# Metabolomics: A window into the mechanism of lifestyle exposures and metabolic health

## How diet and body composition relate to the metabolome and metabolic health

Doctoral thesis

to obtain a doctorate (PhD)

from the Faculty of Medicine

of the University of Bonn

## Christian Brachem

from Sögel, Germany

2023

Written with authorization of

the Faculty of Medicine of the University of Bonn

First reviewer:        Prof. Dr. Ute Nöthlings

Second reviewer:     Prof. Dr. Matthias Schmid

Day of oral examination: 25.09.2023

From the Institute of Nutrition and Food Sciences

Director: Prof. Dr. Matthias Wüst

**Table of Content**

# List of Abbreviations

DONALD DOrtmund Nutrition and Anthropometry Longitudinally Designed

BMI      Body Mass Index

BF       Body Fat Percentage

# 1 Abstract

**Background:** Modifiable risk factors, specifically in this work body composition and habitual diet, are of great importance in epidemiology. The metabolome is a rich resource to explore biological pathways from modifiable risk factors to markers of metabolic health and by extension to health conditions. In three studies, we aimed to identify associations between the metabolome of adolescents and young adults and body composition (study one) or habitual diet (study two). Additionally, we aimed to replicate previous results on associations between the metabolome and markers of metabolic health in adolescents and young adults and test the newly discovered associations as mediators in the relationship of modifiable risk factors and markers of metabolic health (study three).

**Methods:** All three studies were performed in a sub-sample of the DONALD open cohort study using untargeted metabolome measurements of the urine (all studies) and blood metabolome (studies two and three). We used linear and random forest regression (study one), as well as partial least squares regression (study two) to discover new associations between the modifiable risk factors and the metabolome. We utilized linear regression to replicate metabolites associated in two independent studies from our systematic literature review and causal mediation analysis to search for mediators (study three). Where applicable, we held the false discovery rate at 5% to correct for multiple testing. We used imputation to address missing values and stratified all our analyses by sex.

**Results:** We identified 30 metabolites associated with body composition (study one), ten of which were associated with both body mass index (BMI) and body fat percentage (BF). We identified six metabolites as putative biomarkers of habitual dietary intake and replicated one additional marker previously reported by others (study two). We replicated ten metabolites associated with markers of metabolic health and identified no potential mediators (study three). We observed few overlaps between the sexes suggesting strong sex differences in all three studies.

**Conclusion:** Our work demonstrated the great potential of exploring the metabolome as a rich resource and tool in epidemiology. We additionally proposed that sex-specific investigations should be considered in future studies.

# 2 Introduction and aims with references

## 2.1 Introduction

Modifiable risk factors are of great importance in epidemiology (Barbaresko et al., 2018; Murray et al., 2020). It has been reported in several studies that many modifiable risk factors are important determinants of non-communicable chronic diseases. These are in turn the most pressing concern for health care research in Western countries, as their impact on population health continues to grow (Yach et al., 2004). Two important modifiable risk factors are body composition and habitual diet.

Body composition (specifically high or very low body fat percentages) has been implicated for years as a major contributor to increased incidences of a variety of chronic diseases, for example diabetes mellitus type 2 or cardio-vascular diseases (Bendor et al., 2020; Guh et al., 2009; Murray et al., 2020). However, in recent years it has become clear that this influence is not as simple to explain as increased measures of body composition result in increased rates of disease. Evidence shows that it is important to note where fat accumulation occurs in the body, with visceral fat being far more disadvantageous compared to subcutaneous fat (Power & Schulkin, 2008; Zhang et al., 2015). Independent of location, fat tissue is an active metabolic and hormonal agent in the homeostasis of the human body (Henry & Clarke, 2008).

Habitual diet is very similar to body composition regarding its early implication in the etiology of different diseases (Barbaresko et al., 2018; Murray et al., 2020). There are compelling reasons for this: What humans eat is bound to matter, as we introduce a lot of different substances to our body that can have beneficial or harmful effects in the body. However, diet differs greatly from body composition in terms of the apparent complexity of its measurement. Habitual diet is notoriously hard to measure because it is on one hand a very mundane, daily ritual, and on the other hand very flexible and always changing according to our needs and (seasonal) availability of foods (Conrad et al., 2018; Naska et al., 2017). Measuring precisely how much, when and what someone eats, while avoiding observation bias is accordingly a complex challenge. Diet is not only difficult to measure

but additionally interwoven with other modifiable risk factors, for example body composition.

With the emergence of -Omics technology, like genomics or proteomics, we are continuously improving our understanding of the interactions of the intricate biochemical systems in the body (Johnson et al., 2016; Rangel-Huerta et al., 2019; Scalbert et al., 2014). A relatively recent addition to the -Omics Panel, metabolomics, can give us a snapshot of the phenotype expressed at certain times. The metabolome is the collection of all molecules or metabolites in a defined bio sample, similar to how the genome is the collection of all genes. The metabolome is influenced heavily by environmental stimuli, therefore it can have quick responses to these stimuli (Johnson et al., 2016). Some metabolites change very rapidly and in the short term, others in a slower fashion. These differences represent the changing expression of biochemical pathways due to a continuous stimulus, for example body composition or habitual diet (Rangel-Huerta et al., 2019; Scalbert et al., 2014).

We have established that body composition and habitual diet are important facets of lifestyle that may influence future health outcomes. However, we are not yet able to establish how these influences are formulated, more specifically which pathways are affected, by how much and when. In order to answer these questions, metabolomics can be a useful tool (Naska et al., 2017; Rangel-Huerta et al., 2019; Scalbert et al., 2014). As explained above, the metabolome is influenced by external stimuli like habitual diet and body composition. For researchers to understand which parts of the metabolome are affected by habitual diet and body composition the first step is to understand how these changes may help preventing and/or delaying the occurrence of non-communicable diseases. A macro level example of this is insulin's role in the development of type two diabetes mellitus (Saini, 2010). Similar mechanisms are bound to be present in the etiology of many chronic diseases. An additional benefit for habitual diet is in the ability to discover dietary intake biomarkers (Scalbert et al., 2014). As introduced earlier, diet is challenging to measure well. However, if dietary assessment can be transitioned into a more objective measurement, for example through integrating biomarkers into present assessment methods, we might be able to disentangle the complex relationship between

health and diet more easily (Conrad et al., 2018; Conrad & Nöthlings, 2017; Scalbert et al., 2014).

We conducted our research in an adolescent and young adult population. In adolescence biochemical and physiological processes change in dramatic fashion (Viner et al., 2015). Additionally, many adolescence lifestyle habits are carried into adulthood. Because of this the period of adolescence in life is specifically of interest to the study of the metabolome and its connections to lifestyle. Associations present in adolescence that are replicated in adults can further improve our understanding of small changes that have important impacts over long periods of time. Adolescence is, however, not studied nearly as much as adulthood. This is probably due to the increased burden placed on this research: participants are harder to recruit, because not only the participant but parents need to be involved and interested, ethical approval is more complex, as children and adolescents are under special protections, etc.

Biological sex has an enormous influence on the biochemical processes that make up the metabolites of the metabolome (Clegg & Mauvais-Jarvis, 2018). We can see these biochemical differences in many aspects of the literature, be it in metabolic processes associated with drugs (Whitley & Lindsey, 2009) or simply in the concentrations of long-established circulating markers for physiological health (Lew et al., 2017). However, current metabolomic literature is not well stratified by sex, most likely to preserve power because sample sizes are still small. These sex differences are additionally not only present in biochemical realities, but also in our lifestyle, as gender roles and stereotypes continue to have a large influence on our behaviors (Esteban-Gonzalo et al., 2020; Sloan et al., 2009). Therefore, we stratified all our analyses by sex.

To summarize, small changes in modifiable risk factors may have a large impact on health later in life. We can already see the effects of these changes in conventional markers for metabolic health. For these markers we often already have a hypothesis for the mechanisms that result in associations. For example, we have a well-defined though not complete understanding of the mechanism of blood pressure on the incidence of heart disease (Oh & Cho, 2020). We additionally know that both diet (Ndanuko et al., 2016) and body composition (Landi et al., 2018) can influence blood pressure levels. The larger knowledge gap we have today is in the mechanism between the modifiable exposures

and the metabolic health markers. Metabolomics could be an important resource in the drive to understand these relationships as well.

## 2.2  Aims

To this end, our overarching aim was to identify and confirm metabolites associated with modifiable risk factors and their relationship to markers of chronic disease. To achieve this, we defined the following sub-targets:

1. Identify new metabolites associated with the modifiable risk factors body composition (see 3.1 Paper 1) and diet (see 3.2 Paper 2).

2. Find and replicate putative biomarkers of habitual dietary intake (see 3.2 Paper 2).

3. Investigate the presence of mediation in these metabolites on the path from lifestyle to conventional systemic markers for chronic disease (see 3.3 Paper 3).

4. Confirm previous associations between the metabolome and conventional markers of metabolic health in an adolescent population (see 3.3 Paper 3).

## 2.3  References

Barbaresko, J., Rienks, J., & Nöthlings, U. (2018). Lifestyle Indices and Cardiovascular Disease Risk: A Meta-analysis. *American Journal of Preventive Medicine*, *55*(4), 555–564 https://doi.org/10.1016/j.amepre.2018.04.046

Bendor, C. D., Bardugo, A., Pinhas-Hamiel, O., Afek, A., & Twig, G. (2020). Cardiovascular morbidity, diabetes and cancer risk among children and adolescents with severe obesity. *Cardiovascular Diabetology*, *19*(1), 79 https://doi.org/10.1186/s12933-020-01052-1

Clegg, D. J., & Mauvais-Jarvis, F. (2018). An integrated view of sex differences in metabolic physiology and disease. *Molecular Metabolism*, *15*, 1–2 https://doi.org/10.1016/j.molmet.2018.06.011

Conrad, J., Koch, S. A. J., & Nöthlings, U. (2018). New approaches in assessing food intake in epidemiology. *Current Opinion in Clinical Nutrition and Metabolic Care*, *21*(5), 343–351 https://doi.org/10.1097/MCO.0000000000000497

Conrad, J., & Nöthlings, U. (2017). Innovative approaches to estimate individual usual dietary intake in large-scale epidemiological studies. *The Proceedings of the Nutrition Society*, *76*(3), 213–219 https://doi.org/10.1017/S0029665116003025

Esteban-Gonzalo, S., Sik Ying Ho, P., Aparicio-García, M. E., & Esteban-Gonzalo, L. (2020). Understanding the Meaning of Conformity to Feminine Norms in Lifestyle Habits and Health: A Cluster Analysis. *International Journal of Environmental Research and Public Health*, *17*(4), 1370 https://doi.org/10.3390/ijerph17041370

Guh, D. P., Zhang, W., Bansback, N., Amarsi, Z., Birmingham, C. L., & Anis, A. H. (2009). The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis. *BMC public health*, *9*, 88 https://doi.org/10.1186/1471-2458-9-88

Henry, B. A., & Clarke, I. J. (2008). Adipose Tissue Hormones and the Regulation of Food Intake. *Journal of Neuroendocrinology*, *20*(6), 842–849 https://doi.org/10.1111/j.1365-2826.2008.1730.x

Johnson, C. H., Ivanisevic, J., & Siuzdak, G. (2016). Metabolomics: beyond biomarkers and towards mechanisms. *Nature Reviews Molecular Cell Biology*, *17*(7), 451–459 https://doi.org/10.1038/nrm.2016.25

Landi, F., Calvani, R., Picca, A., Tosato, M., Martone, A. M., Ortolani, E., Sisto, A., D'Angelo, E., Serafini, E., Desideri, G., Fuga, M. T., & Marzetti, E. (2018). Body mass index is strongly associated with hypertension: Results from the longevity check-up 7+ study. *Nutrients*, *10*(12), 1976 https://doi.org/10.3390/nu10121976

Lew, J., Sanghavi, M., Ayers, C. R., McGuire, D. K., Omland, T., Atzler, D., Gore, M. O., Neeland, I., Berry, J. D., Khera, A., Rohatgi, A., & Lemos, J. A. de. (2017). Sex-based differences in cardiometabolic biomarkers. *Circulation*, *135*(6), 544–555 https://doi.org/10.1161/CIRCULATIONAHA.116.023005

Murray, C. J. L., Aravkin, A. Y., Zheng, P., Abbafati, C., Abbas, K. M., Abbasi-Kangevari, M., Abd-Allah, F., Abdelalim, A., Abdollahi, M., Abdollahpour, I., Abegaz, K. H., Abolhassani, H., Aboyans, V., Abreu, L. G., Abrigo, M. R. M., Abualhasan, A., Abu-Raddad, L. J., Abushouk, A. I., Adabi, M., … Lim, S. S. (2020). Global burden of 87 risk factors in 204 countries and territories, 19902019: A systematic analysis for the global burden of disease study 2019. *The Lancet*, *396*(10258), 1223–1249 https://doi.org/10.1016/S0140-6736(20)30752-2

Naska, A., Lagiou, A., & Lagiou, P. (2017). Dietary assessment methods in epidemiological research: Current state of the art and future prospects. *F1000Research*, *6*, 926 https://doi.org/10.12688/f1000research.10703.1

Ndanuko, R. N., Tapsell, L. C., Charlton, K. E., Neale, E. P., & Batterham, M. J. (2016). Dietary patterns and blood pressure in adults: A systematic review and meta-analysis of randomized controlled Trials12. *Advances in Nutrition*, *7*(1), 76–89 https://doi.org/10.3945/an.115.009753

Oh, G. C., & Cho, H.-J. (2020). Blood pressure and heart failure. *Clinical Hypertension*, *26*(1), 1 https://doi.org/10.1186/s40885-019-0132-x

Power, M. L., & Schulkin, J. (2008). Sex differences in fat storage, fat metabolism, and the health risks from obesity: possible evolutionary origins. *British Journal of Nutrition*, *99*(5), 931–940 https://doi.org/10.1017/S0007114507853347

Rangel-Huerta, O. D., Pastor-Villaescusa, B., & Gil, A. (2019). Are we close to defining a metabolomic signature of human obesity? A systematic review of metabolomics studies. *Metabolomics: Official Journal of the Metabolomic Society*, *15*(6), 93 https://doi.org/10.1007/s11306-019-1553-y

Saini, V. (2010). Molecular mechanisms of insulin resistance in type 2 diabetes mellitus. *World Journal of Diabetes*, *1*(3), 68–75 https://doi.org/10.4239/wjd.v1.i3.68

Scalbert, A., Brennan, L., Manach, C., Andres-Lacueva, C., Dragsted, L. O., Draper, J., Rappaport, S. M., Hooft, J. J. J. van der, & Wishart, D. S. (2014). The food metabolome: a window over dietary exposure. *The American Journal of Clinical Nutrition*, *99*(6), 1286–1308 https://doi.org/10.3945/ajcn.113.076133

Sloan, C., Gough, B., & Conner, M. (2009). Healthy masculinities? How ostensibly healthy men talk about lifestyle, health and gender. *Psychology and Health*, *25*(7), 783–803 https://doi.org/10.1080/08870440902883204

Viner, R. M., Ross, D., Hardy, R., Kuh, D., Power, C., Johnson, A., Wellings, K., McCambridge, J., Cole, T. J., Kelly, Y., & Batty, G. D. (2015). Life course epidemiology: recognising the importance of adolescence. *Journal of Epidemiology and Community Health*, *69*(8), 719–720 https://doi.org/10.1136/jech-2014-205300

Whitley, H. P., & Lindsey, W. (2009). Sex-based differences in drug activity. *American Family Physician*, *80*(11), 1254–1258 https://www.aafp.org/afp/2009/1201/p1254.html

Yach, D., Hawkes, C., Gould, C. L., & Hofman, K. J. (2004). The global burden of chronic DiseasesOvercoming impediments to prevention and control. *JAMA*, *291*(21), 2616–2622 https://doi.org/10.1001/jama.291.21.2616

Zhang, M., Hu, T., Zhang, S., & Zhou, L. (2015). Associations of Different Adipose Tissue Depots with Insulin Resistance: A Systematic Review and Meta-analysis of Observational Studies. *Scientific Reports*, *5*, 18495 https://doi.org/10.1038/srep18495

# 3 Publications

## 3.1 Publication 1: Associations of BMI and Body Fat with Urine Metabolome in Adolescents Are Sex-Specific: A Cross-Sectional Study

*metabolites*

**MDPI**

*Article*

# Associations of BMI and Body Fat with Urine Metabolome in Adolescents Are Sex-Specific: A Cross-Sectional Study

**Christian Brachem** [1,*], **Julia Langenau** [1], **Leonie Weinhold** [2], **Matthias Schmid** [2], **Ute Nöthlings** [1,†] **and Kolade Oluwagbemigun** [1,†]

1    Unit of Nutritional Epidemiology, Department of Nutrition and Food Sciences,
     Rheinische Friedrich-Wilhelms-University Bonn, 53115 Bonn, Germany; langenau@uni-bonn.de (J.L.);
     noethlings@uni-bonn.de (U.N.); koluwagb@uni-bonn.de (K.O.)
2    Institute for Medical Biometry, Informatics and Epidemiology (IMBIE), University Hospital Bonn,
     53127 Bonn, Germany; weinhold@imbie.meb.uni-bonn.de (L.W.);
     matthias.schmid@imbie.uni-bonn.de (M.S.)
*    Correspondence: cbrachem@uni-bonn.de; Tel.: +49-(0)-228-73-3989
†    These authors contributed equally to this work.

check for updates

**Abstract:** Epidemiologic studies examining the relationship between body composition and the urine metabolome may improve our understanding of the role of metabolic dysregulation in body composition-related health conditions. Previous studies, mostly in adult populations, have focused on a single measure of body composition, body mass index (BMI), and sex-specific associations are rarely explored. We investigate sex-specific associations of two measures of body composition—BMI and body fat (BF)—with the urine metabolome in adolescents. In 369 participants (age 16–18, 49% female) of the Dortmund Nutritional and Anthropometric Longitudinally Designed (DONALD) study, we examined sex-specific associations of these two measures of body composition, BMI and BF, and 1407 (467 unknown) 24 h urine metabolites analyzed by untargeted metabolomics cross-sectionally. Missing metabolites were imputed. We related metabolites (dependent variable) to BMI and BF (independent variable) separately using linear regression. The models were additionally adjusted for covariates. We found 10 metabolites associated with both BMI and BF. We additionally found 11 metabolites associated with only BF, and nine with only BMI. None of these associations was in females. We observed a strong sexual dimorphism in the relationship between body composition and the urine metabolome.

**Keywords:** metabolomics; adolescents; body composition; sex-specific; body mass index; body fat

## 1. Introduction

Overweightness (including obesity) has reached epidemic proportions. Approximately 39% of the adult human population is overweight (BMI (body mass index) $\geq 25$ to $< 30$ kg/m$^2$) or obese (BMI $\geq 30$ kg/m$^2$) [1]. The global prevalence of overweightness (BMI $> +1$ standard deviation above the median) among adolescents aged 10 to 19 years has increased steadily over the last 40 years, from 4.3% in 1975 to 17.3% in 2016 [1,2]. Current research suggests that overweightness and obesity contribute to the increasing risk of chronic diseases [1]. The global burden of disease study estimated that in 2015, roughly 7% of deaths from any cause and roughly 5% of disability-adjusted life-years globally were due to high BMI [3]. Metabolic dysregulation, in addition to inflammation and insulin resistance, may mediate the link between overweightness and many chronic diseases, like Type 2 diabetes or cardiovascular diseases. There is mounting evidence that these links are already present in

15

adolescents [1,4,5], implying an increase in risk of future incidence of chronic diseases. Therefore, it is important to find the metabolic changes already present in adolescence, and to understand the link between overweightness and disease progression in later life.

Being overweight is known to be related to metabolic changes—for example, through body fat functioning as an endocrine organ, producing adipokines like leptin or visfatin [6,7]. Additionally, past studies have shown that overweightness is likely to be a causal influence on the metabolome phenotype [8]. A recent review [7] that summarizes the current knowledge of the metabolomic signature of adult obesity concluded that many metabolite groups are altered, including sexual steroids, amino acids, and acylcarnitines, among others. Interestingly, only a few epidemiological studies have explored the relationship between body composition and the metabolome in adolescents [9,10]. Cho et al. [9] quantitatively measured the global metabolic repertoire in adolescents, and showed that endogenous metabolites and inflammation-related metabolites are related to body composition. Saner et al. [10] investigated metabolomic profiles in obese children and adolescents (ages 6 to 18), and found associations in post-pubertal males of several metabolites, including fatty acids, triglycerides, isoleucine, leucine, and glycoprotein with obesity measures. However, overall evidence is scarce calling for more studies profiling the adiposity metabolome, preferably by untargeted methods.

It is well-known that the body composition of adolescents is sex-specific [5]. While BMI tends to be comparable between males and females, body fat in females is physiologically higher starting in late puberty (Tanner stages IV and V). In addition, a sexual dimorphism in metabolism is well recognized [11]. Thus, investigating sex differences may reveal pathophysiologically relevant variations, with potential implications for overweightness- or obesity-related health conditions.

We decided to investigate two different measures of body composition to increase our confidence in the metabolite–body composition associations that are present for both measures. We used BMI, as it is the most widely used measure for body composition in observational studies [5,12]. It is well-understood that BMI is a good marker for body composition on the population level [13]. Specifically, in an adolescent population it has been demonstrated that BMI categories correctly identify children with excess body fat in roughly 85% of cases [5]. However, it has well-documented shortcomings regarding body fat distribution [14,15]. To address these shortcomings, we also used body fat percentage, as estimated with skinfold measurements.

Here, we explored the a priori, sex-stratified relationship between these two measures of body composition, BMI and BF, and the urine metabolome cross-sectionally among adolescents.

## 2. Results

### 2.1. Basic Characteristics

The basic characteristics of the participants (180 females, 189 males) are shown in Table 1. Females had a higher BF, were less physically active, consumed fewer calories, and were less likely to be overweight than males. Males were less likely to be current alcohol consumers, and more frequently their mothers were employed and of higher educational status. Roughly 20% of males and 13% of females were overweight (BMI ≥ 25).

**Table 1.** Characteristics of 369 Dortmund Nutritional and Anthropometric Longitudinally Designed (DONALD) study participants aged 16 to 18 years.

| Variable | *n* | Total | Male | Female |
|---|---|---|---|---|
| | | *n* = 369 | *n* = 189 | *n* = 180 |
| Age (years) | 369 | 17.3 (1.0) | 17.2 (1.0) | 17.4 (1.0) |
| Body Fat Percent | 368 | 21.8 (8.0) | 16.6 (5.6) | 27.3 (6.4) |
| BMI (kg/m$^2$) | 369 | 22.2 (3.7) | 22.5 (4.0) | 21.9 (3.2) |
| Overweight (BMI ≥ 25): Yes | 369 | 62 (16.8%) | 39 (20.6%) | 23 (12.8%) |
| Metabolic Equivalent of Task-Hours (met-h/week) | 207 | 41.3 (37.0) | 45.9 (43.3) | 36.6 (28.8) |
| Calories (kcal) | 364 | 2189.2 (616.7) | 2545.8 (565.4) | 1816.5 (415.4) |
| Protein (%kcal) | 364 | 13.9 (2.8) | 14.1 (2.7) | 13.7 (2.8) |
| Fat (%kcal) | 364 | 33.7 (6.6) | 33.9 (7.0) | 33.6 (6.1) |
| Carbohydrates (%kcal) | 364 | 50.5 (6.9) | 49.8 (7.4) | 51.2 (6.3) |
| Smoking Status | 118 | | | |
| Never | | 56 (15.2%) | 28 (14.8%) | 28 (15.6%) |
| Former | | 35 (9.5%) | 13 (6.9%) | 22 (12.2%) |
| Current | | 27 (7.3%) | 12 (6.3%) | 15 (8.3%) |
| Alcohol Status | 155 | | | |
| Never | | 8 (2.2%) | 5 (2.6%) | 3 (1.7%) |
| Former | | 11 (3%) | 5 (2.6%) | 6 (3.3%) |
| Current | | 136 (36.9%) | 65 (34.4%) | 71 (39.4%) |
| Maternal Occupation: Working (full or part-time) | 364 | 222 (60.2%) | 122 (64.6%) | 100 (55.6%) |
| Maternal Education: >12 Years of Education | 365 | 190 (51.5%) | 101 (53.4%) | 89 (49.4%) |
| Breastfeeding Duration (weeks) | 363 | 25.0 (18.3) | 24.3 (18.9) | 25.7 (17.7) |
| Maternal Gestational Weight Gain (kg) | 348 | 12.8 (4.1) | 12.7 (4.2) | 13.0 (4.1) |
| Maternal BMI (kg/m$^2$) (kg/m$^2$) | 358 | 23.7 (3.7) | 23.8 (3.5) | 23.7 (3.9) |
| Smoking Household: Yes | 265 | 86 (23.3%) | 43 (22.8%) | 43 (23.9%) |

Data are presented as mean (standard deviation) for continuous measures and *n* (column percent) for categorical measures. Available *n* values differ because of missing data.

*2.2. Linear Regression Models*

2.2.1. Summarizing Metabolites into Groups Using Independent Component Analysis

We kept the first seven independent components (IC), according to the scree plot. The composition of the extracted components are recorded in Table A1. In our sample, no IC was associated with BMI or BF for either sex. A table of β-estimates with confidence limits can be found in Table A2.

2.2.2. Metabolites Associated with Both BMI and BF

There were 10 metabolites (0.8% of metabolites analyzed) significantly associated with both BMI and BF in males, and zero metabolites in females (Figure 1). A table of β-estimates with confidence limits can be found in Table A3. The estimates presented here are back-transformed from the log-scale.

There were four amino acids associated significantly with BMI and BF: guanidinosuccinate (negative, BMI: 0.97 (0.96 to 0.99), BF: 0.98 (0.97 to 0.99)), isobutyrylglycine (C4) (negative, BMI: 0.97 (0.95 to 0.98), BF: 0.98 (0.97 to 0.99)), isovalerylglycine (negative, BMI: 0.96 (0.95 to 0.98), BF: 0.97 (0.96 to 0.98)), and tigloylglycine (negative, BMI: 0.97 (0.96 to 0.99), BF: 0.97 (0.97 to 0.98)).

In the super-pathway of cofactors and vitamins, nicotinamide N-oxide (positive, BMI: 1.05 (1.02 to 1.08), BF: 1.04 (1.03 to 1.06)) was associated. Additionally, the xenobiotic succinimide (negative, BMI: 0.98 (0.97 to 0.99), BF: 0.99 (0.98 to 0.99)) was associated with both BMI and BF significantly.

Furthermore, we found significant associations with both BMI and BF for the partially characterized molecule glucuronide of $C_{10}H_{18}O_2$ (12) (positive, BMI: 1.05 (1.03 to 1.07), BF: 1.03 (1.02 to 1.04)) as well as the unknown metabolites X-21851 (positive, BMI: 1.04 (1.02 to 1.06), BF: 1.02 (1.01 to 1.04)), X-24469 (positive, BMI: 1.03 (1.02 to 1.05), BF: 1.02 (1.01 to 1.03)), and X-24801 (positive, BMI: 1.03 (1.02 to 1.05), BF: 1.02 (1.01 to 1.03)).
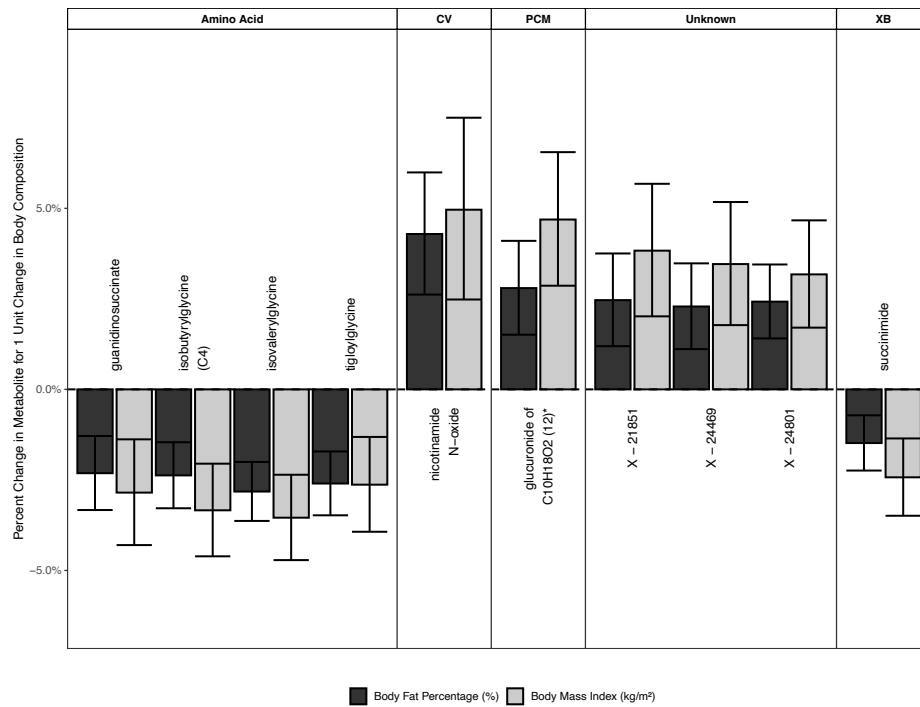
**Figure 1.** Metabolites associated with body mass index (BMI) and body fat (BF). Estimates are back-transformed linear regression beta coefficients, regressing metabolites on body composition (BMI or BF). BMI is measured in kg/m$^2$ and body fat in percent. Abbreviations: CV, cofactor and vitamins; PCM: partially characterized molecules; XB: xenobiotics.

2.2.3. Metabolites Associated with Either BMI or BF

There were 20 metabolites (1.6% of metabolites analyzed) significantly associated with either BMI or BF. Of these, 11 were associated with BF and nine with BMI. All 20 associations were in males, none in females. A graphical representation of these results is presented in Figure 2. In Table A4, we present β-estimates and confidence intervals for all metabolites. The estimates presented here are back-transformed from the log-scale.
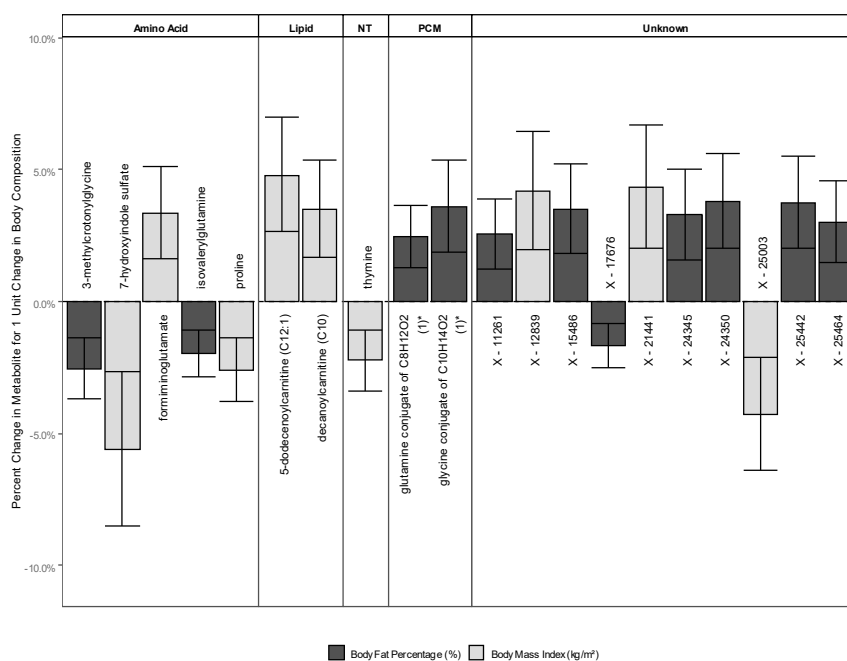
18

**Figure 2.** Metabolites associated with either BMI or BF. Estimates are back-transformed linear regression beta coefficients, regressing metabolites on body composition (BMI or BF). BMI units are kg/m$^2$ and body fat units are percent. Abbreviations: NT, nucleotide; PCM, partially characterized molecules.

*2.3. Metabolites Associated with BMI*

The amino acids formiminoglutamate (positive, BMI: 1.03 (1.02 to 1.05)), 7-hydroxyindole sulfate (negative, BMI: 0.94 (0.92 to 0.97)), and proline (negative, BMI: 0.97 (0.96 to 0.99)) were associated with BMI. Additionally, the nucleobase thymine (BMI: 0.98 (0.97 to 0.99)) was associated negatively with BMI. Two lipids were significantly associated with BMI: decanoylcarnitine (C10) (positive, BMI: 1.04 (1.02 to 1.05)) and 5-dodecenoylcarnitine (C12:1) (positive, BMI: 1.05 (1.03 to 1.07)). Three unknown metabolites (X-12839 (positive, BMI: 1.04 (1.02 to 1.06)), X-21441 (positive, BMI: 1.04 (1.02 to 1.07)), and X-25003 (negative, BMI: 0.96 (0.94 to 0.98))) were associated with BMI.

*2.4. Metabolites Associated with BF*

The amino acids 3-methylcrotonylglycine (negative, BF: 0.97 (0.96 to 0.99)) and isovalerylglutamine (negative, BF: 0.98 (0.97 to 0.99)) were significantly associated with BF. The energy metabolite malate (negative, BF: 0.97 (0.96 to 0.99)) was significantly associated with BF as well. Additionally, there were two partially characterized molecules (glutamine conjugate of $C_8H_{12}O_2$ (1) (positive, BF: 1.02 (1.01 to 1.04)) and glycine conjugate of $C_{10}H_{14}O_2$ (1) (positive, BF: 1.04 (1.02 to 1.05)) and seven unknown metabolites (X-11261 (positive, BF: 1.03 (1.01 to 1.04)), X-15486 (positive, BF: 1.04 (1.02 to 1.05)), X-17676 (negative, BF: 0.98 (0.97 to 0.99)), X-24345 (positive, BF: 1.03 (1.02 to 1.05)), X-24350 (positive, BF: 1.04 (1.02 to 1.06)), X-25442 (positive, BF: 1.04 (1.02 to 1.06)), and X-25464 (positive, BF: 1.03 (1.01 to 1.05))) that were significantly associated.

## 3. Discussion

The current study explores the sex-specific cross-sectional associations of two measures of body composition, BMI and BF, and the urine metabolome and urine metabolite patterns (ICs) in adolescent boys and girls. Approximately 2.4% of the urine metabolome was associated with body composition in boys; no association was seen in girls. Our results underscore the presence of changes in the urine metabolome in relation to body composition already in adolescence. To our knowledge, this is the first study to relate two measures of body composition to the urine metabolome in adolescents. Our results strongly suggests sex-specificity in associations.

We advise the reader that the results of the present study were exploratory, and therefore should not be overemphasized. Any interpretation we give here in relation to the biological process may only be seen as one of many possible explanations for the reported associations. In fact, many of the reported compounds have not been reported in conjunction with body composition before. A more in-depth investigation of these single compounds is, however, outside of the scope of this study. We found 10 metabolites that related to both measures of body composition in males, and none in females. These metabolites were guanidinosuccinate, isobutyrylglycine (C4), isovalerylglycine, tigloylglycine, nicotinamide N-oxide, glucuronide of $C_{10}H_{18}O_2$ (12), X-21851, X-24469, X-24801, and succinimide. Nicotinamide N-oxide [1] and tigloylglycine [2] have been associated with BMI in prior studies. The other eight molecules are reported in association with body composition here for the first time. Additionally, we found 20 metabolites associated with either BMI or BF. When metabolites are significantly associated with both measures of body composition, we should have higher confidence in their association. As both measures have their own unrelated measurement error while measuring different aspects of the same concept (body composition), a significant association with both BMI and BF should indicate that it is more likely related to this underlying concept. The metabolites associated with only BMI or BF, however, were all associated in the same direction with the other body composition measurement. Additional discussion of these metabolites can be found in Table A5.

In general, our results reinforced the idea of sexual dimorphism in metabolism. The stronger association in males is consistent with previous studies in mice [3], adults [4,5], and adolescents [6], as well as our own recent findings within this study population [7]. One potential explanation is that sex hormones might modify the relationship between body composition and the urine metabolome. Specifically, prior studies have shown changes in the type of body composition and overall obesity in relation to sex hormones and displaying sexual dimorphism in their mode of effect [8–11]. Furthermore, the sexual dimorphism in the urine metabolome is well-documented [4,12–15]. As sex hormones play an important role in many metabolic pathways, e.g., they have been shown to regulate the liver energy homeostasis [16], an interaction between sex hormones, body composition, and the urine metabolome is plausible. Another explanation, as was shown for urine cortisol levels [5], is that sex differences relate to other factors of metabolism, such as enzyme activity. Wang et al. [17] showed that lipid and lipoprotein metabolism is in fact independent of sex hormone administration, even though there are significant sex differences; however, the mechanism remains to be elucidated. The specific mechanism of sex difference in metabolism might therefore differ for different pathways, and deserves to be studied further. Our results may help to explain sex differences in weight-related health conditions.

We used independent component analysis (ICA) to summarize metabolites into fewer components in the current analysis. We chose ICA because the components are statistically independent, and their interpretation in biological processes allows for the mixture of different pathways and processes that contribute to the living system. Because metabolomics takes a snapshot of these processes and systems, these components hold a large value for understating of processes. In the current study, none of the ICs we retained were associated with body composition. This suggests that body composition influences specific metabolic pathways, and not a mixture of different pathways captured by the ICA.

Guanidinosuccinate is produced by the oxidation of argininosuccinic acid, and was associated with higher measures of body composition in males. The oxidation of guanidinosuccinate occurs favorably with increased levels of urea, and results in a decline of hepatic levels of arginine [18]. It is

20

well-known that the urea cycle is dysregulated with higher adiposity [19]; therefore, reduced renal function compared to the average adolescent may partly explain our findings. Guanidinosuccinate may be a marker of the kidneys' ability to eliminate urea, particularly in males.

Isobutyrylglycine (C4) is a short-chain acylglycine in the catabolism of leucine, isoleucine, and valine. In newborn screenings, elevated levels of this metabolite are used to diagnose isobutyryl-CoA dehydrogenase deficiency [20]. Since isobutrylglycine levels decrease with higher BMI and BF, isobutyryl-CoA dehydrogenase might be upregulated with elevated measures of body composition. Alternatively, smaller amounts of leucine, isoleucine, and valine might be catabolized in individuals with abnormal adiposity. However, the present association was independent of these metabolites.

Isovalerylglycine is an acyl glycine that is produced in the catabolism of leucine [18]. Higher BMI and BF are associated with the metabolism of leucine in rats [21]; however, no study to date exists in humans. This metabolite has also been suggested as a biomarker for cheese consumption [22]. Although we did not specifically adjust for cheese intake, the fact that we adjusted for macronutrient intake suggests that our finding is independent of cheese intake.

Tigloylglycine is an acylglycine that is an intermediate of the isoleucine catabolism [18]. Like isovalerylglycine, it was suggested as a biomarker for the consumption of cheese [18]. Again, we adjusted for nutrition, so an association because of cheese consumption is unlikely. Urinary acylglycine decreases with higher BMI have been documented before [2]. Similar to other leucine, isoleucine, and valine metabolites, the enzyme metabolizing this compound might be upregulated, or the overarching pathway of branched-chain amino acid (BCAA) catabolism might be dysregulated.

BCAAs have a well-documented association with higher markers of body composition: increased blood levels of BCAAs correlated with higher levels of body composition [19]. A recent study by Elliot et al. [23] reported associations between increased urine levels of leucine, isoleucine, and valine and BMI. Additionally, they reported lower levels of ketoleucine with higher BMI. Ketoleucine is the first metabolic product in the energy use of leucine [24]. The metabolites we found that decreased with higher measures of body composition are downstream metabolites of BCAAs, which are produced through similar processes as ketoleucine from leucine, namely when their respective BCAA is used for energy in skeletal muscle. As BCAAs are not the first energy source muscles use in response to physical activity, increased blood levels of BCAAs and decreased levels of their energy pathway downstream products are in line with decreased physical activity and overabundance of other energy sources in persons with higher measures of body composition.

Nicotinamide N-oxide is a precursor of nicotinamide adenine dinucleotide (NAD) and a catabolite of nicotinamide [18,25]. Increased urine nicotinamide N-oxide is associated with high-fat, diet-induced obesity in mice [26]. In humans, serum levels of another nicotinamide was positively associated with BMI and waist circumference [1]. This finding suggests that in individuals with higher measures of body composition, there is a nicotinamide overload, or enzymes catabolizing nicotinamide to nicotinamide N-oxide are overexpressed or hyper-activated. However, our result is independent of nicotinamide, which favors the latter explanation.

Succimide is commonly found in anticonvulsant drugs [18]. The fact that a common side effect of anticonvulsant drugs are changes in weight [27] might provide a potential explanation for the association with adiposity.

Additionally, there are no available data on the relationship between the unknown metabolites X-21851, X-24469, and X-24801, or the partially characterized metabolite glucuronide of C10H18O2 (12) and body composition. Besides, since they are without biochemical identities, or only partially characterized, it is difficult to provide explanations. Nevertheless, with the rapidly developing field of metabolomics, the identification of these metabolites should not be far from sight.

The present study has some notable strengths. We investigated the associations between body composition and the urine metabolome using two measures of body composition, in order to achieve

a more comprehensive relationship between body composition and alteration in urine metabolites. The sex-specific investigation defined a priori also ensures that sex-specific relationships are well explored. Additionally, we used 24 h urine samples in a comparatively large study population to study the urine metabolome with an untargeted approach. To limit the possibility of false positives that untargeted approaches entail, we controlled for multiple testing by holding the False Discovery Rate (FDR) at 5%.

However, we acknowledge several limitations to the study. First, our participants are all Caucasians (Germans), residing in a large city (Dortmund) and surroundings, mostly from a high socioeconomic background. Thus, the generalizability of our findings is limited. Further, our study sample had very few individuals in the extremes of body composition, namely in the underweight (BMI < 18.5) and the obese (BMI ≥ 30) classifications, our findings may only be generalizable to individuals with normal and overweight body composition status. More associations of metabolites with BF as compared to BMI may also be due to BF having a larger variation in our study sample. Additionally, we cannot rule out residual confounding by either unknown or unmeasured (for example, genetic influences) factors. Lastly, because we only had one measurement of the urine metabolome, we were not able to establish a relationship of body composition and variability in the urine metabolome.

Future research should try to identify the unknown or partially characterized molecules that were associated in this study, as they have potential to help elucidate the biological mechanisms of the relationship between body composition and metabolic function on the pathway to health outcomes. Additionally, more studies are needed that stratify their metabolomic analysis by sex, in order to increase our understanding of the physiological differences in metabolism between males and females. Furthermore, future studies should try to replicate our findings in an independent adolescent population, and try to extend the analysis to a longitudinal design to elucidate the temporal relation of body composition with urine metabolome. Additionally, it would be interesting to evaluate differences between the blood and urine metabolome in a similar study setting, preferably in the same participants. Overall, metabolomics would benefit greatly from more unified data analysis approaches to facilitate meta-analysis of different cohorts. Lastly, a similar analysis carried out in a cohort with a larger proportion of overweight and obese participants would help to disentangle the gradient relationship between body composition and the urine metabolome.

## 4. Materials and Methods

### 4.1. Study Design

The present analysis is conducted within the Dortmund Nutritional and Anthropometric Longitudinally Designed (DONALD) study. Briefly, the DONALD study is a longitudinal open cohort study with the aim of analyzing detailed data on diet, growth, development, and metabolism from infancy to adulthood [28]. All study participants were invited to the study center on a regular basis, every 3 months until their first birthday, biennially in their second year, and annually thereafter. The anthropometric measurements are conducted by experienced nurses [28]. Data collected includes demographic, family, and socioeconomic characteristics, as well as anthropometric measurements, such as height, body weight, skin fold thickness, and 3 day weighed dietary records [28]. Informed written consent was obtained from parents and from participants themselves on reaching adolescence. The ethics committee of the University of Bonn, Germany (project identification: 098/06) approved the study.

### 4.2. Study Participants

The current study sample were DONALD participants from a previous study that explored BMI trajectories [29]. These 689 individuals are singletons, full-term (37 to 42 weeks of gestation), and had a birth weight of at least 2500 g. Of these, 369 participants had a 24 h urine samples between the ages of 16 to 18, from which the urine metabolome were profiled by an untargeted metabolomics approach.

22

*4.3. Variable Assessment*

4.3.1. Outcome: Untargeted Metabolomic Profiling of the Urine Metabolome

Ultra-high-performance liquid chromatography–tandem mass spectroscopy (UPLC-MS/MS) was used to identify metabolites in the 24 h urine samples. Peak identification was done in the propriety Laboratory Information Management System (LIMS) of Metabolon Inc. (Morrisville, NC, USA). Compounds were identified by comparison of their retention time/index (RI), mass-to-charge ratio (*m/z*), and chromatographic data (e.g., MS/MS spectral data) to library standards. Structurally unknown biochemicals were identified by occurrence. Peaks were quantified using area-under-the-curve and normalized with block correction corrected for inter-day instrument tuning differences. Further details on the methodology of the metabolic profiling have been reported elsewhere [30]. This analysis resulted in 1407 annotated features used in this analysis.

4.3.2. Exposure: Body Composition Measures

Body composition parameters were examined at every follow-up by experienced nursing staff. BMI was calculated using height and weight. BF was calculated from four skin-fold thickness measurements (biceps, triceps, iliaca, and scapula), using age, puberty status, and sex-specific equations from Deurenberg et al. [31].

4.3.3. Covariates

We constructed a directed acyclic graph (DAG; cf. cf. Supplemental Figure S1) to assess the minimally sufficient sets of variables to use for covariate adjustment in the analysis of the present data.

Family and socioeconomic characteristics around birth were assessed at the first study visit in the DONALD study, at around three months after birth. We included maternal employment (full- or part-time employment vs. no employment), maternal education (>12 vs. ≤12 years of education), smoking in the household (yes vs. no), maternal BMI (kg/m$^2$), and duration of breast-feeding (weeks). Dietary intake was assessed annually by three-day weighted dietary records. We calculated individual means of daily calorie and macronutrient intake, using our continuously updated in-house food composition database LEBTAB [32]. Macronutrient intake was estimated as percent of calories consumed. Using the method described by Schofield [33], the metabolic equivalents of task-hours (met-h) were determined from basal rates. The expanded met-h per week were subsequently calculated from the Adolescent Physical Activity Recall Questionnaire (APARQ) [34]. Alcohol and smoking status were assessed via questionnaire. Participants were grouped into current, former, and never drinkers (or smokers). Missing values were filled backwards for never drinkers (or smokers) and forwards for current and former drinkers (or smokers). Backwards filling means "never" consumption was used later to fill in missing values at time points prior to the non-missing answer, e.g., "never" consumption at age 20 was used to fill a missing value at age 16, since the participant was never a consumer. Forward filling refers to the same concept but forwards in time, e.g., current alcohol consumption at age 15 was used to fill a missing value for the current alcohol consumer at age 17.

*4.4. Statistical Analysis*

Statistical Analysis was performed using SAS software (Version 9.4 of the SAS System for Windows, copyright 2002–2012 SAS Institute Inc., Cary, NC, USA) and R software (Version 3.6) [35]. All analyses were a priori and were stratified by sex.

4.4.1. Data Pre-Treatment

Metabolite concentration were normalized by urine osmolality and rescaled to set the median equal to 1. Because the distributions of most metabolites were skewed, we performed natural

log-transformation. We excluded metabolites with more than 20% of missing values (*n* = 140) to keep the data quality acceptable. Missing values for the remaining 1267 metabolites were imputed.

### 4.4.2. Imputation of Missing Values

Missing values for all variables were imputed with multiple imputations (as implemented in the R package "mice" [36]) with 10 imputations and five iterations. We used the random forest method built into the "mice" package. The variables used as predictors for imputation for each "to be imputed" variable were selected according to the suggestions of the authors of the mice package [37]: variables were selected as predictors in the imputation if they had at least a correlation of *r* = 0.35 with the "to be imputed" variable, and at least 70% of the observations used for imputation had complete data. Imputation was stratified by sex.

### 4.4.3. Summarizing Metabolites into Groups Using the Independent Component Analysis

To reduce intercorrelation among metabolites, we summarized them into fewer interpretable components using independent component analysis (ICA). We used the "icafast" function from the "ica" Package [38] to perform the ICA, stratified by sex. We used the mean across all imputed datasets to extract the component model. The number of components to include in the regression analysis was selected by visual inspection of the scree plot. We then calculated the component scores for each imputed dataset according to the mean model. We characterized the components by their correlation with the metabolites, using the top 20 most correlated metabolites.

### 4.4.4. Linear Regression Model

In order to model the associations between BMI and BF and the urine metabolome, we fitted a linear regression model for each of the 1177 log-transformed metabolites and seven ICs as dependent variables, and either BMI or BF as independent variables. The models were additionally adjusted for the minimally sufficient set suggested by the DAG, which are physical activity (met-h/week), age, alcohol and smoking status, nutrition (total energy (kcal), protein (%kcal), fat (%kcal), and carbohydrates (%kcal)), smoking household, maternal occupation at study entry, maternal education, breastfeeding duration, and maternal BMI at study entry. If not otherwise specified, variables were measured during the same follow-up as the 24 h urine sample was taken. We performed multiple testing corrections by controlling the false discovery rate at five percent with the Benjamini–Hochberg procedure [39]. Metabolites associated with both BMI and BF were considered a signature of body composition.

## 5. Conclusions

In conclusion, 10 metabolites (10 in males, none in females) were associated with both measures of body composition, which could collectively be considered a metabolic signature of body composition. The sexual dimorphism in the relationship between body composition and the urine metabolome may explain sex differences in body composition-related health conditions.

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

## Appendix A : Additional Information on Independent Components

**Table A1.** Names of the Independent Components.

| IC | Constructed by |
|---|---|
| | Sex: female |
| IC1 | Amino acid (10), unknown (5), lipid (2), nucleotide (2), and xenobiotics (1) |
| IC2 | Amino acid (6), xenobiotics (5), unknown (5), partially characterized molecules (3), and peptide (1) |
| IC3 | Unknown (10), xenobiotics (4), amino acid (3), peptide (2), and lipid (1) |
| IC4 | Unknown (9), xenobiotics (5), amino acid (3), and lipid (3) |
| IC5 | Amino acid (6), lipid (5), unknown (5), partially characterized molecules (2), nucleotide (1), and xenobiotics (1) |
| IC6 | Unknown (12), amino acid (3), xenobiotics (2), lipid (1), partially characterized molecules (1), and peptide (1) |
| IC7 | Amino acid (5), xenobiotics (5), unknown (5), lipid (2), nucleotide (1), partially characterized molecules (1), and peptide (1) |
| | Sex: male |
| IC1 | Unknown (9), xenobiotics (4), amino acid (3), nucleotide (2), energy (1), and lipid (1) |
| IC2 | Amino acid (5), lipid (5), nucleotide (3), xenobiotics (2), unknown (2), carbohydrate (1), partially characterized molecules (1), and peptide (1) |
| IC3 | Unknown (8), amino acid (5), xenobiotics (3), lipid (2), carbohydrate (1), and energy (1) |
| IC4 | Xenobiotics (13), unknown (6), and lipid (1) |
| IC5 | Xenobiotics (7), unknown (6), partially characterized molecules (4), amino acid (1), lipid (1), and nucleotide (1) |
| IC6 | Unknown (10), lipid (4), partially characterized molecules (4), amino acid (1), and xenobiotics (1) |
| IC7 | Xenobiotics (8), unknown (6), carbohydrate (4), lipid (1), and partially characterized molecules (1) |

**Table A2.** Regression Coefficients for the ICs and body composition.

| IC | β Body Mass Index | 95% CI | *p* (FDR) | β Body Fat Percent | 95% CI | *p* (FDR) |
|----|----|----|----|----|----|----|
| | | | Sex: Female | | | |
| IC1 | 1.025 | 0.890 to 1.180 | 0.991 | 0.996 | 0.914 to 1.086 | 0.991 |
| IC2 | 0.974 | 0.832 to 1.141 | 0.991 | 1.030 | 0.943 to 1.125 | 0.991 |
| IC3 | 0.979 | 0.838 to 1.143 | 0.991 | 1.024 | 0.946 to 1.109 | 0.991 |
| IC4 | 0.945 | 0.829 to 1.076 | 0.991 | 1.014 | 0.942 to 1.093 | 0.991 |
| IC5 | 0.885 | 0.733 to 1.070 | 0.991 | 1.054 | 0.942 to 1.179 | 0.991 |
| IC6 | 0.930 | 0.814 to 1.064 | 0.991 | 1.057 | 0.979 to 1.141 | 0.991 |
| IC7 | 0.928 | 0.799 to 1.078 | 0.991 | 1.034 | 0.949 to 1.126 | 0.991 |
| | | | Sex: Male | | | |
| IC1 | 1.012 | 0.903 to 1.135 | 1.000 | 0.986 | 0.893 to 1.088 | 1.000 |
| IC2 | 0.969 | 0.880 to 1.067 | 1.000 | 1.001 | 0.923 to 1.086 | 1.000 |
| IC3 | 1.045 | 0.940 to 1.162 | 1.000 | 0.974 | 0.891 to 1.064 | 1.000 |
| IC4 | 0.980 | 0.877 to 1.095 | 1.000 | 1.053 | 0.958 to 1.157 | 1.000 |
| IC5 | 0.993 | 0.896 to 1.100 | 1.000 | 1.039 | 0.959 to 1.127 | 1.000 |
| IC6 | 0.956 | 0.872 to 1.048 | 1.000 | 1.015 | 0.943 to 1.093 | 1.000 |
| IC7 | 0.991 | 0.889 to 1.104 | 1.000 | 0.952 | 0.870 to 1.041 | 1.000 |

Estimates were generated from linear regression models, with independent components as the dependent variable and body mass index or body fat percent as the dependent variable. Multiple testing adjustments were performed by controlling the false discovery rate at 5%. Estimates are bac-transformed. The models were additionally adjusted for age, calories consumed, protein consumed, carbohydrates consumed, fat consumed, maternal gestational weight gain, maternal BMI, breastfeeding duration, smoking status, smoking household, alcohol status, maternal occupation, maternal education, metabolic equivalent of task-hours. Abbreviations: IC, independent component; FDR, False Discovery Rate

## Appendix B : Regression Coefficient Tables

**Table A3.** Metabolites associated with both body composition measures.

| Biochemical | Sub Pathway | β Body Mass Index | 95% CI | *p* (FDR) | β Body Fat Percent | 95% CI | *p* (FDR) |
|---|---|---|---|---|---|---|---|
| | | Super-pathway: Amino Acid | | | | | |
| guanidinosuccinate | Guanidino and acetamido metabolism | 0.971 | 0.957 to 0.986 | 0.046 | 0.977 | 0.967 to 0.987 | 0.014 |
| isobutyrylglycine (C4) | Leucine, isoleucine, and valine metabolism | 0.967 | 0.954 to 0.979 | 0.002 | 0.976 | 0.967 to 0.985 | 0.001 |
| isovalerylglycine | Leucine, isoleucine, and valine metabolism | 0.965 | 0.953 to 0.976 | 0.000 | 0.972 | 0.964 to 0.980 | 0.000 |
| tigloylglycine | Leucine, isoleucine, and valine metabolism | 0.974 | 0.961 to 0.987 | 0.036 | 0.974 | 0.965 to 0.983 | 0.000 |
| | | Super-pathway: CV | | | | | |
| Nicotinamide N-oxide | Nicotinate and nicotinamide metabolism | 1.050 | 1.025 to 1.075 | 0.030 | 1.043 | 1.026 to 1.060 | 0.001 |
| | | Super-pathway: PCM | | | | | |
| Glucuronide of C10H18O2 (12) * | Partially characterized molecules | 1.047 | 1.029 to 1.066 | 0.001 | 1.028 | 1.015 to 1.041 | 0.016 |
| | | Super-pathway: Unknown | | | | | |
| X-21851 | | 1.038 | 1.020 to 1.057 | 0.021 | 1.025 | 1.012 to 1.038 | 0.044 |
| X-24469 | | 1.035 | 1.018 to 1.052 | 0.025 | 1.023 | 1.011 to 1.035 | 0.044 |
| X-24801 | | 1.032 | 1.017 to 1.047 | 0.016 | 1.024 | 1.014 to 1.034 | 0.004 |
| | | Super-pathway: XB | | | | | |
| Succinimide | Chemical | 0.976 | 0.965 to 0.986 | 0.011 | 0.985 | 0.978 to 0.993 | 0.046 |

Estimates were generated from linear regression models with natural log-transformed biochemicals as the dependent variables, and body mass index or body fat percent as the dependent variable. Multiple testing adjustment was performed by controlling the false discovery rate at 5%. Estimates are back-transformed. The models were additionally adjusted for age, calories consumed, protein consumed, carbohydrates consumed, fat consumed, maternal gestational weight gain, maternal BMI, breastfeeding duration, smoking status, smoking household, alcohol status, maternal occupation, maternal education, metabolic equivalent of task-hours. * Compound was not identified by a standard, but we are confident in its identity. Abbreviations: CV, cofactors and vitamins; PCM, partially characterized molecules; XB, xenobiotics; FDR, False Discovery Rate.

**Table A4.** Metabolites associated with either BMI or BF.

| Biochemical | Sub-Pathway | Body Mass Index | 95% CI | *p* (FDR) | Body Fat Percent | 95% CI | *p* (FDR) |
|---|---|---|---|---|---|---|---|
| | | **Super-Pathway: Amino Acid** | | | | | |
| Formiminoglutamate | Histidine metabolism | **1.033** | **1.016 to 1.051** | **0.041** | 1.019 | 1.007 to 1.031 | 0.162 |
| 3-methylcrotonylglycine | Leucine, isoleucine, and valine metabolism | 0.971 | 0.954 to 0.988 | 0.091 | **0.975** | **0.963 to 0.986** | **0.021** |
| Isovalerylglutamine | Leucine, isoleucine, and valine metabolism | 0.978 | 0.966 to 0.991 | 0.120 | **0.980** | **0.972 to 0.989** | **0.016** |
| 7-hydroxyindole sulfate | Tryptophan metabolism | **0.944** | **0.915 to 0.973** | **0.050** | 0.961 | 0.940 to 0.982 | 0.060 |
| proline | Urea cycle; arginine and proline metabolism | **0.974** | **0.962 to 0.986** | **0.023** | 0.988 | 0.979 to 0.997 | 0.281 |
| | | **Super-Pathway: Lipid** | | | | | |
| decanoylcarnitine (C10) | Fatty acid metabolism (acyl carnitine, medium chain) | **1.035** | **1.017 to 1.054** | **0.046** | 1.023 | 1.010 to 1.036 | 0.083 |
| 5-dodecenoylcarnitine (C12:1) | Fatty acid metabolism (acyl carnitine, monounsaturated) | **1.048** | **1.027 to 1.070** | **0.009** | 1.025 | 1.010 to 1.041 | 0.110 |
| | | **Super-Pathway: Nucleotide** | | | | | |
| Thymine | Pyrimidine metabolism, thymine containing | **0.978** | **0.966 to 0.989** | **0.046** | 0.987 | 0.979 to 0.995 | 0.169 |
| | | **Super-Pathway: PCM** | | | | | |
| Glutamine conjugate of C8H12O2 (1) * | Partially characterized molecules | 1.030 | 1.013 to 1.047 | 0.065 | **1.025** | **1.013 to 1.036** | **0.021** |
| Glycine conjugate of C10H14O2 (1) * | Partially characterized molecules | 1.044 | 1.019 to 1.069 | 0.071 | **1.036** | **1.019 to 1.054** | **0.023** |
| | | **Super-Pathway: Unknown** | | | | | |
| X-11261 | | 1.032 | 1.013 to 1.051 | 0.111 | **1.025** | **1.012 to 1.039** | **0.045** |
| X-12839 | | **1.042** | **1.020 to 1.065** | **0.048** | 1.024 | 1.009 to 1.040 | 0.175 |
| X-15486 | | 1.039 | 1.015 to 1.064 | 0.118 | **1.035** | **1.018 to 1.052** | **0.021** |
| X-17676 | | 0.981 | 0.969 to 0.993 | 0.142 | **0.983** | **0.975 to 0.991** | **0.034** |
| X-21441 | | **1.043** | **1.020 to 1.067** | **0.047** | 1.029 | 1.013 to 1.046 | 0.078 |
| X-24345 | | 1.040 | 1.015 to 1.065 | 0.123 | **1.033** | **1.016 to 1.050** | **0.044** |
| X-24350 | | 1.040 | 1.014 to 1.067 | 0.156 | **1.038** | **1.020 to 1.056** | **0.020** |
| X-25003 | | **0.957** | **0.936 to 0.979** | **0.044** | 0.976 | 0.960 to 0.992 | 0.192 |
| X-25442 | | 1.041 | 1.016 to 1.067 | 0.120 | **1.038** | **1.020 to 1.055** | **0.015** |
| X-25464 | | 1.039 | 1.017 to 1.062 | 0.076 | **1.030** | **1.015 to 1.046** | **0.037** |

Estimates were generated from linear regression models, with natural log-transformed biochemicals as the dependent variable and body mass index or body fat percent as the dependent variable. Multiple testing adjustments were performed by controlling the false discovery rate at 5%. Estimates are back-transformed. The models were additionally adjusted for age, calories consumed, protein consumed, carbohydrates consumed, fat consumed, maternal gestational weight gain, maternal BMI, breastfeeding duration, smoking status, smoking household, alcohol status, maternal occupation, maternal education, metabolic equivalent of task-hours. * Compund was not identified by a standard, but we are confident in its identity. Abbreviations: CV, cofactors and vitamins; PCM, partially characterized molecules; XB, xenobiotics. Associations in females are highlighted in italic. The significant association for each metabolite is highlighted in bold.

## Appendix C : Additional Discussion for Metabolites Associated with Either BMI or BF

**Table A5.** Additional discussion for metabolites associated with either BMI or BF.

| Biochemical | Sub-Pathway | Sex | Body Composition | Discussion |
|---|---|---|---|---|
| **Super-Pathway: Amino Acid** | | | | |
| 7-hydroxyindole sulfate | Tryptophan metabolism | male | BMI | Part of the serotonin-related pathway of tryptophan [18]. A relationship to mood and depression, which has been documented to be influenced by weight and the perception thereof in adolescents [40], is a possible explanation for this association. |
| Formiminoglutamate | Histidine metabolism | male | BMI | Measurements in urine after oral application of histidine are used to determine folate deficiency [18]. Higher levels of this metabolite in the urine of individuals with higher adiposity might point to an increased need for folate. In fact, overweightness was previously shown to be associated with decreased levels of folate [41]. |
| Proline | Urea cycle; arginine and proline metabolism | male | BMI | Proline was inversely associated with adiposity in our study. This is in agreement with findings in children, in which lower levels of the metabolite have been observed in overweight children [42], but is in contrast to findings in adults [43,44]. This suggests that the relationship of adiposity with proline varies with the developmental stage of life. |
| 3-methylcrotonylglycine | Leucine, isoleucine, and valine metabolism | male | BF | A catabolite of leucine. Elevated levels of this metabolite in urine are usually found in patients with a deficiency of 3-methylcrotonyl-CoA carboxylase, an inborn error of the metabolism [18]. Decreased levels in our sample could be explained by hyperactivation of 3-methylcrotonyl-CoA carboxylase or disruption of the leucine metabolism. |
| Isovalerylglutamine | Leucine, isoleucine and valine metabolism | male | BF | No information |

**Table A5.** *Cont.*

| Biochemical | Sub-Pathway | Sex | Body Composition | Discussion |
|---|---|---|---|---|
| | | **Super Pathway: Lipid** | | |
| 5-dodecenoylcarnitine (C12:1) | Fatty acid metabolism (acyl carnitine, monounsaturated) | male | BMI | Medium-chain acylcarnitines (MCACs), see decanoylcarnitine (C10) |
| Decanoylcarnitine (C10) | Fatty acid metabolism (acyl carnitine, medium chain) | male | BMI | Decanoylcarnitine (C10) is a medium-chain fatty acid acylcarnitine that was significantly associated with higher measures of body composition. In fact, urine decanoylcarnitine has been shown to differentiate young men with normal weight from those with obesity [45], and differentiates individuals with metabolically healthy obesity from those with metabolically abnormal obesity [46]. Additionally, it is among a group of acylcarnitines that is positively related to fat oxidation [47]. It was suggested previously that high levels of medium-chain acylcarnitines (MCACs) reflect distal β- oxidation for energy use. C6 and C10 in particular are used as markers for MCAC flux [48]. Higher levels of MCAC have also been related to a disrupted branched-chain amino acid (BCAA) metabolism [49,50]. Additionally, increased levels of MCAC were suggested as markers for insulin resistance in overweight and obese individuals [51]. Increased levels of C10 in our sample are in line with the findings of previous studies in adults and children, reporting either higher levels of closely related acylcarnitines or C10 exactly [19]. However, most of these were using different tissues (e.g., blood or muscle fiber) as their biospeciminen [19]. |
| | | **Super Pathway: Nucleotide** | | |
| Thymine | Pyrimidine metabolism, thymine containing | male | BMI | Change within increasing adiposity is in line with cytosine per thymine change present in a single-nucleotide polymorphism (SNP) that is associated with BMI and BF [52]. This supports evidence that adiposity has a genetic component. Future studies should explore the relation between this SNP and adiposity |

**Table A5.** *Cont.*

| Biochemical | Sub-Pathway | Sex | Body Composition | Discussion |
|---|---|---|---|---|
| | | **Super-Pathway: PCM** | | |
| Glutamine conjugate of C8H12O2 (1) * | Partially characterized molecules | male | BF | No information |
| Glycine conjugate of C10H14O2 (1) * | Partially characterized molecules | male | BF | No information |
| | | **Super-Pathway: Unknown** | | |
| X-12839 | | male | BMI | No information |
| X-21441 | | male | BMI | No information |
| X-25003 | | male | BMI | No information |
| X-11261 | | male | BF | No information |
| X-15486 | | male | BF | No information |
| X-17676 | | male | BF | No information |
| X-24345 | | male | BF | No information |
| X-24350 | | male | BF | No information |
| X-25442 | | male | BF | No information |
| X-25464 | | male | BF | No information |

* Compound was not identified by a standard, but we are confident in its identity. Abbreviations: BF, body fat percentage; BMI, body mass index; CV, cofactors and vitamins; PCM, partially characterized molecules; XB, xenobiotics.

31

## References

1. Liu, M.; Li, L.; Chu, J.; Zhu, B.; Zhang, Q.; Yin, X.; Jiang, W.; Dai, G.; Ju, W.; Wang, Z.; et al. Serum N(1)-Methylnicotinamide Is Associated With Obesity and Diabetes in Chinese. *J. Clin. Endocrinol. Metab.* **2015**, *100*, 3112–3117. [CrossRef]

2. Thévenot, E.A.; Roux, A.; Xu, Y.; Ezan, E.; Junot, C. Analysis of the Human Adult Urine metabolome Variations with Age, Body Mass Index, and Gender by Implementing a Comprehensive Workflow for Univariate and OPLS Statistical Analyses. *J. Proteome Res.* **2015**, *14*, 3322–3335. [CrossRef] [PubMed]

3. Won, E.-Y.; Yoon, M.-K.; Kim, S.-W.; Jung, Y.; Bae, H.-W.; Lee, D.; Park, S.G.; Lee, C.-H.; Hwang, G.-S.; Chi, S.-W. Gender-Specific Metabolomic Profiling of Obesity in Leptin-Deficient ob/ob Mice by 1H NMR Spectroscopy. *PLoS ONE* **2013**, *8*, e75998. [CrossRef] [PubMed]

4. Xie, G.; Ma, X.; Zhao, A.; Wang, C.; Zhang, Y.; Nieman, D.; Nicholson, J.K.; Jia, W.; Bao, Y.; Jia, W. The Metabolite Profiles of the Obese Population Are Gender-Dependent. *J. Proteome Res.* **2014**, *13*, 4062–4073. [CrossRef] [PubMed]

5. Raven, P.W.; Taylor, N.F. Sex differences in the human metabolism of cortisol. *Endocr. Res.* **1996**, *22*, 751–755. [CrossRef]

6. Saner, C.; Harcourt, B.E.; Pandey, A.; Ellul, S.; McCallum, Z.; Kao, K.-T.; Twindyakirana, C.; Pons, A.; Alexander, E.J.; Saffery, R.; et al. Sex and puberty-related differences in metabolomic profiles associated with adiposity measures in youth with obesity. *Metabolomics* **2019**, *15*, 75. [CrossRef]

7. Oluwagbemigun, K.; Anesi, A.; Ulaszewska, M.M.; Clarke, G.; Alexy, U.; Schmid, M.; Roden, M.; Herder, C.; Mattivi, F.; Nöthlings, U. Longitudinal relationship of amino acids and indole metabolites with long-term body mass index and cardiometabolic risk markers in young individuals. *Sci. Rep.* **2020**, *10*, 1–13. [CrossRef]

8. Kim, C.; Dabelea, D.; Kalyani, R.R.; Christophi, C.A.; Bray, G.A.; Pi-Sunyer, X.; Darwin, C.H.; Yalamanchi, S.; Barrett-Connor, E.; Golden, S.H.; et al. Changes in Visceral Adiposity, Subcutaneous Adiposity, and Sex Hormones in the Diabetes Prevention Program. *J. Clin. Endocrinol. Metab.* **2017**, *102*, 3381–3389. [CrossRef]

9. Rauschert, S.; Uhl, O.; Koletzko, B.; Mori, T.A.; Beilin, L.J.; Oddy, W.H.; Hellmuth, C. Sex differences in the association of phospholipids with components of the metabolic syndrome in young adults. *Biol. Sex Differ.* **2017**, *8*, 10. [CrossRef]

10. Brown, L.M.; Gent, L.; Davis, K.; Clegg, D.J. Metabolic impact of sex hormones on obesity. *Brain Res.* **2010**, *1350*, 77–85. [CrossRef]

11. Mongraw-Chaffin, M.L.; Anderson, C.A.M.; Allison, M.A.; Ouyang, P.; Szklo, M.; Vaidya, D.; Woodward, M.; Golden, S.H. Association between sex hormones and adiposity: Qualitative differences in women and men in the multi-ethnic study of atherosclerosis. *J. Clin. Endocrinol. Metab.* **2015**, *100*, E596–E600. [CrossRef] [PubMed]

12. Clegg, D.J.; Mauvais-Jarvis, F. An integrated view of sex differences in metabolic physiology and disease. *Mol. Metab.* **2018**, *15*, 1–2. [CrossRef] [PubMed]

13. Darst, B.F.; Koscik, R.L.; Hogan, K.J.; Johnson, S.C.; Engelman, C.D. Longitudinal plasma metabolomics of aging and sex. *Aging (Albany Ny)* **2019**, *11*, 1262–1282. [CrossRef] [PubMed]

14. Rist, M.J.; Roth, A.; Frommherz, L.; Weinert, C.H.; Krüger, R.; Merz, B.; Bunzel, D.; Mack, C.; Egert, B.; Bub, A.; et al. Metabolite patterns predicting sex and age in participants of the Karlsruhe Metabolomics and Nutrition (KarMeN) study. *PLoS ONE* **2017**, *12*, e0183228. [CrossRef]

15. Kalyani, R.R.; Franco, M.; Dobs, A.S.; Ouyang, P.; Vaidya, D.; Bertoni, A.G.; Gapstur, S.M.; Golden, S.H. The association of endogenous sex hormones, adiposity, and insulin resistance with incident diabetes in postmenopausal women. *J. Clin. Endocrinol. Metab.* **2009**, *94*, 4127–4135. [CrossRef]

16. Shen, M.; Shi, H. Sex Hormones and Their Receptors Regulate Liver Energy Homeostasis. *Int. J. Endocrinol.* **2015**, *2015*, 294278. [CrossRef]

17. Wang, X.; Magkos, F.; Mittendorfer, B. Sex differences in lipid and lipoprotein metabolism: It's not just about sex hormones. *J. Clin. Endocrinol. Metab.* **2011**, *96*, 885–893. [CrossRef]

18. Wishart, D.S.; Feunang, Y.D.; Marcu, A.; Guo, A.C.; Liang, K.; Vázquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C.; Karu, N.; et al. HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Res.* **2018**, *46*, D608–D617. [CrossRef]

19. Rangel-Huerta, O.D.; Pastor-Villaescusa, B.; Gil, A. Are we close to defining a metabolomic signature of human obesity? A systematic review of metabolomics studies. *Metabolomics* **2019**, *15*, 93. [CrossRef]

32

20. Oglesbee, D.; He, M.; Majumder, N.; Vockley, J.; Ahmad, A.; Angle, B.; Burton, B.; Charrow, J.; Ensenauer, R.; Ficicioglu, C.; et al. Development of a newborn screening follow-up algorithm for the diagnosis of isobutyryl-CoA dehydrogenase deficiency. *Genet. Med.* **2007**, *9*, 108–116. [CrossRef]

21. Zampieri, T.T.; Torres-Leal, F.L.; Campaña, A.B.; Lima, F.B.; Donato, J. l-Leucine Supplementation Worsens the Adiposity of Already Obese Rats by Promoting a Hypothalamic Pattern of Gene Expression that Favors Fat Accumulation. *Nutrients* **2014**, *6*, 1364–1373. [CrossRef] [PubMed]

22. Hjerpsted, J.B.; Ritz, C.; Schou, S.S.; Tholstrup, T.; Dragsted, L.O. Effect of cheese and butter intake on metabolites in urine using an untargeted metabolomics approach. *Metabolomics* **2014**, *10*, 1176–1185. [CrossRef]

23. Elliott, P.; Posma, J.M.; Chan, Q.; Garcia-Perez, I.; Wijeyesekera, A.; Bictash, M.; Ebbels, T.M.D.; Ueshima, H.; Zhao, L.; Van Horn, L.; et al. Urinary metabolic signatures of human adiposity. *Sci. Transl. Med.* **2015**, *7*, 285ra62. [CrossRef] [PubMed]

24. Holeček, M. Branched-chain amino acids in health and disease: Metabolism, alterations in blood plasma, and as supplements. *Nutr. Metab.* **2018**, *15*, 33. [CrossRef]

25. Menon, R.M.; González, M.A.; Adams, M.H.; Tolbert, D.S.; Leu, J.H.; Cefali, E.A. Effect of the rate of niacin administration on the plasma and urine pharmacokinetics of niacin and its metabolites. *J. Clin. Pharmacol.* **2007**, *47*, 681–688. [CrossRef]

26. Jung, J.-Y.; Kim, I.Y.; Na Kim, Y.; Kim, J.-S.; Shin, J.H.; Jang, Z.-H.; Lee, H.-S.; Hwang, G.-S.; Seong, J.K. 1H NMR-based metabolite profiling of diet-induced obesity in a mouse mode. *Bmb Rep.* **2012**, *45*, 419–424. [CrossRef]

27. Pickrell, W.O.; Lacey, A.S.; Thomas, R.H.; Smith, P.E.M.; Rees, M.I. Weight change associated with antiepileptic drugs. *J. Neurol. Neurosurg. Psychiatry* **2013**, *84*, 796–799. [CrossRef]

28. Kroke, A.; Manz, F.; Kersting, M.; Remer, T.; Sichert-Hellert, W.; Alexy, U.; Lentze, M.J. The DONALD Study. History, current status and future perspectives. *Eur. J. Nutr.* **2004**, *43*, 45–54. [CrossRef]

29. Oluwagbemigun, K.; Buyken, A.E.; Alexy, U.; Schmid, M.; Herder, C.; Nöthlings, U. Developmental trajectories of body mass index from childhood into late adolescence and subsequent late adolescence-young adulthood cardiometabolic risk markers. *Cardiovasc. Diabetol.* **2019**, *18*, 9. [CrossRef]

30. Evans, A.M.; Bridgewater, B.R.; Liu, Q.; Mitchell, M.W.; Robinson, R.J.; Dai, H.; Stewart, S.J.; DeHaven, C.D.; Miller, L.A.D. High Resolution Mass Spectrometry Improves Data Quantity and Quality as Compared to Unit Mass Resolution Mass Spectrometry in High-Throughput Profiling Metabolomics. *Metabolomics* **2014**, *4*, 1–7. [CrossRef]

31. Deurenberg, P.; Pieters, J.J.L.; Hautvast, J.G.A.J. The assessment of the body fat percentage by skinfold thickness measurements in childhood and young adolescence. *Br. J. Nutr.* **1990**, *63*, 293. [CrossRef] [PubMed]

32. Sichert-Hellert, W.; Kersting, M.; Chahda, C.; Schäfer, R.; Kroke, A. German food composition database for dietary evaluations in children and adolescents. *J. Food Compos. Anal.* **2007**, *20*, 63–70. [CrossRef]

33. Schofield, W.N. Predicting basal metabolic rate, new standards and review of previous work. *Hum Nutr. Clin. Nutr.* **1985**, *39*, 5–41. [PubMed]

34. Booth, M.L.; Okely, A.D.; Chey, T.N.; Bauman, A. The reliability and validity of the Adolescent Physical Activity Recall Questionnaire. *Med. Sci. Sports Exerc.* **2002**, *34*, 1986–1995. [CrossRef]

35. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Core Team: Vienna, Austria, 2019.

36. van Buuren, S.; Groothuis-Oudshoorn, K. Mice: Multivariate Imputation by Chained Equations in R. *J. Stat. Soft.* **2011**, 1–68. [CrossRef]

37. van Buuren, S. *Flexible Imputation of Missing Data, Second edition*; Chapman and Hall/CRC: Boca Raton, FL, USA; London, UK; New York, NY, USA, 2018.

38. Independent Component Analysis. 1.0-2; R Package; Helwig, N.E. 2018.

39. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B (Methodol.)* **1995**, *57*, 289–300. [CrossRef]

40. Martyn-Nemeth, P.; Penckofer, S.; Gulanick, M.; Velsor-Friedrich, B.; Bryant, F.B. The relationships among self-esteem, stress, coping, eating behavior, and depressive mood in adolescents. *Res. Nurs. Health* **2009**, *32*, 96–109. [CrossRef]

41. Bird, J.K.; Ronnenberg, A.G.; Choi, S.-W.; Du, F.; Mason, J.B.; Liu, Z. Obesity is associated with increased red blood cell folate despite lower dietary intakes and serum concentrations. *J. Nutr.* **2015**, *145*, 79–86. [CrossRef]

42. Guevara-Cruz, M.; Vargas-Morales, J.M.; Méndez-García, A.L.; López-Barradas, A.M.; Granados-Portillo, O.; Ordaz-Nava, G.; Rocha-Viggiano, A.K.; Gutierrez-Leyte, C.A.; Medina-Cerda, E.; Rosado, J.L.; et al. Amino acid profiles of young adults differ by sex, body mass index and insulin resistance. *Nutr. Metab. Cardiovasc. Dis.* **2018**, *28*, 393–401. [CrossRef]

43. Bagheri, M.; Djazayery, A.; Farzadfar, F.; Qi, L.; Yekaninejad, M.S.; Aslibekyan, S.; Chamari, M.; Hassani, H.; Koletzko, B.; Uhl, O. Plasma metabolomic profiling of amino acids and polar lipids in Iranian obese adults. *Lipids Health Dis.* **2019**, *18*, 1–9. [CrossRef]

44. Wahl, S.; Yu, Z.; Kleber, M.; Singmann, P.; Holzapfel, C.; He, Y.; Mittelstrass, K.; Polonikov, A.; Prehn, C.; Römisch-Margl, W.; et al. Childhood obesity is associated with changes in the serum metabolite profile. *Obes. Facts* **2012**, *5*, 660–670. [CrossRef] [PubMed]

45. Wang, C.; Feng, R.; Sun, D.; Li, Y.; Bi, X.; Sun, C. Metabolic profiling of urine in young obese men using ultra performance liquid chromatography and Q-TOF mass spectrometry (UPLC/Q-TOF MS). *J. Chromatogr. B* **2011**, *879*, 2871–2876. [CrossRef] [PubMed]

46. Chen, H.-H.; Tseng, Y.J.; Wang, S.-Y.; Tsai, Y.-S.; Chang, C.-S.; Kuo, T.-C.; Yao, W.-J.; Shieh, C.-C.; Wu, C.-H.; Kuo, P.-H. The metabolome profiling and pathway analysis in metabolic healthy and abnormal obesity. *Int. J. Obes.* **2015**, *39*, 1241–1248. [CrossRef]

47. Butte, N.F.; Liu, Y.; Zakeri, I.F.; Mohney, R.P.; Mehta, N.; Voruganti, V.S.; Goring, H.; Cole, S.A.; Comuzzie, A.G. Global metabolomic profiling targeting childhood obesity in the Hispanic population. *Am. J. Clin. Nutr.* **2015**, *102*, 256–267. [CrossRef] [PubMed]

48. Baker, P.R.; Boyle, K.E.; Koves, T.R.; Ilkayeva, O.R.; Muoio, D.M.; Houmard, J.A.; Friedman, J.E. Metabolomic analysis reveals altered skeletal muscle amino acid and fatty acid handling in obese humans. *Obesity* **2015**, *23*, 981–988. [CrossRef]

49. Newgard, C.B.; An, J.; Bain, J.R.; Muehlbauer, M.J.; Stevens, R.D.; Lien, L.F.; Haqq, A.M.; Shah, S.H.; Arlotto, M.; Slentz, C.A.; et al. A Branched-Chain Amino Acid-Related Metabolic Signature that Differentiates Obese and Lean Humans and Contributes to Insulin Resistance. *Cell Metab.* **2009**, *9*, 311–326. [CrossRef]

50. Schooneman, M.G.; Vaz, F.M.; Houten, S.M.; Soeters, M.R. Acylcarnitines: Reflecting or inflicting insulin resistance? *Diabetes* **2013**, *62*, 1–8. [CrossRef]

51. Makrecka-Kuka, M.; Sevostjanovs, E.; Vilks, K.; Volska, K.; Antone, U.; Kuka, J.; Makarova, E.; Pugovics, O.; Dambrova, M.; Liepinsh, E. Plasma acylcarnitine concentrations reflect the acylcarnitine profile in cardiac tissues. *Sci. Rep.* **2017**, *7*, 17528. [CrossRef]

52. Moleres, A.; Martinez, J.A.; Marti, A. Genetics of Obesity. *Curr Obes Rep.* **2013**, *2*, 23–31. [CrossRef]

## 3.2 Publication 2: Exploring the Association between Habitual Food Intake and the Urine and Blood Metabolome in Adolescents and Young Adults: A Cohort Study

**RESEARCH ARTICLE**

# Exploring the Association between Habitual Food Intake and the Urine and Blood Metabolome in Adolescents and Young Adults: A Cohort Study

*Christian Brachem,\* Kolade Oluwagbemigun, Julia Langenau, Leonie Weinhold, Ute Alexy, Matthias Schmid, and Ute Nöthlings*

**Scope:** Habitual diet may be reflected in metabolite profiles that can improve accurate assessment of dietary exposure and further enhance our understanding of their link to health conditions. The study aims to explore the relationship of habitual food intake with blood and urine metabolites in adolescents and young adults.

**Methods:** The study population comprises 228 participants (94 males and 134 females) of the DONALD study. Dietary intake is assessed by yearly repeated 3d-food records. Habitual diet is estimated as the average consumption of 23 food groups in adolescence. Using an untargeted metabolomics approach, the study quantifies 2638 metabolites in plasma and 1407 metabolites in urine. In each sex, unique diet–metabolite associations using orthogonal projection to latent structures (oPLS) and random forests (RF) is determined.

**Results:** Six metabolites in agreement between oPLS and RF in urine, one in female (vanillylmandelate to processed/other meat) and five in males (indole-3-acetamide, and N6-methyladenosine to eggs; hippurate, citraconate/glutaconate, and X − 12111 to vegetables) are observed. No association in blood in agreement is observed.

**Conclusion:** A limited reflection of habitual food group intake by single metabolites in urine and not in blood is observed. The explored biomarkers should be confirmed in additional studies.

## 1. Background

Dietary intakes are generally obtained by self-reports using instruments like food records, 24-h recalls, or food frequency questionnaires. However, these methods are costly and prone to measurement error and improvements in dietary intake assessment are sought.[1] Multiple administrations of short-term instruments or using a combination of different instruments help to capture habitual food intake.[1,2a,b,c,3] These combinations still carry the risk of measurement error, even though diminished.[4]

Dietary biomarkers can improve the assessment of habitual food intake either replacing or strengthening self-reports.[3] With the advent of –omics technology in health sciences, it has become more feasible to explore the vast amount of data contained in the human metabolome to search for candidate biomarkers.[3,5] The metabolome provides a rich resource of potential diet biomarkers. Prior studies have already suggested putative biomarkers for habitual or recent consumption of meat,[6,7] cocoa,[8,9] wine,[7,10] cabbage,[11a,b] citrus,[12,13] and more. Andersen et al.[14] and Rebholz et al.[15] also showed that it is possible to distinguish more complex dietary patterns with markers in the metabolome. In a previous paper from our group, we conducted a systematic review of habitual food intake—blood metabolite associations replicated at least twice and identified 82 putative biomarkers,[16] 44 of which

C. Brachem, K. Oluwagbemigun, J. Langenau, U. Nöthlings
Nutritional Epidemiology
Department of Nutrition and Food Sciences
Rheinische Friedrich-Wilhelms-University Bonn
53115 Bonn, Germany
E-mail: cbrachem@uni-bonn.de

L. Weinhold, M. Schmid
Institute for Medical Biometry
Informatics and Epidemiology (IMBIE)
University Hospital Bonn
53127 Bonn, Germany

U. Alexy, U. Nöthlings
Nutritional Epidemiology
Department of Nutrition and Food Sciences
Rheinische Friedrich-Wilhelms-University Bonn
DONALD Study
Heinstück 11, 44225 Dortmund, Germany

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/mnfr.202200023

we were able to test in this study. We replicated 26 of these 44 associations.

Praticò et al.[17] proposed eight criteria to validate putative biomarkers for food intake: plausibility, dose–response, time–response, robustness, reliability, stability, analytical performance, and reproducibility. However, before the putative biomarker can be validated, candidates for validation need to be identified. For habitual intake, this can be achieved with observational studies assessing both habitual food intake and the metabolome, and exploring their associations. In order to improve the chance to validate putative biomarkers, discovery studies should already implement sensible constraints, for example uniqueness of associations to facilitate better performance in one or more validation criteria. For conclusive validation, there is a need for independent studies ideally designed specifically to evaluate one or multiple validity criteria. Additionally, biomarker discovery is not only a matter of repeated association in observational studies but should rather also be founded in biological validity of the association. Randomized controlled trials (RCTs) are an essential part of this process as well. Only the interplay of discovery in observational studies and validation in independent and well designed RCTs may lead to valid and robust markers fit to enhance or replace current assessment methods.[18a,b]

In adolescence, behaviors such as smoking, physical activity, and dietary patterns are often initiated and can be tracked into adulthood.[19] However, there is a substantial lack of studies exploring food metabolite associations in adolescents and young adults. Current diet-metabolomics studies in adolescents explored the effect of dietary counseling,[20] inadequate nutrient intake,[21] or the specific diet needed for phenylketonuria.[22a,b] However, no study to date has explored the relationship of habitual food intakes in adolescence with the metabolome. Adolescence is the phase with the largest physiological developments in human life,[19] thus intake biomarkers or the strength of the relationship between habitual food intake and the metabolome may differ from adults. Thus, our exploration of the habitual food intake–metabolome associations in adolescence will help to understand further the use of metabolomics in nutritional epidemiology.

While exploring the metabolome for markers of habitual food intake the choice of bio specimen—commonly used are blood or urine—has a large influence.[23] Both blood and urine were shown to reflect short term food intake (∼3 h after food consumption),[24] and both have been successfully used in longitudinal settings to study habitual food intake.[8,12,25,26] The choice of bio specimen often depends on facets of study designs and practical implications for future use. In order to facilitate most use cases, we investigated both 24-h urine samples and plasma blood samples for our study. The aim of the current project was to explore potential urine and blood metabolite biomarkers of habitual food intake among adolescents and young adults.

## 2. Experimental Section

### 2.1. Study Design

The present analysis was conducted in the DOrtmund Nutritional and Anthropometric Longitudinally Designed (DONALD) study.[27,28] Informed written consent was obtained from par-

ents and from participants themselves on reaching adolescence. The ethics committee at the medical faculty of the Rheinische Friedrich-Wilhelms-University Bonn, Germany (project identification: 098/06) approved the study. Briefly, DONALD was an ongoing longitudinal open cohort study situated in Dortmund, Germany, with the goal of analyzing detailed data on diet, growth, development, and metabolism between infancy and adulthood.[27,28] Participants were first examined at the age of 3 months and returned for three more visits in the first year of life, two in the second, and annually thereafter until the age of 18, when examinations started following a 5-year cycle. Examinations included 3-day weighed dietary records (3d-WDR), anthropometric measurements, collection of 24-h urine samples (starting at age 3–4), collection of blood samples (starting at age 18), and interviews on lifestyle and medical examinations. Further details on the study design had been published elsewhere.[27,28]

### 2.2. Study Participants

The current study included DONALD participants that were singletons, full term births (37–42 weeks of gestation), and had a birth weight of at least 2500 g. For the current analysis participants had to have a measurement of both the urine and blood metabolome, as well as at least two 3d-WDR in adolescence (age 10–18). Overall, 228 participants were eligible for the current study.

### 2.3. Variable Assessment

#### 2.3.1. Assessment of Usual Dietary Intake

Multiple annually applied 3d-WDR to assess dietary intake was used. Participants had to have at least two 3d-WDR. This study defined habitual intake during adolescence as the mean intake across all available 3d-WDRs between the ages 10 and 18.

Habitual food group intake was standardized to grams per 1000 kcal to account for inter- and intra-individual differences in daily calorie consumption. This study grouped single food items to 23 food groups according to our in house food composition database LEBTAB.[29] Using the database, recipes of commercial food products had been decomposed to their ingredients before food grouping. The following 23 food groups were constructed for the current analysis: dairy, cheese, eggs, beef, pork, poultry, processed and other meat, fish, animal fat, vegetable fat, cereals, pasta, potatoes, vegetables, vegetable juice, legumes, fruit, fruit juice, nuts, water, alcohol, instant beverages, and sweets.

#### 2.3.2. Untargeted Metabolomic Profiling of the Metabolome

Metabolon Inc. (Morrisville, NC, USA) performed an untargeted metabolomics assay with lipodomics on plasma and an untargeted metabolomics assay on urine samples. For both the plasma and urine untargeted assays, Metabolon used ultra-high performance liquid chromatography-tandem mass spectroscopy (UPLC-MS/MS) to identify metabolites in the samples. Peak identification was done in their propriety Laboratory Information Management System. Peaks were identified by comparison

of their retention time/index (RI), mass to charge ratio ($m/z$), and chromatographic data (e.g., MS/MS spectral data) to Metabolons library. Metabolon maintained a library of authenticated standards with over 3300 commercially available purified standard compounds as well as information on unnamed biochemical identified by recurring occurrence. Peaks were quantified using area-under-the-curve and normalized with block correction correcting for inter-day instrument tuning differences. Further details on the metabolic profiling had been reported elsewhere[30] and the full analytical report provided by Metabolon could be found in Additional File 1, Supporting Information. Both blood and urine untargeted assays were performed in this fashion. Both urine and blood metabolite values were corrected for inter day tuning differences by scaling their median to one (block correction). Urine metabolite values were additionally corrected for urine osmolality post-acquisition. Both urine and blood metabolite values were log transformed, centered to a mean of zero and scaled to a standard deviation of one prior to analysis.

Metabolon quantified 1042 (811 known and 231 unknown) and 1407 (940 known and 467 unknown) in blood and urine, respectively.

### 2.3.3. Complex Lipid Platform Measurement

Lipids were extracted from samples in methanol: dichloromethane in the presence of internal standards. The extracts were concentrated under nitrogen and reconstituted in 0.25 mL of 10 mM ammonium acetate dichloromethane: methanol (50:50). The extracts were transferred to inserts and placed in vials for infusion-MS analysis, performed on a Shimazdu LC with nano PEEK tubing and the Sciex SelexIon-5500 QTRAP. The samples were analyzed via both positive and negative mode electrospray. The 5500 QTRAP scan was performed in MRM mode with the total of more than 1100 MRMs. Individual lipid species were quantified by taking the peak area ratios of target compounds and their assigned internal standards, then multiplying by the concentration of internal standard added to the sample. Lipid class concentrations were calculated from the sum of all molecular species within a class, and fatty acid compositions were determined by calculating the proportion of each class comprised by individual fatty acids. Nine hundred sixty six lipid species in 14 classes as well as 265 fatty acids were identified.

### 2.3.4. Covariates

Body weight and height were measured at every follow-up by experienced nursing staff. Body mass index (BMI) was calculated using height (meters) and weight (kilograms) with the formula $BMI = \frac{weight}{height^2}$. This study calculated the mean age at diet examination as the mean age of the available protocols.

### 2.4. Statistical Analysis

Statistical analysis was performed using R (Version 4.0.3).[31] All analyses were stratified by sex, because in both diet and metabolome large sex differences were known.[32a,b,c,d,e,f]

#### 2.4.1. Data Analysis

First, both male and female datasets were splitted into a training (70%) and a testing (30%) set. Orthogonal projection was used to latent structures (oPLS) to explore the association between food groups and the metabolome in the training data (discovery model). The oPLS procedure used in this study was implemented for application in metabolomics[33] according to the original method by Trygg and Wold.[34] Briefly, oPLS extended PLS (projection to latent structures) by removing the non-correlated systematic variation of the descriptor variables (in this case: the metabolites). The resulting components were easier to interpret. PLS models extracted components (latent variables) on which the descriptor variables were loaded, similar to PCA (principal component analysis), with the major difference that the components were built supervised to maximally explain the variance in the $Y$-variable (in this case food group consumption). The loading on the component could be read as the strength of association between $X$ and $Y$ (metabolites and food group consumption, respectively). The 'opls()'-function from the 'ropls' R package was used to calculate the models.[33] Sevenfold cross-validation (in the training dataset) was used to select the ideal number of orthogonal components for each model. Then, the performance of the oPLS model was evaluated in the test data (Additional File 2, Supporting Information).

To validate the above findings, the study also explored the metabolome associations using random forests (RF) regression with the ''ranger'' R package[35] in the training data on default tuning parameters (validation model). Random forest was an ensemble machine learning tool that was based on combining the results of multiple decision tree models. Splits were made according to the Gini impurity, measuring how well a potential split was separating the samples of the classes (food groups) in this particular node. Food groups were used as the dependent variable and all metabolites as the independent variables. Again, in order to evaluate overfitting the model to predict group assignment in the test dataset was used (Additional File 2, Supporting Information). The results of the RF were presented as the rank of the variable importance measure, i.e., how important each variable was for making accurate predictions. The gini impurity importance measure was used.

The results from the discovery models were also used to try to replicate the 82 putative blood biomarkers identified in at least two independent study populations in a recent systematic literature search conducted by our group[16] in this study population. The oPLS model results were searched without additional criteria for the presence of these food–metabolite associations.

For the food–metabolite associations from above, this study additionally estimated their correlations adjusting for BMI, time difference between habitual diet examination and metabolome measurement (difference = age at sample collection – mean age at diet examination), and age at the sample collection of the respective bio specimen.

Metabolites were discussed further as putative biomarkers when they were 1) in the top 50 most loaded (absolute loading) on a component for only one food group in the oPLS model and 2) in the top 50 most important metabolites in the random forest model. The ranking was performed for each model individually, meaning separately for each food group and again stratified

**Table 1.** Characteristics of 228 adolescents in the DONALD study.

| | Overall, N = 228[a] | Male, N = 94[a] | Female, N = 134[a] | p-value[b] |
|---|---|---|---|---|
| Age at blood draw | 18.2 [5.9] | 18.1 [5.4] | 18.4 [6.0] | 0.2 |
| BMI at blood draw | 22.3 [4.3] | 23.2 [4.9] | 21.9 [3.8] | 0.004 |
| Age at urine collection | 18.09 [1.00] | 18.09 [1.04] | 18.09 [0.87] | 0.3 |
| BMI at urine collection | 21.9 [3.7] | 22.1 [3.8] | 21.9 [3.7] | 0.4 |
| Median age at diet examination | 14.07 [0.57] | 14.06 [0.59] | 14.07 [0.51] | 0.5 |
| Smoking status at urine collection | | | | 0.6 |
| Never | 33 (41%) | 11 (41%) | 22 (42%) | |
| Former | 25 (31%) | 7 (26%) | 18 (34%) | |
| Current | 22 (28%) | 9 (33%) | 13 (25%) | |
| Unknown | 148 | 67 | 81 | |

BMI and Age had no missing values. Mean age at diet examination is calculated as the mean age from all available 3d-WDRs; [a] Median [IQR]; n (%); [b] Wilcoxon rank sum test; Pearson's Chi-squared test.

by sex. For these metabolites, this study additionally computed partial correlations between the food group intake and the metabolite value, adjusting for BMI at sample collection, time difference between habitual diet examination and metabolome measurement, and age at sample collection. The 'pcor()' function was used from the 'ppcor' package to calculate the partial correlations.

### 2.4.2. Missing Values

Because every participant had at least two 3d-WDR, this study did not impute the missing dietary records for participants with less than their maximum possible 3d-WDR.

Metabolites were excluded from the analysis when more than 70% of data were missing. Based on this, the study excluded 91 and 67 metabolites in female blood and urine, respectively, and 87 and 74 metabolites in male blood and urine, respectively.

Missing values in the oPLS models were handled by applying the NIPALS ("Nonlinear Iterative Partial Least Squares") algorithm for the estimation of components. The method was first described in 1966 by Wold[36] for PCA and had been expanded to many similar methods, including oPLS. The advantage of NI-PALS was, that it did not impute missing values, but rather the algorithm could iterate with missing values in the data, with only computation time rising with larger percentages of missing values. For the random forest model, this study performed a single imputation with the 'missRanger()' function, using 10 trees with a maximum depth of six and three non-missing candidate values for predictive mean matching.

For partial correlation coefficients this study used pairwise complete cases, of metabolite, food group, BMI at sample collection, and age at sample collection.

## 3. Results

Males and females in our study population were similar in many characteristics. Females had slightly lower BMI (−1.3 and −0.2 median BMI, at blood and urine collection, respectively) (**Table 1**). Participants had a median age of 18.20 [IQR: 5.50] years at blood draw, and 18.09 [IQR: 1.00] years at urine collection.

**Table 2.** Median habitual food group consumption of 228 males and females in the DONALD study.

| Food group b) | Overall, N = 228[a] | Male, N = 94[a] | Female, N = 134[a] |
|---|---|---|---|
| Dairy | 88 [70] | 96 [77] | 82 [62] |
| Cheese | 9 [8] | 10 [8] | 9 [8] |
| Eggs | 4.24 [3.61] | 4.13 [3.54] | 4.35 [3.82] |
| Beef | 3.2 [4.2] | 3.5 [4.6] | 2.7 [4.0] |
| Pork | 5.2 [5.8] | 6.8 [6.5] | 4.6 [5.4] |
| Poultry | 3.1 [4.7] | 3.4 [5.2] | 2.9 [4.3] |
| Other meat | 6.9 [5.9] | 7.6 [7.5] | 6.2 [4.9] |
| Fish | 2.02 [3.87] | 2.22 [3.86] | 1.75 [3.60] |
| Animal fat | 4.58 [3.13] | 4.15 [2.67] | 5.06 [3.25] |
| Vegetable fat | 8.3 [4.7] | 8.0 [3.8] | 8.7 [5.3] |
| Cereals | 133 [42] | 132 [40] | 136 [41] |
| Pasta | 0.07 [1.01] | 0.14 [1.22] | 0.00 [0.85] |
| Potatoes | 18 [13] | 18 [13] | 18 [13] |
| Vegetables | 29 [22] | 24 [21] | 32 [24] |
| Vegetable juice | 0.00 [0.03] | 0.00 [0.04] | 0.00 [0.02] |
| Legumes | 0.69 [1.91] | 0.58 [1.82] | 0.73 [2.05] |
| Fruit | 34 [30] | 27 [30] | 41 [33] |
| Fruit Juice | 35 [38] | 37 [41] | 35 [37] |
| Nuts | 2.38 [1.98] | 2.47 [2.45] | 2.37 [1.62] |
| Water | 289 [186] | 288 [188] | 295 [169] |
| Alcohol | 0.1 [2.1] | 0.2 [7.8] | 0.1 [0.7] |
| Beverages, instant | 0.00 [0.02] | 0.00 [0.01] | 0.00 [0.03] |
| Sweets | 40 [18] | 39 [20] | 40 [17] |

[a] Median [IQR]; [b] Consumption is measured in $\frac{g}{1000\ kcal}$, averaged over all available protocols before urine collection and in adolescence (age 10–20) for each participant.

Notably different intakes across sexes were the higher dairy, fruit juice, and alcohol consumption in males, as well as the higher fruit and higher vegetable consumption in females. For the other food groups, males and females reported similar median consumption amounts (**Table 2**), though energy intake differed significantly, as expected. Males had a median daily caloric

**Table 3.** Male—metabolites associated uniquely in oPLS and important in RF.

| Bio specimen | Biochemical | Super pathway | Sub pathway | Loading[a] | RF rank[b] | Partial correlation[c] |
|---|---|---|---|---|---|---|
| Eggs | | | | | | |
| Urine | Indole-3-acetamide | Xenobiotics | Chemical | −0.09 | 28 | −0.25 |
| Urine | N6-methyladenosine | Nucleotide | Purine metabolism, adenine containing | −0.10 | 31 | −0.14 |
| Vegetables | | | | | | |
| Urine | Hippurate | Xenobiotics | Benzoate metabolism | 0.07 | 21 | 0.06 |
| Urine | Citraconate/glutaconate | Energy | TCA cycle | 0.07 | 31 | 0.14 |
| Urine | X - 12111 | | | 0.07 | 35 | −0.12 |

[a] Loading is generated from oPLS Models; [b] RF Rank is the rank according to the importance measure in random forest; [c] Partial correlation is partial for BMI, time difference between habitual diet and metabolome measurement, and age.

intake of 2329 kcal and females had 1923 kcal. Participants completed on average 7.76 of the eight possible 3d-WDRs. Males completed slightly fewer protocols on average (males = 7.69 as compared to females = 7.81).

The complete model results, without additional constraints, are available in Additional File 2, Supporting Information. The cross-validation search in the oPLS models for number of components suggested three components for the combination of eggs and male urine as well as vegetables and male urine. For other and processed meats and male blood, female urine, and male urine, fruit juice and female urine, pork and male urine, and cheese and male blood the search suggested two components. No other model surpassed the cross-validation threshold to select a component (Q2Y improvement), suggesting that the predictive value of these models is too low.

In the oPLS models, 34 (of the top 50) metabolites were uniquely loaded on components for other and processed meat across both male and female sex in urine (Additional File 3, Supporting Information). No metabolite in the top 50 was associated uniquely with a food group in both urine and blood (across bio specimen). In the RF model alcohol (seven metabolites), other meat (seven metabolites), dairy (six metabolites), fruit juice (six metabolites), pork (six metabolites), eggs (five metabolites), pasta (five metabolites), legumes (five metabolites), water (five metabolites), potatoes (four metabolites), cereals (three metabolites), fish (three metabolites), vegetables (three metabolites), vegetable juice (three metabolites), poultry (three metabolites), vegetable fat (two metabolites), nuts (two metabolites), fruit (two metabolites), beef (one metabolite), beverages, instant (one metabolite), animal fat (one metabolite), cheese (one metabolite) and, sweets (one metabolite) were ranked in the top 50 in both sexes in blood. In urine processed and other meat (10), poultry (eight metabolites), cereals (four metabolites), beef (three metabolites), animal fat (three metabolites), dairy (three metabolites), vegetables (three metabolites), fruit (three metabolites), fruit juice (three metabolites), pork (three metabolites), alcohol (two metabolites), eggs (two metabolites), vegetable fat (two metabolites), fish (two metabolites), legumes (two metabolites), sweets (two metabolites), water (two metabolites), cheese (one metabolite), potatoes (one metabolite), and pasta (one metabolite) had metabolites ranked in the top 50 in both sexes (Additional File 4, Supporting Information). In the RF models, there is no metabolite ranked in the top 50 in both bio specimen as well. We additionally provide

spearman correlation coefficients for all metabolite–food group combinations in Additional File 5, Supporting Information. As shown in **Table 3**, in male participants, we found five unique food–metabolite associations across methods in urine and none in blood. The associations were eggs with indole-3-acetamide, and N6-methyladenosine, as well as vegetables with hippurate, citraconate/glutaconate, and X – 12111.

As shown in **Table 4**, in females, we found one unique food–metabolite association across methods in urine and none in blood. The metabolite vanillylmandelate (VMA) is associated with processed and other meat.

Of the 82 prior reported blood metabolites,[16] we have measurements of 31 in blood and 32 in urine. Of these, 28 are measured both in urine and in blood. Combining this with roughly similar food groups, we were able to search our models for eight possible replications, where both food group and metabolite were available in our data. We were able to replicate one association in male urine: pork consumption and the unknown metabolites X – 11381. The metabolite was additionally associated with eggs and vegetables in our models (**Table 5**). The metabolite has a partial correlation with habitual pork consumption of $r = -0.10$.

## 4. Discussion

This study among 228 adolescents and young adults suggests that habitual food intake is reflected to a limited extent in the metabolome when focusing on single metabolites. A limited number of associations were found in urine, but not in blood. The associations we found were also sex dependent. In our discovery analysis, we identified six uniquely associated putative biomarkers in urine. We observed more and different associations in males compared to females. To our knowledge, this is the first study to relate habitual consumption of multiple food groups assessed with food records to both the untargeted urine and the blood metabolome in adolescents and young adults. We additionally were able to replicate one association between metabolites and food groups that were associated with more than one food group.

One of our findings relates to hippurate or hippuric acid which is an acyl glycine formed in the conjugation of benzoic acid with glycine.[37] It was associated with increased vegetable consumption in male urine. It is a regular component of urine and increased levels were previously associated with increased

**Table 4.** Females—metabolites associated uniquely in oPLS and important in RF.

| Bio specimen | Biochemical | Super pathway | Sub pathway | Loading[a] | RF rank[b] | Partial correlation[c] |
|---|---|---|---|---|---|---|
| Processed and other meat | | | | | | |
| Urine | Vanillylmandelate (VMA) | Amino acid | Tyrosine metabolism | 0.06 | 42 | −0.00 |

[a] Loading is generated from oPLS Models; [b] RF Rank is the rank according to the importance measure in random forest; [c] Partial correlation is partial for BMI, time difference between habitual diet and metabolome measurement, and age.

**Table 5.** Metabolites replicated from literature.

| Sex | Food group | Food group (Literature) | Biochemical | Other associated food Groups | Loading[a] | RF rank[b] | Partial correlation[c] |
|---|---|---|---|---|---|---|---|
| Urine | | | | | | | |
| Male | Pork | Red meat | X - 11381 | Eggs, vegetables | −0.05 | 884 | −0.10 |

[a] Loading is generated from the oPLS Model; [b] RF Rank is the rank according to the importance measure in random forest; [c] Partial correlation is partial for BMI, time difference between habitual diet and metabolome measurement, and age.

consumption of phenolic compounds such as tea, wine, fruits, or vegetables.[37] Previous investigations in a different subpopulation of our study cohort documented an association between hippuric acid and fruit and vegetable consumption[38] and later suggested urinary hippuric acid excretion as a biomarker for dietary flavonoid intake from fruit and vegetables.[39] This association was recently also replicated in adults.[40] This suggests that hippuric acid may be a good candidate biomarker for vegetable intake, especially for high polyphenol containing vegetables.

Indole-3-acetamide is a 3-alkylindole[37] and is also known as 2-(3-indolyl)acetamide or IAM. In our analysis, its urine levels were associated negatively with habitual egg consumption in males. It was previously detected in several foods, like butternut squash, pineapples, or the common pea.[37] IAM is one of the bacterial tryptophan catabolites[41] produced in the gut and eggs are among the richest sources of dietary tryptophan.[42] Additionally, peas are a common replacement for soy in chicken feed,[43] therefore this association might be due to changes in the food chain.

N6-Methyladenosine (m6A) is an endogenous urinary nucleoside product of the degradation of tRNA. It was associated with decreased egg consumption in males in our sample. Unfortunately, there are no prior studies relating urine levels to dietary intake. However, one of the dietary factors that regulate m6A-RNA-methylation, betaine,[44] was associated in plasma with egg consumption,[45] which may be the pathway for the association between m6A and eggs.

Vanillylmandelate (VMA) was associated negatively with processed and other meat in female urine. VMA is a chemical intermediate in the synthesis of artificial vanilla flavorings and an end-stage metabolite of the catecholamines.[37] Processed meat has been previously associated with the catecholamine metabolism.[46] It may be associated in females specifically because of different consumption habits within processed and other meat. Females in our sample ate more game, less sausages, and less meat dishes compared to males. However, most of the consumption in the processed and other meat group for both sexes is due to sausages. VMA may be associated specifically with processed and other meats due to different levels of catecholamines contained in the meats used in the production of the processed meat products. It was additionally asso-

ciated with an increased risk for colorectal cancer,[47] which is known to be associated with higher consumption of processed meats.[48]

Citraconate/glutaconate is an energy metabolite in the TCA cycle. In our analysis, its urine levels were associated with vegetable consumption in males. It has been previously reported in association with habitual coffee-consumption in postmenopausal women[25] and with a short term diet intervention of whole grain wheat intake versus refined grain wheat intake.[49] In our sample, this metabolite was associated with coffee consumption as well, though only few participants were consuming coffee. In further analysis, we observed that the association between vegetable intake and citraconate/glutaconate was independent of coffee intake. This suggests that the association between vegetable intake and citraconate/glutaconate is unlikely to be influenced by coffee intake.

The unknown metabolite X – 12111 in male urine was associated with vegetable consumption in our study. It was previously associated with habitual consumption of total citrus and juices.[50] However, this metabolite is not structurally identified so that further discussion is not possible, yet.

Furthermore, we replicated the association of the unknown metabolite X – 11381 with pork in the urine of males. As before, this metabolite is not structurally identified, therefore identification is needed prior to additional discussion is possible.

There are other well-documented diet–metabolite associations that we were not able to replicate in our study, because we did not have a measure for this specific metabolite (e.g., proline betaine). This is due to the untargeted nature of our metabolomics analysis. Additionally, we used multiple 3d-WDRs to estimate habitual food intake, compared to most other studies using food frequency questionnaires.

Our approach to combine two sophisticated statistical methods and apply rigid criteria for further discussion proved successful in this context. With hippurate, we identified a very promising putative biomarker that other studies with different contexts suggested previously as well. This gives us further confidence that the stringent criteria succeeded to suggest associations that are unlikely to be spurious. However, discovery of markers of habitual intake in population-based studies is a complex multi-step

**ADVANCED**
SCIENCE NEWS
www.advancedsciencenews.com

**Molecular Nutrition**
Food Research
www.mnf-journal.com

process, ideally including explorative studies (like ours) and more targeted analyses focusing in on single food groups. No universal standard analysis approach has emerged, yet, which would greatly increase the replicability of studies. The low correlation between the food and metabolites suggests that linear association may be a suboptimal measure of the biological effect of the current foods on their associated metabolites in this study population.

Though successful in our more homogeneously collected urine metabolome data, our approach suggested molecules only for a limited number of food groups and with varying degrees of plausibility. In the more heterogeneously collected blood metabolome data the combination approach was not successful, as the different methods selected different kinds of metabolites—oPLS suggested mostly lipid-based metabolites while RF did not select the lipid-based metabolites even in the top 500 most relevant metabolites. Based on these findings we conducted a sensitivity analysis separating blood metabolites and blood lipodomics, resulting in some blood metabolites associated in females according to our criteria and no lipodomics associated. We included these results in Additional File 6, Supporting Information but emphasize the large impact of excluding half the variables from a model is bound to have as well as the deviation from our analysis protocol due to results. These results should be considered only for exploratory analysis. Besides, the choice of statistical methods seems to be have a huge impact on the results of metabolomics analysis.

Additionally it is important to remember that, though the metabolome measurement in our study is untargeted it is not "complete." The human metabolome database contains roughly 220,000 human metabolites,[37] while single studies can identify thousands and more metabolites in their sample this still represent only a fraction of all available metabolites. Furthermore, food groups are very heterogeneously defined among studies, for example, dairy might include cheeses in some but not all studies. These differences might be an additional hurdle to overcome in the replication of food–metabolite associations. A single study reporting few or no associations is therefore not a suggestion that there is nothing to find, but rather that in this specific study population with its characteristics, food groups, and measured metabolites there was no association. Replications in other studies are necessary to confirm our findings.

### 4.1. Strengths and Limitations

The present study has some notable strengths. We investigated the associations between the habitual group food intake, estimated by repeated food records, and the global metabolome from two biofluids to achieve a more comprehensive overview of the metabolome response to habitual food intake. We used multiple 3d-WDR as the measurement of habitual food intake, as well as 24 h urine samples and blood plasma in a comparatively large study population with an untargeted metabolomics approach. We employed two sophisticated statistical analysis and stringent criteria to identify putative biomarkers in order to increase the reliability of our results. Identification of putative intake biomarkers in explorative analysis depend on the methods for measurement of the metabolome, the information on diet, the size of the study population, and the statistical analysis.[51] By combining parametric and non-parametric statistical methods, we aimed to decrease the influence of method choice on our results and in turn increase reproducibility. Additionally, we constrained novel associations to be uniquely associated with only one food group to increase plausibility, reliability, and reproducibility. We adapted our food group definitions to better fit food groups in the literature and increase the chance of reproducibility. However, because there is no universal standard to define food groups used across studies and different assessment instruments, heterogeneity in the food group definition remains.

On the other hand, we acknowledge several limitations to the study. First, our participants are all Caucasians (Germans), residing in a large city (Dortmund) and surroundings, and are mostly from a high socioeconomic background. This may limit the generalizability of our findings. Further, because we did not select participants based on their diet, the ranges of food consumption amounts were not as large in all food groups as in studies designed with one comparison in mind. We used one type of dietary assessment, 3-day food records, albeit repeated. Including another dietary assessment instrument such as the FFQ may improve the estimation of habitual dietary intake. The current method of estimation may underestimate infrequently consumed food groups and over estimate food groups consumed in large quantities during the observation period. The time between 3d-WDR and blood collection differed greatly between participants. We accounted for this by adjusting for the time difference, yet heterogeneity in these data remain. Additionally, we cannot rule out residual confounding by either unknown or unmeasured factors such as genetic influences or cooking method. Lastly, because we only had one measurement of urine and blood metabolites the temporal reproducibility of these findings is unknown.

### 4.2. Future Research

Future research should try to validate the putative biomarkers we have identified. Additional research should further aim to identify more biological pathways of known diet–metabolite associations. The unknown or partially characterized molecules that were associated in this study should be identified, as they have potential to help elucidate the biological mechanisms of the relationship between diet and the metabolome as well as and may function as possible biomarkers. Additionally, more studies are needed that stratify their metabolomic analysis by sex, in order to increase our understanding of the physiological differences in metabolism between males and females. Furthermore, future studies should try to replicate our findings in an independent population and extend this analysis to repeated metabolome measurements. Further investigation of the identified compounds as potential biomarkers, for example, in feeding intervention trials, or their exact metabolic pathway would be especially valuable as well. Future studies would benefit greatly from an integrated model where intervention studies and population-based studies inform each other, while using the exact same metabolome measurement. Overall, metabolomics would benefit greatly from both a more unified data analysis approach as well as a unified measurement approach to better facilitate meta-analysis and replication of results from different studies.

Future studies could aim to focus more on single food groups, utilizing controlled-feeding trials and observational studies where participants are selected to maximize the difference between intakes of a single food group. For explorative studies like our present study, a combination of two or more statistical methods seems to deliver promising results.

## 5. Conclusion

We observed a limited reflection of habitual diet in the urine metabolome, identifying six putative biomarkers of habitual food intake in urine of adolescents and young adults in two independent statistical methods. Additionally, we replicated one association from the literature in the urine of male adolescents. We propose to consider these metabolites as biomarkers of habitual intake in future validation studies.

## Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

## Acknowledgements

## Conflict of Interest

The authors declare no conflict of interest.

## Author Contributions

Conceptualization, C.B., K.O. and U.N.; methodology C.B., K.O., L.W., M.S., and U.N.; software, C.B.; validation, J.L. and K.O.; formal analysis, C.B.; investigation, C.B.; resources, K.O., U.A. and U.N.; data curation, K.O., U.A. and U.N.; writing—original draft preparation, C.B.; writing—review and editing, C.B., J.L., L.W., M.S., U.A., K.O. and U.N.; visualization, C.B.; supervision, K.O., M.S. and U.N.; project administration, K.O. and U.N.; funding acquisition, U.N. All authors have read and agreed to the final version of the manuscript.

## Data Availability Statement

The datasets generated and/or analyzed during the current study are not publicly available due data protection concerns for sensitive data, but are available on reasonable request and approval of the principal investigator. All model results are available in the supplement to this article (Additional File 1).

## Keywords

biomarker, blood metabolome, DONALD study, habitual food intake, metabolomics, plasma, urine metabolome, 24-h urine

[1] J. Conrad, U. Nöthlings, *Proc. Nutr. Soc.* **2017**, *76*, 213.
[2] a) G. H. C. Laureano, V. B. L. Torman, S. P. Crispim, A. L. M. Dekkers, S. A. Camey, *Nutrients* **2016**, *8*, 166; b) J. Conrad, S. A. J. Koch, U. Nöthlings, *Curr. Opin. Clin. Nutr. Metab. Care* **2018**, *21*, 343; c) A. Naska, A. Lagiou, P. Lagiou, *F1000 Res.* **2017**, *6*, 926.
[3] A. Scalbert, L. Brennan, C. Manach, C. Andres-Lacueva, L. O. Dragsted, J. Draper, S. M. Rappaport, J. J. J. van der Hooft, D. S. Wishart, *Am. J. Clin. Nutr.* **2014**, *99*, 1286.
[4] a) R. J. Carroll, D. Midthune, A. F. Subar, M. Shumakovich, L. S. Freedman, F. E. Thompson, V. Kipnis, *Am. J. Epidemiol.* **2012**, *175*, 340; b) A.-K. Illner, U. Harttig, G. Tognon, D. Palli, S. Salvini, E. Bower, P. Amiano, T. Kassik, A. Metspalu, D. Engeset, E. Lund, H. Ward, N. Slimani, M. Bergmann, K. Wagner, H. Boeing, *Public Health Nutr.* **2011**, *14*, 1055.
[5] a) L. O. Dragsted, Q. Gao, G. Praticò, C. Manach, D. S. Wishart, A. Scalbert, E. J. M. Feskens, *Genes Nutr.* **2017**, *12*, 24; b) A. J. Lloyd, M. Beckmann, S. Haldar, C. Seal, K. Brandt, J. Draper, *Am. J. Clin. Nutr.* **2013**, *97*, 377.
[6] a) J. Xu, S. Yang, S. Cai, J. Dong, X. Li, Z. Chen, *Anal. Bioanal. Chem.* **2010**, *396*, 1451; b) C. Stella, B. Beckwith-Hall, O. Cloarec, E. Holmes, J. C. Lindon, J. Powell, F. van der Ouderaa, S. Bingham, A. J. Cross, J. K. Nicholson, *J. Proteome Res.* **2006**, *5*, 2780.
[7] S. S. Heinzmann, C. A. Merrifield, S. Rezzi, S. Kochhar, J. C. Lindon, E. Holmes, J. K. Nicholson, *J. Proteome Res.* **2012**, *11*, 643.
[8] M. Garcia-Aloy, R. Llorach, M. Urpi-Sarda, O. Jáuregui, D. Corella, M. Ruiz-Canela, J. Salas-Salvadó, M. Fitó, E. Ros, R. Estruch, C. Andres-Lacueva, *Mol. Nutr. Food Res.* **2015**, *59*, 212.
[9] R. Llorach, M. Urpi-Sarda, O. Jauregui, M. Monagas, C. Andres-Lacueva, *J. Proteome Res.* **2009**, *8*, 5060.
[10] R. Zamora-Ros, M. Urpí-Sardà, R. M. Lamuela-Raventós, R. Estruch, M. A. Martínez-González, M. Bulló, F. Arós, A. Cherubini, C. Andres-Lacueva, *FreeRadic. Biol. Med.* **2009**, *46*, 1562.
[11] a) D. H. May, S. L. Navarro, I. Ruczinski, J. Hogan, Y. Ogata, Y. Schwarz, L. Levy, T. Holzman, M. W. McIntosh, J. W. Lampe, *Br. J. Nutr.* **2013**, *110*, 1760; b) W. M. B. Edmands, O. P. Beckonert, C. Stella, A. Campbell, B. G. Lake, J. C. Lindon, E. Holmes, N. J. Gooderham, *J. Proteome Res.* **2011**, *10*, 4513.
[12] E. Pujos-Guillot, J. Hubert, J.-F. Martin, B. Lyan, M. Quintana, S. Claude, B. Chabanas, J. A. Rothwell, C. Bennetau-Pelissero, A. Scalbert, B. Comte, S. Hercberg, C. Morand, P. Galan, C. Manach, *J. Proteome Res.* **2013**, *12*, 1645.
[13] S. S. Heinzmann, I. J. Brown, Q. Chan, M. Bictash, M.-E. Dumas, S. Kochhar, J. Stamler, E. Holmes, P. Elliott, J. K. Nicholson, *Am. J. Clin. Nutr.* **2010**, *92*, 436.
[14] M.-B. S. Andersen, Å. Rinnan, C. Manach, S. K. Poulsen, E. Pujos-Guillot, T. M. Larsen, A. Astrup, L. O. Dragsted, *J. Proteome Res.* **2014**, *13*, 1405.
[15] C. M. Rebholz, A. H. Lichtenstein, Z. Zheng, L. J. Appel, J. Coresh, *Am. J. Clin. Nutr.* **2018**, *108*, 243.
[16] J. Langenau, K. Oluwagbemigun, C. Brachem, W. Lieb, R. Di Giuseppe, A. Artati, G. Kastenmüller, L. Weinhold, M. Schmid, U. Nöthlings, *Metabolites* **2020**, *10*, 468.
[17] G. Praticò, Q. Gao, A. Scalbert, G. Vergères, M. Kolehmainen, C. Manach, L. Brennan, S. H. Pedapati, L. A. Afman, D. S. Wishart, R. Vázquez-Fresno, C. Andres-Lacueva, M. Garcia-Aloy, H. Verhagen, E. J. M. Feskens, L. O. Dragsted, *Genes Nutr.* **2018**, 13.
[18] a) J. de Vries, J.-M. Antoine, T. Burzykowski, A. Chiodini, M. Gibney, G. Kuhnle, A. Méheust, L. Pijls, I. Rowland, *Eur. J. Nutr.* **2013**, *52*,

1685; b) P. J. Aggett, J.-M. Antoine, N.-G. Asp, F. Bellisle, L. Contor, J. H. Cummings, J. Howlett, D. J. G. Müller, C. Persin, L. T. J. Pijls, G. Rechkemmer, S. Tuijtelaars, H. Verhagen, *Eur. J. Nutr.* **2005**, *44*, i5.

[19] R. M. Viner, D. Ross, R. Hardy, D. Kuh, C. Power, A. Johnson, K. Wellings, J. McCambridge, T. J. Cole, Y. Kelly, G. D. Batty, *J. Epidemiol. Community Health* **2015**, *69*, 719.

[20] M. Lehtovirta, K. Pahkala, H. Niinikoski, A. J. Kangas, P. Soininen, H. Lagström, J. S. A. Viikari, T. Rönnemaa, A. Jula, M. Ala-Korpela, P. Würtz, O. T. Raitakari, *J. Pediatr.* **2018**, *195*, 190.e3.

[21] K. J. Li, N. Jenkins, G. Luckasen, S. Rao, E. P. Ryan, *PLoS ONE* **2018**, *13*, e0205899.

[22] a) B. M. Stroup, N. Nair, S. G. Murali, K. Broniowska, F. Rohr, H. L. Levy, D. M. Ney, *J. Nutr.* **2018**, *148*, 194; b) D. M. Ney, S. G. Murali, B. M. Stroup, N. Nair, E. A. Sawin, F. Rohr, H. L. Levy, *Mol. Genet. Metab.* **2017**, *121*, 96.

[23] J.-L. Sébédio, *Adv. Food Nutr. Res.* **2017**, *82*, 83.

[24] S. Primrose, J. Draper, R. Elsom, V. Kirkpatrick, J. C. Mathers, C. Seal, M. Beckmann, S. Haldar, J. H. Beattie, J. K. Lodge, M. Jenab, H. Keun, A. Scalbert, *Br. J. Nutr.* **2011**, *105*, 1277.

[25] Y. Wang, S. M. Gapstur, B. D. Carter, T. J. Hartman, V. L. Stevens, M. M. Gaudet, M. L. McCullough, *J. Nutr.* **2018**, *148*, 932.

[26] T. Pallister, A. Jennings, R. P. Mohney, D. Yarand, M. Mangino, A. Cassidy, A. MacGregor, T. D. Spector, C. Menni, *PLOS ONE* **2016**, *11*, e0158568.

[27] A. Kroke, F. Manz, M. Kersting, T. Remer, W. Sichert-Hellert, U. Alexy, M. J. Lentze, *Eur. J. Nutr.* **2004**, *43*, 45.

[28] A. E. Buyken, U. Alexy, M. Kersting, T. Remer, *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* **2012**, *55*, 875.

[29] W. Sichert-Hellert, M. Kersting, C. Chahda, R. Schäfer, A. Kroke, *J. Food Compos. Anal.* **2007**, *20*, 63.

[30] A. Evans, B. R. Bridgewater, Q. Liu, M. W. Mitchell, R. J. Robinson, H. Dai, S. J. Stewart, C. D. DeHaven, L. A. Miller, *Metabolomics* **2014**, *04*, 1000132.

[31] R Core Team, *R: A Language and Environment for Statistical Computing*, Vienna, Austria **2019**.

[32] a) J. Krumsiek, K. Mittelstrass, K. T. Do, F. Stückler, J. Ried, J. Adamski, A. Peters, T. Illig, F. Kronenberg, N. Friedrich, M. Nauck, M. Pietzner, D. O. Mook-Kanamori, K. Suhre, C. Gieger, H. Grallert, F. J. Theis, G. Kastenmüller, *Metabolomics* **2015**, *11*, 1815; b) G. Sergi, M. Dianin, A. Bertocco, B. M. Zanforlini, C. Curreri, M. Mazzochin, L. A. Simons, E. Manzato, C. Trevisan, *Nutr. Metab. Cardiovasc. Dis.* **2020**, *30*, 1452; c) S. Siddiqui, H. Zainal, S. N. Harun, S. M. Sheikh Ghadzi, S. Ghafoor, *Diabetes Metab. Syndr.* **2020**, *14*, 1243; d) M. J. Rist, A. Roth, L. Frommherz, C. H. Weinert, R. Krüger, B. Merz, D. Bunzel, C. Mack, B. Egert, A. Bub, B. Görling, P. Tzvetkova, B. Luy, I. Hoffmann, S. E. Kulling, B. Watzl, *PLOS ONE* **2017**, *12*, e0183228; e) D. J. Clegg,

F. Mauvais-Jarvis, *Mol. Metab.* **2018**, *15*, 1; f) B. F. Darst, R. L. Koscik, K. J. Hogan, S. C. Johnson, C. D. Engelman, *Aging (Albany NY)* **2019**, *11*, 1262.

[33] E. A. Thévenot, A. Roux, Y. Xu, E. Ezan, C. Junot, *J. Proteome Res.* **2015**, *14*, 3322.

[34] J. Trygg, S. Wold, *J. Chemometrics* **2002**, *16*, 119.

[35] M. N. Wright, A. Ziegler, *J. Stat. Soft.* **2017**, *77*. DOI: https://doi.org/10.18637/jss.v077.i01.

[36] H. Wold in *Multivariate analysis; proceedings* (Ed.: P. R. Krishnaiah), Academic Press, New York, **1966**, pp. S. 394–420.

[37] D. S. Wishart, Y. D. Feunang, A. Marcu, A. C. Guo, K. Liang, R. Vázquez-Fresno, T. Sajed, D. Johnson, C. Li, N. Karu, Z. Sayeeda, E. Lo, N. Assempour, M. Berjanskii, S. Singhal, D. Arndt, Y. Liang, H. Badran, J. Grant, A. Serra-Cayuela, Y. Liu, R. Mandal, V. Neveu, A. Pon, C. Knox, M. Wilson, C. Manach, A. Scalbert, *Nucleic Acids Res.* **2018**, *46*, D608.

[38] D. Krupp, N. Doberstein, L. Shi, T. Remer, *J. Nutr.* **2012**, *142*, 1314.

[39] K. J. Penczynski, D. Krupp, A. Bring, K. Bolzenius, T. Remer, A. E. Buyken, *Eur. J. Nutr.* **2017**, *56*, 757.

[40] E. D. Clarke, M. E. Rollo, C. E. Collins, L. Wood, R. Callister, M. Philo, P. A. Kroon, R. L. Haslam, *Nutrients* **2020**, *12*, 3431.

[41] a) T. D. Hubbard, I. A. Murray, G. H. Perdew, *Drug Metab. Dispos.* **2015**, *43*, 1522; b) A. Agus, J. Planchais, H. Sokol, *Cell Host Microbe.* **2018**, *23*, 716.

[42] Y. A. Attia, M. A. Al-Harthi, M. A. Korish, M. H. Shiboob, *Animals (Basel)* **2020**, *10*, 406.

[43] a) F. A. Igbasan, W. Guenter, *Poultry Sci.* **1997**, *76*, 331; b) F. Fru-Nji, E. Niess, E. Pfeffer, *J. Poult. Sci.* **2007**, *44*, 34.

[44] J. Wu, Y. Li, J. Yu, Z. Gan, W. Wei, C. Wang, L. Zhang, T. Wang, X. Zhong, *Front. Pharmacol.* **2020**, *11*, 568006.

[45] C. Zhu, L. Sawrey-Kubicek, A. S. Bardagjy, H. Houts, X. Tang, R. Sacchi, J. M. Randolph, F. M. Steinberg, A. M. Zivkovic, *Nutr. Res.* **2020**, *78*, 36.

[46] W. H. A. de Jong, G. Eisenhofer, W. J. Post, F. A. J. Muskiet, E. G. E. de Vries, I. P. Kema, *J. Clin. Endocrinol. Metab.* **2009**, *94*, 2841.

[47] M. L. McCullough, R. A. Hodge, P. T. Campbell, V. L. Stevens, Y. Wang, *Metabolites* **2021**, *11*.

[48] R. L. Santarelli, F. Pierre, D. E. Corpet, *Nutr. Cancer* **2008**, *60*, 131.

[49] Y. Zhu, P. Wang, W. Sha, S. Sang, *Sci. Rep.* **2016**, *6*, 36278.

[50] M. C. Playdon, J. N. Sampson, A. J. Cross, R. Sinha, K. A. Guertin, K. A. Moy, N. Rothman, M. L. Irwin, S. T. Mayne, R. Stolzenberg-Solomon, S. C. Moore, *Am. J. Clin. Nutr.* **2016**, *104*, 776.

[51] J. Antonelli, B. L. Claggett, M. Henglin, A. Kim, G. Ovsak, N. Kim, K. Deng, K. Rao, O. Tyagi, J. D. Watrous, K. A. Lagerborg, P. V. Hushcha, O. V. Demler, S. Mora, T. J. Niiranen, A. C. Pereira, M. Jain, S. Cheng, *Metabolites* **2019**, *9*, 143.

## 3.3 Publication 3: Replication and mediation of the association between the metabolome and clinical markers of metabolic health in an adolescent cohort study

# scientific reports

Check for updates

OPEN

# Replication and mediation of the association between the metabolome and clinical markers of metabolic health in an adolescent cohort study

Christian Brachem[1✉], Leonie Weinhold[2], Ute Alexy[1], Matthias Schmid[2], Kolade Oluwagbemigun[1,3] & Ute Nöthlings[1,3]

Metabolomics-derived metabolites (henceforth metabolites) may mediate the relationship between modifiable risk factors and clinical biomarkers of metabolic health (henceforth clinical biomarkers). We set out to study the associations of metabolites with clinical biomarkers and a potential mediation effect in a population of young adults. First, we conducted a systematic literature review searching for metabolites associated with 11 clinical biomarkers (inflammation markers, glucose, blood pressure or blood lipids). Second, we replicated the identified associations in a study population of n = 218 (88 males and 130 females, average age of 18 years) participants of the DONALD Study. Sex-stratified linear regression models adjusted for age and BMI and corrected for multiple testing were calculated. Third, we investigated our previously reported metabolites associated with anthropometric and dietary factors mediators in sex-stratified causal mediation analysis. For all steps, both urine and blood metabolites were considered. We found 41 metabolites in the literature associated with clinical biomarkers meeting our inclusion criteria. We were able to replicate an inverse association of betaine with CRP in women, between body mass index and C-reactive protein (CRP) and between body fat and leptin. There was no evidence of mediation by lifestyle-related metabolites after correction for multiple testing. We were only able to partially replicate previous findings in our age group and did not find evidence of mediation. The complex interactions between lifestyle factors, the metabolome, and clinical biomarkers warrant further investigation.

**Abbreviations**
| | |
|---|---|
| CVD | Cardio vascular disease |
| T2DM | Type 2 diabetes mellitus |
| BMI | Body mass index |
| BF | Body fat percentage |
| SLR | Systematic literature search |
| CRP | C-reactive protein |
| IL-6 | Interleukin 6 |
| IL-18 | Interleukin 18 |
| HDL | High density lipoprotein |
| LDL | Low density lipoprotein |
| DONALD | Dortmund nutritional and anthropometric longitudinally designed |

[1]Nutritional Epidemiology, Department of Nutrition and Food Sciences, Rheinische Friedrich-Wilhelms-University Bonn, Meckenheimer Allee 166a, 53115 Bonn, Germany. [2]Institute for Medical Biometry, Informatics and Epidemiology (IMBIE), University Hospital Bonn, 53127 Bonn, Germany. [3]These authors jointly supervised this work: Kolade Oluwagbemigun and Ute Nöthlings. ✉email: cbrachem@uni-bonn.de

nature portfolio

1

| 3d-WDR | Three day weighed dietary record |
| UPLC-MS/MS | Ultra-high performance liquid chromatography-tandem mass spectroscopy |
| RI | Retention time/index |
| m/z | Mass to charge ratio |
| BP | Blood pressure |
| ACME | Average causal mediation effect |
| SD | Standard deviation |
| PLS | Partial least squares |

Chronic diseases such as cardiovascular disease (CVD), type 2 diabetes mellitus (T2DM), and cancers are among the largest public health burdens modern societies' face[1,2]. Important for the prevention of these diseases are key modifiable risk factors such as body composition and dietary intake.

Metabolomics is a rich resource in the process of elucidating the etiology of diseases[3,4]. To realize the potential of the metabolome it is important to validate putative biomarkers (henceforth called "metabolites") and replicate their associations across studies and settings[5]. Well-established clinical biomarkers of metabolic health (henceforth called "clinical biomarkers"), for example cholesterol as a clinical biomarker for CVD[6,7], HBa1C for T2DM[8], or inflammation markers (e.g. CRP, IL-8)[4] appear to be intricately linked with metabolites[3,5,9,10]. However, study findings are largely inconsistent, and might differ by sex and age groups[11–16] calling for in depth confirmation and replication across sexes and age groups.

Modifiable lifestyle factors including body composition and food intake are linked to a number of chronic diseases such as type 2 diabetes[17–20], CVD[7,21–23] or cancer types[24–27] through alterations in the human metabolome. With respect to prevention, a life course approach elucidates preventive potential in younger age groups, e.g. early adulthood, which has been shown to be relevant[28,29]. In these age groups, clinical biomarkers are of importance to evaluate chronic disease risk. While the relationship of body composition and dietary intake with clinical biomarkers is well reported, less is known on potential mediation through the metabolome. We recently reported associations between body composition and the metabolome (19 metabolites for body mass index (BMI) and 20 for body fat (BF) in urine[30], as well as between habitual food intake (in food groups) and the metabolome (6 metabolites) in urine and blood[31]). The association of body composition and dietary intake with clinical biomarkers may be linked via some of these metabolites.

To investigate this complex relationship, the aims of the current study were first to identify associations of metabolites with clinical biomarkers based on a systematic literature review (SLR), second to replicate these associations in our study population, and third to evaluate whether our previously reported body composition- and habitual food intake-associated metabolites mediate the association of body composition and habitual food intake with clinical biomarkers. Of note, we focused on the age groups of adolescents and young adults as a particular time window of relevance for prevention.

## Methods

**Systematic literature review.**   We first conducted a SLR of studies indexed in PubMed, separate for each clinical biomarker, to identify relationships between metabolites and clinical biomarkers to be replicated in our study. A detailed description of the search terms and flow-charts can be found in Additional File S1. Briefly, we included studies that reported associations between Inflammation markers (C-reactive protein (CRP), Interleukin-6 (IL-6), Interleukin-18 (IL-18), Adiponectin, and Leptin), glucose, blood pressure (BP) (systolic blood pressure, diastolic blood pressure, and Hypertension) and blood lipids (high-density lipoprotein (HDL), low-density lipoprotein (LDL), total triglycerides) and the human blood or urine metabolome. We developed a search term for each of these clinical biomarkers. The review was conducted by CB only.

We included all studies where at least one association was reported. We identified additional studies through screening of citations and literature reviews. Information about associations of metabolites and clinical biomarkers was finally extracted from each included study. Of these, only associations reported in at least two independent studies were considered "consistent" and further used in the current study.

**Study design.**   Both, the confirmation and mediation analyses were conducted in a subpopulation of the DOrtmund Nutritional and Anthropometric Longitudinally Designed (DONALD) study[32,33]. Briefly, the DONALD Study is an ongoing longitudinal open cohort study in Dortmund, Germany, with the goal of analyzing detailed data on diet, growth, development, and metabolism between infancy and adulthood[32,33]. Participants are first examined at the age of 3 months and return for three more visits in the first year of life, two in the second and annually thereafter until the age of 18, when examinations start following a five-year cycle. Examinations include 3-day weighed dietary records (3d-WDR), anthropometric measurements, collection of 24-h urine samples (starting at age 3–4), collection of blood samples (starting at age 18), and interviews on lifestyle and medical examinations. Further details on the study design have been published elsewhere[32,33].

**Study participants.**   We included all DONALD study participants that were singletons, full term births (37–42 weeks of gestation) and had a birth weight of at least 2500 g. For the current analysis participants had to have a measurement of both the urine and blood metabolome, as well as at least one measurement of each clinical biomarker. Overall, 218 participants were eligible for the current study.

**Variable assessment.**   *Assessment of clinical biomarkers.*   Inflammation markers (C-reactive protein (CRP), Interleukin-6 (IL-6), Interleukin-18 (IL-18), Adiponectin, and Leptin), glucose, and blood lipids (high-density lipoprotein (HDL), low-density lipoprotein (LDL), total triglycerides) were measured in non-fasted

blood plasma. Measurements in blood were always at the same follow-up and from the same sample as metabolome measurement. Blood measurement was always at the same follow-up visit or later than urine metabolome measurement.

Blood pressure (mmHg) was measured multiple times by experienced nursing staff. We used the mean of two repeated measurements for both systolic and diastolic blood pressure. We chose the blood pressure measurement closest after the corresponding metabolome measurement for analysis of the respective participant, which was always at the next study visit.

*Untargeted metabolomic profiling.* The metabolome measurement was already described elsewhere[31]. Briefly, Metabolon Inc. (Morrisville, NC, USA) performed an untargeted metabolomics assay with lipidomics on plasma and an untargeted metabolomics assay on urine samples. For both the plasma and urine untargeted assays, Metabolon used ultra-high performance liquid chromatography-tandem mass spectroscopy (UPLC-MS/MS) to identify metabolites in the samples. Peak identification was done in their propriety Laboratory Information Management System. Compounds were identified by comparison of their retention time/index (RI), mass to charge ratio (m/z) and chromatographic data (e.g. MS/MS spectral data) to library standards. Metabolon maintains a library of authenticated standards with over 3300 commercially available purified standard compounds. Structurally unnamed biochemicals were identified by occurrence. Peaks are quantified using area-under-the-curve and normalized with block correction correcting for inter-day instrument tuning differences. Further details on the metabolic profiling have been reported elsewhere[34]. Both blood and urine untargeted assays were performed in this fashion. Urine metabolite values were additionally normalized to urine osmolality to account for differences in metabolite levels due to differences in the amount of material present in each sample. Metabolon quantified 1042 (811 known and 231 unknown) and 1407 (940 known and 467 unknown) in blood and urine, respectively. A deeper explanation of the metabolomics methods can be found in Additional File S2.

*Complex lipid platform measurement.* Lipids were extracted from samples in methanol:dichloromethane in the presence of internal standards. The extracts were concentrated under nitrogen and reconstituted in 0.25 mL of 10 mM ammonium acetate dichloromethane:methanol (50:50). The extracts were transferred to inserts and placed in vials for infusion-MS analysis, performed on a Shimazdu LC with nano PEEK tubing and the Sciex SelexIon-5500 QTRAP. The samples were analyzed via both positive and negative mode electrospray. The 5500 QTRAP scan was performed in MRM mode with the total of more than 1100 MRMs. Individual lipid species were quantified by taking the peak area ratios of target compounds and their assigned internal standards, then multiplying by the concentration of internal standard added to the sample. Lipid class concentrations were calculated from the sum of all molecular species within a class, and fatty acid compositions were determined by calculating the proportion of each class comprised by individual fatty acids. We identified 966 lipid species in 14 classes as well as 265 fatty acids. A deeper explanation of the lipidomics methods can be found in Additional File S2.

*Body composition and habitual dietary intake.* Body weight and height were measured at every follow-up by experienced nursing staff. Body mass index (BMI) was calculated using height (m) and weight (kg) with the formula $BMI = \frac{weight}{height^2}$. Body fat percent (BF) was calculated from four skin/fold thickness measurements (biceps, triceps, iliaca, and scapula), using age, puberty status, and sex/specific equations from Deurenberg et al.[35]. Previous associations with BMI used in the mediation analysis and further details on body composition assessment were reported in Brachem et al.[30].

We used multiple annually applied 3d-WDRs to assess habitual food intake on the food group level. Participants had to have at least two 3d-WDR in adolescence (according to the WHO definition, age 10–19). We defined habitual intake as the mean intake across all available records in adolescence. To account for differences in consumed calories, we standardized intake to grams per 1000 kcal. Previous associations with habitual food intake used in the mediation analysis and further details on dietary assessment were reported in Brachem et al.[31].

**Statistical analysis.** Statistical analysis was performed using R software (Version 4.0.3)[36]. All analyses were stratified by sex.

*Metabolomics data pre-treatment.* Both urine and blood metabolite values were log transformed, centered to a mean of zero and scaled to a standard deviation of one prior to analysis.

*Replication.* We used ordinary least squares regression to replicate associations between the metabolites and clinical biomarkers in the DONALD study. The clinical biomarkers were used as the dependent variables and metabolites as the independent variables. We adjusted the models for BMI and age, both at sample collection. Data was split into training (70%) and testing (30%) data to evaluate overfitting. We trained the model on the training data and used these models to predict clinical biomarker values in the test data. Results from the test data were used only to evaluate the model quality. We additionally accounted for multiple testing by holding the false discovery rate at 5%[37].

*Mediation analysis.* We used causal mediation analysis to evaluate whether our previously reported body composition-[30] and habitual food intake-related metabolites[31] mediate the association of body composition and habitual food intake with clinical biomarkers. For the first, BMI and BF were the exposure and the clinical biomarker (BP, IL-6, IL-18, CRP, Adiponectin, leptin, total cholesterol, HDL, LDL, and triglyceride levels) were the

outcomes. The 19 (*5-dodecenoylcarnitine (C12:1)*, *7-hydroxyindole* sulfate, *decanoylcarnitine (C10)*, *formimino-glutamate*, *glucuronide of C10H18O2 (12)*, *guanidinosuccinate*, *isobutyrylglycine (C4)*, *isovalerylglycine*, *nicotinamide N-oxide*, *proline*, *succinimide*, *thymine*, *tigloylglycine*, *X—12839*, *X—21441*, *X—21851*, *X—24469*, *X—24801*, and *X—25003*) BMI-associated metabolites and 20 (*3-methylcrotonylglycine*, *glucuronide of C10H18O2 (12)*, *glutamine conjugate of C8H12O2 (1)*, *glycine conjugate of C10H14O2 (1)*, *guanidinosuccinate*, *isobutyrylglycine (C4)*, *isovalerylglutamine*, *isovalerylglycine*, *nicotinamide N-oxide*, *succinimide*, *tigloylglycine*, *X—11261*, *X—15486*, *X—17676*, *X—21851*, *X—24345*, *X—24350*, *X—24469*, *X—24801*, *X—25442*, and *X—25464*) BF-associated metabolites were considered as mediators. For the second, habitual food intake was the exposure, the aforementioned clinical biomarker markers were the outcomes, and the six (eggs: *indole-3-acetamide*, *N6-methyladenosine*; vegetables: *hippurate*, *citraconate/glutaconate*, *X—12111*; processed and other meat: *vanillylmandelate (VMA)*) food group-associated metabolites were considered as mediators. We used the 'mediate()'-function in the R package 'mediation'[38] for the analysis. We used 1000 simulations (the recommended default) and quasi-Bayesian approximation to estimate the standard errors. We used the model-based approach[38]. The mediator model is the linear regression model that regresses the metabolites on BMI, BF, or habitual food intake adjusted for age at sample collection. The habitual food intake models were additionally adjusted for BMI at sample collection. The outcome model is a linear regression model that regresses clinical biomarker on BMI, BF, or habitual food intake, the mediator (metabolites), and adjustment variables. From these models the causal mediation analysis is performed as described by Imai et al.[39]. Briefly, the model estimates the average causal mediation effect (ACME), which is a numeric measure of how much influence the presence of the mediator has on the total effect of the exposure-outcome association, as well as the average direct effect, the average total effect, and the proportion mediated. We corrected for multiple testing by holding the false discovery rate at 5%.

*Missing values.* We excluded metabolites from the analysis when more than 70% of data was missing. Based on this we excluded 91 and 67 metabolites in female blood and urine, respectively, and 87 and 74 metabolites in male blood and urine, respectively.

For the mediation analysis and the regression models, we performed a single imputation with the "missRanger" package, using 10 trees with a maximum depth of six and three non-missing candidate values for predictive mean matching. We used random forest imputation, as it is recommended for imputation of missing metabolomics data[40].

*Sensitivity analysis.* We performed sensitivity analysis on the choice of the missing data threshold in the imputation approach, repeating the complete study protocol excluding metabolites with more than 30% missing data (instead of 70% in the main analysis). In males we additionally excluded 103 metabolites and 106 metabolites in blood and urine, respectively, while in females we excluded 123 and 108 additional metabolites in blood and urine, respectively.

**Ethics approval and consent to participate.** Informed written consent was obtained from parents and from participants themselves on reaching adolescence. The ethics committee of the University of Bonn, Germany (project identification: 098/06) approved the study. We confirm that all methods were performed in accordance with relevant guidelines and in accordance with the Declaration of Helsinki.

## Results

In the SLR, we found metabolites associated with blood pressure and CRP in at least two independent studies (Table 1). Six metabolites (*4-hydroxyhippurate, Androsterone sulfate, Glutamine, Isoleucine, Phenylalanine,* and *Tryptophan)* for blood pressure and four metabolites (*Betaine, Glutamine, Isoleucine,* and *Tryptophan*) for CRP

| Metabolite | Sources |
|---|---|
| Blood pressure | |
| 4-Hydroxyhippurate | Zheng et al.[41,42] |
| Androsterone sulfate | Zheng et al.[41,42] |
| Glutamine | Goïta et al.[43], Le Wang et al.[44] |
| Isoleucine | Liu et al.[45], Le Wang et al.[44] |
| Phenylalanine | Hao et al.[46], Wawrzyniak et al.[47], Goïta et al.[43], Meyer et al.[48], Øvrehus et al.[49] |
| Tryptophan | Liu et al.[45], Le Wang et al.[44] |
| CRP | |
| Betaine | Jutley et al.[50], Pietzner et al.[51] |
| Glutamine | Jutley et al.[50], Pietzner et al.[51] |
| Isoleucine | Jutley et al.[50], Oluwagbemigun et al.[52] |
| Tryptophan | Jutley et al.[50], Kosek et al.[53], Oluwagbemigun et al.[52] |

**Table 1.** Metabolites associated with conventional systemic markers of chronic disease risk in at least two independent observational studies. According to systematic search in PubMed. Metabolites without a match in our metabolites and those we did not replicated are available in Additional File S3.

were present in more than two studies. The full metabolite list we identified in at least one study with their corresponding references can be found in Additional File S3.

In Table 2, we present characteristics of the DONALD study population. Aside from the BMI at blood sampling, there were no differences for basic characteristics between the sexes. Except for diastolic blood pressure at urine collection, IL-6, IL-18, and total blood triglycerides, all clinical biomarkers were significantly different between the sexes though directions differed. Blood pressure (diastolic at blood draw and systolic at both blood draw and urine collection) and blood glucose were higher in males, while CRP, leptin, adiponectin, total cholesterol, HDL, LDL, and triglycerides were higher in females.

We were able to test associations for 41 of the 46 metabolites associated with clinical biomarkers. We were able to replicate metabolite associations for 10 out of the 41 metabolites we found (Table 3), four out of five for CRP and six out of 36 for blood pressure. We discovered six significant associations in female participants, while there were five significant associations in male participants. Across bio specimens, there were more associations present in urine (six) compared to blood (five). One metabolite, phenylalanine, was associated across sexes with systolic blood pressure. Another metabolite, glutamine, was associated with both diastolic and systolic blood pressure. However, the association between glutamine and diastolic blood pressure for male participants was positive, while the association between glutamine and systolic blood pressure for females was inverse. Across sexes and bio specimen more metabolites (six) were associated negatively. When correcting for multiple testing only the association between betaine and CRP in females remained significant. The complete model results, including direct sex comparisons, can be found in Additional File S4. The test set model metrics can be found in Additional File S5.

We found no metabolite significantly mediating the relationship of either body composition or habitual food intake and clinical biomarkers after correcting for multiple testing (Table 4). However, we observed two significant total effects, both in male urine. One between CRP and BMI, CRP is estimated to increase by 0.5 standard deviations (SD) as BMI increases by one unit (p-Value (FDR) < 0.0001) and one between leptin and BF, leptin is estimated to increase by 0.62 standard deviations as BF increases by 1 unit (p-Value (FDR) = 0.040). The full model results are available in Additional File S6. The test set model metrics can be found in Additional File S5.

| | N | Overall, N = 218[1] | Male, N = 88[1] | Female, N = 130[1] | p-value[2] |
|---|---|---|---|---|---|
| BMI [kg/m$^2$] at blood draw | 218 | 22.30 [20.65, 24.91] | 23.25 [21.21, 26.16] | 21.89 [20.31, 24.09] | 0.005 |
| BMI [kg/m$^2$] at urine collection | 218 | 21.88 [19.96, 23.63] | 22.06 [20.38, 23.54] | 21.85 [19.88, 23.68] | 0.4 |
| Age [years] at blood draw | 218 | 18.00 [18.00, 23.00] | 18.00 [18.00, 23.00] | 18.00 [18.00, 23.75] | 0.5 |
| Age [years] at urine collection | 218 | 18.00 [17.00, 18.00] | 18.00 [16.00, 18.00] | 18.00 [17.00, 18.00] | 0.7 |
| Age difference [years] between last dietary record and blood draw | 218 | 1.00 [0.00, 6.75] | 1.00 [0.00, 4.00] | 1.50 [0.00, 7.00] | 0.7 |
| Age difference [years] between last dietary record and urine collection | 218 | 0.00 [0.00, 0.00] | 0.00 [0.00, 0.00] | 0.00 [0.00, 0.00] | 0.4 |
| 3-Methylcrotonylglycine | 206 | 0.99 [0.71, 1.25] | 1.00 [0.83, 1.25] | 0.94 [0.61, 1.24] | 0.075 |
| 5-Dodecenoylcarnitine (c12:1) | 218 | 0.93 [0.58, 1.38] | 1.18 [0.73, 1.66] | 0.84 [0.50, 1.26] | 0.001 |
| 7-Hydroxyindole sulfate | 211 | 0.92 [0.54, 1.62] | 1.25 [0.81, 1.97] | 0.68 [0.46, 1.11] | < 0.001 |
| Citraconate/glutaconate | 218 | 1.00 [0.63, 1.80] | 1.09 [0.72, 2.00] | 0.88 [0.57, 1.68] | 0.042 |
| Decanoylcarnitine (c10) | 217 | 0.95 [0.64, 1.44] | 0.92 [0.64, 1.58] | 1.00 [0.64, 1.39] | 0.7 |
| Formiminoglutamate | 218 | 0.97 [0.64, 1.33] | 1.01 [0.76, 1.35] | 0.85 [0.54, 1.31] | 0.020 |
| Glucuronide of c10h18o2 (12) | 217 | 1.00 [0.68, 1.58] | 1.03 [0.79, 1.44] | 0.98 [0.66, 1.65] | 0.4 |
| Glutamine conjugate of c8h12o2 (1) | 218 | 0.98 [0.68, 1.54] | 0.98 [0.72, 1.45] | 0.99 [0.63, 1.58] | 0.7 |
| Glycine conjugate of c10h14o2 (1) | 218 | 0.95 [0.59, 1.68] | 1.00 [0.68, 1.92] | 0.87 [0.59, 1.63] | 0.4 |
| Guanidinosuccinate | 218 | 0.96 [0.70, 1.30] | 1.10 [0.81, 1.31] | 0.83 [0.61, 1.23] | 0.005 |
| Hippurate | 218 | 0.99 [0.69, 1.44] | 0.92 [0.69, 1.35] | 1.07 [0.70, 1.50] | 0.3 |
| Indole-3-acetamide | 198 | 0.93 [0.58, 1.87] | 1.06 [0.57, 2.03] | 0.88 [0.61, 1.78] | 0.5 |
| Isobutyrylglycine (c4) | 218 | 0.98 [0.67, 1.31] | 1.08 [0.82, 1.33] | 0.87 [0.61, 1.28] | 0.011 |
| Isovalerylglutamine | 218 | 0.99 [0.66, 1.36] | 1.15 [0.96, 1.44] | 0.83 [0.56, 1.16] | < 0.001 |
| Isovalerylglycine | 218 | 1.00 [0.69, 1.44] | 1.21 [0.82, 1.51] | 0.86 [0.62, 1.34] | < 0.001 |
| N6-methyladenosine | 215 | 0.99 [0.71, 1.42] | 1.06 [0.77, 1.44] | 0.96 [0.69, 1.41] | 0.14 |
| Nicotinamide n-oxide | 214 | 1.01 [0.60, 1.59] | 1.00 [0.65, 1.50] | 1.05 [0.57, 1.78] | 0.7 |
| Proline | 218 | 0.99 [0.69, 1.27] | 1.09 [0.78, 1.42] | 0.90 [0.66, 1.13] | 0.004 |
| Succinimide | 217 | 0.98 [0.70, 1.35] | 1.14 [0.90, 1.60] | 0.86 [0.60, 1.22] | < 0.001 |
| Thymine | 217 | 1.00 [0.75, 1.33] | 1.09 [0.82, 1.42] | 0.90 [0.64, 1.18] | 0.004 |
| Tigloylglycine | 218 | 0.93 [0.71, 1.28] | 1.11 [0.83, 1.41] | 0.82 [0.68, 1.23] | 0.004 |
| Vanillylmandelate (vma) | 218 | 0.99 [0.79, 1.36] | 1.11 [0.86, 1.44] | 0.93 [0.76, 1.27] | 0.007 |

**Table 2.** Characteristics and markers of metabolic health of 218 DONALD participants. [1]Median [IQR]. [2]Wilcoxon rank sum test Males VS Females. Unknown metabolites are not displayed.

| Metabolite | Sex | Bio specimen | Super pathway | Sub pathway | β | 95% CI | p-value | p-value (FDR) |
|---|---|---|---|---|---|---|---|---|
| Clinical biomarker: CRP | | | | | | | | |
| *Betaine* | *Female* | *Blood* | *Amino acid* | *Glycine, serine and threonine metabolism* | *− 0.40* | *− 0.61 to − 0.19* | *0.0002* | *0.0220* |
| Glutamine | Male | Urine | Amino acid | Glutamate metabolism | − 0.39 | − 0.63 to − 0.15 | 0.0022 | 0.1008 |
| Isoleucine | Male | Urine | Amino acid | Leucine, isoleucine and valine metabolism | − 0.29 | − 0.53 to − 0.04 | 0.0218 | 0.2060 |
| Tryptophan | Male | Urine | Amino acid | Tryptophan metabolism | − 0.38 | − 0.63 to − 0.13 | 0.0033 | 0.1137 |
| Clinical biomarker: diastolic blood pressure | | | | | | | | |
| Glutamine | Male | Blood | Amino acid | Glutamate metabolism | 0.25 | 0.02–0.48 | 0.0337 | 0.2682 |
| Clinical biomarker: systolic blood pressure | | | | | | | | |
| 4-Hydroxyhippurate | Female | Urine | Xenobiotics | Benzoate metabolism | 0.28 | 0.08–0.48 | 0.0072 | 0.1374 |
| Androsterone sulfate | Female | Blood | Lipid | Androgenic steroids | − 0.17 | − 0.35 to − 0.00 | 0.0496 | 0.3076 |
| Glutamine | Female | Blood | Amino Acid | Glutamate metabolism | − 0.24 | − 0.41 to − 0.07 | 0.0064 | 0.1368 |
| Phenylalanine | Female | Urine | Amino Acid | Phenylalanine metabolism | 0.19 | 0.00–0.38 | 0.0477 | 0.3076 |
| | Male | Blood | Amino Acid | Phenylalanine metabolism | 0.25 | 0.02–0.48 | 0.0328 | 0.2682 |
| Tryptophan | Female | Urine | Amino Acid | Tryptophan metabolism | 0.19 | 0.00–0.38 | 0.0466 | 0.3076 |

**Table 3.** Replicated risk-markers-metabolites association for CRP, diastolic blood pressure, and systolic blood pressure. Estimates are generated from linear regression. Models were adjusted for age and BMI, both at sample collection. Metabolites were log-transformed prior to analysis. Estimates and 95% CI are on the log scale. We controlled the false discovery rate (FDR) at 5% to account for multiple testing. Metabolites significant after correction for multiple testing are marked in italics.

| Bio specimen | Sex | Clinical biomarker | Mediating metabolite | Total effect | | | ACME | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Estimate | 95% CI | p-value[1] | Estimate | 95% CI | p-value[1] |
| Exposure: BMI | | | | | | | | | |
| Urine | Male | CRP | 5-Dodecenoyl-carnitine (C12:1) | 0.51 | 0.254–0.747 | 0.000 | − 0.03 | − 0.134 to 0.035 | 0.983 |
| Exposure: body fat (%) | | | | | | | | | |
| Urine | Male | Leptin | Glucuronide of C10H18O2 (12)* | 0.62 | 0.203–1.040 | 0.040 | − 0.10 | − 0.285 to 0.018 | 0.983 |

**Table 4.** Metabolites mediating the association of body composition and habitual food intake with clinical biomarkers. Estimates and confidence intervals are in standard deviations. *ACME* average causal mediation effect, *CRP* C-reactive Protein. [1]p-values are corrected for multiple testing by holding the false discovery rate at 5%.

In our sensitivity analysis on the amount of missing data we observed one additional significant association after correcting for multiple testing, between glutamine and CRP in male urine (p (FDR) = 0.046, ß = − 0.39, − 0.63 to − 0.15). Additionally, we observed one additional significant total effect in the mediation analysis, between BMI and leptin in male urine and no mediators. The full results tables for the sensitivity analysis can be found in Additional File S7.

## Discussion

In the present study, we conducted an SLR identifying 41 metabolite- clinical biomarker associations (36 for BP, 5 for CRP) that were reported in at least two independent studies. Of these 41, we were able to replicate 10 associations, in our own study population one of which was significant after multiple testing correction. Additionally, we found no evidence of a metabolite mediating the association between body composition or habitual food intake and clinical biomarkers.

**Systematic literature review.** We identified 41 metabolites associated with clinical biomarker variables in at least two studies. Interestingly, these were distributed only between two clinical biomarker variables: blood pressure and CRP. Most of the metabolites (36 of 41) were associated with blood pressure. The methods applied to investigate the relationships of the metabolome and the clinical biomarker variables were very heterogeneous. They ranged from correlation analysis (e.g.[54]), through regression (e.g.[46]) to advanced machine learning methods like random forests (e.g.[55]) or PLS (partial least squares) variants (e.g.[56]). Additionally, it would be very useful for future SLR to have an easier format to export all study results in an appendix. Because of the large number of associations usually present in metabolomics studies, this would greatly increase the possibility for future studies to build on. Another important observation in the SLR is that only four out of 50 studies[43,52,57,58] stratified by sex, with two additional studies having cohorts restricted to either males[59] or females[60], though many additional studies adjusted for or matched by sex. Given how strong the influence[11–13,15,16,30,31,52] of sex is on many different aspects of the metabolome, a better and ideally unified strategy to account for these influences in future studies is needed. Most studies included in the SLR were in exclusively adult study populations. Three studies studied children[53,61,62] and two studies[52,57] adolescents and young adults. Age is another influential factor in the composition of the metabolome that may need additional adjustment strategies in the long term[63].

**Replication.** We were able to replicate 10 out of the 41 metabolites testable in our study sample. We found more metabolites replicated in females compared to males (six and five, respectively) and only one metabolite, *phenylalanine*, associated with systolic blood pressure across sexes. In this replication analysis, only one association, the negative association between *betaine* measured in the blood of males and CRP remained significant after correction for multiple testing. *Betaine* is an essential osmolyte derived from either diet or by oxidation of choline[64,65]. Insufficiencies of *betaine* have been associated with many chronic diseases, such as metabolic syndrome, T2DM or vascular diseases[65]. Additionally, *betaine* is considered as an anti-oxidant[64] and fulfills anti-inflammatory functions[66]. The inverse association between *betaine* in male blood and CRP we observed is therefore in line with the literature.

*Phenylalanine* was not significantly associated with systolic blood pressure after correction for multiple testing but it is interesting. It is the only metabolite associated across sexes and indirectly across bio specimen. It's association with higher blood pressure is in line with previous literature, that reported a strong association with infant pulmonary hypertension[67] and more generally elevated cardiovascular risk[68]. Furthermore, it was elevated in metabolically unhealthy obese (compared to metabolically healthy obese)[69]. Because it is a precursor to catecholamines an increase in blood pressure even has a known physiological pathway already[64]. More studies are needed to discern the causal order and exact mechanism of phenylalanine on blood pressure.

**Mediation.** We did not identify any metabolite as potential mediator of the relationship between either body composition or habitual food intake and clinical biomarkers.

While we did not identify any mediators in our sample, we still believe there will be mediators identified in the future. Mediators are notoriously hard to identify, as their study requires many association tests (which in turn requires a correction for multiple testing), a large study population and large effect sizes. All three of which were limiting factors within our study.

**Sensitivity analysis.** We performed sensitivity analysis on the amount of missing data permitted in the metabolites prior to imputation. We excluded over 100 additional metabolites, but the results did not change in meaning. In the replication analysis, as was expected by reducing the number of metabolites and therefore statistical tests, the metabolite closest to significance in the main analysis was statistically significant in the sensitivity analysis. However, the point estimates of the metabolites remained the same. In the mediation analysis, one additional total effect remained significant after correcting for multiple testing but no mediating effects, the same as the main analysis. Therefore, interpretation of the results was not depended on the choice of missingness permitted in the metabolites prior to imputation.

**Future research.** Future research should take the sex differences we reported into consideration in their own study design, ideally by stratification, in order to further our understanding of the physiological differences in metabolism between males and females. A study evaluating the metabolites associated with metabolic health markers as mediators to lifestyle factors would be a great continuation of the present study, ideally in a larger cohort. Lastly, metabolomics would greatly benefit from both a more unified data analysis approach as well as a unified measurement approach to better facilitate meta-analysis and ease the burden of replicating results from different studies.

**Strengths and limitations.** The present study has some notable strengths. We used results from our own previous studies to investigate mediation and conducted a SLR to facilitate replication of previously reported associations in the literature. We were able to use global measurements of the urine and blood metabolome in the same participants for both analyses in a comparatively (for metabolomics) large study population. Though the number of statistical tests required for metabolomics in relation to the available data in our study is high, therefore sampling power may be a reason for few total associations found. We employed state of the art statistical analysis and machine learning to investigate both the mediation and the replication. However, we acknowledge several limitations to the study. Our participants are Caucasians (Germans), residing in a large city (Dortmund) and surrounding area and are mostly from a high socio-economic background. This may limit the generalizability of our findings. We used non-fasted plasma samples, which increases the variability of inter and intra par-

ticipant variability of measurements introducing non-differential measurement error. We constructed habitual diet from multiple measurements in adolescence, which increases the time difference between diet measurement and metabolome assessment. This limits results to more long-term markers but increases the effect size needed to detect a signal. Additionally, we cannot rule out residual confounding by either unknown or unmeasured confounders or related factors such as genetics. In our mediation analysis, we had, compared to other mediation analysis, a relatively small sample size. Lastly, we only have one measurement of the metabolome available, so the temporal reproducibility of these findings is unknown.

## Conclusions

In summary, we identified 41 metabolites associated in at least two independent studies with clinical biomarker and replicated ten associations in our own data, only one of which was significant after multiple testing correction. Additionally, there was no metabolite mediating the relationship between body composition or habitual diet and clinical biomarker. The intricate interplay between lifestyle factors, the metabolome, and metabolic health warrants further investigation.

## Data availability

The datasets generated and/or analyzed during the current study are not publicly available due data protection concerns for sensitive data but are available on reasonable request and approval of the principal investigator. Requests can be sent to *epi@uni-bonn.de*. All model results are available in the supplement to this article.

## References

1. Global Burden of Disease Collaborative Network. *Global Burden of Disease Study 2019 (GBD 2019) Reference Life Table* (Institute for Health Metrics and Evaluation (IHME), 2021).
2. Mozaffarian, D. *et al.* Heart disease and stroke statistics-2015 update: A report from the American Heart Association. *Circulation* **131**, e29-322 (2015).
3. Zhang, A., Sun, H., Yan, G., Wang, P. & Wang, X. Metabolomics for biomarker discovery: moving to the clinic. *Biomed Res. Int.* **2015**, 354671 (2015).
4. Furman, D. *et al.* Chronic inflammation in the etiology of disease across the life span. *Nat. Med.* **25**, 1822–1832 (2019).
5. López-López, Á., López-Gonzálvez, Á., Barker-Tejeda, T. C. & Barbas, C. A review of validated biomarkers obtained through metabolomics. *Expert Rev. Mol. Diagn.* **18**, 557–575 (2018).
6. Millán, J. *et al.* Lipoprotein ratios: Physiological significance and clinical usefulness in cardiovascular prevention. *Vasc. Health Risk Manag.* **5**, 757–765 (2009).
7. Upadhyay, R. K. Emerging risk biomarkers in cardiovascular diseases and disorders. *J. Lipids* **2015**, 971453 (2015).
8. Bennett, C. M., Guo, M. & Dharmage, S. C. HbA(1c) as a screening tool for detection of type 2 diabetes: A systematic review. *Diabetic Med. J. Br. Diabetic Assoc.* **24**, 333–343 (2007).
9. Susan, C. *et al.* Potential impact and study considerations of metabolomics in cardiovascular health and disease: A scientific statement from the american heart association. *Circ. Cardiovasc. Genet.* **10**, e000032 (2017).
10. Iida, M., Harada, S. & Takebayashi, T. Application of metabolomics to epidemiological studies of atherosclerosis and cardiovascular disease. *J. Atheroscler. Thromb.* **26**, 747–757 (2019).
11. Peters, S. A. E., Singhateh, Y., Mackay, D., Huxley, R. R. & Woodward, M. Total cholesterol as a risk factor for coronary heart disease and stroke in women compared with men: A systematic review and meta-analysis. *Atherosclerosis* **248**, 123–131 (2016).
12. Leuzzi, C., Sangiorgi, G. M. & Modena, M. G. Gender-specific aspects in the clinical presentation of cardiovascular disease. *Fundam. Clin. Pharmacol.* **24**, 711–717 (2010).
13. Isacco, L. & Miles-Chan, J. L. Gender-specific considerations in physical activity, thermogenesis and fat oxidation: Implications for obesity management. *Obesity Rev.* **19**(Suppl 1), 73–83 (2018).
14. Darst, B. F., Koscik, R. L., Hogan, K. J., Johnson, S. C. & Engelman, C. D. Longitudinal plasma metabolomics of aging and sex. *Aging* **11**, 1262–1282 (2019).
15. Clegg, D. J. & Mauvais-Jarvis, F. An integrated view of sex differences in metabolic physiology and disease. *Mol. Metab.* **15**, 1–2 (2018).
16. Krumsiek, J. *et al.* Gender-specific pathway differences in the human serum metabolome. *Metabolomics* **11**, 1815–1833 (2015).
17. Fretts, A. M. *et al.* Associations of circulating very-long-chain saturated fatty acids and incident type 2 diabetes: A pooled analysis of prospective cohort studies. *Am. J. Clin. Nutr.* **109**, 1216–1223 (2019).
18. Guasch-Ferré, M. *et al.* Metabolomics in prediabetes and diabetes: A systematic review and meta-analysis. *Diabetes Care* **39**, 833–846 (2016).
19. Park, J.-E., Lim, H. R., Kim, J. W. & Shin, K.-H. Metabolite changes in risk of type 2 diabetes mellitus in cohort studies: A systematic review and meta-analysis. *Diabetes Res. Clin. Pract.* **140**, 216–227 (2018).
20. Sun, Y., Gao, H.-Y., Fan, Z.-Y., He, Y. & Yan, Y.-X. Metabolomics signatures in type 2 diabetes: A systematic review and integrative analysis. *J. Clin. Endocrinol. Metab.* **105**, 1000–1008 (2020).
21. Deng, Y., Huang, C., Su, J., Pan, C.-W. & Ke, C. Identification of biomarkers for essential hypertension based on metabolomics. *Nutr. Metab. Cardiovasc. Dis.* **31**, 382–395 (2021).
22. Dhingra, R. & Vasan, R. S. Biomarkers in cardiovascular disease. *Trends Cardiovasc. Med.* **27**, 123–133 (2017).
23. Ruiz-Canela, M. *et al.* Comprehensive metabolomic profiling and incident cardiovascular disease: A systematic review. *J. Am. Heart Assoc.* **6**, e005705 (2017).
24. Mallafré-Muro, C. *et al.* Comprehensive volatilome and metabolome signatures of colorectal cancer in urine: A systematic review and meta-analysis. *Cancers* **13** (2021).
25. Long, N. P. *et al.* Metabolomics-guided global pathway analysis reveals better insights into the metabolic alterations of breast cancer. *J. Pharm. Biomed. Anal.* **202**, 114134 (2021).
26. Campi, R. *et al.* Novel liquid biomarkers and innovative imaging for kidney cancer diagnosis: What can be implemented in our practice today? A systematic review of the literature. *Eur. Urol. Oncol.* **4**, 22–41 (2021).
27. Ahmed-Salim, Y. *et al.* The application of metabolomics in ovarian cancer management: A systematic review. *Int. J. Gynecol. Cancer* **31**, 754–774 (2021).

28. GBD 2015 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 79 behavioural, environmental, and occupational, and metabolic risks or clusters of risks, 1990–2015: A systematic analysis for the global burden of disease study 2015. *Lancet (London, England).* **388**, 1659–1724 (2015).

29. Rippe, J. M. Lifestyle strategies for risk factor reduction, prevention, and treatment of cardiovascular disease. *Am. J. Lifestyle Med.* **13**, 204–212 (2018).

30. Brachem, C. *et al.* Associations of BMI and body fat with urine metabolome in adolescents are sex-specific: A cross-sectional study. *Metabolites* **10**, 330 (2020).

31. Brachem, C. *et al.* Exploring the association between habitual food intake and the urine and blood metabolome in adolescents and young adults: A cohort study. *Mol. Nutr. Food Res.* **66**, e2200023 (2022).

32. Kroke, A. *et al.* THE DONALD Study. History, current status and future perspectives. *Eur. J. Nutr.* **43**, 45–54 (2004).

33. Buyken, A. E., Alexy, U., Kersting, M. & Remer, T. Die DONALD Kohorte. Ein aktueller Überblick zu 25 Jahren Forschung im Rahmen der Dortmund Nutritional and Anthropometric Longitudinally Designed Study. *Bundesgesundheitsblatt Gesundheits-forschung Gesundheitsschutz* **55**, 875–884 (2012).

34. Evans, A. *et al.* High resolution mass spectrometry improves data quantity and quality as compared to unit mass resolution mass spectrometry in high-throughput profiling metabolomics. *Metabolomics* **04**, 1–7 (2014).

35. Deurenberg, P., Pieters, J. J. & Hautvast, J. G. The assessment of the body fat percentage by skinfold thickness measurements in childhood and young adolescence. *Br. J. Nutr.* **63**, 293–303 (1990).

36. R Core Team. *R: A Language and Environment for Statistical Computing* (2019).

37. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. Ser. B (Methodol.)* **57**, 289–300 (1995).

38. Tingley, D., Yamamoto, T., Hirose, K., Keele, L. & Imai, K. Mediation: R package for causal mediation analysis. *J. Stat. Soft.* **59**, 1–38 (2014).

39. Imai, K., Keele, L. & Tingley, D. A general approach to causal mediation analysis. *Psychol. Methods* **15**, 309–334 (2010).

40. Kokla, M., Virtanen, J., Kolehmainen, M., Paananen, J. & Hanhineva, K. Random forest-based imputation outperforms other methods for imputing LC-MS metabolomics data: A comparative study. *BMC Bioinform.* **20**, 492 (2019).

41. Zheng, Y. *et al.* Metabolomics and incident hypertension among African Americans: The atherosclerosis risk in communities (ARIC) study. *Hypertension* **62**, 398–403 (2013).

42. Zheng, Y. *et al.* Associations between metabolomic compounds and incident heart failure among African Americans: The ARIC study. *Am. J. Epidemiol.* **178**, 534–542 (2013).

43. Goïta, Y. *et al.* Sexual dimorphism of metabolomic profile in arterial hypertension. *Sci. Rep.* **10**, 7517 (2020).

44. Wang, Le. *et al.* Reconstruction and analysis of correlation networks based on GC-MS metabolomics data for young hypertensive men. *Anal. Chim. Acta* **854**, 95–105 (2015).

45. Liu, Y. *et al.* An ultrasonication-assisted extraction and derivatization protocol for GC/TOFMS-based metabolite profiling. *Anal. Bioanal. Chem.* **400**, 1405–1417 (2011).

46. Hao, Y. *et al.* A nested case-control study of association between metabolome and hypertension risk. *Biomed. Res. Int.* **2016**, 7646979 (2016).

47. Wawrzyniak, R. *et al.* Untargeted metabolomics provides insight into the mechanisms underlying resistant hypertension. *Curr. Med. Chem.* **26**, 232–243 (2019).

48. de Meyer, T. *et al.* NMR-based characterization of metabolic alterations in hypertension using an adaptive, intelligent binning algorithm. *Analyt. Chem.* **80**, 3783–3790 (2008).

49. Øvrehus, M. A. *et al.* Gene expression studies and targeted metabolomics reveal disturbed serine, methionine, and tyrosine metabolism in early hypertensive nephrosclerosis. *Kidney Int. Rep.* **4**, 321–333 (2019).

50. Jutley, G. S. *et al.* Relationship between inflammation and metabolism in patients with newly presenting rheumatoid arthritis. *Front. Immunol.* **12**, 676105 (2021).

51. Pietzner, M. *et al.* Comprehensive metabolic profiling of chronic low-grade inflammation among generally healthy individuals. *BMC Med.* **15**, 210 (2017).

52. Oluwagbemigun, K. *et al.* Longitudinal relationship of amino acids and indole metabolites with long-term body mass index and cardiometabolic risk markers in young individuals. *Sci. Rep.* **10**, 6399 (2020).

53. Kosek, M. N. *et al.* Plasma tryptophan and the kynurenine-tryptophan ratio are associated with the acquisition of statural growth deficits and oral vaccine underperformance in populations with environmental enteropathy. *Am. J. Trop. Med. Hyg.* **95**, 928–937 (2016).

54. Thévenot, E. A., Roux, A., Xu, Y., Ezan, E. & Junot, C. Analysis of the human adult urinary metabolome variations with age, body mass index, and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses. *J. Proteome Res.* **14**, 3322–3335 (2015).

55. Dietrich, S. *et al.* Identification of serum metabolites associated with incident hypertension in the european prospective investigation into cancer and nutrition-potsdam study. *Hypertension (Dallas, Tex.: 1979)* **68**, 471–477 (2016).

56. Ameta, K. *et al.* Essential hypertension: A filtered serum based metabolomics study. *Sci. Rep.* **7**, 2153 (2017).

57. Kumar, A. A. *et al.* Plasma leptin level mirrors metabolome alterations in young adults. *Metabolomics* **16**, 87 (2020).

58. ter Horst, R. *et al.* Sex-specific regulation of inflammation and metabolic syndrome in obesity. *Arterioscler. Thromb. Vasc. Biol.* **40**, 1787–1800 (2020).

59. van Deventer, C. A. *et al.* Use of metabolomics to elucidate the metabolic perturbation associated with hypertension in a Black South African Male Cohort: The SABPA study. *J. Am. Soc. Hypertension JASH* **9**, 104–114 (2015).

60. Mokkala, K. *et al.* GlycA, a novel marker for low grade inflammation, reflects gut microbiome diversity and is more accurate than high sensitive CRP in reflecting metabolomic profile. *Metabolomics* **16**, 76 (2020).

61. Kirchberg, F. F. *et al.* Metabolomics reveals an entanglement of fasting leptin concentrations with fatty acid oxidation and gluco-neogenesis in healthy children. *PLoS ONE* **12**, e0183185 (2017).

62. Li, K. J., Jenkins, N., Luckasen, G., Rao, S. & Ryan, E. P. Plasma metabolomics of children with aberrant serum lipids and inadequate micronutrient intake. *PLoS ONE* **13**, e0205899 (2018).

63. Rist, M. J. *et al.* Metabolite patterns predicting sex and age in participants of the Karlsruhe Metabolomics and Nutrition (KarMeN) study. *PLoS ONE* **12**, e0183228 (2017).

64. Wishart, D. S. *et al.* HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Res.* **46**, D608–D617 (2018).

65. Lever, M. & Slow, S. The clinical significance of betaine, an osmolyte with a key role in methyl group metabolism. *Clin. Biochem.* **43**, 732–744 (2010).

66. Zhao, G. *et al.* Betaine in inflammation: Mechanistic aspects and applications. *Front. Immunol.* **9**, 1070 (2018).

67. Steurer, M. A. *et al.* Altered metabolites in newborns with persistent pulmonary hypertension. *Pediatr. Res.* **84**, 272–278 (2018).

68. Würtz, P. *et al.* Metabolite profiling and cardiovascular event risk: A prospective study of 3 population-based cohorts. *Circulation* **131**, 774–785 (2015).

69. Cheng, D., Zhao, X., Yang, S., Cui, H. & Wang, G. Metabolomic signature between metabolically healthy overweight/obese and metabolically unhealthy overweight/obese: A systematic review. *Diabetes Metab. Syndrome Obesity Targets Therapy* **14**, 991–1010 (2021).

### Competing interests
The authors declare no competing interests.

### Additional information
**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-30231-9.

**Correspondence** and requests for materials should be addressed to C.B.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

# 4 Discussion and conclusion with references

## 4.1 Discussion

We published three research articles exploring the metabolome and replicating metabolite associations in adolescents and young adults. In the first research article (see 3.1), we explored the urine metabolome for associations with body composition factors BMI and BF. We identified 30 metabolites, 10 of which were associated with both BMI and BF. In the second research article (see 3.2), we explored the urine and blood metabolome of the same participants for associations with a focus on putative biomarkers, as well as searching for replicates from a previous literature search. We identified six metabolites as putative biomarkers and replicated one additional marker from the literature search. In the third research article (see 3.3), we conducted a systematic literature search showing 41 metabolites associated in at least two independent studies with markers of metabolic health and replicated ten of these metabolites in our own cohort. Additionally, we extended the existing results with a mediation analysis identifying no mediating metabolites in the relationship between lifestyle and risk markers.

In all three research articles we found a strong sexual dimorphism of the metabolome. In both blood and urine we saw stronger and/or more associations in one sex depending on the context. For example, in our investigation of body composition we found no metabolite associated with female body composition in our sample, even though statistical reasons are unlikely (as sample size and distribution of variables were approximately equal). In the association with systemic risk markers for chronic disease, however, we found more replicated associations with females compared to males. Only one association was present in both males and females. These results reinforce previous findings both in our own study population (Oluwagbemigun et al., 2020; Oluwagbemigun et al., 2021), in mice (Won et al., 2013), and in independent adult (Darst et al., 2019; Krumsiek et al., 2015; Rist et al., 2017) and adolescent (Saner et al., 2019) study populations. There are multiple hypotheses as to why these sex differences exist. The complete mechanism, however, remains to be elucidated. One popular hypothesis is that sex hormones may be responsible for these differences. Sex hormones play important roles in many different metabolic pathways, like liver energy metabolism (Shen & Shi, 2015) or enzyme activity in the cortisol metabolism (Raven & Taylor, 1996). However, sex hormones are not the

only differentiating factor for the metabolome and its associated pathways. The lipid and lipoprotein metabolism for example shows large sex differences, which were independent of sex hormone administration in a study by Wang et al. (Wang et al., 2011). In summary, there is no one clear explanation for sex differences in the metabolome. Fortunately, there are simple statistical methods, like stratification, that are efficient and appropriate for such stark differences. Unfortunately, to date not many metabolomics studies have stratified their populations by sex (or accounted for the differences in another way). Future studies should consider these marked differences in all steps of their study planning, especially in the design of the statistical analysis.

In our research on body composition, we identified both novel and replicated metabolites. This may suggest that adolescents and young adults differ slightly from adult populations, though some associations were consistent. However, these differences might also stem from other factors related to body composition. For example, the location of fatty tissue (subcutaneous, visceral, or even more specific in the liver) may play a major role in the metabolomic signature of body composition (Rangel-Huerta et al., 2019). Additionally, further factors like current activity level or muscle mass may influence the biochemical pathways that are activated in relationship to body composition (Rangel-Huerta et al., 2019). Future studies should consider these factors, as we did with the use of multiple measures of body composition as opposed to only BMI.

Habitual diet is an important, yet exceedingly complex, exposure in epidemiology (Conrad et al., 2018; Conrad & Nöthlings, 2017). In our second research article, we identified new putative biomarkers and replicated an association from a previous literature review (Langenau et al., 2020). Our results were in line with previous studies (for example for hippuric acid (Clarke et al., 2020; Penczynski et al., 2015)) and we have identified promising metabolites for further studies. Future studies should test and replicate these markers in different settings to understand the pathways to association and their possible application as biomarkers of habitual intake. However, it is important to acknowledge that many well replicated associations were not present in our study. This shows the complexity of diet research in observational studies. Additionally, methodological aspects continue to have a large impact on diet results as well – in our case, the untargeted measurement of the metabolome (compared to more common targeted measurements)

and the use of weighed dietary records to estimate habitual diet (compared to food frequency questionnaires). The efforts to improve and standardize diet assessment are ongoing. Future studies should integrate multiple state-of-the-art measurements of diet. In metabolome-diet associations the use of integrated designs combining both intervention and observational studies that inform each other while using the same tools for metabolome and diet measurement may greatly increase the potential knowledge gain in this complex setting.

In the third paper we were able to confirm multiple associations between metabolites and markers of metabolic health from the literature and identified no mediators. We confirmed 10 of the 41 metabolite-metabolic health marker associations identified in two independent study populations in our systematic literature search. We identified no mediators for either body composition or food intake. We still believe there will be mediators identified in the future because any mediation effects in a complex system like the metabolome seem probable. Mediators are notoriously hard to identify, as their study requires many association tests (which in turn requires a correction for multiple testing), a large study population, and large effect sizes, all three of which were limiting factors within the presented study (see 3.3). Though we were able to contribute important results with this study, showing that association between metabolites and risk markers for chronic disease are already present at an early age, we were not able to show associations to disease outcomes. Accordingly, future studies should extend our research and analysis methods to specific disease outcomes. This would elucidate specific pathways from lifestyle exposures through the metabolome and known risk markers of disease-to-disease outcomes. These results in turn may be instrumental in the long-term understanding of disease etiology and might help to develop new approaches to therapy or drug discovery.

In our three papers we applied multi-faceted methods and stringent criteria to identify associations. We chose this approach in order to gain more actionable, robust results from exploratory analysis. In our first analysis we used multiple independent measures (BMI and body fat) of the same concept (body composition) and concentrated on agreement between both measures. In our second paper we used two exploratory methods and applied stringent criteria to select putative biomarkers of intake. In our third research article we used the results of the previous studies as well as already established

associations found in the literature. In all three cases this approach was successful, as we found and replicated relevant results that fit well with the literature and established hypotheses in addition to some newly discovered associations. There are, however, also some limitations to this approach. While the average relevance of our results may be high, we cannot completely rule out false positives. Additionally, false negatives are more likely with more restrictions placed on the models, as power decreases and sample size is limited. Both of these concerns can be alleviated through proper replication of presented findings in future studies and good documentation and publication of full models and analysis strategies, so that borderline results may still contribute to future literature reviews and meta-analysis.

The metabolomics literature is very heterogeneous. In all three research articles we explored the literature and in the third article we conducted a systematic literature review. It was clear throughout that methods, reporting, and interpretation of results are not standardized. First, there are several initiatives to standardize certain aspects, for example the Metabolomic Standards Reporting Initiative (Sumner et al., 2007), as well as tools to facilitate analysis, for example MetaboAnalyst (Pang et al., 2021). However, these initiative and tools are not widespread yet and no common ground has emerged. This is even visible in our own studies, as we applied four different methods (linear regression, random forest regression, orthogonal projection to latent structures, and causal mediation analysis) in the different contexts of the studies. Even though we based our choices on available literature and simulation studies (see the papers for more details) and took concentrated efforts to use standard implementations and widely used methods, very few other studies are available that would allow for a meta-analysis of any of the three main focuses of our studies. Second, a meta-analysis of the current literature on metabolomics would be nearly impossible because the measurements of the metabolome are widely different between untargeted and targeted metabolomics, but even within scope kits may differ greatly in accuracy and number of metabolites detected. The applied methods form an additional hurdle, as evident in two popular methods: partial least squares regression and random forest. The results of these methods cannot be aggregated together in a meta-analysis. Third, studies often do not publish full model results or full data sets. An increased drive for standardization of the measurement of the metabolome, on the

analysis of metabolomics datasets, and their reporting might greatly increase the quality of knowledge gained from metabolomics in the future.

## 4.2  Conclusion

We identified over 30 new and replicated metabolites associated with the modifiable risk factors body composition or habitual diet. The seven (six new, one replicated) metabolites associated with habitual diet are putative biomarkers of habitual dietary intake. We were not able to detect mediation on the path from lifestyle to conventional systemic markers of chronic disease. Ten previous associations between the metabolome and conventional markers of metabolic health in the DONALD study were confirmed. In all research contexts we observed a strong sexual dimorphism.

Given these results, metabolomics presents itself as a valuable tool and a rich resource to understand the complex interactions of biochemical pathways with exposures and their relationship to disease etiology. However, additional efforts in standardization and replication of results is required to exploit its full potential in the context of modifiable risk factors in epidemiology.

## 4.3 References

Clarke, E. D., Rollo, M. E., Collins, C. E., Wood, L., Callister, R., Philo, M., Kroon, P. A., & Haslam, R. L. (2020). The Relationship between Dietary Polyphenol Intakes and Urinary Polyphenol Concentrations in Adults Prescribed a High Vegetable and Fruit Diet. *Nutrients*, *12*(11), 3431 https://doi.org/10.3390/nu12113431

Conrad, J., Koch, S. A. J., & Nöthlings, U. (2018). New approaches in assessing food intake in epidemiology. *Current Opinion in Clinical Nutrition and Metabolic Care*, *21*(5), 343–351 https://doi.org/10.1097/MCO.0000000000000497

Conrad, J., & Nöthlings, U. (2017). Innovative approaches to estimate individual usual dietary intake in large-scale epidemiological studies. *The Proceedings of the Nutrition Society*, *76*(3), 213–219 https://doi.org/10.1017/S0029665116003025

Darst, B. F., Koscik, R. L., Hogan, K. J., Johnson, S. C., & Engelman, C. D. (2019). Longitudinal plasma metabolomics of aging and sex. *Aging*, *11*(4), 1262–1282 https://doi.org/10.18632/aging.101837

Krumsiek, J., Mittelstrass, K., Do, K. T., Stückler, F., Ried, J., Adamski, J., Peters, A., Illig, T., Kronenberg, F., Friedrich, N., Nauck, M., Pietzner, M., Mook-Kanamori, D. O., Suhre, K., Gieger, C., Grallert, H., Theis, F. J., & Kastenmüller, G. (2015). Gender-specific pathway differences in the human serum metabolome. *Metabolomics: Official Journal of the Metabolomic Society*, *11*(6), 1815–1833 https://doi.org/10.1007/s11306-015-0829-0

Langenau, J., Oluwagbemigun, K., Brachem, C., Lieb, W., Giuseppe, R. di, Artati, A., Kastenmüller, G., Weinhold, L., Schmid, M., & Nöthlings, U. (2020). Blood Metabolomic Profiling Confirms and Identifies Biomarkers of Food Intake. *Metabolites*, *10*(11), 468 https://doi.org/10.3390/metabo10110468

Oluwagbemigun, K., Anesi, A., Clarke, G., Schmid, M., Mattivi, F., & Nöthlings, U. (2021). An Investigation into the Temporal Reproducibility of Tryptophan Metabolite Networks Among Healthy Adolescents. *International journal of tryptophan research: IJTR*, *14*, 11786469211041376 https://doi.org/10.1177/11786469211041376

Oluwagbemigun, K., Anesi, A., Ulaszewska, M., Clarke, G., Alexy, U., Schmid, M., Roden, M., Herder, C., Mattivi, F., & Nöthlings, U. (2020). Longitudinal relationship of amino acids

and indole metabolites with long-term body mass index and cardiometabolic risk markers in young individuals. *Scientific Reports*, *10*(1), 6399 https://doi.org/10.1038/s41598-020-63313-z

Pang, Z., Chong, J., Zhou, G., de Lima Morais, D. A., Chang, L., Barrette, M., Gauthier, C., Jacques, P.-É., Li, S., & Xia, J. (2021). MetaboAnalyst 5.0: Narrowing the gap between raw spectra and functional insights. *Nucleic Acids Research*, *49*(W1), W388–W396 https://doi.org/10.1093/nar/gkab382

Penczynski, K. J., Krupp, D., Bring, A., Bolzenius, K., Remer, T., & Buyken, A. E. (2015). Relative validation of 24-h urinary hippuric acid excretion as a biomarker for dietary flavonoid intake from fruit and vegetables in healthy adolescents. *European Journal of Nutrition*, *56*(2), 757–766 https://doi.org/10.1007/s00394-015-1121-9

Rangel-Huerta, O. D., Pastor-Villaescusa, B., & Gil, A. (2019). Are we close to defining a metabolomic signature of human obesity? A systematic review of metabolomics studies. *Metabolomics: Official Journal of the Metabolomic Society*, *15*(6), 93 https://doi.org/10.1007/s11306-019-1553-y

Raven, P. W., & Taylor, N. F. (1996). Sex differences in the human metabolism of cortisol. *Endocrine Research*, *22*(4), 751–755 https://doi.org/10.1080/07435809609043772

Rist, M. J., Roth, A., Frommherz, L., Weinert, C. H., Krüger, R., Merz, B., Bunzel, D., Mack, C., Egert, B., Bub, A., Görling, B., Tzvetkova, P., Luy, B., Hoffmann, I., Kulling, S. E., & Watzl, B. (2017). Metabolite patterns predicting sex and age in participants of the Karlsruhe Metabolomics and Nutrition (KarMeN) study. *PloS One*, *12*(8), e0183228 https://doi.org/10.1371/journal.pone.0183228

Saner, C., Harcourt, B. E., Pandey, A., Ellul, S., McCallum, Z., Kao, K.-T., Twindyakirana, C., Pons, A., Alexander, E. J., Saffery, R., Burgner, D. P., Juonala, M., & Sabin, M. A. (2019). Sex and puberty-related differences in metabolomic profiles associated with adiposity measures in youth with obesity. *Metabolomics: Official Journal of the Metabolomic Society*, *15*(5), 75 https://doi.org/10.1007/s11306-019-1537-y

Shen, M., & Shi, H. (2015). Sex Hormones and Their Receptors Regulate Liver Energy Homeostasis. *International Journal of Endocrinology*, *2015*, 294278 https://doi.org/10.1155/2015/294278

Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., Fan, T. W.-M., Fiehn, O., Goodacre, R., Griffin, J. L., Hankemeier, T., Hardy, N., Harnly, J., Higashi, R., Kopka, J., Lane, A. N., Lindon, J. C., Marriott, P., Nicholls, A. W., … Viant, M. R. (2007). Proposed minimum reporting standards for chemical analysis chemical analysis working group (CAWG) metabolomics standards initiative (MSI). *Metabolomics : Official Journal of the Metabolomic Society*, *3*(3), 211–221 https://doi.org/10.1007/s11306-007-0082-2

Wang, X., Magkos, F., & Mittendorfer, B. (2011). Sex differences in lipid and lipoprotein metabolism: It's not just about sex hormones. *The Journal of Clinical Endocrinology and Metabolism*, *96*(4), 885–893 https://doi.org/10.1210/jc.2010-2061

Won, E.-Y., Yoon, M.-K., Kim, S.-W., Jung, Y., Bae, H.-W., Lee, D., Park, S. G., Lee, C.-H., Hwang, G.-S., & Chi, S.-W. (2013). Gender-Specific Metabolomic Profiling of Obesity in Leptin-Deficient ob/ob Mice by 1H NMR Spectroscopy. *PLOS ONE*, *8*(10), e75998 https://doi.org/10.1371/journal.pone.0075998

## Acknowledgements

I would like to acknowledge some people without whom this thesis would not have been possible.

First, I would like to thank my wonderful partner Marvin for his unrelenting support and unwavering ability to invoke joy even in the most frustrating of times. Second, my parents, Gertrud and Hermann, who gave their all to make all my wishes and goals possible and offered never-ending support. Third, my siblings, Elisabeth, Johannes, and Kathrin, who all helped and listened in uncountable ways. Fourth, my dog Sherlock who forced me to take breaks and head out into nature despite being locked in during a pandemic.

Furthermore, I would like to acknowledge my wonderful colleagues at the Institute for Food and Nutrition Sciences in the Nutritional Epidemiology Department, especially Julia and Kolade. All throughout I knew I had people who shared my experiences and were there to listen and give advice when needed.

Last, but not least, I would like to acknowledge the DONALD study participants. Without their commitment to share their experiences and give their time this dissertation would not have been possible.