 'a probabilistic multiparametric description of an organism in a given physiological state based on analysis of its cell types, biofluids or tissues'. Later, metabotyping was described in several studies as the 'process of grouping similar individuals based on their metabolic or phenotypic characteristics[6,9–13]. These wide and general definitions of metabotypes allow the inclusion of all studies establishing subgroups based on (1) healthy or sick people (thus also in the diagnosis or subgrouping of patients), (2) the fasting state or response to interventions, (3) a few or a variety of metabolites and (4) specifically selected single metabolite subclasses like lipoproteins, diverse metabolite subclasses or the addition of other variables like underlying causes for differences in metabolism – for example, genetic, epigenetic or gut microbiome information.

The selection of variables plays an important role in the identification and separation of metabotypes. Grouping of individuals based on a few variables or single specific metabolite classes provides a restricted definition of metabotypes, as only a small part of human metabolism is taken into account. However, for the establishment of plasma lipoprotein clusters in the studies by van Bochove *et al.*[29] and Frazier-Wood *et al.*[33], or of plasma fatty acid patterns in the study by Li *et al.*[39], restriction to the respective lipid variables seemed to be sufficient for subclassification. Likewise, Wang *et al.*[50] considered only the plasma carotenoid levels after a dietary intervention with

carotenoids. The same was the case in the study by Morris *et al.*[9] considering only blood glucose levels, measured at several points in time, to identify groups with differential glucose responses to an oral glucose-tolerance test. This is of course in accordance with the current clinical practice for classification of type 2 diabetes based on the plasma kinetics of glucose. In diagnosing or subgrouping patients, the restriction of variables to disease-related parameters could also be sufficient for subclassification. For example, Arguelles *et al.*[58] established subgroups of the metabolic syndrome patients based on the standard criteria for disease description, namely waist circumference, systolic and diastolic blood pressure, HDL, TAG, fasting glucose and medication use. The grouping in other studies using plasma fatty acids for the description of the metabolic syndrome[53] and fasting incretins for the subgrouping of diabetes[56] could be probably refined by the consideration of additional disease-related variables.

There is no consensus yet on a uniform use of the term 'metabotype', thus it is subjectively applied, usually based on the respective study objectives. In this review, the definitions of metabotypes differed considerably; they ranged between general fasting metabotypes, more specific fasting parameter subgroups like plasma lipoprotein[29,33] or fatty acid clusters[39] and response groups to defined meal challenges or dietary interventions according to the number and type of the selected clustering variables. Although an accepted definition of metabotype seems attractive, there is also the view that there is no need for a strict metabotype definition. On the one hand, it may be argued that a metabotype has by its nature a wide definition and should not be restricted. On the other hand, a better comparability of studies could be achieved using a stricter definition. Even if a strict general definition appears implausible or unrealistic, more precise sub-definitions of metabotypes could be developed, for example for lipid and carbohydrate (glucose) metabolism. Thus, metabolic variables restricted to specific metabolic pathways like to those of lipoproteins may be sufficient depending on the respective study objective.

However, it is assumed that the inclusion of various metabolites originating from different pathways as well as additional information from anthropometry or that obtained by including genetics, epigenetics or the gut microbiome in the process of metabotyping provides a more precise characterisation of individuals and, thus, the establishment of more refined and generally valid metabotypes[70]. This can be achieved through the use of '-omics' data such as metabolomics, genomics and epigenomics, where research is growing rapidly[2,71,72]. Thus, it may be wise to suggest a stricter definition of generally valid metabotypes in healthy subjects or population-based samples by at least the use of variables originating from different metabolic pathways, preferably the use of targeted or untargeted metabolomics data.

Further, there is no agreement as to whether the definition of metabotypes should be based on fasting data (see Table 1) or rather on metabolic response data to interventions (see Table 2), for which we identified only five studies that met the inclusion criteria. An argument for the use of metabolic response data to interventions is the increase of variation between individuals as some metabolic differences are only visible through challenges and would remain undetected using fasting blood values[73]. However, the establishment of metabotypes by means of fasting data allows extensive measurements of larger study populations and is thus more feasible in the general population. It is important to note that intra-individual variations of metabolite concentrations may also occur because of diurnal time, stress, latent diseases as well as by measurement and storage conditions of the samples[5,64,74,75]. However, these differences were shown to be smaller than inter-individual differences, suggesting that individual metabotypes are relatively robust[76].

Differences in statistical analyses

As a variety of statistical methods are available for the establishment of metabotypes[70], there is an on-going discussion on which statistical methods should be used to obtain the best spread between subgroups. The preprocessing of variables is especially dependent on the structure of the variables and the requirements of the subsequent clustering methods. Thus, the implementation of outlier exclusion and data transformation has to be decided individually. If the number of clustering variables exceeds one per ten observations, application of data-reduction analyses like PCA or multiple-correspondence analysis must be considered to avoid over adjustment[77]. In many studies included in this review, standardisation has been applied to the cluster variables to avoid bias from different scales and units in the grouping analysis[78,79]. The most commonly used method is z-standardisation $\left(z = \frac{X - \text{mean}}{\text{SD}}\right)$.

Concerning the different clustering methods[78–82], k-means cluster analysis and hierarchical cluster analysis were applied most commonly. Each clustering method has its own advantages and disadvantages and must be selected depending on the characteristics of the respective data set (e.g. depending on the scale level or the sample size). k-Means cluster analysis seems to be more suitable for large data sets than hierarchical clustering. However, the number of clusters has to be specified in advance for k-means cluster analysis, whereas hierarchical clustering does not need the number of clusters to be determined[82]. In addition, there are novel clustering techniques available in the field of bioinformatics, for example the so-called machine learning methods[83].

The selection of validation criteria like statistical tests and clustering indices is also dependent on the structure of the data. The reproducibility of metabotypes should be tested in a validation data set to confirm the results and to prove their generalisability.

Differences in the main findings

The aim of most studies was to examine metabolic differences between the established metabotypes and to test associations with certain diseases. However, the application of metabotypes, especially the development of targeted interventions for responsive subgroups, is rather limited in the literature. In addition, intervention by supplementation may increase serum levels in all subgroups but with possibly either larger effects in some subgroups or attainment of a threshold concentration

considered to be within the normal range. Thus, responsiveness to an intervention does not necessarily mean benefit and, therefore, outcome parameters also need to be properly defined to evaluate the benefit of interventions, which so far has been rare in previous studies. Only few studies investigated the responsiveness of the established metabotypes to dietary interventions with regard to a specific disease. O'Sullivan et al.[30] identified a subgroup with a positive response to vitamin-D supplementation concerning the metabolic syndrome; Vázquez-Fresno et al.[32] detected a subgroup of patients at cardiovascular risk responsive to red wine polyphenols; and Moazzami et al.[47] identified individuals with reduced insulin sensitivity after consumption of bread. There is only one study that developed tailored dietary recommendations for subgroups using a decision-tree approach[31]. Until now, the established metabotypes have not been transferred to larger populations for specific, tailored interventions.

Conclusion

In conclusion, this literature review shows that metabotyping can help identify metabolically similar subpopulations or patient subgroups responding differently to defined nutritional interventions. Consequently, better tailored and, thus, more precise dietary recommendations than generalised advice may be provided to whole populations at a metabotype group level. The aim of future studies should be the refinement of the definition of generally valid metabotypes in large samples, especially with a possibly more precise phenotype description of individuals based on different '-omics' data, particularly metabolomics data. Another aim should be the development of stricter definitions of specific metabotypes for metabolic pathways. The metabotypes should then be tested for differential reactions to diverse dietary factors with regard to properly defined outcome parameters. On the basis of such results, populations can be better stratified in order to provide effective tailored prevention and intervention programs. The implementation of these recommendations in populations may become a future task. Finally, individual health benefits may be improved and the rising costs in the health-care system originating from obesity and other diet-related metabolic diseases may be better controlled.

Acknowledgements

References

1. Nicholson JK (2006) Global systems biology, personalized medicine and molecular epidemiology. *Mol Syst Biol* **2**, 52.
2. Beger RD, Dunn W, Schmidt MA, *et al.* (2016) Metabolomics enables precision medicine: 'A White Paper, Community Perspective'. *Metabolomics* **12**, 149.
3. de Roos B (2013) Personalised nutrition: ready for practice? *Proc Nutr Soc* **72**, 48–52.
4. Nicholson JK & Holmes E (2006) Global systems biology and personalized healthcare solutions. *Discov Med* **6**, 63–70.
5. Holmes E, Wilson ID & Nicholson JK (2008) Metabolic phenotyping in health and disease. *Cell* **134**, 714–717.
6. O'Donovan CB, Walsh MC, Gibney MJ, *et al.* (2016) Can metabotyping help deliver the promise of personalised nutrition? *Proc Nutr Soc* **75**, 106–114.
7. Celis-Morales C, Livingstone KM, Marsaux CF, *et al.* Effect of personalized nutrition on health-related behaviour change: evidence from the Food4me European randomized controlled trial. *Int J Epidemiol*, (Epublication ahead of print version 14 August 2016).
8. Livingstone KM, Celis-Morales C, Navas-Carretero S, *et al.* (2016) Effect of an Internet-based, personalized nutrition randomized trial on dietary changes associated with the Mediterranean diet: the Food4Me Study. *Am J Clin Nutr* **104**, 288–297.
9. Morris C, O'Grada C, Ryan M, *et al.* (2013) Identification of differential responses to an oral glucose tolerance test in healthy adults. *PLOS ONE* **8**, e72890.
10. Nicholson JK, Holmes E, Kinross JM, *et al.* (2012) Metabolic phenotyping in clinical and surgical environments. *Nature* **491**, 384–392.
11. Kaput J (2008) Nutrigenomics research for personalized nutrition and medicine. *Curr Opin Biotechnol* **19**, 110–120.
12. Brennan L (2008) Session 2: personalised nutrition. Metabolomic applications in nutritional research. *Proc Nutr Soc* **67**, 404–408.
13. Brennan L (2016) Use of metabotyping for optimal nutrition. *Curr Opin Biotechnol* **44**, 35–38.
14. Pontoizeau C, Fearnside JF, Navratil V, *et al.* (2011) Broad-ranging natural metabotype variation drives physiological plasticity in healthy control inbred rat strains. *J Proteome Res* **10**, 1675–1689.
15. Gavaghan CL, Holmes E, Lenz E, *et al.* (2000) An NMR-based metabonomic approach to investigate the biochemical consequences of genetic strain differences: application to the C57BL10J and Alpk:ApfCD mouse. *FEBS Lett* **484**, 169–174.
16. Clayton TA, Lindon JC, Cloarec O, *et al.* (2006) Pharmaco-metabonomic phenotyping and personalized drug treatment. *Nature* **440**, 1073–1077.
17. Holmes E, Nicholls AW, Lindon JC, *et al.* (2000) Chemometric models for toxicity classification based on NMR spectra of biofluids. *Chem Res Toxicol* **13**, 471–478.
18. Gavaghan CL, Wilson ID & Nicholson JK (2002) Physiological variation in metabolic phenotyping and functional genomic studies: use of orthogonal signal correction and PLS-DA. *FEBS Lett* **530**, 191–196.
19. Gavaghan McKee CL, Wilson ID & Nicholson JK (2006) Metabolic phenotyping of nude and normal (Alpk:ApfCD, C57BL10J) mice. *J Proteome Res* **5**, 378–384.
20. Fearnside JF, Dumas ME, Rothwell AR, *et al.* (2008) Phylometabonomic patterns of adaptation to high fat diet feeding in inbred mice. *PLoS ONE* **3**, e1668.
21. Bollard ME, Holmes E, Lindon JC, *et al.* (2001) Investigations into biochemical changes due to diurnal variation and estrus cycle in female rats using high-resolution (1)H NMR spectroscopy of urine and pattern recognition. *Anal Biochem* **295**, 194–202.

22. Calvani R, Brasili E, Praticò G, et al. (2014) Fecal and urinary NMR-based metabolomics unveil an aging signature in mice. *Exp Gerontol* **49**, 5–11.

23. Garcia-Perez I, Villasenor A, Wijeyesekera A, et al. (2012) Urinary metabolic phenotyping the slc26a6 (chloride-oxalate exchanger) null mouse model. *J Proteome Res* **11**, 4425–4435.

24. Qiao Q, Li T, Sun J, et al. (2011) Metabolomic analysis of normal (C57BL/6J, 129S1/SvImJ) mice by gas chromatography-mass spectrometry: detection of strain and gender differences. *Talanta* **85**, 718–724.

25. Plumb R, Granger J, Stumpf C, et al. (2003) Metabonomic analysis of mouse urine by liquid-chromatography-time of flight mass spectrometry (LC-TOFMS): detection of strain, diurnal and gender differences. *Analyst* **128**, 819–823.

26. Yang Y, Liu Y, Zheng L, et al. (2014) Serum metabonomic analysis of apoE(-/-) mice reveals progression axes for atherosclerosis based on NMR spectroscopy. *Mol Biosyst* **10**, 3170–3178.

27. JPT Higgins & Green S (editors) (2011) *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. ■: The Cochrane Collaboration. http://handbook.cochrane.org/ (accessed November 2016).

28. World Health Organization (2003) *Diet, Nutrition and the Prevention of Chronic Diseases. Joint WHO/FAO Expert Consultation. WHO Technical Report Series*, no. 916. Geneva: WHO.

29. van Bochove K, van Schalkwijk DB, Parnell LD, et al. (2012) Clustering by plasma lipoprotein profile reveals two distinct subgroups with positive lipid response to fenofibrate therapy. *PLOS ONE* **7**, e38072.

30. O'Sullivan A, Gibney MJ, Connor AO, et al. (2011) Biochemical and metabolomic phenotyping in the identification of a vitamin D responsive metabotype for markers of the metabolic syndrome. *Mol Nutr Food Res* **55**, 679–690.

31. O'Donovan CB, Walsh MC, Nugent AP, et al. (2015) Use of metabotyping for the delivery of personalised nutrition. *Mol Nutr Food Res* **59**, 377–385.

32. Vázquez-Fresno R, Llorach R, Perera A, et al. (2016) Clinical phenotype clustering in cardiovascular risk patients for the identification of responsive metabotypes after red wine polyphenol intake. *J Nutr Biochem* **28**, 114–120.

33. Frazier-Wood AC, Glasser S, Garvey WT, et al. (2011) A clustering analysis of lipoprotein diameters in the metabolic syndrome. *Lipids Health Dis* **10**, 237.

34. Zubair N, Kuzawa CW, McDade TW, et al. (2012) Cluster analysis reveals important determinants of cardiometabolic risk patterns in Filipino women. *Asia Pac J Clin Nutr* **21**, 271–281.

35. Zubair N, Kuzawa CW, Lee NR, et al. (2014) Clustering and determinants of cardiometabolic risk factors among Filipino young adults. *Asia Pac J Clin Nutr* **23**, 148–158.

36. Wilcox MA, Wyszynski DF, Panhuysen CI, et al. (2003) Empirically derived phenotypic subgroups – qualitative and quantitative trait analyses. *BMC Genet* **4**, Suppl. 1, S15.

37. Wilcox M, Li Q, Sun Y, et al. (2009) Genome-wide association study for empirically derived metabolic phenotypes in the Framingham Heart Study offspring cohort. *BMC Proc* **3**, Suppl. 7, S53.

38. Tzeng CR, Chang YC, Chang YC, et al. (2014) Cluster analysis of cardiovascular and metabolic risk factors in women of reproductive age. *Fertil Steril* **101**, 1404–1410.

39. Li K, Brennan L, McNulty BA, et al. (2016) Plasma fatty acid patterns reflect dietary habits and metabolic health: a cross-sectional study. *Mol Nutr Food Res* **60**, 2043–2052.

40. Bermúdez V, Rojas J, Salazar J, et al. (2015) Sensitivity and specificity improvement in abdominal obesity diagnosis using cluster analysis during waist circumference cut-off point selection. *J Diabetes Res* **2015**, 750265.

41. Micciolo R (1992) Clustering of hemodynamic and metabolic abnormalities and of anthropometric characteristics in men and women, aged 21–60 years. *Hum Biol* **64**, 539–566.

42. Baumgartner RN, Siervogel RM & Roche AF (1989) Clustering of cardiovascular risk factors in association with indices of adiposity and adipose tissue distribution in adults. *Am J Hum Biol* **1**, 43–52.

43. Huang RC, Mori TA, Burke V, et al. (2009) Synergy between adiposity, insulin resistance, metabolic risk factors, and inflammation in adolescents. *Diabetes Care* **32**, 695–701.

44. Andreeva-Gateva PA, Simeonov VD, Georrgieva-Nikolova RT, et al. (2014) Cluster analysis of the components of the metabolic syndrome in subjects with an increased risk for diabetes mellitus type 2. *Endocrinologya* **19**, 168–185.

45. Ventura AK, Loken E & Birch LL (2006) Risk profiles for metabolic syndrome in a nonclinical sample of adolescent girls. *Pediatrics* **118**, 2434–2442.

46. Bucci CM, Legnani WE & Armentano RL (2016) Clinical values dataset processing through cluster analysis to find cardiovascular risk. *J Phys Conf Ser* **705**, 012033.

47. Moazzami AA, Shrestha A, Morrison DA, et al. (2014) Metabolomics reveals differences in postprandial responses to breads and fasting metabolic characteristics associated with postprandial insulin demand in postmenopausal women. *J Nutr* **144**, 807–814.

48. Qureshi W, Wagenknecht L, Watkins S, et al. (2014) Abstract 23: metabolomic features associated with incident hypertension – results from Insulin Resistance Atherosclerosis Study. *Circulation* **129**, Suppl. 1, A23.

49. Krishnan S, Newman JW, Hembrooke TA, et al. (2012) Variation in metabolic responses to meal challenges differing in glycemic index in healthy women: Is it meaningful? *Nutr Metab (Lond)* **9**, 26.

50. Wang TT, Edwards AJ & Clevidence BA (2013) Strong and weak plasma response to dietary carotenoids identified by cluster analysis and linked to beta-carotene 15,15'-mono-oxygenase 1 single nucleotide polymorphisms. *J Nutr Biochem* **24**, 1538–1546.

51. Bouwman J, Vogels JT, Wopereis S, et al. (2012) Visualization and identification of health space, based on personalized molecular phenotype and treatment response to relevant underlying biological processes. *BMC Med Genomics* **5**, 1.

52. Chua EC, Shui G, Lee IT, et al. (2013) Extensive diversity in circadian regulation of plasma lipids and evidence for different circadian metabolic phenotypes in humans. *Proc Natl Acad Sci U S A* **110**, 14468–14473.

53. Zák A, Burda M, Vecka M, et al. (2014) Fatty acid composition indicates two types of metabolic syndrome independent of clinical and laboratory parameters. *Physiol Res* **63**, Suppl. 3, S375–S385.

54. Schader LM (2015) Comparison of genetic risk factors between two type II diabetes subtypes. Bachelor Thesis, University of Arizona.

55. Li L, Cheng WY, Glicksberg BS, et al. (2015) Identification of type 2 diabetes subgroups through topological analysis of patient similarity. *Sci Transl Med* **7**, 311ra174.

56. Amato MC, Pizzolanti G, Torregrossa V, et al. (2016) Phenotyping of type 2 diabetes mellitus at onset on the basis of fasting incretin tone: results of a two-step cluster analysis. *J Diabetes Investig* **7**, 219–225.

57. Frei F, Lessa Bde S, Nogueira JC, et al. (2013) Cluster analysis for classification of patients submitted to Fobi-Capella bariatric surgery. *Arq Bras Cir Dig* **26**, Suppl. 1, 33–38.

58. Arguelles W, Llabre MM, Sacco RL, et al. (2015) Characterization of metabolic syndrome among diverse Hispanics/Latinos living in the United States: latent class analysis from

the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). *Int J Cardiol* **184**, 373–379.

59. Kim E, Oh W, Pieczkiewicz DS, *et al.* (2014) Divisive hierarchical clustering towards identifying clinically significant pre-diabetes subpopulations. *AMIA Annu Symp Proc* **2014**, 1815–1824.

60. Mäkinen VP, Soininen P, Forsblom C, *et al.* (2008) 1H NMR metabonomics approach to the disease continuum of diabetic complications and premature death. *Mol Syst Biol* **4**, 167.

61. Botelho PB, Fioratti CO, Abdalla DS, *et al.* (2010) Classification of individuals with dyslipidaemia controlled by statins according to plasma biomarkers of oxidative stress using cluster analysis. *Br J Nutr* **103**, 256–265.

62. Holmes E, Loo RL, Stamler J, *et al.* (2008) Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* **453**, 396–400.

63. Lenz EM, Bright J, Wilson ID, *et al.* (2004) Metabonomics, dietary influences and cultural differences: a 1H NMR-based study of urine samples obtained from healthy British and Swedish subjects. *J Pharm Biomed Anal* **36**, 841–849.

64. Dumas ME, Maibaum EC, Teague C, *et al.* (2006) Assessment of analytical reproducibility of 1H NMR spectroscopy based metabonomics for large-scale epidemiological research: the INTERMAP Study. *Anal Chem* **78**, 2199–2208.

65. Chaleckis R, Murakami I, Takada J, *et al.* (2016) Individual variability in human blood metabolites identifies age-related differences. *Proc Natl Acad Sci U S A* **113**, 4252–4259.

66. Bernini P, Bertini I, Luchinat C, *et al.* (2009) Individual human phenotypes in metabolic space and time. *J Proteome Res* **8**, 4264–4271.

67. Yousri NA, Kastenmüller G, Gieger C, *et al.* (2014) Long term conservation of human metabolic phenotypes and link to heritability. *Metabolomics* **10**, 1005–1017.

68. Krumsiek J, Mittelstrass K, Do KT, *et al.* (2015) Gender-specific pathway differences in the human serum metabolome. *Metabolomics* **11**, 1815–1833.

69. Kochhar S, Jacobs DM, Ramadan Z, *et al.* (2006) Probing gender-specific metabolism differences in humans by nuclear magnetic resonance-based metabonomics. *Anal Biochem* **352**, 274–281.

70. Bartel J, Krumsiek J & Theis FJ (2013) Statistical methods for the analysis of high-throughput metabolomics data. *Comput Struct Biotechnol J* **4**, e201301009.

71. Valdes AM, Glass D & Spector TD (2013) Omics technologies and the study of human ageing. *Nat Rev Genet* **14**, 601–607.

72. Kurland IJ, Accili D, Burant C, *et al.* (2013) Application of combined omics platforms to accelerate biomedical discovery in diabesity. *Ann N Y Acad Sci* **1287**, 1–16.

73. Krug S, Kastenmüller G, Stückler F, *et al.* (2012) The dynamic range of the human metabolome revealed by challenges. *FASEB J* **26**, 2607–2619.

74. Rezzi S, Martin FP, Alonso C, *et al.* (2009) Metabotyping of biofluids reveals stress-based differences in gut permeability in healthy individuals. *J Proteome Res* **8**, 4799–4809.

75. Ghini V, Saccenti E, Tenori L, *et al.* (2015) Allostasis and resilience of the human individual metabolic phenotype. *J Proteome Res* **14**, 2951–2962.

76. Assfalg M, Bertini I, Colangiuli D, *et al.* (2008) Evidence of different metabolic phenotypes in humans. *Proc Natl Acad Sci U S A* **105**, 1420–1424.

77. Jain AK, Duin RPW & Mao J (2000) Statistical pattern recognition: a review. *IEEE Trans Pattern Anal Mach Intell* **22**, 4–37.

78. Everitt BS, Landau S, Leese M, *et al.* (2011) In *Cluster Analysis*, 5th ed. pp. ■–■ [DJ Balding, NAC Cressie, GM Fitzmaurice, *et al.*, editors]. London: John Wiley & Sons, Ltd.

79. Kaufman L & Rousseeuw PJ (2005) *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken: John Wiley & Sons, Inc.

80. Jain AK, Topchy A, Law MHC, *et al.* (2004) Landscape of clustering algorithms. *Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)*, Vol. **1**, pp. 260–263.

81. Steinley D (2006) K-means clustering: a half-century synthesis. *Br J Math Stat Psychol* **59**, 1–34.

82. Bacher J, Pöge A & Wenzig K (2010) *Clusteranalyse – Anwendungsorientierte Einführung in Klassifikationsverfahren (Cluster Analysis – Application-Oriented Introduction to Classification Methods)*, 3rd ed. Munich: Oldenbourg Wissenschaftsverlag GmbH.

83. Dua S & Chowriappa P (2013) *Data Mining for Bioinformatics*. Boca Raton, FL: CRC Press Taylor & Francis Group.