

# Learning to predict cancer-associated skeletal muscle wasting from $^1\text{H}$ -NMR profiles of urinary metabolites

Roman Eisner · Cynthia Stretch · Thomas Eastman · Jianguo Xia ·  
David Hau · Sambasivarao Damaraju · Russell Greiner · David S. Wishart ·  
Vickie E. Baracos

Received: 18 February 2010 / Accepted: 17 July 2010 / Published online: 1 August 2010  
© Springer Science+Business Media, LLC 2010

**Abstract** Cancer-associated muscle wasting is associated with reduction in functional status, in response to treatment and in life expectancy. Methods currently used to assess muscle loss involve diagnostic imaging techniques such as computed tomography (CT), which are costly, inconvenient, invasive, time consuming and have limited ability to detect early or slowly evolving wasting. We present a novel approach using single time-point urinary metabolite profiles to determine whether a patient is experiencing muscle wasting. We analyzed 93 random urine samples from patients with cancer using  $^1\text{H}$ -NMR. Using two successive CT images we assessed their lumbar skeletal muscle area ( $\text{cm}^2$ ) to estimate the rate of muscle change (% loss or gain over time) for each patient. The average muscle change over time was  $-4.71\%/100$  days in the muscle-losing group and  $+3.91\%/100$  days in the comparator group. Bivariate statistics identified metabolites related with muscle loss, including constituents and

metabolites of muscle (creatine, creatinine, 3-OH-isovalerate), amino acids (Leu, Ile, Val, Ala, Thr, Tyr, Gln, Ser) and intermediary metabolites. We also applied machine-learning techniques to identify patterns of urinary metabolites that identify which patients are likely to lose muscle mass. We evaluated the predictive performance of 8 machine-learning approaches using fivefold cross validation and permutation testing, and found that SVM provided the best generalization accuracy (82.2%). These results suggest that  $^1\text{H}$ -NMR analysis of a single random urine sample may be a fast, cheap, safe and inexpensive tool to screen and monitor muscle loss, and that useful classifiers for predicting related metabolic conditions are possible with the methodology presented.

**Keywords** NMR · Muscle wasting · Cancer · Urine · Machine learning

Roman Eisner and Cynthia Stretch contributed equally to this work.

R. Eisner · T. Eastman · R. Greiner · D. S. Wishart  
Department of Computing Science, University of Alberta,  
Edmonton, AB, Canada

J. Xia · D. Hau · D. S. Wishart  
Department of Biological Sciences, University of Alberta,  
Edmonton, AB, Canada

C. Stretch · V. E. Baracos (✉)  
Department of Oncology, Division of Palliative Care Medicine,  
University of Alberta, 11560 University Avenue, Edmonton,  
AB T6G 1Z2, Canada  
e-mail: vickie.baracos@ualberta.ca

S. Damaraju  
Department of Laboratory Medicine and Pathology,  
University of Alberta, Edmonton, AB, Canada

## 1 Introduction

Metabolic, neuronal and hormonal controls normally ensure that body weight and composition are maintained constant during adult life. Involuntary weight gains or losses are significant perturbations of this precise control. The focus of our research is cancer-associated muscle wasting. Muscle depletion is associated with poor functional status, treatment toxicity and shorter life expectancy (Antoun et al. 2010; Prado et al., 2007, 2008, 2009; Tan et al. 2009). Prado et al. (2008), and others have shown that muscle loss may occur independently of changes in fat mass, and that muscle wasting may be an early or occult phenomenon that is difficult to detect against the background of overall body weight and body weight change, especially in overweight or obese individuals. A recent consensus definition of cachexia

(Evans et al. 2008) makes a distinction between the behavior of skeletal muscle and adipose tissue: “cachexia is a complex metabolic syndrome associated with underlying illness and characterized by loss of muscle with or without loss of fat mass...”. Muscle wasting may go unnoticed in its early stages, if progressing slowly, or if it is obscured by changes in other tissues. Improved approaches for detecting the onset and evolution of muscle wasting would help manage wasting syndromes and facilitate early intervention. Wasting has a cumulative nature. For example, muscle loss at a rate of 0.07% in 1 day appears trivial, but if sustained equals 7% loss over 100 days and 25% loss over a year—a quantity that would have important physiological consequences for most individuals. Dual energy X-ray absorptiometry (DXA), computed tomography (CT) and magnetic resonance imaging (MRI) are considered the most precise measures of adipose and muscle tissues currently available (Heymsfield et al. 1997; Mitsiopoulos et al. 1998; Shen et al. 2004a; Mourtzakis et al. 2008), but have several limitations. Images must be repeated over time to detect loss, access and cost may be limitations and their analysis may be time-consuming and labor-intensive and DEXA and CT expose patients to radiation. Clinicians are keen to find new diagnostic approaches for identifying and monitoring muscle loss that are faster, cheaper, safer and more accessible.

We hypothesized that metabolites produced from tissue breakdown are likely to be a sensitive indicator of muscle wasting and may provide a new diagnostic approach. Muscle breakdown generates amino acids and their various catabolites, as well as urea and creatinine. Several of these end products are detectable in physiological fluids using NMR spectroscopy (Slupsky et al. 2007; Wishart 2008). Coupled with recent advances in machine learning and multivariate statistics, metabolomic approaches have led to the identification of biomarkers for several diseases (e.g. celiac disease, prostate cancer) (Bertini et al. 2009; Mahadevan et al. 2008). Based on these ideas we investigated whether we could detect muscle wasting using metabolomic data from urine samples from patients with cancer. Urine was selected as the biofluid of choice, since several end products of muscle catabolism (e.g. creatinine, methylhistidine) are specifically excreted in the urine. We applied machine-learning techniques to identify patterns of urinary metabolic profiles that discriminate the condition of muscle loss.

## 2 Materials and methods

### 2.1 Study design and sample collection

This study was approved by the Alberta Cancer Board Research Ethics Board. All participants provided written informed consent and had advanced cancer of the colon or

lung, defined as locally recurrent or metastatic. Patients with prior radiation to the kidneys or malignancy of the kidney or urinary tract were excluded as these independently alter the ability to concentrate urine normally. Urine was selected as the biofluid of choice. Patients donated a urine sample taken at random (e.g. not controlled for time of day, or food intake) during a routine visit to the cancer center. We did not undertake a 24 h urine collection, as the patients’ ages and medical conditions (life-limiting illness) limit their inclination to commit to this additional burden. Preliminary data from our group suggests that urine volumes do not differ between cancer patients similar to those studied here ( $n = 17$ , mean 24 h urine volume  $0.025 \pm 0.009$  l/kg body weight) and age and sex matched healthy controls ( $n = 25$ ,  $0.024 \pm 0.011$  l/kg body weight), C Stretch unpublished observations. After adding sodium azide to a final concentration of 0.02% to prevent bacterial growth, samples were stored frozen at  $-80^{\circ}\text{C}$  until NMR data acquisition.

### 2.2 CT image analysis

We compared our predictive model, obtained by applying machine learning methods to urine metabolites, with muscle loss quantified in serial CT images (Ross 2003). Images were analyzed for total skeletal muscle tissue cross-sectional area ( $\text{cm}^2$ ) at the 3rd lumbar vertebra using Slice-O-Matic software V4.3 (Tomovision, Montreal, Canada). Further details of image analysis can be found elsewhere (Shen et al. 2004a, b; Mourtzakis et al. 2008) and in our prior studies (Liefers et al. 2009; Prado et al. 2008). During routine clinical care tumor progression is assessed by CT at intervals of  $\sim 100$  days. Two scans (preceding and following the urine sample) were selected. Muscle area in the CT image preceding the urine collection was used as a reference (baseline) to compute the % loss or gain. We expressed this rate as % change per 100 days, to take into account minor variation in the number of days between scans for different individuals.

### 2.3 NMR spectra acquisition and targeted profiling

Urine samples were prepared by adding 65  $\mu\text{l}$  of an internal chemical shift standard (supplied by Chenomx Inc., Edmonton, Canada consisting of 5 mM sodium 2,2-dimethyl-2-silapentane-5-sulfonate- $\text{d}_6$  (DSS- $\text{d}_6$ ) and 0.2% sodium azide in 99%  $\text{D}_2\text{O}$ ) to 585  $\mu\text{l}$  of urine. Using small amounts of NaOH or HCl, the sample was adjusted to  $\text{pH } 6.75 \pm 0.05$ . A 600  $\mu\text{l}$  aliquot of prepared sample was placed in a 5-mm NMR tube (Wilmad, Buena, NJ). All one-dimensional NMR spectra of urine samples were acquired using the first increment of the standard NOESY pulse sequence using a 600 MHz Varian INOVA NMR spectrometer (Varian Inc.,

Palo Alto, CA) equipped with a triple axis-gradient 5-mm HCN probe.

We used a *targeted profiling* or quantitative metabolomic approach (Wishart 2008; Weljie et al. 2006) to identify and quantify metabolites from the resulting NMR spectra using Chenomx NMRSuite 4.6 (Chenomx Inc. Edmonton, Canada). Quantitative approaches are more interpretable than spectral binning and are also more robust with respect to compound overlap, and variability in solution conditions (e.g. pH and ionic strength) (Saude and Sykes 2007; Wishart 2008). Two analysts (DDH, CS) independently analyzed the spectra and we included only those compounds and concentrations agreed upon by both analysts. Compound spiking with authentic standards from the Human Metabolome Library (Wishart et al. 2007) was used to confirm the identity of difficult-to-assign compounds. As a further check, additional (non-NMR) laboratory analyses were conducted to verify creatinine concentrations and amino acid peak assignments and concentrations. Creatinine was determined colorimetrically (SpectraMAX 190 using SoftMax Pro V5 software) with two different commercial kits based on Jaffé's basic picrate method (Stanbio Creatinine Liquicolor Kit, Cat No. SB 0420-250 and Cedarlane, Creatinine Assay Kit, Cat. No. 500701-480). Amino acid assignments and concentrations were verified by a spike-in experiment with a solution containing Ala, Asn, Gln, Gly, His, Ile, Leu, Lys, Phe, Ser, Tau, Thr, Trp, Tyr, Val, 1-Methylhistidine and 3-Methylhistidine (3-MH). Spiked samples were quantified by NMR as described above and by reverse-phase HPLC using Waters pico-tag® method (Waters Co., MA, USA) (Bidlingmeyer et al. 1984).

## 2.4 Statistical methods

### 2.4.1 Data preprocessing

Many statistical procedures assume that variables are normally distributed and a significant violation of this assumption can seriously increase the chances of making an analytical error. Data can appear non-normal if some values are extreme outliers relative to the rest of the sample. This frequently happens in urine samples as metabolite concentrations can vary up to several hundred-fold. To correct for this problem, we transformed the data by taking the natural log of the concentration values.

Water intake during the day can alter concentration of metabolites in urine. We employed three approaches to correct for this effect, including (a) *normalization to creatinine concentration* in each sample (Holmes et al. 1994), (b) *normalization by total peak area of each sample*; this assumes that the integrated area under an NMR spectrum is a linear function of the detectable metabolite

concentrations in the samples (Bollard et al. 2005; Craig et al. 2006) and (c) *probability quotient normalization* (Dieterle et al. 2006), which calculates a most probable dilution factor (the median) by examining the distribution of the quotients of the amplitudes of a test spectrum by those of a reference spectrum.

### 2.4.2 Development of a classifier

Metabolomic researchers (Slupsky et al. 2007; Wishart 2007; Bertini et al. 2009) compute how each individual compound correlates with the outcome—e.g. muscle loss or muscle gain. While such bivariate analyses typically provide valuable biological insights, they do not directly help clinicians who are primarily interested in making a diagnosis. As our primary goal was to develop a tool that could predict whether a patient is losing muscle based on their urine metabolite concentrations, we considered different analytical tools. For diagnostic applications, it is useful to have a *classifier* that returns a prediction: given the metabolic profile  $m_r$  of patient  $r$ , determine whether this patient  $r$  is losing muscle or not,  $c_r \in \{\text{losing, gaining}\}$ . A classifier can base its prediction on a potentially complicated combination of all metabolite concentrations.

A sample of historical data (e.g. a collection of patient metabolic profiles, along with their respective muscle loss/gain values  $\{[m_r, c_r]\}_r$  over a set of patients), is used as a starting point. We can use the machine learning approach to computationally *learn* a classifier, from this historical data. The classifier can then be used to predict the status of future patients. We summarize below a number of machine learning approaches.

### 2.4.3 Classifiers considered

We examined classification performance using 8 different standard statistical and machine-learning approaches:

- Naïve Bayes—a Bayesian classifier that assumes that metabolite concentrations are all independent of each other, for each of the two classes  $C \in \{\text{losing, gaining}\}$  (Witten and Frank 2005).
- Tree-augmented Naïve Bayes (TAN)—a Bayesian classifier allowing a simplified set of conditional dependencies between pairs of metabolites (forming a tree structure), for the overall distribution (Friedman et al. 1997).
- Multi-TAN—identical to TAN except that the tree structure for the two classes is allowed to differ from one another (Friedman et al. 1997).
- Full dependence model—a Bayesian network classifier allowing any metabolite concentration to depend on any other metabolite concentration.

- (e) Partial Least Squares-Discriminant Analysis (PLS-DA)—a common approach in metabolomics studies that uses an eigenvalue-based approach to create a classifier.
- (f) Decision Trees (also called recursive partitioning systems) (Quinlan 1993) sequentially decide which feature to examine, based on the observed values of the features already examined, until having enough information to return a class value (Witten and Frank 2005).
- (g) Support Vector Machines (Mahadevan et al. 2008) view each instance as a vector in  $k$ -dimensional space, and seek the maximally separating hyperplane between the classes in this space (Witten and Frank 2005). We use a SVM with a linear kernel.
- (h) Pathway Informed Analysis (PIA)—A Bayesian classifier using biological knowledge in the form of metabolic pathways, extracted from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa et al. 2008). PIA treats each pathway as a graph structure, similar to the “substrate graph” structure of Wagner and Fell (2001), where each node of the graph represents a specific metabolite and each edge connects a pair of metabolites that participate in the same reaction (e.g., malate and fumarate) (Eastman 2010). Incorporating pathway knowledge into classifiers represents a confluence of statistical and biological expertise that could improve predictive power (note that none of the other 7 learning algorithms use biological information).

#### 2.4.4 Prediction accuracy of classifiers

We used cross-validation and permutation tests to assess the accuracy of our classifiers. The quality of a classifier is defined by how well it performs on novel test instances that were not part of the training set. Such an evaluation could be based on an external validation set—e.g., new data that the learner has not previously seen. Here, we used cross-validation (Hastie et al. 2001) to approximate an external validation set. This involves partitioning the training data into  $k = 5$  subsets; then  $k$  times we first produce a classifier based on  $(k - 1)$  subsets of the data, which we then test on the remaining subset. We then use these  $k$  evaluation scores to estimate the mean and variance of the accuracy (on novel data) that we would obtain using a classifier built using the entire training set.

Permutation tests are particularly useful for confirming robustness of the classifier and for ensuring it has not been over-trained (Pesarin 2001). We first randomly permute the labels (muscle loss status) for the training data, then run the entire cross-validation evaluation process on this newly

re-labeled data. As permutation removes any correlation between data and label, we should get just “noise” on the permuted data. We then compared the diagnostic accuracy on the original un-permuted data, with the distribution of the accuracy obtained using the various permuted datasets. This allowed us to estimate the likelihood that results from un-permuted data were due to chance.

#### 2.4.5 Bivariate analysis

A standard approach to analyzing quantitative metabolomic data is bivariate analysis—e.g. finding the degree to which the primary outcome depends on each individual metabolite. Each highly-dependent metabolite is a feature that is associated with the biological process of interest (here, muscle wasting). We focus on binary classification, labeling each patient as either losing or gaining muscle and use *mutual information* (Cover and Thomas 2006) to quantify the dependence between the binary class outcome  $C \in \{\text{losing, gaining}\}$  and the real-valued concentration of each of the 63 metabolites  $M \in \{\text{fumarate, malate, oxaloacetate ...}\}$ , which we assume follows a Gaussian distribution. This involves computing:

$$\text{MI}(M, C) = \sum_c \int_{-\infty}^{+\infty} p(m, c) \log \frac{p(m, c)}{p(m)P(c)} dm$$

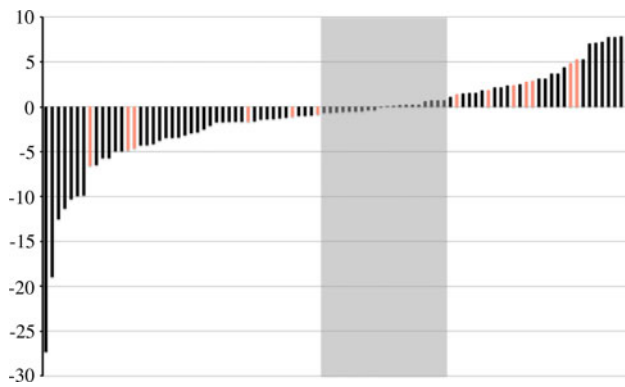
where  $P(c)$  is the probability that the class  $C = c$  (here, we set  $P(C = \text{losing})$  to be the observed frequency of “muscle losing” patients in the data sample) and  $p(m) = 1/\sqrt{2\pi\sigma_m^2} \exp\left(-(m - \mu_m)^2/(2\sigma_m^2)\right)$  is the Gaussian probability density function, which is based on the mean  $\mu_m$  and variance  $\sigma_m^2$  estimated from the data sample. We use a similar function for  $p(m, c)$ , using estimated mean and variance that depends on whether the class is  $C = \text{losing}$  or  $C = \text{gaining}$ . Notice that  $\text{MI}(M, C)$  is 0 if the metabolite  $M$  is completely independent of  $C$ ; larger values indicate a higher degree of correlation.

### 3 Results and discussion

#### 3.1 Muscle loss continuum in advanced cancer patients

Figure 1 shows the distribution of muscle loss and gain for the 93 samples in our study. Patients in the 2 classes (Table 1) did not differ in age, sex, cancer site or stage and these features were uncorrelated with one another and also uncorrelated with muscle loss/gain. Because the measurements of muscle change are only precise to about 1.5%/100 days, we adopted the following simple classification rule: patients were designated as not losing muscle





**Fig. 1** Percentage muscle change continuum in cancer patients as determined by computed tomography image analysis. The *boxed region* highlights the patients excluded from analysis. *Red (light colored) columns* indicate those samples that were misclassified by the SVM during cross-validation. Using serial computed tomography images, patients' muscle change (loss or gain) was computed. The boxed region indicates patients whose muscle change fell within a minimal margin of  $\pm 0.75\%/100$  days. These patients were not included in the analysis as their calculated muscle change is within the precision error of the imaging method. The remaining patients were classified as losing or not losing muscle. These two groups represent the distal ends of the muscle change continuum and statistically different from each other ( $P < 0.001$ ) (Color figure online)

**Table 1** Characteristics of study participants

	Patients with muscle gain >0.75%/100 days	Patients with muscle loss >0.75%/100 days	All patients
Number	29	44	93
Age (mean $\pm$ SD)	64 $\pm$ 11	62 $\pm$ 10	63 $\pm$ 10
Gender			
% Male	48	66	59
Cancer type			
% Lung	28	23	25
% Colorectal	72	77	75
Cancer stage			
% Stage 1	0	0	0
% Stage 2	0	2	1
% Stage 3	31	25	27
% Stage 4	69	73	71

(anabolic) if the change in muscle mass exceeded  $+0.75\%/100$  days; patients were designated as losing muscle (catabolic) if muscle was lost over time and exceeded  $-0.75\%/100$  days. Using this classification scheme we excluded the 20 patients whose change was between  $-0.75\%$  and  $+0.75\%/100$  days (boxed region in Fig. 1). We classified 44 patients as muscle losing (Mean  $-4.71\%/100$  days; SD = 5.13) and 29 patients as not losing muscle (Mean  $+3.91\%/100$  days; SD = 2.33). These two groups of patients with known muscle change status (loss or gain) were used to build predictive models using urinary metabolites.

### 3.2 Metabolites detected and used in statistical approaches

We assigned and quantified 71 metabolites in each sample. Creatinine concentrations assessed by NMR were within 95% (95% confidence interval of 91–97) of the values confirmed by laboratory tests. Spike-in experiments provided positive confirmation of peak assignments for Ala, Asn, Gln, Gly, His, Ile, Leu, Lys, Phe, Ser, Tau, Thr, Trp, Tyr, Val, 1-Methylhistidine and 3-Methylhistidine. We excluded drug metabolites or drug vehicle constituents (ibuprofen, acetaminophen, salicylurate, propionate, propylene glycol, mannitol) from statistical analyses. Methanol, a microbial (non human) metabolite was excluded as is unlikely to be related to muscle loss. Urea was excluded since suppression of the NMR signal by pre-saturation may lead to resonant suppression of the urea peak due to proton exchange with water, thereby making its quantification unreliable (Saude and Sykes 2007). The remaining 63 metabolites were used in subsequent analyses.

Median concentration and concentration ranges for these 63 metabolites are shown (Table 2). Owing to the specific nature of urine, metabolites associated with amino acid metabolism, urea cycle, intermediary metabolism (glycolysis, TCA cycle, 1-carbon metabolism) and creatine metabolism were prominent. Numerous individual metabolites, as well as the total concentration (e.g. the sum of all 63 metabolites), were increased in the patients losing muscle (Mann–Whitney test). Levels of creatinine were higher in patients with muscle loss ( $P < 0.001$ ).

None of the methods of data normalization (by creatinine concentration, by total peak area, probability quotient normalization) proved helpful and all three methods *reduced* the predictive accuracy of the classifiers, compared with no data normalization. Normalization by creatinine did not perform well, most likely because it assumes that creatinine concentration is only related to urine dilution, which is not true in our situation as urinary creatinine originates in muscle and its excretion is known to be raised when muscle is broken down (Akçay et al. 2001). The use of a log transformation was ultimately found to be the only preprocessing step for the metabolite concentrations that improved the predictors' performance compared with raw concentration values.

### 3.3 A classifier for muscle loss based on urinary metabolites

Of all tested algorithms (Table 3), SVM was the most accurate classifier, and predicted muscle loss status with a (fivefold cross validation) accuracy of 82.2% ( $\sigma = 7.45\%$ ). Although PIA produced a classifier with the same accuracy, we focus on the SVM model because it is more

**Table 2** Median concentration and concentration ranges of 63 urine metabolites included in the statistical analyses

Metabolite	<sup>1</sup> H chemical shift (ppm) and coupling <sup>a</sup>	Median concentration (range) (μM)		
		Patients with muscle gain >0.75%/100 days	Patients with muscle loss >0.75%/100 days	All patients
1,6-Anhydro-β-D-glucose	3.53(m), 3.67(m), 3.69(m), 3.75(dd), 4.09(dd), 4.62(dd), 5.45(m)	34.7 (9.4–191.9)	98.5 (4.7–687.3)**	47.5 (4.7–942.6)
1-Methylnicotinamide	4.47(s), 8.17(t), 8.88(d), 8.95(d), 9.26(s)	18.2 (6.4–1036)	51.7 (6.9–474.1)**	36.4 (4.6–1036)
2-Aminobutyrate	0.97(t), 1.89(m), 3.71(dd)	7.4 (1.3–28.9)	15.3 (2.1–173)**	10.5 (1.3–173)
2-Hydroxyisobutyrate	1.35(s)	18.8 (4.9–66.2)	42.1 (7.8–85.4)***	33.7 (4.5–193.4)
2-Oxoglutarate	2.44(t), 3.00(t)	25.3 (5.6–987.2)	69.3 (5.5–2467)*	63 (5.5–2467)
3-Aminoisobutyrate	1.19(d), 2.60(m), 3.03(dd), 3.1(dd)	21.3 (3.1–209.2)	30.7 (2.6–1481)	29.8 (2.6–1481)
3-Hydroxybutyrate	1.19(d), 2.29(dd), 2.40(dd), 1.14(m)	6.5 (2.2–34.3)	25.6 (1.7–176.6)***	11.8 (1.7–176.6)
3-Hydroxyisovalerate	1.26(s), 2.36(s)	5.3 (0.9–57.6)	21.3 (2.5–164.4)***	13 (0.9–359.7)
3-Indoxylsulfate	7.18(m), 7.26(m), 7.36(s), 7.49(d), 7.70(d), 10.10(s)	104.9 (27.8–613.8)	202.4 (34.9–1038)***	165.6 (27.8–1038)
4-Hydroxyphenylacetate	3.44(s), 6.85(m), 7.16(m)	48.2 (15.5–799.6)	93 (17.6–430.9)**	70.1 (15.5–799.6)
Acetate	1.91(s)	16.2 (3.5–202.5)	71.5 (9.9–410.6)***	34.9 (3.3–410.6)
Acetone	2.22(s)	6.8 (2.3–23.8)	8.2 (2.3–206.5)	7.6 (2.1–206.5)
Adipate	1.54(m), 2.19(m)	6.2 (1.6–19.2)	16.1 (3.1–325.6)***	11.1 (1.6–325.6)
Alanine	1.47(d), 3.78(qt)	78.6 (16.8–601)	320.1 (26.8–1314)***	195.7 (13.9–1447)
Asparagine	2.86(dd), 2.95(dd), 3.99(dd), 6.91(s), 7.62(s)	29.2 (6.7–152.6)	64.4 (8–272.9)***	42.3 (6.7–272.9)
Betaine	3.26(s), 3.89(s)	27.3 (2.3–312.4)	112.5 (4.1–391.7)***	63.8 (2.3–788.8)
Carnitine	2.41(dd), 2.45(dd), 3.22(s), 3.40(m), 3.43(m), 4.56(m)	19 (2.7–206.5)	31.7 (4.5–488.1)**	24.6 (2.7–488.1)
Citrate	2.53(d), 2.69(d)	1014 (59.6–4214)	2336 (80.9–13636)**	1790 (59.6–13636)
Creatine	3.02(s), 3.92(s)	18.6 (2.8–393.6)	87.4 (7.9–1863)***	48.8 (1.1–1862.7)
Creatinine	3.03(s), 4.05(s)	3616 (1003–15116)	10003 (1256–33944)***	8032 (868.2–33944)
Dimethylamine	2.72(s)	148.8 (41.3–496.2)	370.6 (46.9–1562)***	306.2 (27.4–1562)
Ethanolamine	3.14(m), 3.82(m)	113.9 (21.5–907.8)	270.7 (16.1–1439)**	212.4 (16.1–1439)
Formate	8.45(s)	61.4 (6.4–294.4)	136.1 (27.7–1476)**	91.7 (6.4–1476)
Fucose	1.20(d), 1.24(d), 3.44(dd), 3.56(t), 3.63(dd), 3.74(dd), 3.76(dd), 3.79(m), 3.80(m), 3.85(dd), 3.86(m), 3.87(m), 3.97(dd), 4.01(m), 4.03(m), 4.07(dd), 4.18(m), 4.55(d), 5.20(d), 5.22(d), 5.27(d)	37.7 (5.7–196.2)	90.2 (13.6–408.4)***	68.4 (5.7–408.4)
Fumarate	6.51	3.1 (0.8–36.2)	6.6 (1.1–96.6)***	4.2 (0.8–96.6)
Glucose	3.24(dd), 3.40(m), 3.41(m), 3.47(m), 3.49(m), 3.53(dd), 3.71(m), 3.72(dd), 3.74(m), 3.82(m), 3.84(m), 3.90(dd), 4.64(d), 5.23(d)	92.9 (26.9–337.5)	397 (43.9–8724.8)***	190 (26.9–8724.8)
Glutamine	2.12(m), 2.15(m), 2.43(m), 2.47(m), 3.76(t), 6.87(s), 7.58(s)	112.9 (23.3–862.1)	401.3 (26.8–1684)***	226.4 (15.1–1684)
Glycine	3.56(s)	382.2 (38.3–2281)	690.2 (52.6–5073)**	560 (38.3–18195)
Glycolate	3.95(s)	66 (5.4–439.9)	179.8 (10.9–682.8)**	126.3 (5.4–885.5)
Guanidoacetate	3.79(s)	45.6 (7–301.1)	96.1 (18.2–563.5)**	72.1 (4.6–563.5)
Hippurate	3.96(d), 7.54(m), 7.55(m), 7.63(t), 7.82(m), 7.83(m), 8.52(s)	574.8 (122.7–6667)	21816 (93.1–19263)***	1274 (93.1–19263)
Histidine	3.15(dd), 3.25(dd), 3.99(qt), 7.11(s), 7.92(s)	78.4 (16.3–616.1)	326 (14.1–1869)***	182.8 (14–1868.8)
Hypoxanthine	8.18(s), 8.20(s)	31.8 (3.8–161.7)	44.8 (4.2–265.3)	42.6 (3.7–265.3)
Isoleucine	0.93(t), 1.00(d), 1.25(m), 1.46(m), 1.97(m), 3.66(d)	4.2 (1.8–18)	8.4 (2–40.2)*	7.7 (1.8–117.5)
Lactate	1.32(d), 4.11(d)	39.4 (7.3–199.3)	110.4 (17.5–3659)***	87.3 (7.3–3659)

**Table 2** continued

Metabolite	<sup>1</sup> H chemical shift (ppm) and coupling <sup>a</sup>	Median concentration (range) (μM)		
		Patients with muscle gain >0.75%/100 days	Patients with muscle loss >0.75%/100 days	All patients
Leucine	0.95(d), 0.96(d), 1.67(m), 1.70(m), 1.73(m), 3.73(m)	9 (2.5–31.4)	24.3 (3.5–103.8)***	19.1 (2.5–103.8)
Lysine	1.43(m), 1.50(m), 1.72(m), 1.88(m), 1.91(m), 3.02(t), 3.75(t)	34.9 (10.5–787.5)	106.7 (15.2–464.6)***	75.6 (4.3–787.5)
Methylamine	2.6(s)	5.1 (1.5–44.6)	20.4 (1.8–52.3)***	15.5 (1.5–108.8)
Methylguanidine	2.83(s)	6.8 (1.7–36.6)	10.4 (2.1–141.6)	8.8 (1.7–141.6)
<i>N,N</i> -Dimethylglycine	2.92(s), 3.72(s)	9.2 (1.2–52.5)	30.4 (3.4–119.9)***	21.3 (1.2–169.5)
<i>O</i> -Acetylcarnitine	2.14(s), 2.50(dd), 2.63(dd), 3.19(s), 3.60(dd), 3.84(dd), 5.59(m)	6.1 (1.2–43.9)	14.2 (1.6–254.6)**	11.6 (1.2–254.6)
Pantothenate	0.89(s), 0.92(s), 2.41(t), 3.39(d), 3.43(qt), 3.44(qt), 3.51(d), 3.98(s), 8.00(dd)	14.4 (3.1–691.4)	26.3 (2.6–187.5)*	22.6 (1.7–691.4)
Pyroglutamate	2.03(m), 2.38(m), 2.41(m), 2.50(m), 4.17(dd)	82.7 (21.4–442.1)	251.7 (37.6–1066)***	155.5 (18–1066)
Pyruvate	2.37(s)	6.5 (0.9–66.6)	21.4 (1.8–184.8)***	15.4 (0.9–184.8)
Quinolate	7.46(dd), 8.00(d), 8.45(d)	26.7 (5.2–163.6)	76 (16.2–260.7)***	51 (5.2–260.7)
Serine	3.84(dd), 3.94(dd), 3.98(dd)	90.5 (16.2–269.8)	218.3 (32.6–1245)***	136.8 (16.2–1245)
Succinate	2.4(s)	8.6 (1.7–221)	50.2 (6.4–587.8)***	29.3 (1.2–587.8)
Sucrose	3.47(t), 3.55(dd), 3.66(s), 3.68(m), 3.76(m), 3.80(m), 3.82(m), 3.83(m), 3.84(m), 3.88(m), 4.04(t), 4.21(d), 5.40(d)	19.2 (6.5–600.6)	67.6 (10.2–2081)***	41.1 (6.5–2081)
Tartrate	4.33(s)	10.7 (2.2–273.1)	16.3 (3–834.5)	12.8 (2.2–834.5)
Taurine	3.25(t), 3.41(t)	176.2 (17.9–1513)	407.5 (55.3–4285)**	280.3 (13.5–4284)
Threonine	1.32(d), 3.59(d), 4.26(m)	39.1 (9.1–250.5)	102.6 (8.2–448.5)***	67.6 (4.4–448.5)
Trigonelline	4.43(s), 8.07(dd), 8.82(m), 8.83(m), 9.11(s)	74.6 (10.1–564.5)	190.2 (10.2–2257)**	97.2 (10.1–2257)
Trimethylamine <i>N</i> -oxide	3.26(s)	243 (55.7–1533)	542.8 (66.8–5460)**	403.2 (14.9–5460)
Tryptophan	3.30(dd), 3.47(dd), 4.05(q), 7.19(m), 7.27(m), 7.31(s), 7.52(d), 7.72(d)	21.3 (10.5–185.4)	82.1 (9.9–260.3)***	56.7 (9.9–260.3)
Tyrosine	3.05(dd), 3.19(dd), 3.94(q), 6.88(m), 7.20(m)	23.9 (4.2–180)	86.8 (14–537.2)***	58.5 (4.2–537.2)
Uracil	5.79(d), 7.52(d)	20.2 (3.1–138)	29.5 (4.2–179.2)	28.1 (3.1–179.2)
Valine	0.98(d), 1.03(d), 2.26(m), 3.61(d)	13.2 (4.1–53.3)	39.8 (4.3–160.1)***	30.8 (4.1–160.1)
Xylose	3.22(dd), 3.31(dd), 3.43(t), 3.52(dd), 3.60(m), 3.62(m), 3.65(m), 3.67(m), 3.69(m), 3.92(dd), 4.57(d), 5.19(d)	32.8 (10.1–259.4)	71.3 (16.6–2158)***	51.2 (9.8–2155)
<i>Cis</i> -aconitate	3.12(d), 5.75(m)	54.1 (12.9–298.1)	235.1 (15.1–1862)***	128.7 (12.9–1862)
Myo-Inositol	3.27(t), 3.53(dd), 3.62(dd), 4.05(m)	30.5 (11.6–315.5)	131.9 (22–850.4)***	78.2 (8.1–850.4)
<i>Trans</i> -aconitate	3.44(s), 3.59(s)	13.6 (4.9–181.2)	45.3 (7.9–216.3)***	26.9 (4.9–639.3)
1-Methylhistidine	3.21(dd), 3.29(dd), 3.74(s), 3.95(dd), 7.13(s), 8.10(s)	73 (11.4–1186)	245 (16.6–2694)**	199.8 (11.4–2694)
3-Methylhistidine	3.08(dd), 3.16(dd), 3.70(s), 3.96(dd), 7.03(s), 7.70(s)	29.7 (8.6–184.6)	82.8 (8–317)***	71.4 (8–317)
Total metabolites		29.2 (0.8–15116)	82.8 (1.1–33944)***	56.7 (0.8–33944)

\*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P \leq 0.001$   $P$  values were obtained using Mann–Whitney nonparametric statistical analysis comparing patients with versus without muscle loss

<sup>a</sup> s Singlet, d doublet, dd doublet–doublet, t triplet, q quartet, m multiplet

familiar. Figure 1 also identifies the patients who were misclassified by the SVM classifier. While 6 misclassified patients had muscle losses/gains of less than 2%, there

were 7 misclassified patients with losses/gains up to 5%. It is not obvious why these patients were misclassified. It could be due to some undetected underlying condition

**Table 3** Predictive performance for muscle loss of 8 machine learning approaches, averaged over the fivefolds of cross-validation

Classifier	Average accuracy (%)	Standard deviation
Support vector machines	82.2	7.5
Pathway informed analysis	82.2	7.2
Multi-tree augmented naïve Bayes	76.8	8.0
PLS-DA	76.7	14.8
Tree augmented naïve Bayes	76.7	9.6
Full dependence model	75.1	6.8
Naïve Bayes model	75.1	9.7
J48 decision tree	65.8	9.1
Random permutation test (SVM)	53.8	6.5

(e.g. kidney dysfunction) or to inherent limitations of the data, given that the minimum interval over which gain or loss can be detected by CT (months) was represented by a single point in time sample.

In 1,000 permutation tests (Pesarin 2001) on the SVM model, which produced an average accuracy of 53.8%, no permutation performed better than the SVM classifier in Table 3. This non-parametric test suggests that chance alone would not have produced this SVM result—i.e., the probability that a classifier (here SVM) could produce a score of 82.2%, if there was no real pattern in the data, is  $P < 0.001$ . This supports our claim that SVM is finding a meaningful pattern within the data. SVM performance was superior to PLS-DA, a method often used in metabolomic studies (Westerhuis et al. 2008). PLS-DA reduces the dimensionality of the data in a way that increases the separation with respect to the topic of interest (e.g., a disease state, or other variable under study). While PLS-DA can overfit (Westerhuis et al. 2008), our results show that PLS-DA performed competitively with the top predictors.

### 3.4 Urine metabolites related to muscle loss

Different metabolites appear in urine by processes of passive diffusion, active transport and reuptake and are not a representative sample of all of intermediary metabolism. Nevertheless, that certain urinary metabolites are related to muscle loss (Table 4) does suggest some of the underlying biology of muscle wasting. Several of the metabolites are constituents, breakdown products or metabolites formed in muscle. Creatinine is a degradation product of creatine, a phosphorylated molecule specific to muscle energy metabolism; both of these compounds were related to muscle loss. Muscle proteins contain a higher proportion of branched chain amino acids compared with proteins in

**Table 4** Bivariate analysis: top 30 urine metabolites related to skeletal muscle loss

Metabolite	Mutual information <sup>a</sup>
Adipate	0.213524065
Glucose	0.203986953
Quinolate	0.196946368
Myo-inositol	0.17310194
Valine	0.165888472
Succinate	0.164968427
Betaine	0.16477065
Leucine	0.16401522
<i>N,N</i> -dimethylglycine	0.157109136
3-Hydroxyisovalerate	0.15366801
Creatine	0.147918127
Acetate	0.14713014
Alanine	0.145640343
Pyroglutamate	0.142823206
3-Hydroxybutyrate	0.142709618
Glutamine	0.141498304
<i>Cis</i> -aconitate	0.134050157
Methylamine	0.130789947
Tryptophan	0.130351855
Dimethylamine	0.126891503
Xylose	0.125333817
Creatinine	0.125318846
Formate	0.123810009
Tyrosine	0.116765473
<i>Trans</i> -aconitate	0.110643466
Lactate	0.109694272
Sucrose	0.108986623
2-Hydroxyisobutyrate	0.107381302
Serine	0.106181068
Threonine	0.101972559

<sup>a</sup> *Mutual information* is a way to quantify dependence between two variables. We computed the mutual information between each of the 63 metabolites the class outcome, *C*. Here we have a binary outcome variable (losing muscle vs. not losing muscle) and a continuous metabolite concentration variable that we assume follows a Gaussian distribution. Mutual Information computed as described under statistical methods and yields unit-less values, larger values indicate a higher degree of dependence

other tissues, and muscle is the predominant site of their catabolism. Thus the association of valine, leucine and of a decarboxylation product of leucine, 3-OH-isovalerate, with muscle loss is not surprising. This is not exclusive to the branched chain amino acids; during muscle protein breakdown, all of its constituent amino acids enter oxidative pathways. Increased levels of several metabolites in urine are possibly indicative of increased flux of amino acids (Leu, Ile, Val, Ala, Thr, Tyr, Gln, Ser), and of amino acid carbon through intermediary metabolism (succinate,



*trans*-aconitate) and 1-carbon metabolism (betaine, trigonelline). Finally, in relation to the suggestion that insulin resistance may be a prominent feature of cancer-associated muscle wasting (Asp et al. 2009; Wang et al. 2006), urinary glucose was elevated in patients with muscle loss (397  $\mu$ M) compared with patients who were not losing muscle (93  $\mu$ M) ( $P < 0.001$ ). Elevated blood and urine glucose levels are associated with insulin-resistant states, and while the urine glucose levels in patients with muscle wasting were below levels typical of diabetes, this may be an early sign of insulin resistance.

### 3.5 First steps towards diagnostic markers of cancer-associated skeletal muscle wasting using metabolic profiling

Human populations are variable with respect to age, gender, ethnicity, diet, drug intake, and health status and some, but not all, of these features can be controlled in research studies. Against this background, we tried to determine whether urine could be used to diagnose patients with skeletal muscle wasting, an important physiological component of the negative nitrogen balance characteristic of wasting syndromes. Muscle wasting may go unnoticed in its early stages, if progressing slowly, or if it is obscured by changes in other tissues. This is why a robust classifier that can achieve an accuracy of 82%, using metabolites in a single randomly collected urine sample could be considered a significant advance compared with a standard approach that requires several months and the acquisition of at least 2 diagnostic images.

We can envisage an even better predictor and more precise diagnostic test based on an extended metabolite profile. In NMR spectra, it is typically possible to detect only those compounds with concentrations  $> 1 \mu$ M. Analysis of blood plasma or more sensitive or comprehensive metabolomic methods (e.g. MS-based methods) may reveal additional metabolites related to muscle loss. Different analytic approaches could also permit detection of compounds involved in lipid metabolism to shed light on fat loss and gain. Furthermore, serial urine sampling and CT image analysis over time would take advantage of repeated measures within individuals and would enable the ability to define biochemical changes early and late in disease and pathways implicated in disease pathogenesis. Finally, it will be important to account for some of the presently unexplained sources of variation, which may limit the precision of this diagnostic test. Variation in metabolite concentrations may relate to time of day and meals (Walsh et al. 2006) and these effects could be reduced by limiting sample collection to morning after a standardized time without food, as is done for many laboratory tests.

## 4 Concluding remarks

Our work is the first attempt to use metabolomics to diagnose muscle wasting occurring as a result of cancer cachexia in humans. We developed a single time-point urine test using concentrations of 63 urinary metabolites to diagnose muscle wasting. This minimally invasive test is rapid, robust, quite accurate (82.2%), and able to detect a small but physiologically relevant rate of muscle loss (outside of  $\pm 0.75\%$  loss/100 days). Metabolites related to muscle wasting include a variety of compounds likely to originate from catabolism of this tissue and may also shed some light on the underlying metabolic aberrations that lead to muscle loss.

**Acknowledgments** The authors wish to thank the Alberta Cancer Foundation (ACF), the Cross Cancer Institute, the Alberta Ingenuity Fund (AIF), the Alberta Advanced Education and Technology (AAET), the Canadian Institutes for Health Research (CIHR) and the Alberta Ingenuity Centre for Machine Learning (AICML) for financial support.

**Financial or material support** This work was supported by grants from the Alberta Cancer Board and Alberta Cancer Foundation, the Alberta Ingenuity Fund, the Alberta Ingenuity Centre for Machine Learning, the Natural Sciences and Engineering Research Council of Canada, and Genome Canada.

**Author disclosures** R. Eisner, C. Stretch, T. Eastman, J. Xia, D. Hau, S. Damaraju, R. Greiner, D.S. Wishart and V.E. Baracos have no conflicts of interest.

## References

- Akçay, M. N., Akçay, G., Solak, S., Balik, A. A., & Aylu, B. (2001). The effect of growth hormone on 24-h urinary creatinine levels in burned patients. *Burns*, 27, 42–45.
- Antoun, S., Baracos, V. E., Birdsell, L., Escudier, B., & Sawyer, M. B. (2010). Low body mass index and sarcopenia associated with dose-limiting toxicity of sorafenib in patients with renal cell carcinoma. *Annals of Oncology*, 21, 1594–1598.
- Asp, M. L., Tian, M., Wendel, A. A., & Belury, M. A. (2009). Evidence for the contribution of insulin resistance to the development of cachexia in tumor-bearing mice. *International Journal of Cancer*, 126, 756–763.
- Bertini, I., Calabro, A., De Carli, V., Luchinat, C., Nepi, S., Porfirio, B., et al. (2009). The metabonomic signature of celiac disease. *J Proteome Research*, 8, 170–177.
- Bidlingmeyer, B. A., Cohen, S. A., & Tarvin, T. L. (1984). Rapid analysis of amino acids using pre-column derivatization. *J Chromatography*, 336, 93–104.
- Bollard, M. E., Stanley, E. G., Lindon, J. C., Nicholson, J. K., & Holmes, E. (2005). NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition. *NMR in Biomedicine*, 18, 143–162.
- Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory*. Hoboken, NJ: Wiley-Interscience.
- Craig, A., Cloarec, O., Holmes, E., Nicholson, J. K., & Lindon, J. C. (2006). Scaling and normalization effects in NMR spectroscopic metabonomic data sets. *Analytical Chemistry*, 78, 2262–2267.

- Dieterle, F., Ross, A., Schlotterbeck, G., & Senn, H. (2006). Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in  $^1\text{H}$  NMR metabonomics. *Analytical Chemistry*, 78, 4281–4290.
- Eastman, T. (2010). *A disease classifier for metabolic profiles based on metabolic pathway knowledge*. MSc Thesis, University of Alberta.
- Evans, W. J., Morley, J. E., Argiles, J., Bales, C., Baracos, V., Guttridge, D., et al. (2008). Cachexia: A new definition. *Clinical Nutrition*, 27, 793–799.
- Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning*, 29, 131–163.
- Hastie, T., Tibshirani, R., & Friedman, J. H. (2001). *The elements of statistical learning: Data mining, inference, and prediction: With 200 full-color illustrations*. New York: Springer.
- Heymsfield, S. B., Wang, Z., Baumgartner, R. N., & Ross, R. (1997). Human body composition: Advances in models and methods. *Annual Review of Nutrition*, 17, 527–558.
- Holmes, E., Foxall, P. J., Nicholson, J. K., Neild, G. H., Brown, S. M., Beddell, C. R., et al. (1994). Automatic data reduction and pattern recognition methods for analysis of  $^1\text{H}$  nuclear magnetic resonance spectra of human urine from normal and pathological states. *Analytical Biochemistry*, 220, 284–296.
- Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., et al. (2008). KEGG for linking genomes to life and the environment. *Nucleic Acids Research*, 36, D480–D484.
- Lieffers, J. R., Mourtzakis, M., Hall, K. D., Mccargar, L. J., Prado, C. M., & Baracos, V. E. (2009). A viscerally driven cachexia syndrome in patients with advanced colorectal cancer: Contributions of organ and tumor mass to whole-body energy demands. *American Journal of Clinical Nutrition*, 89, 1173–1179.
- Mahadevan, S., Shah, S. L., Marrie, T. J., & Slupsky, C. M. (2008). Analysis of metabolomic data using support vector machines. *Analytical Chemistry*, 80, 7562–7570.
- Mitsiopoulos, N., Baumgartner, R. N., Heymsfield, S. B., Lyons, W., Gallagher, D., & Ross, R. (1998). Cadaver validation of skeletal muscle measurement by magnetic resonance imaging and computerized tomography. *Journal of Applied Physiology*, 85, 115–122.
- Mourtzakis, M., Prado, C. M., Lieffers, J. R., Reiman, T., Mccargar, L. J., & Baracos, V. E. (2008). A practical and precise approach to quantification of body composition in cancer patients using computed tomography images acquired during routine care. *Applied Physiology, Nutrition and Metabolism*, 33, 997–1006.
- Pesarin, F. (2001). *Multivariate permutation tests: With applications in biostatistics*. Chichester, New York: J. Wiley.
- Prado, C. M., Baracos, V. E., Mccargar, L. J., Mourtzakis, M., Mulder, K. E., Reiman, T., et al. (2007). Body composition as an independent determinant of 5-fluorouracil-based chemotherapy toxicity. *Clinical Cancer Research*, 13, 3264–3268.
- Prado, C. M., Lieffers, J. R., Mccargar, L. J., Reiman, T., Sawyer, M. B., Martin, L., et al. (2008). Prevalence and clinical implications of sarcopenic obesity in patients with solid tumours of the respiratory and gastrointestinal tracts: A population-based study. *Lancet Oncology*, 9, 629–635.
- Prado, C. M., Baracos, V. E., Mccargar, L. J., Reiman, T., Mourtzakis, M., Tonkin, K., et al. (2009). Sarcopenia as a determinant of chemotherapy toxicity and time to tumor progression in metastatic breast cancer patients receiving capecitabine treatment. *Clinical Cancer Research*, 15, 2920–2926.
- Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. San Mateo, CA: Morgan Kaufmann Publishers.
- Ross, R. (2003). Advances in the application of imaging methods in applied and clinical physiology. *Acta Diabetologica*, 40(Suppl 1), S45–S50.
- Saude, E. J., & Sykes, B. D. (2007). Urine stability for metabolomic studies: Effects of preparation and storage. *Metabolomics*, 3, 19–27.
- Shen, W., Punyanitya, M., Wang, Z., Gallagher, D., St-Onge, M. P., Albu, J., et al. (2004a). Total body skeletal muscle and adipose tissue volumes: Estimation from a single abdominal cross-sectional image. *Journal of Applied Physiology*, 97, 2333–2338.
- Shen, W., Punyanitya, M., Wang, Z., Gallagher, D., St-Onge, M. P., Albu, J., et al. (2004b). Visceral adipose tissue: Relations between single-slice areas and total volume. *American Journal of Clinical Nutrition*, 80, 271–278.
- Slupsky, C. M., Rankin, K. N., Wagner, J., Fu, H., Chang, D., Weljie, A. M., et al. (2007). Investigations of the effects of gender, diurnal variation, and age in human urinary metabolomic profiles. *Analytical Chemistry*, 79, 6995–7004.
- Tan, B. H., Birdsell, L. A., Martin, L., Baracos, V. E., & Fearon, K. C. (2009). Sarcopenia in an overweight or obese patient is an adverse prognostic factor in pancreatic cancer. *Clinical Cancer Research*, 15(22), 6973–6979.
- Wagner, A., & Fell, D. A. (2001). The small world inside large metabolic networks. *Proceedings of the Royal Society of Biological Science*, 268, 1803–1810.
- Walsh, M. C., Brennan, L., Malthouse, J. P., Roche, H. M., & Gibney, M. J. (2006). Effect of acute dietary standardization on the urinary, plasma, and salivary metabolomic profiles of healthy humans. *American Journal of Clinical Nutrition*, 84(3), 531–539.
- Wang, X., Hu, Z., Hu, J., Du, J., & Mitch, W. E. (2006). Insulin resistance accelerates muscle protein degradation: Activation of the ubiquitin-proteasome pathway by defects in muscle cell signaling. *Endocrinology*, 147, 4160–4168.
- Weljie, A. M., Newton, J., Mercier, P., Carlson, E., & Slupsky, C. M. (2006). Targeted profiling: Quantitative analysis of  $^1\text{H}$  NMR metabolomics data. *Analytical Chemistry*, 78, 4430–4442.
- Westerhuis, J. A., Hoefsloot, H. C. J., Smit, S., Vis, D. J., Smilde, A. K., Van Velzen, E. J. J., et al. (2008). Assessment of PLSDA cross validation. *Metabolomics*, 4, 81–89.
- Wishart, D. S. (2007). Current progress in computational metabolomics. *Briefings in Bioinformatics*, 8, 275–284.
- Wishart, D. S. (2008). Quantitative metabolomics using NMR. *Trends in Analytical Chemistry*, 27, 228–237.
- Wishart, D. S., Tzur, D., Knox, C., Eisner, R., Guo, A. C., Young, N., et al. (2007). HMDB: The Human Metabolome Database. *Nucleic Acids Research*, 35(Database issue), D521–D526.
- Witten, I. H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques*. Boston, MA: Morgan Kaufman.