RESEARCH ARTICLE

Discovery of urinary biomarkers of whole grain rye intake in free-living subjects using nontargeted LC-MS metabolite profiling

Kati Hanhineva[1], Carl Brunius[2]*, Agneta Andersson[3], Matti Marklund[4], Risto Juvonen[5], Pekka Keski-Rahkonen[5], Seppo Auriola[5] and Rikard Landberg[2,6]*

[1] Department of Clinical Nutrition, Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland
[2] Department of Food Science, Uppsala BioCenter, Swedish University of Agricultural Sciences, Uppsala, Sweden
[3] Department of Food, Nutrition and Dietetics, Uppsala University, Uppsala, Sweden
[4] Department of Public Health and Caring Sciences, Clinical Nutrition and Metabolism, Uppsala Univeristy, Uppsala, Sweden
[5] School of Pharmacy, University of Eastern Finland, Kuopio, Finland
[6] Nutritional Epidemiology Unit, Institute of Environmental Medicine, Karolinska Insitutet, Stockholm, Sweden

Received: June 2, 2015
Revised: August 5, 2015
Accepted: August 6, 2015

Scope: Whole grain (WG) intake is associated with decreased risk of developing colorectal cancer, type 2 diabetes, and cardiovascular disease and its comorbidities. However, the role of specific grains is unclear. Moreover, intake of specific WG is challenging to measure accurately with traditional dietary assessment methods. Our aim was to use nontargeted metabolite profiling to discover specific urinary biomarkers for WG rye to objectively reflect intake under free-living conditions.

Methods and results: WG rye intake was estimated by weighed food records, and 24 h urine collections were analyzed by LC-MS. Multivariate modeling was undertaken by repeated double cross-validated partial least squares regression against reported WG rye intake, which correlated well with multivariate prediction estimates ($r = 0.67$–0.80, $p < 0.001$), but not with intakes of WG wheat or oats. Hydroxyhydroxyphenyl acetamide sulfate, 3,5-dihydroxyphenylpropionic acid sulfate, caffeic acid sulfate, and hydroxyphenyl acetamide sulfate were among the 20 features that had the greatest potential as intake biomarkers of WG. In addition, three compounds exhibited MS/MS fragmentation of carnitine structures.

Conclusion: With this nontargeted approach, we confirmed the specificity of alkylresorcinol metabolites as biomarkers for WG rye intake, but also discovered other compounds that should be evaluated as putative biomarkers in future studies.

Keywords:
Biomarker / Metabolite profiling / Metabolomics / Rye / Whole grain

1 Introduction

Diet plays an important role in modifying the risk of chronic diseases [1] but there is often a lack of consistency in the findings for specific food items or categories from observational studies as well as from human intervention studies [2]. One reason is the difficulty to accurately measure dietary exposures. This includes the relatively large measurement errors associated with traditional methods typically used in large

Correspondence: Rikard Landberg
E-mail: rikard.landberg@slu.se

Abbreviations: **3DWFR**, 3-day weighed food records; **DHPPA**, dihydroxyphenylpropionic acid; **HHPAA**, hydroxyhydroxyphenyl acetamide; **HILIC**, hydrophilic interaction chromatography; **HPAA**, hydroxyphenyl acetamide; **ICC**, intraclass correlation coefficient; **LV**, latent variable; **QC**, quality control; **rdCV-PLS**, repeated double cross-validated partial least squares regression; **RMSEP**, root mean squared error of prediction; **sPLS**, sparse partial least squares regression; **VIP**, variable importance in projection; **WG**, whole grain

*These authors contributed equally to this work.

Colour Online: See the article online to view Fig. 3 in colour.

observational studies, such as the food frequency questionnaires as well as lack of compliance in human intervention studies [3]. Moreover, translations of food intakes into nutrients and bioactive compounds are dependent on food composition tables [1]. To overcome these problems, specific dietary biomarkers, i.e. compounds ingested with a specific food item or metabolites thereof, may better capture exposure and could hence potentially be used instead or as a complement to traditional dietary assessment methods [2]. Unfortunately, there are very few valid biomarkers available [3] and putative biomarkers identified are typically concentration biomarkers, i.e. biomarkers where the concentrations are correlated with the intake, but where other determinates also may affect the intake-biomarker relationship [4]. In most cases, only a small proportion of the variation of these markers is explained by direct intake with other determinants typically being unknown. New approaches are being developed to combine different biomarkers in order to better assess compliance in human intervention studies as well as combining biomarkers with traditional dietary assessment methods to improve ranking of exposure. Such methods have proven useful in alleviating regression dilution bias and improve statistical power in epidemiological studies. This in turn motivates intensified search for new markers of different dietary exposures [2, 4].

Typically, targeted approaches have been used to validate a priori defined compounds as biomarkers of food and nutrient intakes for monitoring of compliance in human intervention studies and for risk prediction in epidemiological studies. Using this approach, biomarkers of whole grain (WG) wheat and rye [5], some fruits and vegetables [6], sugar [7, 8], and milk fat [9, 10] have been developed. Recently, untargeted metabolite profiling, which allows comprehensive analysis of metabolites in biological fluids, has proven a powerful approach to detect metabolites that derive from dietary exposure [11–14]. This approach does not require previous knowledge on the presence of specific compounds in the food exposures of interest and it has successfully been used to discover new putative biomarker of foods such as citrus, coffee, coconuts, and dietary patterns under intervention conditions [15] and to a lesser extent under free-living conditions [11, 15–18].

WG intake is an important dietary component, which has been consistently associated with protective effects on type 2 diabetes, cardiovascular diseases, and colorectal cancer in epidemiological studies [19–21]. WG intake has also shown inconsistent effects on cardiometabolic risk factors in some interventions studies [22–27], but the effects may differ between grains and it is therefore important to assess their intake separately [28]. To our knowledge, no study has used untargeted metabolomics in search of putative biomarkers of WG intake in a free-living population.

WG rye bread is frequently consumed in the Nordic countries, and it has been consistently associated with lower mortality [29]. In the present study, the aim was to apply untargeted LC-qTOF-MS based metabolomics to discover putative biomarkers specific for WG rye intake in 24 h urine collections made under free-living conditions. The aim was further to evaluate the reproducibility of identified markers in samples taken 1–3 months apart.

2 Materials and methods

2.1 Study design and subjects

Study design and recruitment of participants have been thoroughly described elsewhere [30]. Briefly, free-living participants with no diagnosed or perceived gastrointestinal diseases or symptoms were invited and instructed to adhere to their normal diet. From initially 91 recruited participants, completed 3-day weighed food records (3DWFR), morning spot urine, and 24 h urine collection were available for 66 participants (50 females, 16 males) on two separate occasions (period A and B) approximately 2–3 months apart. Samples from 59 subjects were available for the present study. Body weight and height, BMI, and age were recorded on the second occasion. Mean (± SD) age and BMI of the participants were 44 ± 17 years and 24 ± 4 kg/m^2, respectively. This study was conducted according to the guidelines laid down in the Declaration of Helsinki and all procedures involving human subjects were approved by the Regional Ethical Review Board in Uppsala (log. no. 2008:040). Written informed consent was obtained from all subjects.

2.2 Dietary data

Assessment of self-recorded dietary intake was performed as previously described [30]. In brief, nutrient intake was calculated using the food database of the Swedish National Food Administration (PC-kost 2008–03-06) and a computerized calculation program (Diet XP; Kost och Näringsdata AB). WG was defined in accordance with the definition of the American Association of Cereal Chemists [31] and the content was estimated based on product declarations of reported products. In case data on WG content was missing for products reported and where WG from different grains were included in the product, food companies were contacted to get information on the exact content of WG from the different cereals.

2.3 Urine samples

A morning spot urine sample and a 24 h collection were taken during the last day of the 3DWFR. Participants were asked to collect the first morning void and transfer 25 mL into a falcon tube (spot urine sample) and the rest into the uriset container for 24 h collection, which was ended when the subject went to bed. To prevent microbial growth, participants were advised to add 9 mL 20% HCl to each 24 h collection. Urine samples were brought to the clinic within 24 h, upon which volume was recorded and samples were aliquoted and frozen to −20°C immediately and transferred to −80°C freezer

within a week. All samples were then kept at −80°C until thawed on ice bath for analysis.

For metabolomics analysis, 200 μL urine sample was mixed with 200 μL acetonitrile (VWR International) on 96-well filtration plates (0.2 μm Captiva with 1 mL collection plates; Agilent Technologies). After filtration, the well plates were covered with silicone sealing mats and kept refrigerated until analysis. Quality control (QC) samples were prepared using a urine pool that was prepared by combining small aliquots of randomly selected samples ($n = 50$). Each analytical run was initiated with 10–20 priming injections of a QC sample to achieve stable retention times and ion intensity, monitored by overlaying of resulting total ion chromatograms. Analytical samples were analyzed in random order, interspersed after every 12 injections with QC sample injections to monitor instrument performance and sample stability. The complete sample set from the study included 236 samples (spot urine and 24 h collection samples). This study was performed only using samples from the 24 h collection ($n = 118$).

2.4 LC-qTOF-MS and MS/MS analysis

2.4.1 LC-MS analysis

Samples were analyzed by LC-qTOF-MS (Agilent Technologies) consisting of a 1290 Binary LC system, a Jet Stream ESI source, and a 6540 qTOF mass spectrometer. RP and hydrophilic interaction chromatography (HILIC) were employed, using both positive and negative ESI mode. Autosampler tray was kept at 4°C at all times. For the RP analyses, 4 μL of the sample solution was injected on a Zorbax Eclipse XDB-C18 column (2.1 × 100 mm, 1.8 μm; Agilent Technologies). Column temperature was 50°C. Mobile phase flow rate was 0.4 mL/min and consisted of water (eluent A; Milli-Q Gradient, Millipore, Milford, MA) and methanol (eluent B; Sigma-Aldrich), both containing 0.1% v/v of formic acid (Sigma-Aldrich), delivered with the following gradient conditions: 0–10 min: 2 → 100% B, 10–15 min: 100% B, 15–15.1 min: 100 → 2% B; 15.1–18 min: 2% B. For HILIC, the chromatographic column was Acquity UPLC BEH Amide (100 × 2.1 mm, 1.7 μm; Waters). Mobile phase eluents A and B were 50% ACN v/v and 90% ACN v/v, respectively, both containing 20 mM of pH 3.5 ammonium formate (all from Sigma-Aldrich). The gradient was as follows: 0–2.5 min: 100% B, 2.5–10 min: 100 → 0% B, 10–10.01 min: 0 → 100% B; 10.01–14 min: 100% B. Column temperature was 45°C, flow rate 0.6 mL/min, and injection volume 1 μL.

For both chromatographic methods, ESI source was operated using the following conditions: drying gas (nitrogen) temperature 325°C and flow 10 L/min, sheath gas temperature 350°C and flow 11 L/min, nebulizer pressure 45 psi, capillary voltage 3500 V, nozzle voltage 1000 V, and fragmentor voltage 100 V. Data acquisition was performed using 2 GHz extended dynamic range mode across a mass range of m/z 50–1600. Scan rate was 2.5 Hz (RP) or 1.5 Hz (HILIC). Data acquisition was in centroid mode with an abundance threshold of 150 counts. For automatic data-dependent MS/MS analyses, precursor isolation width was 1.3 Da, and from every precursor scan cycle, the four most abundant ions were selected for fragmentation. These ions were excluded after two product ion spectra and released again for fragmentation after a 0.25 min hold. Precursor scan time was based on ion intensity, ending at 20 000 counts or after 300 ms. Product ion scan time was 300 ms. Collision energies were 10, 20, and 40 V in subsequent runs. Continuous mass axis calibration was performed by monitoring two reference ions from an infusion solution throughout the runs (m/z 113.988900 and m/z 966.000725 for negative; m/z 121.050873 and m/z 922.009798 for positive polarity). Data were acquired using MassHunter Acquisition B.04.00 (Agilent Technologies).

2.4.2 Data processing

Preprocessing of the data was performed with MassHunter Qualitative Analysis B.05.00 (Agilent Technologies), using Find Compounds by Molecular Feature algorithm. Ions were collected to compounds exhibiting isotopic peaks, dimers, and common adducts ($[M+Cl]^-$, $[M+Na]^+$) with mass peak threshold at 200 counts, and compound height threshold at 2500 counts with at least two ions per compound. The data were exported in compound exchange format to Mass Profiler Professional (MPP) software (version 2.2, Agilent Technologies) for compound alignment baselining each entity to median across all samples. Features present in less than 5% of the samples were removed from the data resulting in altogether 16 571 features collected from the four LC-MS modes. The identification of the putative biomarker candidates was accomplished by evaluation of MS/MS spectra and comparison with pure compounds, earlier published data, or searching databases (Metlin, Human Metabolome Database, SciFinder).

Combined data from all four LC-MS modes were imported into R (R Core Team 2014) [32] for multivariate analysis. Observations with data lacking from any LC-MS mode ($n = 6$) were removed, resulting in data from 112 complete observations. Any feature with CV <20% ($n = 179$) was considered not relevant and removed from the dataset. Approximately 0.6% of values were missing, with missingness assumed to be predominantly related to low concentration, as indicated by k-nearest neighbor (kNN) imputation ($k = 5$). However, in order not to introduce bias from k-nearest neighbor -imputation, final imputation of missing values was performed by random sampling from a normal distribution between 0 and 2/3 of lowest measured value within feature.

Multivariate modeling was conducted by repeated double cross-validated partial least squares regression (rdCV-PLS) [33, 34] against reported WG rye intake using an in-house developed procedure. The rdCV-PLS was performed on all complete observations ($n = 112$) of both the full remaining

feature set (16 392 features) and on a feature set obtained from sparse PLS [35, 36] (sPLS) variable importance in projection (VIP) prefiltering. Data prefiltering was conducted by direct sPLS modeling (no cross-validation) keeping 1000 features per latent variable (LV) over three LV. In the sPLS model, the feature matrix was used as independent variables and WG rye, WG wheat, and WG oats intake (g/day) as dependent variables. Those features with VIP >1 in the sPLS model with two LV were selected for filtered rdCV-PLS analysis (1133 features).

As opposed to standard cross-validation, the double cross-validation procedure separates cross-validation into an outer "testing" loop and an inner "tuning" (or validation) loop to reduce bias from overfitting models to experimental data. Feature ranking and selection was thus performed within the inner loop, to minimize statistical overfitting, by iteratively tuning over successively fewer features, removing for each step in the loop the least informative features, as measured by VIP. Both full and prefiltered models were run removing either 10 or 25% of features per iteration resulting in four multivariate models: "Full," "qFull," "VIP," and "qVIP," where *q* denotes "quick" model (i.e. with 25% feature removal per iteration) and VIP denotes prefiltering using sPLS VIP-values as described above. Moreover, the "quick" models were subjected to 12 repetitions of the double cross-validation loop, while the other two were repeated 40 times. Root mean squared error of prediction (RMSEP) of inner loop validation sets was used as fitness function for the tuning algorithm. Those features with average VIP ranks <100 across repetitions in the cross-validation scheme were chosen for more detailed investigation (Table 1). Correlations were calculated between continuous WG intake and model predictions according to both Pearson and Spearman to allow the reader the freedom of choice between the two. Permutation analysis was performed by one-tailed test of $RMSEP_{H1} < RMSEP_{H0}$ distributions ($n = 100$ each) of actual model performance (H1) versus randomized WG rye intake (H0) using the "qVIP" rdCV-PLS parameter settings described above.

Multivariate modeling was parallelized using all four cores on a HP Elitebook with an Intel i7–3687U processor, requiring 1297, 142, 93, and 10 min for the "Full," "qFull," "VIP", and "qVIP" models, respectively. Hundred iterations of "qVIP" models and permutations using "qVIP" parameter settings required approximately 32 h. Full R script is available from the authors upon request.

Reproducibility of putative biomarkers was assessed over the period between first and second sample collection, using the intraclass correlation coefficients (ICC). ICC is defined as the interparticipant variance divided by the total variance (i.e. the sum of intra and interparticipant variance), assuming a common mean between replicate measurements. We calculated ICC and intraparticipant variance components using a freely available SAS macro [37]. The macro uses SAS PROC MIXED to estimate the maximum likelihood estimates of intra-participant and inter-participant variance components. A method described by Hankinson et al. was used to calculate 95% CI of the ICC [38]. ICC calculations were made in SAS version 9.3 (SAS Institute).

2.5 Synthesis of hydroxyphenyl acetamide sulfate conjugate

In order to unequivocally identify hydroxyphenyl acetamide (HPAA) sulfate among two candidate biomarkers with identical masses in the data, a sulfated conjugate of HPAA was produced by sulfotransferase-catalyzed reaction using liver cell cytosolic fraction. Pig liver samples were from female control pigs used for practicing surgical procedures, University of Eastern Finland, Kuopio. Liver microsomes and the cytosolic fraction were prepared from pigs' livers as described previously [39]. In brief, 20 μM HPAA (2-acetamidophenol; CAS 614-80-2, Sigma) was sulfonated in 500 μL 50 mM phosphate buffer pH 7.4 containing 5 mM $MgCl_2$, 1 mg cytosolic protein of pig liver, and 100 μM PAPS for 1 h at 37°C. The reaction was terminated by addition of 1.5 mL acetonitrile. The samples were then centrifuged for 10 min at 10 000 × *g* and the supernatant was collected for analysis. The use of liver samples for these syntheses was approved by the Ethics Committee for Animal Experiments, University of Eastern Finland, Kuopio.

3 Results and discussion

3.1 Multivariate modeling

Multivariate modeling of big variable spaces, especially when employing more complex validation schemes such as rdCV, is a computer intensive task and can in the normal case with few observations present huge dimensionality problems [40]. In order to decrease the variable space, various prefiltering techniques are commonly used: univariate based on fold change and Student's *t*-test and/or multivariate based on either of a number of techniques [41]. Using a sparse version of PLS, the feature matrix was related to total WG intake (LV1), WG rye (LV2), and WG wheat/oats (LV3) (Fig. 1). Feature prefiltering for regression against WG rye intake was therefore based on VIP scores (VIP >1) from the first two LVs.

When subjected to rdCV-PLS modeling, the full and the prefiltered datasets generated considerable overlap between selected features (Table 1). All selected top-ranking features from both datasets exhibited medium to strong Pearson correlations with WG rye intake, indicating that they are putative biomarkers that can be used independently to reflect WG rye intake. Interestingly, one unidentified feature exhibited a negative correlation with WG rye intake, suggesting that metabolites that respond negatively to exposure (e.g. due to increased biotransformation) may in fact be of high predictive value. Moreover, three of the features exhibited poor Spearman correlation with WG rye intake, which may indicate that they either do not contain relevant biological

Table 1. Characteristics of features selected from multivariate rdCV-PLS on reported rye consumption for full (16 571 features) and sPLS-VIP prefiltered (1133 features) datasets

Feature				MS/MS fragment ions (*m/z*)	Average VIP ranks				Correlation[e]		Annotation	ICC (95% CI)
Column	ESI	M[a]	rt[b]		Full[c]	qFull[d]	VIP[c]	qVIP[d]	Pearson	Spearman		
HILIC	Positive	283.9974	1.296		6.0	6.8	6.7	6.3	0.62	0.56	Unknown	0.48 (0.29–0.67)
RP	Negative	138.0677	2.737	227.0550, 181.0513, 137.0604, 95.0502	11.7	12.7	14.2	12.0	0.64	0.52	DHPPA derivative (fragment ion)	0.48 (0.29–0.67)
RP	Negative	247.0152	2.225	246.0078, 166.0505, 108.0451, 118.0290, 96.9591	12.6	12.6	10.6	10.7	0.62	0.61	HHPAA sulfate	0.01 (0.00–1.00)
RP	Negative	262.0145	2.111	137.06034, 181.04988, 261.00624, 95.05063, 151.03962, 79.95769, 119.04942	17.9	19.3	14.0	16.9	0.61	0.65	3,5-DHPPA sulfate	0.41 (0.23–0.55)
HILIC	Negative	368.2198	1.122	367.212, 331.194, 287.076, 243.101, 179.037, 150.031, 120.977	21.0	20.6	27.1	33.1	0.58	0.27	Unknown	0.16 (0.03–0.56)
RP	Positive	160.0739	1.254	121.035, 114.941, 97.918, 69.069, 55.043	27.7	28.8	23.0	25.4	0.58	0.44	Pimelic acid	0.44 (0.25–0.80)
RP	Negative	620.1056	2.115	137.06034, 181.04988, 261.00624, 95.05063, 151.03962, 79.95769, 119.04942	27.9	29.3	26.0	24.1	0.59	0.63	3,5-DHPPA sulfate (with adduct)	0.29 (0.11–0.74)
HILIC	Positive	376.1353	0.706	377.135, 279.170, 211.109, 165.089	33.8	35.3	51.9	32.6	0.55	0.56	Unknown	0.44 (0.26–0.65)
RP	Positive	167.0583	2.615	168.065, 110.059, 92.049	37.8	38.2	75.6	29.3	0.54	0.36	HHPAA	0.04 (0.00–0.98)
RP	Positive	325.1535	2.310	326.160, 267.085, 183.065, 165.054, 137.058, 85.029, 60.081	38.1	>100	76.8	59.6	0.55	0.58	Unknown carnitine	0.71 (0.56–0.82)
HILIC	Negative	259.9990	1.453	258.992, 179.034, 130.085	89.4	31.9	27.4	27.8	0.58	0.58	Caffeic acid sulfate	0.39 (0.20–0.62)
RP	Negative	228.0630	2.737	227.0550, 181.0513, 137.0604, 95.0502	>100	>100	36.3	33.0	0.63	0.34	DHPPA derivative	0.50 (0.31–0.69)
RP	Positive	540.2155	6.037		>100	>100	43.2	17.3	0.54	0.04*	Unknown	0[f]
RP	Positive	207.0532	1.196	208.064, 191.037, 144.047, 119.016, 87.045	>100	>100	73.7	62.7	0.51	0.00*	Unknown	0[f]
RP	Negative	109.0527	2.515	230.011, 150.055, 108.045, 96.960	>100	>100	>100	50.5	0.54	0.50	HPAA sulfate (fragment)	0.08 (0.00–0.76)
RP	Negative	259.9992	2.456	258.99245, 179.03557, 135.04276, 128.0351, 130.08568	>100	>100	>100	79.1	0.54	0.60	caffeic acid sulfate	Same as above
RP	Negative	368.2197	6.046	367.213, 331.188, 287.143, 256.629, 183.138, 112.983	>100	>100	>100	79.5	0.51	0.36	Unknown	0.31 (0.13–0.57)
RP	Positive	474.2411	5.918	475.246, 298.200, 253.143, 164.036, 85.029, 60.080	>100	>100	>100	79.7	−0.35	−0.39	Unknown carnitine	0.73 (0.59–0.84)
RP	Negative	438.0623	6.493	437.054, 357.097, 203.973, 124.016, 79.957	>100	>100	>100	88.5	0.53	0.06*	Unknown sulfate	0.16 (0.03–0.55)
RP	Positive	340.1647	3.171	341.173, 323.151, 264.086, 234.115, 136.075, 85.029, 60.081	>100	>100	>100	99.7	0.44	0.42	Unknown carnitine	0.37 (0.18–0.61)

a) Neutral mass of molecular feature.
b) Chromatographic retention time of molecular feature.
c) Model based on removal of 10% lowest ranked features per iteration.
d) Model based on removal of 25% lowest ranked features per iteration.
e) All univariate correlations significant at $p < 0.001$, except *$p > 0.5$.
f) No confidence interval could be calculated.

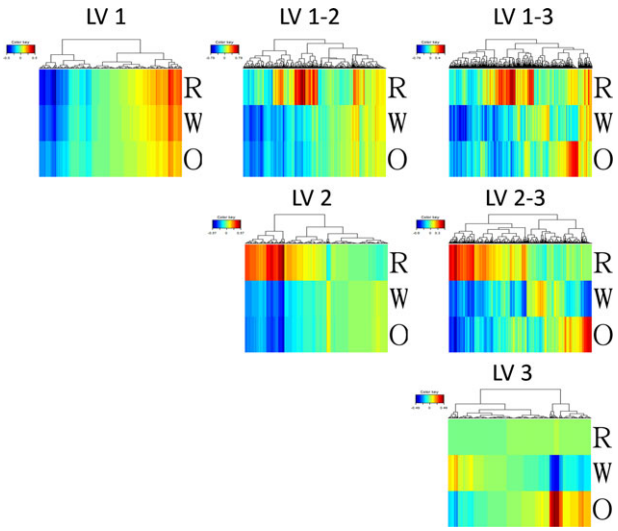

Figure 1. Correlation between WG grain intake (R = rye, W = wheat, O = oats; dependent variables; *y*-axis in subplots) and LC-MS metabolite features (independent variables, *x*-axis in subplots). The correlations are presented as heatmaps viewed for different combinations of LV in the sparse PLS prefiltering model. Separation between rye and the other two cereals occurs mainly in LV 2.

information related to exposure or that they contain other important information that helps to improve overall signal quality.

Multivariate prediction estimates correlated well between models and also with the continuous WG rye intake variable, but not with intake of WG wheat or WG oats (Table 2), indicating a high model specificity to WG rye consumption. Moreover, the metabolite feature correlation with WG rye consumption was higher for the prefiltered models compared to the full models suggesting that prefiltering results in higher predictive power due to noise removal or that prefiltering to a certain degree results in overfitting. Comparing model prediction accuracy to the H0 distribution, average model RMSEP is at the lower end although some random models have even lower values (Fig. 2A). Normal permutation analysis, where single value model performance is compared to the H0 probability density function [42] results in an overall $p = 0.115$ (Fig. 2A), resulting either from the positive skew in the distribution of WG rye intake (data not shown) or indicating model overfitting. However, we would argue that individual model results are representative of an H1 distribution of actual model performance. A permutation test would then investigate H1 < H0 among these two distributions either parametrically or nonparametrically (Fig. 2B). Using this approach, the actual models were shown to vastly outperform random permutations ($p < 10^{-15}$). Prefiltering thus resulted in modeling at considerably lower computational cost, with similar predictive ability as the full models, thus indicating biological validity regardless of choice of hypothesis testing in permutation analysis.

3.2 Identification of putative biomarkers of rye intake

Among the 20 metabolite signals selected from rdCV-PLS modeling, four signals were recognized as fragments or adducts of other compounds included in the list, or represented both in RP and HILIC data. Among the 16 individual compounds, ten were annotated based on the MS/MS fragmentation, four could not be identified in spite of clear MS/MS fragmentation pattern, and for two compounds no MS/MS data were obtained. (Table 1).

3.3 Phenolic compounds

Structural assignments based on MS/MS spectral interpretation was achieved for caffeic acid sulfate and two conjugates of the dihydroxyphenylpropionic acid (DHPPA) (Table 1). The DHPPA compounds included *m/z* 261.006 at 2.11 min and *m/z* 227.055 at 2.74 min. The *m/z* 261 entity exhibited loss of a sulfate unit, resulting in an aglycone with *m/z* 181.049. Fragmentation of the aglycone was compared to the fragmentation of pure 3,5-DHPPA standard exhibiting main MS/MS fragments of *m/z* 137.061 and 95.050, similarly as the metabolite in the sample, and was therefore identified as 3,5-DHPPA sulfate. Notably, the hydroxyl group configuration is suggested to be 3,5 since only this compound gives the clear fragment of *m/z* 95.050. In analogy, the MS/MS spectrum available at Metlin for 3,4-DHPPA (MID:3789) does not show this fragment. The same aglycone fragments (*m/z* 137.061 and 95.050) were observed also for the compound with *m/z* 227.055 at 2.74 min in addition to the neutral loss of 46.05, which could correspond to HCOOH. Although precise identification for this compound could not be achieved, the fragments suggest a conjugate of DHPPA.

The identified phenolic compounds that were specific for WG rye consumption originate mainly from two phytochemical sources; the phenolic acids that are mainly bound to arabinoxylans in the bran, and the phenolic lipids, alkylrecorcinols, found in the outer cuticule of testa/inner curricula of pericarp [43]. The configuration of the hydroxyl groups is different in these two metabolite classes—it is typically 3,4 in phenolic acids, whereas in alkylresorcinols the hydroxyl-groups are in positions 3 and 5 in the phenol ring structure. In this study, representatives of both groups were identified as potential biomarkers; fee and conjugated 3,5-DHPPA are main metabolites from alkylresorcinols [44–46] (1,3-dihydroxy-5-alkylbenzene homologues) found in the outer cuticle of testa/inner curricula of pericarp [43], and are established biomarkers of intake of WG rye, whereas caffeic acid sulfate most likely originates from the bran-bound phenolic acids. Therefore, the identification of small phenolic compounds in untargeted LC-MS analysis is challenging, and definitive assignment and identification can only be performed based on MS/MS spectra from pure standard compounds with careful evaluation of retention times.

Table 2. Correlations between WG rye intake (g/d) and model prediction estimates (Pearson/Spearman correlations above/below diagonal)

	Prediction estimates[a)]				Intake[b)]		
	Full	qFull	VIP	qVIP	WG Rye	WG Wheat	WG Oats
Full		0.99***	0.97***	0.95***	0.68***	0.05	0.10
qFull	0.98***		0.96***	0.93***	0.67***	0.05	0.08
VIP	0.93***	0.92***		0.98***	0.77***	0.06	0.13
qVIP	0.93***	0.90***	0.97***		0.80***	0.09	0.14
WG Rye	0.60***	0.60***	0.70***	0.69***		0.02	0.11
WG Wheat	0.09	0.09	0.11	0.09	−0.03		0.23*
WG Oats	0.11	0.12	0.16	0.13	0.05	0.34***	

a) Multivariate models described in section 2.
b) Based on reported WG intake (g/d).
Levels of statistical significance: $*p < 0.05$; $***p < 0.00$.

3.4 Phenylacetamide compounds

Three of the top-ranked metabolites were annotated as phenylacetamides (Table 1). Hydroxyhydroxyphenyl acetamide (HHPAA) sulfate was identified based on earlier analyses, where it has been included as a urinary biomarker for rye-rich diet [47] as well as a specific biomarker for consumption of breads containing rye bran compared with refined wheat in blood samples from a post prandial study [48]. HHPAA has been suggested to be a metabolite converted from a benzoxazinoid precursor (2,4-dihyxdoxy-1,4-benzoxazin-3-one), mediated by soil bacteria [49, 50]. However, it is not yet known whether dietary phenylacetamides are similarly formed from benzoxazinoid precursors, or at which stages of food processing or digestion such conversion could potentially occur. Additionally, a feature with *m/z* 168.065 was identified as free form of HHPAA based on identical fragmentation and retention time as in earlier analysis [48]. Another ion, namely *m/z* 108.045 eluting at 2.5 min was included among the top-ranked metabolites. This ion is a result of ion source fragmentation of *m/z* 230.013, which is present in the total feature set, however not among top-ranked features. This metabolite has earlier been suggested as another phenylacetamide compound, namely HPAA sulfate [47]. Interestingly, at retention time 1.7 there was another compound with identical mass and fragmentation, which was below the threshold to be included in the top-ranked metabolite list. To positively identify and distinguish these two similar compounds, a sulfate conjugate of the HPAA (2-acetamidophenol, CAS 614-80-2) was synthesized. The synthesized product eluted at 2.5 min, similarly to the metabolite included in the top-ranked list, with identical MS/MS fragmentation (Fig. 3). This metabolite was therefore conclusively identified as HPAA-sulfate. The present findings are in line with recent research showing that metabolized forms of benzoxazinoids such as phenylacetamides and aminophenol sulfate increase in response

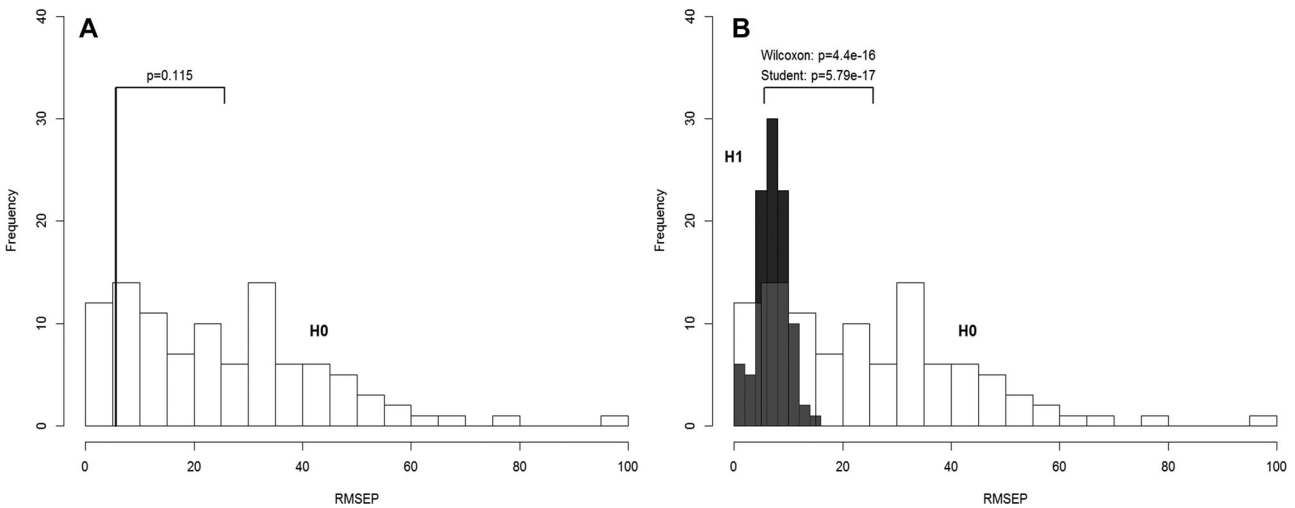

Figure 2. Tests for model performance versus random permutations. (A) Cumulative probability of average model prediction accuracy in Student's *t*-distibution of H0. (B) One-tailed Student's *t*-test and Mann–Whitney–Wilcoxon *U*-test of H1 < H0. H1 and H0 are distributions of prediction accuracies (RMSEP, $n = 100$ each) for actual models and random permutations of WG intake.

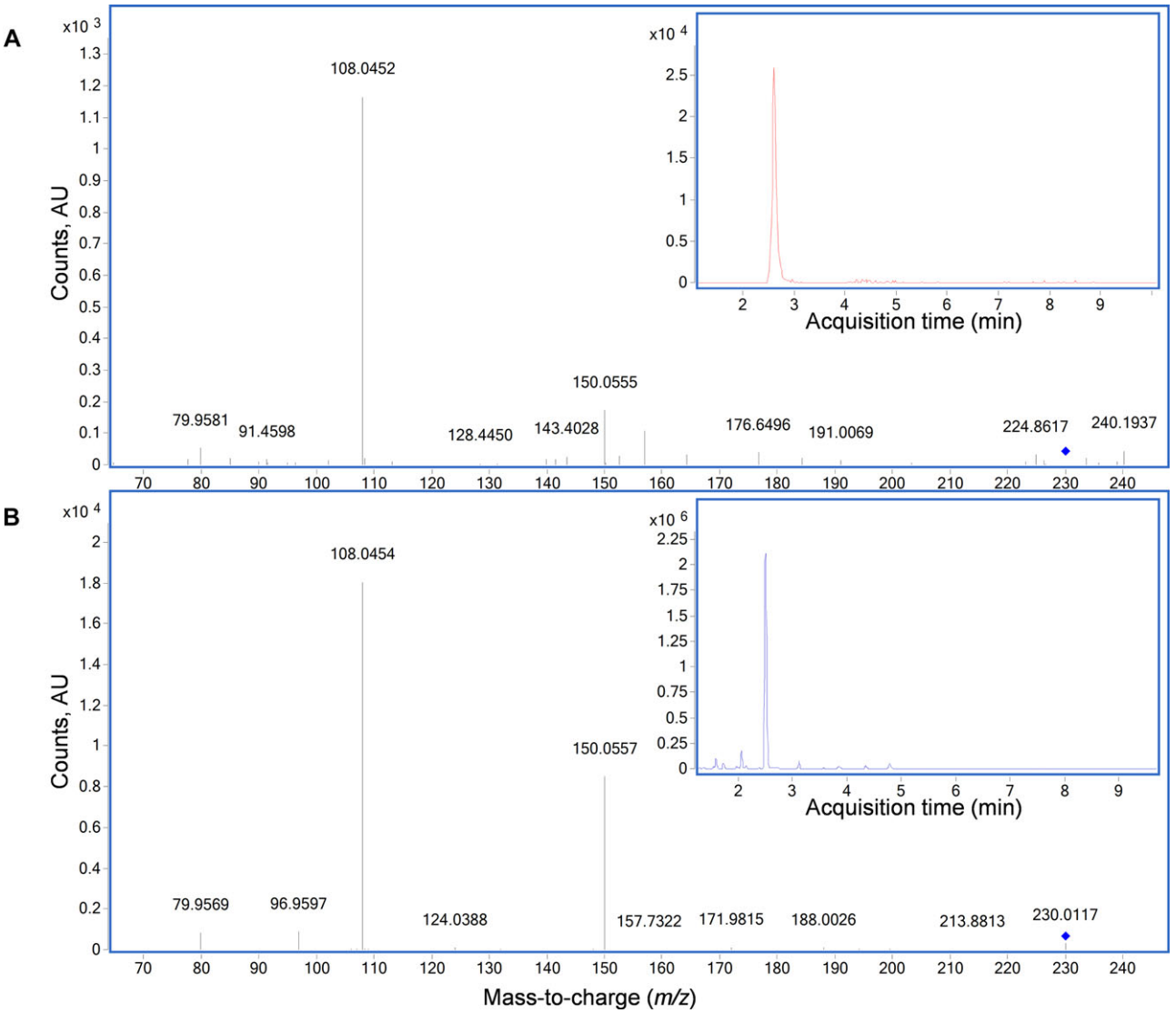

Figure 3. Identification of the HPAA sulfate. Fragmentation spectra obtained in MS/MS with negative ESI with the extracted ion chromatograms for the ion *m/z* 230.012. (A) In vitro synthetized HPAA sulfate, (B) compound in samples.

to recent intake of WG rye products [47, 48, 51]. Evaluation of their stability over time is needed before using them as biomarkers.

3.5 Other metabolites

Among the top biomarker candidates, another compound identified based on MS/MS comparison with the Metlin database was pimelic acid (MID 3280). Interestingly, this compound is a dicarboxylic acid such as azelaic acid, which was previously identified as a urinary biomarker of rye intake [51]. Additionally, a mouse feeding trial has demonstrated increased urinary levels of various dicarboxylic acids following feeding with wheat aleurone enriched diets [52].

Interestingly, three of the biomarker candidates have characteristics of carnitine structures evidenced by ESI(+) MS/MS fragments of *m/z* 85.089 and 60.080 [53], namely *m/z* 326.160, 341.173, and 475.246. However, no matches for carnitine compounds were found in standard public databases (Metlin, HMDB, SciFinder), indicating that these are novel carnitine structures not previously identified. This finding of carnitines in relation to WG rye intake is particularly interesting, as carnitines are linked to metabolism of betaine [53], which is abundant in WG rye. Furthermore, we recently found that addition of rye bran to mouse feed caused increase of various betaine derived compounds, including novel betainized amino acids [54]. These results indicate betaine and carnitine compounds as interesting biomarker candidates in relation to rye, not only as potential biomarkers of intake but

in relation to the metabolic impact of WG rye consumption as well.

3.6 Reproducibility of putative biomarkers for WG rye intake

After identification of the most prominent candidate metabolites indicating consumption of wholegrain rye, the ICC for these metabolites was calculated based on the data collected after the two periods (Table 1). Among the 20 candidate compounds, three showed good reliability (ICC value >0.5) which indicates that these biomarkers do not fluctuate to a great extent and hence better reflect the "average" biomarker concentration over a long time period. Notably, only one of these compounds was tentatively assigned, namely the 3,5-DHPPA derivative eluting at 2.7 min. This biomarker showed good reliability, with an ICC 0.50 similar to what has been shown before for DHPPA aglycone after deconjugation in the same study [55]. Among the other phenolic compounds, 3,5-DHPPA sulfate and caffeic acid sulfate had ICC 0.29 and 0.39, respectively, which indicates modest reliability. All of the compounds tentatively identified as metabolites derived from the benzoxazinoids had relatively low ICCs, which limits their use as biomarkers of habitual WG rye intake.

Interestingly, the highest ICC (0.73) was observed for a compound with inverse correlation to WG rye intake (−0.35). This challenges the assumptions that exposure biomarkers ought to be derived directly from ingested food components or their metabolic degradation products. Moreover, another of the compounds with the highest ICC (0.71) also correlated well with WG rye intake ($r = 0.55$). Both of these compounds belong to the mentioned carnitine class and may be specific and stable biomarkers of WG rye intake. Efforts to identify these compounds are highly warranted. It should be noted that the reproducibility of WG rye intake estimated by 3DWFR in the present study was modest (ICC = 0.43) and it would be unlikely to find many features strongly correlating to WG rye intake that show much higher ICC than that [55].

We recognize that our study have several limitations. For example, the sample size is relatively small and study participants are likely to be more health cautiousness compared to a general Swedish population. Moreover, we used 24 h urine sample collections, which are probably more accurate than spot urine samples, but such samples have seldom been collected in large scale prospective cohort studies. Future studies should evaluate the usefulness of spot urine samples. Our study also has several strengths. For example, dietary intake data was of high quality and participants reported a wide intake range of different WGs, which makes the dataset suitable for dietary biomarker discovery studies. Moreover, the established work flow identified an alkylresorcinol metabolite as one of the biomarkers of WG rye intake, which is a confirmatory finding in addition to several novel biomarker candidates, suggesting that the nontargeted metabolite profiling approach is very useful for biomarker discovery and should be evaluated for other food groups in future studies.

4 Concluding remarks

Our analysis confirmed earlier results on the specificity of alkylresorcinol metabolites as biomarkers for WG rye intake, but also showed several other identified and unidentified compounds that reflected WG rye intake in free-living individuals. Top-ranking metabolites from multivariate models correlated well with reported WG rye intake, but not with the intake of WG oats or WG wheat, suggesting that they are highly specific for WG rye intake. However, only a few of the putative biomarkers showed low variation over time in 24 h urine collections (ICC >0.5), suggesting limited use for assessment of long-term habitual WG rye intake. Those candidates with lowest variation over time (ICC >0.7) remain unidentified but are believed to belong to the carnitine metabolite class. Future efforts are needed to identify these compounds and further assess their long-term stability.

R.L., A.A., and M.M. designed the experiments and collected the samples, K.H. was responsible for the metabolomics study set-up, data preprocessing, and identification of the metabolites. C.B. was responsible for the statistical analysis of the data, S.A. and R.J. performed the synthesis of the sulfate conjugate and their LC-MS analysis, P.K.R. and S.A. were responsible for the LC-MS instrumentation. All the authors participated in the interpretation of the data. K.H., C.B., P.K.R., S.A., and R.L. wrote the manuscript and all the authors contributed to and approved the final manuscript.

The following funding sources are acknowledged for support: Nordforsk Centre of Excellence Programme on Food, Nutrition and Health: HELGA–Nordic Health Whole Grain Food, the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS), The Swedish Research Council- Medicine & Quality Grant from the Swedish University of Agricultural Science, Academy of Finland, Biocenter Finland.

The authors have declared no conflict of interest.

5 References

[1] Scalbert, A., Brennan, L., Manach, C., Andres-Lacueva, C. et al., The food metabolome: a window over dietary exposure. *Am. J. Clin. Nutr.* 2014, *99*, 1286–1306.

[2] Knudsen, M. D., Kyro, C., Olsen, A., Dragsted, L. O. et al., Self-reported whole-grain intake and plasma alkylresorcinol concentrations in combination in relation to the incidence of colorectal cancer. *Am. J. Epidemiol.* 2014, *179*, 1188–1196.

[3] Jenab, M., Slimani, N., Bictash, M., Ferrari, P. et al., Biomarkers in nutritional epidemiology: applications, needs and new horizons. *Hum. Genet.* 2009, *125*, 507–525.

[4] Marklund, M., Magnusdottir, O. K., Rosqvist, F., Cloetens, L. et al., A dietary biomarker approach captures compliance and cardiometabolic effects of a healthy nordic diet in individuals with metabolic syndrome. *J. Nutr.* 2014, *144*, 1642–1649.

[5] Landberg, R., Kamal-Eldin, A., Andersson, A., Vessby, B. et al., Alkylresorcinols as biomarkers of whole-grain wheat and rye intake: plasma concentration and intake estimated from dietary records. *Am. J. Clin. Nutr.* 2008, *87*, 832–838.

[6] Burrows, T. L., Warren, J. M., Colyvas, K., Garg, M. L. et al., Validation of overweight children's fruit and vegetable intake using plasma carotenoids. *Obesity* 2009, *17*, 162–168.

[7] Tasevska, N., Midthune, D., Tinker, L. F., Potischman, N. et al., Use of a urinary sugars biomarker to assess measurement error in self-reported sugars intake in the Nutrition and Physical Activity Assessment Study (NPAAS). *Cancer Epidemiol. Biomarkers Prev.* 2014, *23*, 2874–2883.

[8] Tasevska, N., Runswick, S. A., McTaggart, A., Bingham, S. A., Urinary sucrose and fructose as biomarkers for sugar consumption. *Cancer Epidemiol. Biomarkers Prev.* 2005, *14*, 1287–1294.

[9] Sun, Q., Ma, J., Campos, H., Hu, F. B., Plasma and erythrocyte biomarkers of dairy fat intake and risk of ischemic heart disease. *Am. J. Clin. Nutr.* 2007, *86*, 929–937.

[10] Wolk, A., Vessby, B., Ljung, H., Barrefors, P., Evaluation of a biological marker of dairy fat intake. *Am. J. Clin. Nutr.* 1998, *68*, 291–295.

[11] Guertin, K. A., Moore, S. C., Sampson, J. N., Huang, W. Y. et al., Metabolomics in nutritional epidemiology: identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. *Am. J. Clin. Nutr.* 2014, *100*, 208–217.

[12] Floegel, A., vonRuesten, A., Drogan, D., Schulze, M. B. et al., Variation of serum metabolites related to habitual diet: a targeted metabolomic approach in EPIC-Potsdam. *Eur. J. Clin. Nutr.* 2013, *67*, 1100–1108.

[13] Lloyd, A. J., Beckmann, M., Haldar, S., Seal, C. et al., Data-driven strategy for the discovery of potential urinary biomarkers of habitual dietary exposure. *Am. J. Clin. Nutr.* 2013, *97*, 377–389.

[14] Hanhineva, K., Lankinen, M. A., Pedret, A., Schwab, U. et al., Nontargeted metabolite profiling discriminates diet-specific biomarkers for consumption of whole grains, Fatty fish, and bilberries in a randomized controlled trial. *J. Nutr.* 2015, *145*, 7–17.

[15] Beckmann, M., Lloyd, A. J., Haldar, S., Fave, G. et al., Dietary exposure biomarker-lead discovery based on metabolomics analysis of urine samples. *Proc. Nutr. Soc.* 2013, *72*, 352–361.

[16] Andersen, M. B., Kristensen, M., Manach, C., Pujos-Guillot, E. et al., Discovery and validation of urinary exposure markers for different plant foods by untargeted metabolomics. *Anal. Bioanal. Chem.* 2014, *406*, 1829–1844.

[17] Andersen, M. B., Rinnan, A., Manach, C., Poulsen, S. K. et al., Untargeted metabolomics as a screening tool for estimating compliance to a dietary pattern. *J. Proteome Res.* 2014, *13*, 1405–1418.

[18] Primrose, S., Draper, J., Elsom, R., Kirkpatrick, V. et al., Metabolomics and human nutrition. *Br. J. Nutr.* 2011, *105*, 1277–1283.

[19] Aune, D., Norat, T., Romundstad, P., Vatten, L. J., Whole grain and refined grain consumption and the risk of type 2 diabetes: a systematic review and dose-response meta-analysis of cohort studies. *Eur. J. Epidemiol.* 2013, *28*, 845–858.

[20] Aune, D., Chan, D. S., Lau, R., Vieira, R. et al., Dietary fibre, whole grains, and risk of colorectal cancer: systematic review and dose-response meta-analysis of prospective studies. *BMJ* 2011, *343*, d6617.

[21] Cho, S. S., Qi, L., Fahey, G. C., Jr., Klurfeld, D. M., Consumption of cereal fiber, mixtures of whole grains and bran, and whole grains and risk reduction in type 2 diabetes, obesity, and cardiovascular disease. *Am. J. Clin. Nutr.* 2013, *98*, 594–619.

[22] Andersson, A., Tengblad, S., Karlstrom, B., Kamal-Eldin, A. et al., Whole-grain foods do not affect insulin sensitivity or markers of lipid peroxidation and inflammation in healthy, moderately overweight subjects. *J. Nutr.* 2007, *137*, 1401–1407.

[23] Pereira, M. A., Jacobs, D. R., Jr., Pins, J. J., Raatz, S. K. et al., Effect of whole grains on insulin sensitivity in overweight hyperinsulinemic adults. *Am. J. Clin. Nutr.* 2002, *75*, 848–855.

[24] Landberg, R., Andersson, S. O., Zhang, J. X., Johansson, J. E. et al., Rye whole grain and bran intake compared with refined wheat decreases urinary C-peptide, plasma insulin, and prostate specific antigen in men with prostate cancer. *J. Nutr.* 2010, *140*, 2180–2186.

[25] Juntunen, K. S., Laaksonen, D. E., Poutanen, K. S., Niskanen, L. K. et al., High-fiber rye bread and insulin secretion and sensitivity in healthy postmenopausal women. *Am. J. Clin. Nutr.* 2003, *77*, 385–391.

[26] Tighe, P., Duthie, G., Vaughan, N., Brittenden, J. et al., Effect of increased consumption of whole-grain foods on blood pressure and other cardiovascular risk markers in healthy middle-aged persons: a randomized controlled trial. *Am. J. Clin. Nutr.* 2010, *92*, 733–740.

[27] Brownlee, I. A., Moore, C., Chatfield, M., Richardson, D. P. et al., Markers of cardiovascular risk are not changed by increased whole-grain intake: the WHOLEheart study, a randomised, controlled dietary intervention. *Br. J. Nutr.* 2010, *104*, 125–134.

[28] Andersson, A. A. M., Dimberg, L., Åman, P., Landberg, R., Recent findings on certain bioactive components in whole grain wheat and rye. *J. Cereal Sci.* 2014, *59*, 294–311.

[29] Olsen, A., Egeberg, R., Halkjaer, J., Christensen, J. et al., Healthy aspects of the Nordic diet are related to lower total mortality. *J. Nutr.* 2011, *141*, 639–644.

[30] Andersson, A., Marklund, M., Diana, M., Landberg, R., Plasma alkylresorcinol concentrations correlate with whole grain wheat and rye intake and show moderate reproducibility over a 2- to 3-month period in free-living Swedish adults. *J. Nutr.* 2011, *141*, 1712–1718.

[31] American Association of Cereal Chemists International, *Whole Grain Definition - Cereal Foods World*, *1999*, 45–79.

[32] R Core Team, *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*, R Core Team, Vienna, Austria 2014. http://www.R-project.org/.

[33] Westerhuis, J., Hoefsloot, H. J., Smit, S., Vis, D. et al., Assessment of PLSDA cross validation. *Metabolomics* 2008, *4*, 81–89.

[34] Filzmoser, P., Liebmann, B., Varmuza, K., Repeated double cross validation. *J. Chemom.* 2009, *23*, 160–171.

[35] Chun, H., Keles, S., Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *J. R. Stat. Soc. Series B. Stat. Methodol.* 2010, *72*, 3–25.

[36] LeCao, K. A., Rossouw, D., Robert-Granie, C., Besse, P., A sparse PLS for variable selection when integrating omics data. *Stat. Appl. Genet. Mol. Biol.* 2008, *7*, 35.

[37] Hertzmark, E. S. D., SAS program and documentation. Intraclass correlation coefficients and their 95 percent confidence intervals. cited January 25, 2015. Available from: https://www.hsph.harvard.edu/donna-spiegelman/software/icc9/.

[38] Hankinson, S. E., Manson, J. E., Spiegelman, D., Willett, W. C. et al., Reproducibility of plasma hormone levels in postmenopausal women over a 2–3-year period. *Cancer Epidemiol. Biomarkers Prev.* 1995, *4*, 649–654.

[39] Lang, M. A., Gielen, J. E., Nebert, D. W., Genetic evidence for many unique liver microsomal P-450-mediated monooxygenase activities in heterogeneic stock mice. *J. Biol. Chem.* 1981, *256*, 12068–12075.

[40] Hastie, T., Tibshirani, R., Friedman, J., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, New York 2009.

[41] Christin, C., Hoefsloot, H. C., Smilde, A. K., Hoekman, B. et al., A critical assessment of feature selection methods for biomarker discovery in clinical proteomics. *Mol. Cell. Proteomics* 2013, *12*, 263–276.

[42] Pesarin, F., Salmaso, L., A review and some new results on permutation testing for multivariate problems. *Stat. Comput* 2012, *22*, 639–646.

[43] Landberg, R., Kamal-Eldin, A., Salmenkallio-Marttila, M., Rouau, X. et al., Localization of alkylresorcinols in wheat, rye and barley kernels. *J. Cereal Sci.* 2008, *48*, 401–406.

[44] Ross, A. B., Aman, P., Kamal-Eldin, A., Identification of cereal alkylresorcinol metabolites in human urine-potential biomarkers of wholegrain wheat and rye intake. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 2004, *809*, 125–130.

[45] Marklund, M., McKeown, N. M., Blumberg, J. B., Chen, C.-O., Hepatic biotransformation of alkylresorcinols is mediated via cytochrome P450 and β-oxidation: a proof of concept study. *Food Chem.* 2013, *139*, 925–930.

[46] Marklund, M., Stromberg, E. A., Hooker, A. C., Hammarlund-Udenaes, M. et al., Chain length of dietary alkylresorcinols affects their in vivo elimination kinetics in rats. *J. Nutr.* 2013, *143*, 1573–1578.

[47] Beckmann, M., Lloyd, A. J., Haldar, S., Seal, C. et al., Hydroxylated phenylacetamides derived from bioactive benzoxazinoids are bioavailable in humans after habitual consumption of whole grain sourdough rye bread. *Mol. Nutr. Food Res.* 2013, *57*, 1859–1873.

[48] Hanhineva, K., Keski-Rahkonen, P., Lappi, J., Katina, K. et al., The postprandial Plasma Rye Fingerprint Includes Benzoxazioid-Derived Phenylacetamide Sulfates. *J. Nutr.* 2014, *144*, 1–7.

[49] Zikmundova, M., Drandarov, K., Bigler, L., Hesse, M. et al., Biotransformation of 2-benzoxazolinone and 2-hydroxy-1,4-benzoxazin-3-one by endophytic fungi isolated from *Aphelandra tetragona*. *Appl. Environ. Microbiol.* 2002, *68*, 4863–4870.

[50] Zikmundova, M., Drandarov, K., Hesse, M., Werner, C., Hydroxylated 2-amino-3H-phenoxazin-3-one derivatives as products of 2-hydroxy-1,4-benzoxazin-3-one (HBOA) biotransformation by Chaetosphaeria sp., an endophytic fungus from Aphelandra tetragona. *Z. Naturforsch. C* 2002, *57*, 660–665.

[51] Bondia-Pons, I., Barri, T., Hanhineva, K., Juntunen, K. et al., UPLC-QTOF/MS metabolic profiling unveils urinary changes in humans after a whole grain rye versus refined wheat bread intervention. *Mol. Nutr. Food Res.* 2013, *57*, 412–422.

[52] Pekkinen, J., Rosa, N. N., Savolainen, O. I., Keski-Rahkonen, P. et al., Disintegration of wheat aleurone structure has an impact on the bioavailability of phenolic compounds and other phytochemicals as evidenced by altered urinary metabolite profile of diet-induced obese mice. *Nutr. Metab.* 2014, *11*, 1–7075-11–1.

[53] Pekkinen, J., Olli, K., Huotari, A., Tiihonen, K. et al., Betaine supplementation causes increase in carnitine metabolites in the muscle and liver of mice fed a high-fat diet as studied by nontargeted LC-MS metabolomics approach. *Mol. Nutr. Food Res.* 2013, *57*, 1959–1968.

[54] Pekkinen, J., Rosa-Sibakov, N., Micard, V., Keski-Rahkonen, P. et al., Amino acid-derived betaines dominate as urinary markers for rye bran intake in mice fed high-fat diet?A nontargeted metabolomics study. *Mol. Nutr. Food Res.* 2015, *59*, 1550–1562.

[55] Marklund, M., Landberg, R., Andersson, A., Aman, P. et al., Alkylresorcinol metabolites in urine correlate with the intake of whole grains and cereal fibre in free-living Swedish adults. *Br. J. Nutr.* 2013, *109*, 129–136.

