The Journal of Nutrition
Genomics, Proteomics, and Metabolomics

Prediction of Skeletal Muscle and Fat Mass in Patients with Advanced Cancer Using a Metabolomic Approach[1–3]

Cynthia Stretch,[4] Thomas Eastman,[5] Rupasri Mandal,[6] Roman Eisner,[5] David S. Wishart[5,6] Marina Mourtzakis,[10] Carla M. M. Prado,[7] Sambasivarao Damaraju,[8,9] Ronald O. Ball,[7] Russell Greiner,[5] and Vickie E. Baracos[4,7]*

[4]Department of Oncology, Cross Cancer Institute, [5]Department of Computing Science, [6]Department of Biological Sciences, [7]Alberta Institute for Human Nutrition, [8]PolyomX Program, Cross Cancer Institute, and [9]Department of Laboratory Medicine and Pathology, University of Alberta, Edmonton, AB, Canada; and [10]Department of Kinesiology, University of Waterloo, Waterloo, Ontario, Canada

Abstract

Urine and plasma metabolites originate from endogenous metabolic pathways in different organs and exogenous sources (diet). Urine and plasma were obtained from advanced cancer patients and investigated to determine if variations in lean and fat mass, dietary intake, and energy metabolism relate to variation in metabolite profiles. Patients (n = 55) recorded their diets for 3 d and after an overnight fast they were evaluated by DXA and indirect calorimetry. Metabolites were measured by NMR and direct injection MS. Three algorithms were used [partial least squares discriminant-analysis, support vector machines (SVM), and least absolute shrinkage and selection operator] to relate patients' plasma/urine metabolic profile with their dietary/physiological assessments. Leave-one-out cross-validation and permutation testing were conducted to determine statistical validity. None of the algorithms, using 63 urine metabolites, could learn to predict variations in individual's resting energy expenditure, respiratory quotient, or their intake of total energy, fat, sugar, or carbohydrate. Urine metabolites predicted appendicular lean tissue (skeletal muscle) with excellent cross-validation accuracy (98% using SVM). Total lean tissue correlated highly with appendicular muscle (Pearson r = 0.98; $P < 0.0001$) and gave similar cross-validation accuracies. Fat mass was effectively predicted using the 63 urine metabolites or the 143 plasma metabolites, exclusively. In conclusion, in this population, lean and fat mass variation could be effectively predicted using urinary metabolites, suggesting a potential role for metabolomics in body composition research. Furthermore, variation in lean and fat mass potentially confounds metabolomic studies attempting to characterize diet or disease conditions. Future studies should account or correct for such variation. J. Nutr. 142: 14–21, 2012.

Introduction

Recent progress in high-throughput analytical technologies and bioinformatics now permits simultaneous analysis of hundreds of compounds constituting the metabolome. Metabolomic analyses give complex fingerprints that appear to be characteristic of a given metabolic phenotype or diet. Although many (1–3) have suggested that metabolomic analysis has the potential to change how nutrition research is conducted, much of this potential remains unrealized (4). A surprisingly small number of metabolomic studies have been conducted in human nutrition to date and progress is hampered by a number of unsolved problems, most notably by the lack of well-established, standardized methods for collecting, measuring, analyzing, and reporting data (1,5).

One important prerequisite for effective use of metabolomic approaches is to understand how variability in endogenous (e.g., tissue metabolism) and exogenous (e.g., diet) metabolite sources affects metabolome profiles. A conceptual framework for these contributions includes multiple elements (**Fig. 1**). Food intake may be the largest contribution to diurnal variation in metabolites. Diet is also a source of elements characteristic of specific foods: phytochemicals [e.g., coffee (6), tea (7), cocoa (8), and almonds (9)] or amines [e.g., fish (10)]. Metabolomic assessment in the postabsorptive state would generally limit the immediate influence of meals on substrate flux (11). Individuals also may have widely divergent body proportions of organs (12), fat, and skeletal muscle (13). The condition of sarcopenia (severe muscle depletion) in an individual with otherwise normal body weight (Fig. 1) would

[1] Supported by grants from the Alberta Cancer Board and Alberta Cancer Foundation to V.E.B., D.S.W., R.G., and S.D. R.G., T.E., and R.E. were also funded by Alberta Ingenuity Centre for Machine Learning and Natural Sciences and Engineering Research Council of Canada.

[2] Author disclosures: C. Stretch, T. Eastman, R. Mandal, R. Eisner, D. S. Wishart, M. Mourtzakis, C. M. M. Prado, S. Damaraju, R. O. Ball, R. Greiner, and V. E. Baracos, no conflicts of interest.

[3] Supplemental Tables 1–7 are available from the "Online Supporting Material" link in the online posting of the article and from the same link in the online table of contents at jn.nutrition.org.

* To whom correspondence should be addressed. E-mail: vickie.baracos@ualberta.ca

Manuscript received July 12, 2011. Initial review completed August 17, 2011. Revision accepted October 6, 2011.
First published online December 7, 2011; doi:10.3945/jn.111.147751.

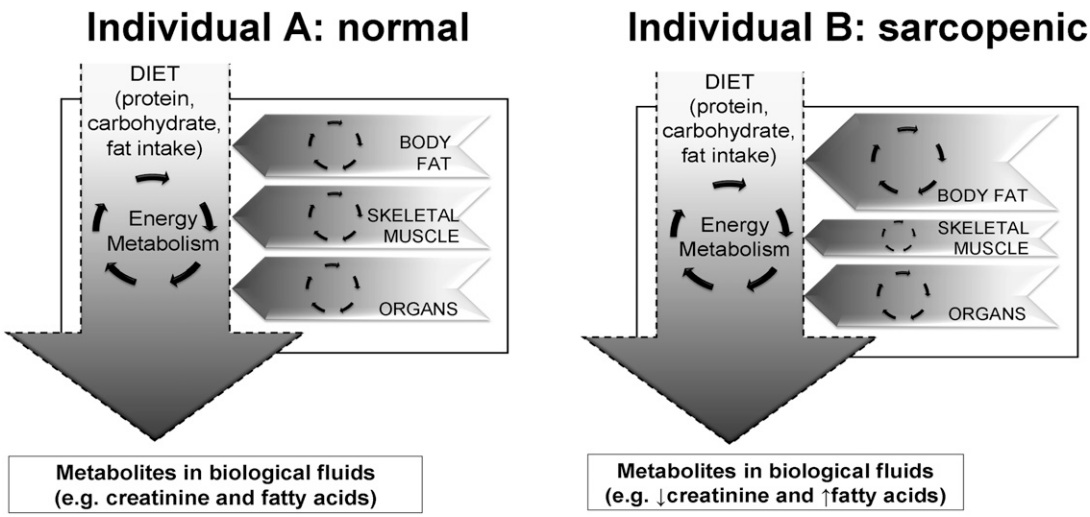

FIGURE 1 A conceptual framework. Contributions of multiple elements (diet, lean and fat mass, energy metabolism) to the metabolome of different biofluids. Considering two individuals of equal body weight and equal macronutrient intake but with different muscle and fat mass [normal (*A*), sarcopenic, i.e., suffering from severe muscle depletion (*B*)], it would be expected that metabolic fluxes in *B* would have a proportionately higher amount of fat-derived metabolites and a lower proportion of muscle-derived metabolites.

result in a disproportionately low contribution of muscle-derived metabolites to the metabolome overall. Beside the well-known relation of urinary creatinine excretion to skeletal muscle mass (14), the amount of lean and fat mass remains an unexplored source of variation in human metabolite profiling studies.

Based on the foregoing, the following hypothesis was explored: in the postabsorptive state, the metabolome is defined in part by the varying mass of tissues (e.g., adipose tissue, skeletal muscle) as these produce tissue-specific metabolites in the course of their turnover/metabolism. For example, adipose tissue is the origin of fatty acids and hence of ketones and creatinine and 3-methylhistidine originate in skeletal muscle.

A population of patients with advanced (IV) cancer was selected for study. Detection of nutritional and metabolic alterations that accompany the progression of cancer is a crucial part of patient care. Patients with advanced cancers are known to have wide variations in lean and fat mass, dietary intake, metabolic rate, and fuel metabolism due to the disease (12,13,15). Metabolomics is a particularly attractive technology to detect such variations in vulnerable patient populations, because it is noninvasive. This is particularly important for detecting variations in lean and fat mass, because these are not routinely measured in the clinic despite the negative outcomes associated with lean tissue wasting (13). Unfortunately, clinically pragmatic methods such as anthropometry, underwater weighing, and bioelectrical impedance analysis may be insensitive (16), associated with high participant burden, or grossly underestimate or overestimate lean mass compared to DXA (17), respectively. To begin to understand the potential utility of metabolomic profiling, plasma and urine metabolites were quantified using two fast and inexpensive profiling platforms: proton NMR and MS. Metabolite patterns were assessed in the postabsorptive state using multivariate statistics and machine learning approaches to detect metabolite signatures of these features.

Methods

Study design. Approval was provided by the Research Ethics Board of the Alberta Cancer Board. Eligible participants were recruited between January 2005 and October 2006 and included men and women with nonsmall cell lung or colorectal cancer who were >18 y of age, able to communicate freely in English, and able to provide informed consent; variation owing to cancer stage was reduced by inclusion only of patients with stage IV disease according to the international classification system TNM. At this advanced disease stage, both lung and colorectal cancer patients have similar characteristics (12,13). Study participants were receiving therapy appropriate to their disease and stage, including chemotherapy, treatments for the side effects of chemotherapy, as well as medications for comorbid conditions. Patients with creatinine clearance <60 mL/min (n = 11), radiation to the kidneys (n = 2), or bladder metastasis resulting in blood in the urine (n = 3) were excluded, because these independently affect urinary excretion, making the total number of included patients 55. Although there is no explicit formula for sample size calculation for metabolomic studies, previous work from our group (18) and others (9,19–21) discriminated dietary or metabolic types with samples of this size.

Assessments. Patients collected diet records under the supervision of a dietician for 3 d, including one weekend day (22). To reduce the confounding acute effect of meals, participants were studied after a 12-h fast spanning the night of the third day to the following morning. Although this fasting could be expected to reduce the influence of prior intake, because protein intake during the day is related to the rate of amino acid oxidation during the night (23), a potential influence of prior protein intake may be expected.

Participants attended the Human Nutrition Research Unit for sampling and metabolic evaluation. Although full details regarding the assessments on these patients have been published (24), a brief summary is provided below.

Height and weight were measured with participants barefoot and in a hospital gown. DXA used a LUNAR Prodigy High Speed Digital Fan Beam X-Ray-Based Densitometer (General Electric) with enCORE 9.20 software for analysis of TFM[10] and LST. ALST was calculated by summing the LST from the limbs (arms and legs) (25) and is a measure of appendicular skeletal muscle (26).

Nutrient intakes were estimated using the Canadian Nutrient File Database (FOOD PROCESSOR II nutrient analysis software, version 9.0; Esha Research). Total energy intake and macronutrient intakes (total protein, fat, carbohydrate, and sugar) were calculated. Because the body weight (range 48–142 kg) and composition (range of percentage of body fat = 13.8–56.2%) was variable, both the absolute or per-kilogram

[10] Abbreviations used: ALST, appendicular lean soft tissue; LASSO, least absolute shrinkage and selection operator; LOOCV, leave-one-out cross validation; LST, lean soft tissue; PLS-DA, partial least squares discriminant analysis; REE, resting energy expenditure; RQ, respiratory quotient; SVM, support vector machine; TFM, total fat mass.

body weight expression of energy intake and expenditure would be difficult to interpret. These data were thus expressed per kilogram LST as assessed by DXA (27). REE and RQ were determined by indirect calorimetry (VMax 29N, SensorMedics) as detailed in Prado et al. (24).

Urine and plasma samples were collected immediately upon arrival to the research unit. Sodium azide was added to urine samples to a final concentration of ~0.02% to prevent bacterial growth. Whole blood was collected and plasma was isolated by a clinical laboratory provider (Dynacare Kasper Medical Laboratories, certified by the College of American Pathologists and College of Physicians and Surgeons of Alberta). Urine and plasma samples were stored at −80°C until ready for analysis.

NMR spectroscopy. Urine samples were prepared and analyzed according to a recently published procedure (18). Plasma was prepared by removing high-molecular weight compounds by ultrafiltration using Nanosep 3kDa microcentrifuge filter tubes. Prior to filtration, microcentrifuge filter tubes were washed using distilled deionized water to remove glycerol used as a preservative in the filters. Ultrafiltrate volumes ranged from 250 to 400 μL. Ultrafiltrates were then brought to volume (585 μL) using distilled deionized water. As with the urine samples, plasma was combined with 65 μL of internal standard (Chenomx, consisting of ~5 mmol/L sodium 2,2-dimethyl-2-silapentane-5-sulfonate d6, 0.2% sodium azide in 99% D_2O) and pH corrected to 6.75 ± 0.05 using small amounts of NaOH or HCl. A 600-μL aliquot of prepared sample was placed in a 5-mm NMR tube for NMR spectral acquisition.

One-dimensional NMR spectra were acquired using the standard NOESY pulse sequence on a 4-channel Varian Inova-600 MHz NMR spectrometer with a triax-gradient 5-mm HCN probe. Quantification of metabolites by targeted profiling was performed using Chenomx NMRSuite 4.6 (Chenomx).

Two analysts independently used Chenomx software to identify metabolite concentrations; only consensus assignments agreed upon by both analysts were used in statistical analysis. Laboratory analyses were also conducted to verify creatinine concentrations and amino acid peak assignments. A more complete description of these additional laboratory analyses has been published (18).

MS. Direct flow injection MS using an AbsoluteIDQ kit (BIOCRATES Life Sciences) was used for the analysis of plasma and urine samples. This kit assay in combination with a 4000 QTrap (Applied Biosystems/MDS Sciex) mass spectrometer permits the identification and quantification of up to 160 metabolites in urine and plasma. Samples were prepared according to manufacturer's instructions. A standard flow injection protocol consisting of two 20-μL injections (one for the positive and one for the negative ion detection mode) was applied for all measurements. Multiple reaction monitoring detection was used for quantification. MetIQ software (BIOCRATES Life Sciences AG, Austria) was used to control assay workflow, including sample registration, and calculation of metabolite concentrations.

Metabolite data preprocessing. Concentration values for the metabolites can range over several orders of magnitude both within and between patients. This is addressed by using the natural log of the concentration values. The issue of different urine dilutions may be addressed by methods such as normalization by creatinine concentration (28), by total peak area (29), and by probability quotient (30). These methods reduced predictive accuracy of various algorithms compared to no data normalization (18). Therefore, our main analysis used only log transformation as the only preprocessing step. In addition, normalization of urine metabolite concentrations to whole body LST mass was also conducted.

Statistical methods. The objective of the statistical analysis is to relate the patient's plasma/urine metabolites with the patient's dietary/physiological assessments (i.e., class). Because there is no single accepted method for statistical treatment of quantitative metabolomic data, several methods previously used were compared. Data were analyzed by PLS-DA (31), SVM (32), and LASSO (33). Each of these algorithms uses a labeled data set (i.e., data describing a set of patients, along with the class label for each patient, e.g., patient has high energy intake vs. low energy intake) to produce a classifier that can predict the class label of a new patient. Here, classes were defined as the distal ranges of values (highest and lowest) for each assessment where an instance (patient) was labeled high if his/her measurement on this assessment was at least 0.5 SD higher than the median and labeled low if the assessment was at least 0.5 SD lower than the median. For example, only individuals with high and low energy intake were included, leaving out an intermediate group representing a band approximately the width of the measurement error of this variable.

The effectiveness of each learning algorithm was assessed, i.e., how accurately the classifier classifies novel patients, by LOOCV and permutation testing. The baseline accuracy rate was compared with the LOOCV accuracy results obtained by PLS-DA, SVM, and LASSO. The baseline accuracy rate is the frequency of the most common class, expressed as a percentage (i.e., if 60 patients are in class A and 40 patients in class B, then the baseline accuracy would be 60%). Note that metabolomic information is not used in calculating the baseline accuracy. Thus, if the metabolic profile data contains any signal with respect to a particular classification task, then it would increase the classification accuracy above the baseline accuracy rate, the maximum of which is 100%. Predictors were subjected to permutation testing (1000 permutations) to determine whether the predictive cross-validation accuracies of these classifiers were significant. An acceptable model was that very few (<50 of 1000, i.e., $P < 0.05$) permuted models outperform the original model.

PLS-DA is commonly used to build predictors using eigenvalues (31,34). LOOCV analysis by PLS-DA was conducted using SIMCA P11.0 (Umetrics). SVM (32) views each instance as a vector in multi-dimensional space and seeks the maximally separating hyperplane between the classes in this space. SVM analysis (with a linear kernel) using LOOCV was conducted using the WEKA machine learning package (35). LASSO is a linear classifier based on a form of regularized regression, which incorporates a penalty into the least squares objective function when learning a set of regression coefficients. LASSO implicitly performs variable selection, because it sets some of the regression coefficients to zero; hence, the associated variables (here, metabolite concentrations) will not contribute to the model. This technique was implemented using R (36) and used the glmnet package to perform LASSO regression using LOOCV (37).

Many researchers interested in nutrition and metabolism may ask, "Which metabolites best discriminate the classes?" This was addressed by using mutual information to quantify the dependence between each metabolite and the class outcome [the two groups for each assessment, e.g., high muscle mass vs. low muscle mass; see (18)]. Mutual information analysis yields unitless values, where larger values indicate a higher degree of dependence.

Pearson correlation analysis and Mann-Whitney nonparametric tests were conducted using SPSS (17.0).

Results

Distribution of measured dietary/physiological assessments in advanced cancer patients. Of the 55 cancer patients included in this study, 58% were male, 45% had lung cancer, and the overall median age was 61 ± 11 y. The median and variation for each measured assessment as well as the characteristics of the classes are shown (**Table 1**).

Urine: metabolites identified and quantified. Using NMR 71 metabolites were quantified. However, 8 metabolites were excluded, because they were drug metabolites or constituents of a vehicle for drug administration (ibuprofen, acetaminophen, salicylurate, propionate, propylene glycol, and mannitol), belonged to microbial metabolism (methanol), or had unreliable quantification (urea) (18). Urea has unreliable quantification, because suppression of the NMR signal by presaturation may lead to resonant suppression of the urea peak due to proton

TABLE 1 Variation in lean and fat mass, energy intake, and energy metabolism for all cancer patients included in the data analysis and by low and high class for each variable[1]

	All patients	Low class[2]		High class	
			n		*n*
Lean and fat mass, *kg*					
TFM	28.0 ± 9.5 (11.4–62.5)	17 ± 3.5 (11.4–23.1)	14	37.5 ± 7.6 (33.0–62.5)	14
LST	46.4 ± 10.6 (29.3–65.8)	35.7 ± 3.2 (29.3–40.5)	21	57.8 ± 4.6 (51.7–65.8)	18
ALST	19.3 ± 4.8 (12.4–28.5)	14.7 ± 1.2 (12.4–16.7)	20	24.8 ± 2.3 (21.9–28.5)	20
Dietary intake					
Total energy intake, $kcal \cdot kg\ LST^{-1}$	46.5 ± 12.8 (27.4–85.1)	32.5 ± 3.8 (27.4–40.1)	11	62.6 ± 10.7 (53.5–85.1)	14
CHO intake, $g \cdot kg\ LST^{-1}$	6.0 ± 2.0 (3.0–11.9)	4.3 ± 0.6 (3.0–4.9)	14	7.8 ± 1.4 (7.0–11.9)	20
Sugar intake, $g \cdot kg\ LST^{-1}$	1.9 ± 0.9 (0.5–4.6)	1.1 ± 0.2 (0.5–1.4)	17	2.9 ± 0.6 (2.3–4.6)	19
Fat intake, $g \cdot kg\ LST^{-1}$	1.7 ± 0.7 (0.2–4.0)	1.0 ± 0.3 (0.2–1.3)	12	2.3 ± 0.5 (2.1–4.0)	15
Protein intake, $g \cdot kg\ LST^{-1}$	1.9 ± 0.6 (1.0–3.6)	1.4 ± 0.2 (1.0–1.5)	14	2.5 ± 0.4 (2.2–3.6)	16
Energy metabolism					
REE, $kcal \cdot kg\ LST^{-1} \cdot d^{-1}$	33 ± 5 (25–46)	29 ± 4 (25–30)	12	38 ± 4 (35–46)	19
RQ	0.8 ± 0.1 (0.7–0.9)	0.7 ± 0 (0.7–0.7)	11	0.8 ± 0 (0.8–0.9)	21

[1] Values are median ± SD (range), $n = 55$ (all) or as shown for each class. ALST, appendicular lean soft tissue; LST, lean soft tissue; REE, resting energy expenditure; RQ respiratory quotient; SVM, support vector machine; TFM, total fat mass.
[2] As detailed in "Methods," low and high classes included patients who had values 0.5 SD lower and higher than the median for each variable, respectively.

exchange with water, thereby making its quantification unreliable (38). A list of the 63 remaining metabolites can be found in **Supplemental Table 1**. NMR-measured concentrations of creatinine were confirmed using laboratory tests (intraclass correlation of 0.95 with a 95% CI of 0.91–0.97). Spike-in experiments provided positive confirmation of peak assignments for amino acids (18) (data not shown).

MS identified 117 metabolites, including 12 amino acids, 37 acyl carnitines, 55 glycerophospholipids, 12 sphingolipids, and the combined concentration of hexose sugars.

Urine: results of statistical analyses. The urine data analysis summary (**Table 2**) presents the most accurate predictive models using all three methods (i.e., SVM, LASSO, and PLS-DA). The most accurate predictors were for appendicular skeletal muscle mass: SVM (LOOCV accuracy = 98%), LASSO (LOOCV accuracy = 90%), and PLS-DA (LOOCV accuracy = 85%) compared with a baseline accuracy rate of 54%. LST (which included skeletal muscle, soft lean tissues, organs, and skin) was also accurately predicted with all three algorithms (Table 2). Similar accuracies of these models may be explained by the high correlation (Pearson $r = 0.98$; $P < 0.0001$) between total lean and appendicular lean tissues. Satisfactory predictive models were achieved for TFM using SVM (LOOCV accuracy = 79%), LASSO (LOOCV accuracy = 82%), and PLS-DA (LOOCV accuracy = 79%) compared with the baseline accuracy rate of 50%. The median concentrations and SD of urinary metabolites quantified by NMR for the two classes (high and low) of LST and fat mass are listed in **Supplemental Tables 2** and **3**, respectively. High- and low-percentage of lean and fat mass did not produce predictive models that were any better than could be obtained by chance.

The distribution of cancer types (lung, colorectal) did not differ between the high- and low-fat mass groups or between the high- and low-lean tissue mass groups. However, the low class for lean tissues was composed mainly of women (95%) and the high class was composed entirely of men (100%). By itself, sex was predicted by urinary metabolite profiles, with a base model

TABLE 2 Most accurate predictive models based on SVM, LASSO, and PLS-DA analysis and permutation testing[1]

	Analysis from urine metabolites[2]			Analysis from plasma metabolites[3]
	LST	ALST	TFM	TFM
Baseline accuracy, *%*	54	51	50	50
SVM				
LOOCV accuracy, *%*	90	98	79	71
Permuted models better than original data, *n(P)*	0 (<0.00001)	0 (<0.00001)	15 (0.015)	62 (0.062)
LASSO				
LOOCV accuracy, *%*	87	90	82	88
Permuted models better than original data, *n(P)*	1 (0.001)	1 (0.001)	10 (0.01)	1 (0.001)
PLS-DA				
LOOCV accuracy, *%*	85	85	79	79
Permuted models better than original data, *n(P)*	0 (<0.00001)	1 (0.001)	17 (0.017)	27 (0.027)

[1] ALST, appendicular lean soft tissue; LASSO, least absolute shrinkage and selection operator; LOOCV, leave-one-out cross validation; LST, lean soft tissue; PLS-DA, partial least squares discriminant analysis; SVM, support vector machine; TFM, total fat mass.
[2] Urine metabolites include only metabolites analyzed by NMR ($n = 63$ metabolites).
[3] Plasma metabolites included only metabolites analyzed by MS ($n = 143$ metabolites).

(58%), SVM, LASSO, and PLS-DA providing 91, 85, and 78% LOOCV accuracy, respectively. However, in accordance with prior publications (39,40), it seems likely that sex was discriminated mainly by the fact that men are generally larger and more muscular than women, leading to differential production of muscle-specific metabolites. This is consistent with the observation that creatine, a muscle-specific metabolite, was in the top two metabolites that contributed to the discrimination of sex in the three statistical analyses as well as the mutual information (see below).

Total energy, carbohydrate, sugar, and fat intake could not be predicted accurately from the NMR urine metabolite data; this was true when the class labels were determined based on absolute (total) intake, intake · kg body weight^{-1}, or intake · kg LST^{-1}. However, protein intake did result in a satisfactory but relatively weak model with a baseline accuracy of 53% and LOOCV accuracies using SVM (70%), LASSO (73%,) and PLS-DA (73%). Protein intake also produced several randomly permuted models that appeared more accurate than the model learned using the original data (48 times for SVM and 28 times for LASSO). To test whether variation in lean tissue was confounding chances of building predictive models based on macronutrient intakes, urine metabolite concentrations were normalized to the whole body LST mass for each individual and reran all of the classifiers. This normalization made no difference in the accuracy of the predictive models for any macronutrient (e.g., the LOOCV for SVM for protein intake was 73% with this normalization compared with 70%; see above). That LST mass and fat mass could be predicted better than dietary intake seems initially surprising taking into account the relationships between intake and body weight in healthy individuals. However, total energy intake explained only 14% of the variation in LST (Pearson r^2 = 0.14; $P < 0.01$) and was unrelated to TFM (Pearson r^2 = 0.005, nonsignificant).

The RQ classes included a group with substantial fat oxidation (RQ = 0.7) and a class with a slightly higher RQ of 0.8 reflecting more mixed oxidation. RQ did not produce a predictive model better than could be obtained by chance, as determined by permutation testing. REE classes (Table 1) were developed for each sex and then aggregated, because this was a sex-dependent variable (41). The median energy intake per kilogram body weight or per kilogram LST for patients in the low and high REE classes did not differ (47 ± 10 kcal · kg LST^{-1}; 29 ± 8 kcal · kg body weight^{-1} for both). REE also did not produce a predictive model any better than could be obtained by chance, regardless of the basis of normalization of this value (total, per kilogram body weight, per kilogram LST).

Classifiers built using urine metabolites measured by MS alone or NMR and MS pooled together resulted in no improvements in LOOCV accuracy and, in fact, for most models the accuracy decreased. Finally, we did not attempt to class patients in relation to their medications, because they were many and varied, with a total of 69 different drugs (mean 4/patient, not including chemotherapy) being used.

Urine: metabolites related to lean and fat mass. Bivariate analysis allowed for the ranking of metabolites according to their mutual information for ALST and TFM. Because LST and ALST share the same top 30 metabolites, albeit in slightly different rank order, **Table 3** shows only ALST. This mutual information analysis for ALST and LST further supports the suggestion that the discrimination of sex is nothing more than the discrimination of lean mass, because 17 of the top 20 metabolites (including creatine) in the list of mutual information for sex were identical, with the metabolites discriminating muscle and lean tissue (not shown).

Plasma: metabolites identified and quantified. MS identified 143 metabolites, including 15 amino acids, 25 acyl carnitines, 87 glycerophospholipids, 15 sphingolipids, and the combined concentration of all hexose sugars (**Supplemental Table 5**). Quantitative NMR analysis identified the 31 metabolites listed in **Supplemental Table 6**.

Plasma: results of data analyses. Plasma NMR data analysis resulted in poor predictive models (i.e., not different from the baseline accuracy rate) for lean and fat mass, percent lean and fat mass, total energy and macronutrient intake, and energy metabolism. Plasma MS data resulted in satisfactory prediction using SVM (71%), LASSO (88%), and PLS-DA (79%) of total body fat compared to the baseline of 50%. The median concentrations and SD of plasma metabolites quantified by MS for the two fat mass groups (high and low) are listed in **Supplemental Table 7**. Predictive models built using plasma metabolites measured by NMR and MS pooled together resulted in no improvements in LOOCV accuracy and, in fact, for most models the accuracy decreased.

Metabolites related to total body fat. Bivariate analysis was used to rank metabolites according to their mutual information for TFM; the top 30 are shown (Table 3). All of the metabolites included in the top 30 are lipid molecules (acylcarnitines, phosphatidylcholines, lysophosphatidylcholines, and sphingomyelins).

Discussion

The present quantitative metabolomic study supports the hypothesis that body lean and fat mass have distinctive metabolic profiles. Broad categories (high and low) of muscle mass quantity were accurately predicted from metabolite concentrations in easily obtained physiological fluids. This level of discrimination lends itself to the identification of occult sarcopenia (i.e., absolute muscle mass >2 SD less than for normal healthy adults), a clinically important condition (13). The above results also suggest the potential of metabolomic approaches to further studies of organ mass per se. These findings suggest that lean/muscle mass can be easily predicted using urinary metabolite profiles, which is advantageous over other means as it can be obtained noninvasively. People who are not candidates for other forms of imaging (too large, CT or MRI contraindications), live in remote locations, or are too frail to undergo assessment could be possible candidate populations for a metabolomics-based screening test.

The variation in lean and fat mass inherent in patients with cancer and possibly other chronic conditions may confound metabolomic studies intended to look at diet, metabolic disorders, or diseases. This variation could be eliminated by assessing patients only within predefined lean and fat mass ranges by ensuring that lean and fat mass were used as the basis for matching participants in different treatment groups, or by explicitly including these factors when building predictors for various conditions.

Sources of variation in metabolite profiles. To eliminate acute effects of dietary intake, samples were collected under standardized conditions of overnight fasting. A diet record completed over the 3 d preceding sampling and lean and fat mass

TABLE 3 Bivariate analysis: top 30 metabolites related to the best predictive models

	Urinary metabolites		Plasma metabolites[1]
Rank[2]	ALST	TFM	TFM
	Metabolite *(mutual information value)*		
1	Fumarate (0.162)	3-Indoxylsulfate (0.167)	lysoPC.a.C26:0 (1.126)
2	Creatine (0.133)	Uracil (0.149)	C8 (1.012)
3	4-Hydroxyphenylacetate (0.074)	Fumarate (0.125)	C10 (0.849)
4	Quinolinate (0.058)	O-Acetylcarnitine (0.122)	C14 (0.716)
5	2-Oxoglutarate (0.056)	Methylamine (0.121)	C5 (0.449)
6	Adipate (0.056)	Acetone (0.099)	C8:1 (0.262)
7	Sucrose (0.048)	Taurine (0.095)	C3 (0.221)
8	Betaine (0.045)	Tartrate (0.088)	C0 (0.21)
9	Trigonelline (0.042)	Glycolate (0.084)	PC.aa.C40:4 (0.188)
10	Formate (0.038)	3-Aminoisobutyrate (0.073)	SM.C24:1 (0.181)
11	Glycolate (0.037)	Hypoxanthine (0.069)	PC.aa.C38:3 (0.177)
12	Taurine (0.03)	Trigonelline (0.058)	PC.aa.C40:5 (0.148)
13	O-Acetylcarnitine (0.03)	Carnitine (0.055)	PC.aa.C38:4 (0.14)
14	1-Methylnicotinamide (0.027)	1-Methylnicotinamide (0.054)	PC.aa.C34:4 (0.133)
15	Xylose (0.026)	Tryptophan (0.052)	C16 (0.124)
16	Glucose (0.024)	Adipate (0.047)	C18:2 (0.12)
17	2-Aminobutyrate (0.022)	Dimethylamine (0.043)	PC.aa.C36:4 (0.115)
18	Guanidoacetate (0.022)	Trimethylamine-N-oxide (0.041)	C4 (0.109)
19	3-Aminoisobutyrate (0.022)	2-Oxoglutarate (0.04)	PC.aa.C30:2 (0.108)
20	Methylguanidine (0.022)	Creatinine (0.039)	SM.C16:1 (0.105)
21	Succinate (0.02)	3-Methylhistidine (0.037)	PC.aa.C38:5 (0.095)
22	Tartrate (0.02)	Hippurate (0.036)	lysoPC.a.C14:0 (0.093)
23	Tryptophan (0.019)	Pantothenate (0.035)	PC.aa.C36:2 (0.087)
24	Lactate (0.019)	2-Hydroxyisobutyrate (0.035)	lysoPC.a.C18:0 (0.084)
25	cis-Aconitate (0.019)	Pyroglutamate (0.033)	PC.aa.C24:0 (0.084)
26	Hippurate (0.018)	3-Hydroxybutyrate (0.032)	PC.aa.C4:3 (0.082)
27	Pyroglutamate (0.018)	Threonine (0.031)	SM.C18:1 (0.081)
28	Tyrosine (0.017)	Ethanolamine (0.028)	C12 (0.074)
29	Acetone (0.016)	Lactate (0.026)	PC.aa.C36:1 (0.074)
30	3-Indoxylsulfate (0.016)	Succinate (0.026)	PC.aa.C36:3 (0.072)

[1] ALST, appendicular lean soft tissue; Cx:y; acylcarnitine (x = number of carbons in acyl chain, y = location of double bond); lysoPC a, lysoPhosphatidylcholine acyl; PC.aa, phosphatidylcholine diacyl; SM, sphingomyelin; TFM, total fat mass.
[2] Mutual information was used to quantify the dependence between two variables (see ''Methods''); this allowed us to rank metabolites according to the degree of dependence with the two different classes (low and high). In this case, mutual information was high when a particular metabolite was highly correlated with ALST or TFM.

measurements provided the estimates of energy and macronutrient intake. The fasting period would appear to have largely eliminated the effects of foods eaten in the 3 d preceding the measurements, because neither total energy intake nor that of any energy macronutrient was associated with a clear-cut metabolite profile. Notably, total caloric intake had very weak relationships with fat and lean mass. This disconnect between intake and body weight/fat or lean mass is a cardinal feature of cancer patients and is attributed to metabolic alterations that are not seen in simple malnutrition (42).

Protein intake resulted in a relatively weak predictive model with LOOCV accuracy of up to 73%. There are many possible reasons for the appearance of a metabolic profile, when stratified for protein intake. All of the elements of macronutrient intake (sugar, fat, total carbohydrate, and protein) were correlated with one another. The individuals within the low-protein class here had protein-energy malnutrition, with median intakes of 0.8 g protein and 24.9 kcal · kg body weight^{-1} · d^{-1} and a protein : energy ratio of 0.034 g protein · kcal^{-1}. This was different ($P <$ 0.001) from the high-protein class with a protein intake of 1.6 g · kg body weight^{-1} · d^{-1}, an energy intake of 37.0 kcal · kg body weight^{-1} · d^{-1}, and a protein : energy ratio of 0.047 g protein · kcal^{-1} and can be compared with recommended values of recommended intakes of 1.2–2.0 g protein and 30–35 kcal · kg body weight^{-1} · d^{-1} for this patient population (43). Thus, in overnight-fasted individuals, the presence of quite severe malnutrition could be detected but with a relatively weak signal. It will be useful to further understand why the patients with the low energy intakes chose and consumed foods with a low protein : energy ratio, because it may not be to their advantage.

Analysis could not produce a satisfactory predictor for REE and RQ. It seems likely that the fasting protocol contributed to a low level of variation in RQ. REE was expressed in relation to whole body LST to limit the variation contributed by variability in body weight and in percent body fat. Here, the mean REE of the two classes (27.7 ± 1.9 vs. 37.2 ± 3.2 kcal · kg LST^{-1} · d^{-1}) can be compared with values for healthy individuals across the entire lifespan. With reference to the meta-analysis reported by Weinsier et al. (41), 27.7 kcal · kg LST^{-1} · d^{-1} is a normal metabolic rate for healthy individuals aged 50–70 y. By contrast, 37.2 kcal · kg LST^{-1} · d^{-1} (30% higher) is hypermetabolic and REE in this range is not expected beyond adolescence. The lack of discrimination of REE in our models suggests that the higher

overall metabolic throughput is not necessarily associated with alterations in patterns of metabolites.

Previously published population-based data demonstrated high variability in muscle and fat mass in cancer patients (44). Individuals with the same body weight and BMI may differ by up to 2.5-fold in the amount of lean and muscle tissue. Previous work also indicated that even patients with metastatic cancers were likely to be overweight or obese (13,44), with severe muscle wasting (sarcopenia) concurrently present in a significant proportion of the population. Like many elderly people, cancer patients may be affected by age, limited mobility, and their primary disease as well as comorbid conditions, each resulting in muscle loss.

There has been interest in testing for metabolic signature of cancer. The present results suggest that, to identify specific metabolic discriminates of cancer, as opposed to cancer-associated variations in lean and fat mass and food intake (which are not specific to this disease), studies should be conducted with the following controls: *1*) participants should be in the fasted state, because it reduces the effect of diet; *2*) participants could be stratified for protein intake or provided a standardized protein intake prior to the measurement to remove its effect (45); and *3*) as suggested above, assess patients only within predefined categories of lean and fat mass or by ensuring that lean and fat mass is used as a basis for matching participants in cancer compared to noncancer groups. Many diseases in addition to cancer are associated with wasting of the lean tissues (chronic obstructive pulmonary disease, chronic heart failure, chronic renal failure, diabetes). In the absence of suitable controls, a metabolic signature apparently due to one of these diseases may merely reflect the presence of wasting (a nonspecific effect) as opposed to a specific disease-related process.

Methodological considerations. Particular attention was given to characterizing the patient population using precise measures of lean and fat mass, supplemented with estimates of metabolic rate, RQ, and macronutrient intake. Empirical results showed that SVM was the most accurate classifier of the three algorithms tested; because this was the second time that superior predictive accuracy was obtained with SVM (18) compared with multiple other methods, it may be a good method for future work.

No single analytical platform captures all metabolites in a biological sample. Proton NMR is less sensitive than MS but easily captures amino acids and their intermediates, tricarboxilic acid cycle intermediates, and other metabolites involved in energy metabolism (e.g., glucose). Both skeletal muscle and organ tissues metabolize compounds producing end products that are ultimately excreted from the body via urine. Thus, it was not surprising that metabolites mainly responsible for the discrimination between low and high lean and muscle tissues includes creatine, which originates mainly from muscle. Several molecules within 1-carbon and amino acid metabolism (such as betaine, trigonelline, guanidoacetate) and intermediary metabolites (2-oxoglutarate, succinate, fumarate, pyruvate) may represent the daily turnover and oxidation of amino acids in the lean tissues, which would be quantitatively higher in individuals with higher lean mass. The sensitivity of MS enabled the quantification of lipid molecules in plasma that was not possible with NMR and proved to be a superior platform to build a predictive model for TFM.

In conclusion, the above results can inform future studies using metabolomic approaches in human nutrition and metabolism. Metabolomics is an emerging field and as unresolved issues are addressed and the technology becomes more widely available, translation into clinical settings will become more realistic. The present study addressed one of these unresolved issues, namely the effect of tissue mass variability on the metabolome. The above findings raise the possibility of a noninvasive test for lean and fat mass based on urine sampling in cancer patients. This may have applications in screening or in individuals who cannot undergo diagnostic imaging by DXA, CT, or MRI. Also, our findings will assist in the design of future studies to assist with minimization of sources of variation that may confound the interpretation of results.

Acknowledgments

C.S. and V.E.B. were responsible for the study design; M.M. and C.M.M.P. were involved with the sample collection; C.S., R.M., and R.B. were involved in sample analysis; C.S., T.E., R.E., and R.G. were responsible for the data analysis; all authors contributed to interpretation of the results and development of the manuscript. All the authors read and approved the final manuscript.

Literature Cited

1. Gibney MJ, Walsh M, Brennan L, Roche HM, German B, van Ommen B. Metabolomics in human nutrition: opportunities and challenges. Am J Clin Nutr. 2005;82:497–503.
2. Zivkovic AM, German JB. Metabolomics for assessment of nutritional status. Curr Opin Clin Nutr Metab Care. 2009;12:501–7.
3. Zeisel SH, Freake HC, Bauman DE, Bier DM, Burrin DG, German JB, Klein S, Marquis GS, Milner JA, Pelto GH, et al. The nutritional phenotype in the age of metabolomics. J Nutr. 2005;135:1613–6.
4. Wishart DS. Metabolomics: applications to food science and nutrition research. Trends Food Sci Technol. 2008;19:482–93.
5. Scalbert A, Brennan L, Fiehn O, Hankemeier T, Kristal BS, van Ommen B, Pujos-Guillot E, Verheij E, Wishart D, Wopereis S. Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. Metabolomics. 2009;5:435–58.
6. Stalmach A, Mullen W, Barron D, Uchida K, Yokota T, Cavin C, Steiling H, Williamson G, Crozier A. Metabolite profiling of hydroxycinnamate derivatives in plasma and urine after the ingestion of coffee by humans: identification of biomarkers of coffee consumption. Drug Metab Dispos. 2009;37:1749–58.
7. van Velzen EJ, Westerhuis JA, van Duynhoven JP, van Dorsten FA, Grun CH, Jacobs DM, Duchateau GS, Vis DJ, Smilde AK. Phenotyping tea consumers by nutrikinetic analysis of polyphenolic end-metabolites. J Proteome Res. 2009;8:3317–30.
8. Llorach R, Urpi-Sarda M, Jauregui O, Monagas M, Andres-Lacueva C. An LC-MS-based metabolomics approach for exploring urinary metabolome modifications after cocoa consumption. J Proteome Res. 2009;8:5060–8.
9. Llorach R, Garrido I, Monagas M, Urpi-Sarda M, Tulipani S, Bartolome B, Andres-Lacueva C. Metabolomics study of human urinary metabolome modifications after intake of almond (Prunus dulcis (Mill.) D. A. Webb) skin polyphenols. J Proteome Res. 2010;9:5859–67.
10. Wang Y, Tang H, Nicholson JK, Hylands PJ, Sampson J, Holmes E. A metabonomic strategy for the detection of the metabolic effects of chamomile (Matricaria recutita L.) ingestion. J Agric Food Chem. 2005;53:191–6.
11. Primrose S, Draper J, Elsom R, Kirkpatrick V, Mathers JC, Seal C, Beckmann M, Haldar S, Beattie JH, Lodge JK, et al. Metabolomics and human nutrition. Br J Nutr. 2011;105:1277–83.
12. Lieffers JR, Mourtzakis M, Hall KD, McCargar LJ, Prado CM, Baracos VE. A viscerally driven cachexia syndrome in patients with advanced colorectal cancer: contributions of organ and tumor mass to whole-body energy demands. Am J Clin Nutr. 2009;89:1173–9.
13. Prado CM, Lieffers JR, McCargar LJ, Reiman T, Sawyer MB, Martin L, Baracos VE. Prevalence and clinical implications of sarcopenic obesity in patients with solid tumours of the respiratory and gastrointestinal tracts: a population-based study. Lancet Oncol. 2008;9:629–35.

14. Heymsfield SB, Arteaga C, McManus C, Smith J, Moffitt S. Measurement of muscle mass in humans: validity of the 24-hour urinary creatinine method. Am J Clin Nutr. 1983;37:478–94.
15. Hutton JL, Martin L, Field CJ, Wismer WV, Bruera ED, Watanabe SM, Baracos VE. Dietary patterns in patients with advanced cancer: implications for anorexia-cachexia therapy. Am J Clin Nutr. 2006;84:1163–70.
16. Orphanidou C, McCargar L, Birmingham CL, Mathieson J, Goldner E. Accuracy of subcutaneous fat measurement: comparison of skinfold calipers, ultrasound, and computed tomography. J Am Diet Assoc. 1994;94:855–8.
17. Ellegård LH, Ahlen M, Korner U, Lundholm KG, Plank LD, Bosaeus IG. Bioelectric impedance spectroscopy underestimates fat-free mass compared to dual energy X-ray absorptiometry in incurable cancer patients. Eur J Clin Nutr. 2009;63:794–801.
18. Eisner R, Stretch C, Eastman T, Xia J, Hau D, Damaraju S, Greiner R, Wishart DS, Baracos VE. Learning to predict cancer-associated skeletal muscle wasting from 1H-NMR profiles of urinary metabolites. Metabolomics. 2011;7:25–34.
19. Chorell E, Moritz T, Branth S, Antti H, Svensson MB. Predictive metabolomics evaluation of nutrition-modulated metabolic stress responses in human blood serum during the early recovery phase of strenuous physical exercise. J Proteome Res. 2009;8:2966–77.
20. Miccheli A, Marini F, Capuani G, Miccheli AT, Delfini M, Di Cocco ME, Puccetti C, Paci M, Rizzo M, Spataro A. The influence of a sports drink on the postexercise metabolism of elite athletes as investigated by NMR-based metabolomics. J Am Coll Nutr. 2009;28:553–64.
21. Walsh MC, Brennan L, Pujos-Guillot E, Sebedio JL, Scalbert A, Fagan A, Higgins DG, Gibney MJ. Influence of acute phytochemical intake on human urinary metabolomic profiles. Am J Clin Nutr. 2007;86:1687–93.
22. Posner BM, Martin-Munley SS, Smigelski C, Cupples LA, Cobb JL, Schaefer E, Miller DR, D'Agostino RB. Comparison of techniques for estimating nutrient intake: the Framingham Study. Epidemiology. 1992;3:171–7.
23. Price GM, Halliday D, Pacy PJ, Quevedo MR, Millward DJ. Nitrogen homeostasis in man: influence of protein intake on the amplitude of diurnal cycling of body nitrogen. Clin Sci (Lond). 1994;86:91–102.
24. Prado CM, Mourtzakis M, Baracos VE, Reiman T, Sawyer MB, McCargar LJ. Overweight and obese patients with solid tumors may have sarcopenia, poor prognosis and early features of cachexia. Int J Body Compos Res. 2010;8:7–15.
25. Mourtzakis M, Prado CM, Lieffers JR, Reiman T, McCargar LJ, Baracos VE. A practical and precise approach to quantification of body composition in cancer patients using computed tomography images acquired during routine care. Appl Physiol Nutr Metab. 2008;33:997–1006.
26. Heymsfield SB, Smith R, Aulet M, Bensen B, Lichtman S, Wang J, Pierson RN Jr. Appendicular skeletal muscle mass: measurement by dual-photon absorptiometry. Am J Clin Nutr. 1990;52:214–8.
27. Cao DX, Wu GH, Zhang B, Quan YJ, Wei J, Jin H, Jiang Y, Yang ZA. Resting energy expenditure and body composition in patients with newly detected cancer. Clin Nutr. 2010;29:72–7.
28. Holmes E, Foxall PJ, Nicholson JK, Neild GH, Brown SM, Beddell CR, Sweatman BC, Rahr E, Lindon JC, Spraul M, et al. Automatic data reduction and pattern recognition methods for analysis of 1H nuclear magnetic resonance spectra of human urine from normal and pathological states. Anal Biochem. 1994;220:284–96.
29. Craig A, Cloarec O, Holmes E, Nicholson JK, Lindon JC. Scaling and normalization effects in NMR spectroscopic metabonomic data sets. Anal Chem. 2006;78:2262–7.
30. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. Anal Chem. 2006;78:4281–90.
31. Westerhuis JA, Hoefsloot HCJ, Smit S, Vis DJ, Smilde AK, van Velzen EJJ, van Duijnhoven JPM, van Dorsten FA. Assessment of PLSDA cross validation. Metabolomics. 2008;4:81–9.
32. Mahadevan S, Shah SL, Marrie TJ, Slupsky CM. Analysis of metabolomic data using support vector machines. Anal Chem. 2008;80:7562–70.
33. Friedman J, Hastie T, Hofling H, Tibshirani R. Pairwise coordinate optimization. Ann Appl Stat. 2007;1:302–32.
34. Gavaghan CL, Wilson ID, Nicholson JK. Physiological variation in metabolic phenotyping and functional genomic studies: use of orthogonal signal correction and PLS-DA. FEBS Lett. 2002;530:191–6.
35. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. ACM SIGKDD Explor. 2009;11:10–8.
36. R Development Core Team. R: a language and environment for statistical computing; 2008 [cited November 3, 2011]. ISBN 3–900051–07–0. Available from: http://www.R-project.org.
37. Friedman J, Hastie T, Tibshirani R. Regularized paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33:1–22.
38. Saude EJ, Sykes BD. Urine stability for metabolomic studies: effects of preparation and storage. Metabolomics. 2007;3:19–27.
39. Slupsky CM, Rankin KN, Wagner J, Fu H, Chang D, Weljie AM, Saude EJ, Lix B, Adamko DJ, Shah S, et al. Investigations of the effects of gender, diurnal variation, and age in human urinary metabolomic profiles. Anal Chem. 2007;79:6995–7004.
40. Psihogios NG, Gazi IF, Elisaf MS, Seferiadis KI, Bairaktari ET. Gender-related and age-related urinalysis of healthy subjects by NMR-based metabonomics. NMR Biomed. 2008;21:195–207.
41. Weinsier RL, Schutz Y, Bracco D. Reexamination of the relationship of resting metabolic rate to fat-free mass and to the metabolically active components of fat-free mass in humans. Am J Clin Nutr. 1992;55:790–4.
42. Fearon K, Strasser F, Anker SD, Bosaeus I, Bruera E, Fainsinger RL, Jatoi A, Loprinzi C, MacDonald N, Mantovani G, et al. Definition and classification of cancer cachexia: an international consensus. Lancet Oncol. 2011;12:489–95.
43. Arends J, Bodoky G, Bozzetti F, Fearon K, Muscaritoli M, Selga G, van Bokhorst-de van der Schueren MA, von Meyenfeldt M, Zurcher G, Fietkau R, et al. ESPEN guidelines on enteral nutrition: non-surgical oncology. Clin Nutr. 2006;25:245–59.
44. Baracos VE, Reiman T, Mourtzakis M, Gioulbasanis I, Antoun S. Body composition in patients with non-small cell lung cancer: a contemporary view of cancer cachexia with the use of computed tomography image analysis. Am J Clin Nutr. 2010;91:S1133–7.
45. Thorpe JM, Roberts SA, Ball RO, Pencharz PB. Prior protein intake may affect phenylalanine kinetics measured in healthy adult volunteers consuming 1 g protein · kg^{-1} · d^{-1}. J Nutr. 1999;129:343–8.