Published in final edited form as:

J Proteome Res. 2008 August ; 7(8): 3091–3101. doi:10.1021/pr070520u.

A STATISTICAL MODEL FOR iTRAQ DATA ANALYSIS

Elizabeth G. Hill^{*,†}, John H. Schwacke[†], Susana Comte-Walters[‡], Elizabeth H. Slate[†], Ann L. Oberg[§], Jeanette E. Eckel-Passow[§], Terry M. Therneau[§], and Kevin L. Schey[‡]

[†] Department of Biostatistics, Bioinformatics and Epidemiology, Medical University of South Carolina

[‡] Department of Cell and Molecular Pharmacology, Medical University of South Carolina

[§] Division of Biostatistics, Mayo Clinic College of Medicine, Rochester, Minnesota

Abstract

We describe biological and experimental factors that induce variability in reporter ion peak areas obtained from iTRAQ experiments. We demonstrate how these factors can be incorporated into a statistical model for use in evaluating differential protein expression and highlight the benefits of using analysis of variance to quantify fold change. We demonstrate the model's utility based on an analysis of iTRAQ data derived from a spikein study.

Keywords

Analysis of variance; ANOVA; iTRAQ; mass spectrometry; differential expression

Introduction

Isobaric tags for relative and absolute quantitation (iTRAQTM) [Applied Biosystems (ABI)-Framingham, MA]^{1, 2} are seeing increased use for differential protein expression analysis. Through the use of amine-specific isobaric tags, the iTRAQ technology facilitates the simultaneous analysis of up to four different conditions of interest in the same iTRAQ experiment. Data normalization and differential expression analyses are typically carried out in a series of steps, with summaries from one step carried forward to the next. This piecemeal approach to data analysis is inefficient (estimation of normalization effects is based on incomplete data) and does not properly account for the loss in degrees of freedom attributable to normalization.³ Furthermore, commercial software does not provide means for combining data from multiple iTRAQ experiments.

iTRAQ data analyses and subsequent interpretations vary substantially in the literature. To identify differentially expressed proteins, one common approach selects upper and lower bounds for ratio thresholds but the criteria for threshold selection are variable. For example, Seshi⁴ arbitrarily declares ratios greater than 1.25 or less than 0.8 as differentially expressed, while Salim et al.⁵ use thresholds of 1.2 and 0.83. Alternatively, Ross et al.¹ constructed thresholds based on the global average ratio and standard deviation within conditions, declaring protein ratios in excess of one standard deviation of the global mean as differentially expressed.

*To whom correspondence should be addressed. Department of Biostatistics, Bioinformatics and Epidemiology, Hollings Cancer Center, Medical University of South Carolina, 86 Jonathan Lucas Street, MSC 955, Charleston SC, 29425. Phone: (843) 876-1115. Fax: (843) 792-4233. Email: hille@musc.edu..

Supporting Information Available. Additional detail describing the development of the mathematical model for iTRAQ reporter ion peak areas is available as supplemental material. This information is available free of charge via the Internet at <http://pubs.acs.org>.

Other authors have based inference on more traditional statistical approaches such as *t*-tests and one-way analysis of variance (ANOVA) models, with the data analysis conducted one protein at a time.⁶⁻⁸

We present an ANOVA analytic approach that combines both normalization (bias removal) and assessment of differential protein expression in a single model fit to the collection of reporter ion peak areas (corrected for isotopic overlap) from all observed tandem mass spectra. Notably, our model allows for analysis of data from multiple iTRAQ experiments, overcoming the constraint of current iTRAQ protein quantitation software that limits analysis to a single experiment. Our approach is similar to that of Kerr et al.³ who conduct normalization and differential expression analysis of microarray data using ANOVA.

Our ANOVA model is derived by considering both biologic and experimental sources of variability that give rise to the observed data. Specifically, we derive our model beginning with mathematical expressions for the differences in protein profiles for two experimental conditions, and successively introduce multiplicative factors reflecting the stages of an iTRAQ experiment that result in the collection of observed reporter ion peak areas. We demonstrate that logarithmic transformation of the multiplicative model gives rise to an additive model that fits within the broader class of ANOVA models, thereby naturally facilitating fold-change estimation and assessment of uncertainty associated with those estimates. For the basic scientist, our derivation motivates the structure of the resulting ANOVA model by building the model through an understanding of the biologic and experimental processes that lead to the observed data. For the statistician, our approach identifies explicitly those factors deemed most likely to explain the variability in the observed reporter ion peak areas, so that one is fully aware of the sources of variation included, grouped or ignored in any given analysis.

We demonstrate the performance of ANOVA on data derived from a single iTRAQ experiment applied to a mixture of known proteins in pre-specified ratios prepared in house to mimic the 2006 Proteomics Research Group Association of Biomolecular Resource Facilities round-robin sample.⁹ We note, however, that our approach accommodates the analysis of multiple iTRAQ experiments. We direct the reader to this manuscript's companion paper by Oberg et al.¹⁰ that demonstrates the application of the ANOVA modeling approach described herein to the analysis of six iTRAQ experiments comparing serum protein profiles in patients across three histologic subtypes of acute cardiomyopathy. We complete the presentation of our analysis with a comparison of our results with those obtained from a more traditional analysis of iTRAQ data using formulas for protein quantitation from the GPS software [GPS Version 3.6 software (ABI)]. We conclude with a discussion of strengths and limitations.

Materials and Methods

Constructing a Mathematical Model for iTRAQ Reporter Ion Peak Areas

As is common in statistical analysis, we first develop a model that relates the quantity of interest and the characteristics of the data to the quantities we observe. In this case, we wish to estimate the relative change in expression for a particular protein (our quantity of interest) by relating the measured areas for each of four reporter ion peaks (our observed quantities) to the characteristics of both the observed spectra and the sample conditions from which the spectra are derived (e.g. treatment, sex, age, iTRAQ tag, fraction). Ideally we wish to account for all effects that influence the observed quantities and eliminate the influence of those effects not associated with the hypothesis of interest. Unfortunately, this is not entirely feasible but we include as many of these effects as practical and include terms for noise (error) that account for these remaining differences. In doing so we allow for an assessment of the uncertainty in our estimates and thereby the confidence we place in the associated findings.

High throughput proteomic technologies, such as iTRAQ, attempt to measure protein-level expression changes in complex mixtures through a complicated process of digestion, labeling, fractionation, mass spectrometry, spectrum data processing, and statistical analysis. As such, the model described here must: (1) relate differences in treatment to relative differences in protein expression; (2) relate protein expression to peptide expression; and (3) relate peptide expression to observed reporter ion peak areas. Model components describing relationships (1) and (2) comprise sources of biologic variability, while those describing (3) comprise experimental sources of variability. Our model captures these relationships using simple multiplicative expressions where, for example, we assume the abundance of a peptide is equal to the product of the abundance of the associated protein and a factor specific to that protein and peptide. The fact that researchers tend to think in fold changes (ratios) on the raw scale is an indication that they believe effects are multiplicative on this scale. For example, the factor relating protein i to peptide j may reflect the efficiency of peptide production during trypsin digestion. We write this multiplicative relationship as $a_j = A_i \cdot L_j$ where i and j index the protein and peptide, respectively, a_j is a number giving the amount (e.g. concentration) of peptide j , A_i is a number giving the amount of the associated protein, and L_j is a number giving the multiplicative factor relating the j th peptide's concentration to its associated protein's concentration.

In the following sections we construct a mathematical model for iTRAQ reporter ion peak areas by considering the contributions of both biologic and experimental components of variation to the observed reporter ion peak areas. We demonstrate how the resulting expression translates into a statistical model that fits within the well-known class of ANOVA models, thereby facilitating estimation of expression ratios and associated measures of uncertainty. For simplicity we assume scientific interest surrounds protein expression for subjects from a condition of interest (e.g. treated or diseased subjects), hereafter referred to as the *treated* condition, relative to those from some control condition (e.g. non-diseased subjects). We note, however, the model generalizes to any number of conditions. Furthermore, the model is more general than required by the application provided here to demonstrate its utility. Indeed, the model is not 'one-size-fits-all', but in general must be tailored to the selected experimental design. We acknowledge that, due to variations in experimental design, not all model elements are identifiable from one application to the next, but we include all terms in the model for completeness. Additionally, the model does not include all sources of error and we discuss limitations in the conclusions.

Population- and subject-level protein profiles—We define the *expected protein expression profile* for a given condition as the collection of proteins and their corresponding amounts in a representative sample (e.g. serum, saliva, tissue supernatant) of uniform size (volume or mass) from a *population* of interest. We use the term *expected* in a statistical sense to represent the true, but unknown, average state of nature for the universe of subjects (or *population*) to whom scientific interest is directed (e.g. all adults with a given disease). Furthermore, we define the *complete proteome* as the set of all proteins encoded in an organism's genome, and the *protein profile* as the list of numbers giving the concentrations for each protein in that proteome (unexpressed proteins have a protein profile concentration of zero). While the complete proteome is fixed, the protein profile can change across conditions, individuals, and even samples from the same individual. Since we are constrained by finite sample sizes, the expected protein profile is a theoretical concept and impossible to determine exactly. It is important, however, in our statistical model as it provides the foundation for the relationships that follow. We refer to this expected protein profile as the *population-level protein profile*.

For the proteome containing I proteins, we express the concentration of the i th protein as P_i and the entire set of numbers giving the expression levels of each of those proteins as $[P_i]$. Let

$R_{i,c}$ represent the relative amount of protein i comparing condition c to the control condition. For c equal to the treated condition, $R_{i,c}$ is the expected protein ratio comparing the treated to control condition (the primary parameter of interest), and the product $P_i \cdot R_{i,c}$ represents the expected amount of protein i in the treated condition. Alternatively, when c identifies the control condition, $R_{i,c} = 1$ so that $P_i \cdot R_{i,c} = P_i$. Thus a unified notation, $[P_i \cdot R_{i,c}]$, represents the *population-level protein profile* for both conditions simultaneously.

As described in the Supporting Information, we allow for differences in total protein across conditions and include in the model the term R_c equal to the relative amounts of total protein comparing the treated to control condition. We follow the same convention stated for the protein by condition factor ($R_{i,c}$), and set $R_c = 1$ when c indexes the control condition. We complete the profile by including this factor and write the population-level protein profile for condition c as $[P_i \cdot R_{i,c} \cdot R_c]$.

Biologic variability induces deviations from the population average across study subjects and has been shown to contribute substantially to the variability observed in iTRAQ data.¹¹ Accordingly, let $D_{k,i}$ be the observed amount of protein i relative to its expected amount for subject k . Then $[P_i \cdot R_{i,c} \cdot R_c \cdot D_{k,i}]$ is the *subject-level protein profile* for subject k in condition c .

Subject-level peptide profile—iTRAQ measurements are made at the peptide level and therefore the model must reflect the relationships between peptide and associated protein expression levels. This association may be ambiguous as some tryptic peptides can be associated with more than one protein. These peptides are said to be degenerate and are often eliminated prior to analysis. In this model we assume that observed peptides are uniquely assigned to a protein. Accordingly, we use function notation in our peptide subscripts, $j(i)$, to indicate the j th peptide is uniquely derived from the i th protein. In statistical parlance, we say that peptides are *nested* within proteins.

Post-translational modifications and/or splice variants can affect individual peptides within a protein in a condition-dependent manner. As such, our model includes terms capturing the effect of condition at the peptide level in addition to the condition-specific protein-level effects discussed previously. Let $F_{j(i)}$ be the ratio of the expected amount of the j th peptide to the expected amount of the i th protein for subjects in the control condition. Furthermore, let $G_{j(i),c}$ be the ratio of the expected amount of peptide j comparing condition c to the control condition. For c equal to the control condition, $G_{j(i),c} = 1$. Then the *subject-level peptide profile* for the k th subject in condition c is given by $[P_i \cdot R_{i,c} \cdot R_c \cdot D_{k,i} \cdot F_{j(i)} \cdot G_{j(i),c}]$.

Observed reporter ion peak areas—Our model development accommodates the joint analysis of data derived from multiple iTRAQ experiments within the same study. For a given iTRAQ experiment, the steps leading from tryptic peptides to observed reporter ion peak areas include: sample loading into iTRAQ channels; peptide derivitization; mixing of labeled samples; fractionation; and mass spectrometry. Let $I_{q,\ell}$ be the proportion of sample loaded into the ℓ th channel ($\ell=114, 115, 116 \text{ or } 117$) for iTRAQ experiment q . Only labeled peptides contribute to the reporter ion cluster and thus, each contributing peptide must have been derivitized in the iTRAQ labeling reaction. For the q th experiment, define $Z_{q,\ell}$ as the labeling efficiency (value between 0 and 1) of the ℓ th reagent indicating the fraction of the peptides successfully derivitized in the labeling step. The labeled iTRAQ samples are then mixed together. Let $M_{q,\ell}$ be the relative amount of iTRAQ sample for the ℓ th reagent added to this mixture.

The mixture of labeled peptides is then separated using two stages of chromatography and fractions are subjected to mass spectrometry. Peptide peaks are selected from the mass spectra

of each fraction and subjected to tandem mass spectrometry (MS/MS). Four reporter ion peaks appear in a small cluster in the low mass range (m/z 114 – 117). These four peak intensities (areas under the peaks corrected for isotopic impurity) are assumed to be proportional to the amount of the given peptide labeled with the appropriate tag. For the q th iTRAQ experiment, we define this constant of proportionality as B_q and include it as a factor in the peptide profile to yield the expected reporter ion peak area profiles.

We relate the expected reporter ion peak areas to the observed reporter ion peak areas through a term that represents the measurement noise associated with each observed peak from MS/MS spectrum s . We assume that the measurement noise, represented by the random quantity E , is distributed such that the mean of the logarithm of E is zero. Under this condition, the error contributes no bias to the reporter ion peak area measurement scale. We assume further that the biological and measurement errors are uncorrelated. The resulting profile of *observed reporter ion peak areas* is

$$\left[P_i \cdot R_{i,c} \cdot R_c \cdot D_{k,i} \cdot F_{j(i)} \cdot G_{j(i),c} \cdot I_{q,\ell} \cdot Z_{q,\ell} \cdot M_{q,\ell} \cdot B_q \cdot E_{q,s,j(i),\ell} \right]. \quad (1)$$

Translation to a Statistical Model

Our purpose is to use Model 1 to address the scientific question of interest, namely to identify differentially expressed proteins across conditions. We focus now on translating the conceptual model described in Model 1 to a statistical model. We begin with two simplifications, the first of which combines the biological error, D , and measurement noise, E , into a single error term, H . We also drop the subject-level subscript from the combined error term, since the subject is uniquely identified from the combination of experiment (q) and tag (114, 115, 116, or 117). The second simplification combines factors associated with sample loading (I), iTRAQ labeling (Z), and sample mixing (M) into a single factor, V . The latter is necessary since these factors are *confounded*, that is to say their contributions can not be estimated separately. Practically, this means that if we detect a uniform shift (bias) in the collection of reporter ion peak areas associated with one label relative to another within a given iTRAQ experiment, that shift could be attributed to inaccuracies in sample loading, differences in labeling efficiencies, or to pipetting or other technical errors. Regardless, there is no way to disentangle these effects. This simplified version of Model 1 becomes

$$\left[P_i \cdot R_{i,c} \cdot R_c \cdot F_{j(i)} \cdot G_{j(i),c} \cdot V_{q,\ell} \cdot B_q \cdot H_{i,j(i),c,q,s,\ell} \right]. \quad (2)$$

To summarize, the collection of observed reporter ion peak areas is described by a product of factors capturing effects due to: protein (P_i); protein by condition ($R_{i,c}$); condition (R_c); peptide ($F_{j(i)}$); peptide by condition ($G_{j(i),c}$); loading, labeling and mixing differences across iTRAQ experiments ($V_{q,\ell}$); iTRAQ experiment (B_q); and the biological and experimental error not captured by the remaining terms ($H_{i,j(i),c,q,s,\ell}$).

Finally, we apply a logarithmic transformation to the collection of observed reporter ion peak areas so that Model 2 becomes additive since ANOVA models operate on an additive scale. To emphasize the relationship between factors in Model 2 and their log-transformed counterparts, we use the corresponding lower-case variable to represent the logarithm of the factor specified in Model 2. Furthermore, let $y_{i,j(i),c,q,s,\ell}$ be the log reporter ion peak area corresponding to protein i , peptide j , condition c , iTRAQ experiment q , tag ℓ , and spectrum s . On a logarithmic scale, the collection of observed reporter ion peak areas is given by the expression

$$y_{i,j(i),c,q,s,\ell} = p_i + r_{i,c} + r_c + f_{j(i)} + g_{j(i),c} + v_{q,\ell} + b_q + h_{i,j(i),c,q,s,\ell}. \quad (3)$$

Note that we now write our model as an additive equation. Here, the expression on the left hand side of Model 3 represents that which we observe (in this case, the log transformed reporter ion peak areas), generically referred to as the *response*. The terms on the right hand side (with the exception of the error term) are the effects we consider meaningful in explaining the observed variability in the response, appropriately referred to as *explanatory* or *predictor* variables. With respect to differential expression, $r_{i,c}$ is the effect of interest while those remaining can be thought of as normalization terms. The error term captures all the unexplained variability in the response.

We complete the translation to a statistical model by introducing a constant (or *intercept*), u , into Model 3 (and described in the paragraph that follows), resulting in

$$y_{i,j(i),c,q,s,\ell} = u + p_i + r_{i,c} + r_c + f_{j(i)} + g_{j(i),c} + v_{q,\ell} + b_q + h_{i,j(i),c,q,s,\ell}. \quad (4)$$

A summary of the terms contained in Model 4 and their interpretation is provided in Table 2. Additionally, we impose the following constraint on all explanatory variables in Model 4: the value of the model parameter corresponding to exactly one level of each predictor is defined as zero. We refer to this ‘zero’ level as that variable’s *reference* level. Thus, a variable with N levels requires estimation of only $N - 1$ values (or *parameters*), since the parameter corresponding to one level is fixed at zero. For example consider the variable ‘condition’ and corresponding model parameters, r_c . In our model development, we assume condition has two levels (treated and control). We set the value of r_c to zero when condition = control, and refer to the control level as the reference level for the condition variable in Model 4. This also is consistent with the development of Model 1 in which R_c equals one for the control condition (the logarithm of one is zero). All other model predictors are parameterized similarly.

What does this transformation accomplish? First, Model 4 is a re-parameterization of Model 3 with the result that the intercept is the average log reporter ion peak area for the group defined by the reference levels of all model explanatory variables, and the parameters corresponding to the remaining levels of each variable describe the change (increase or decrease) in average response relative to the reference level. (Another common re-parameterization sets the intercept equal to the overall mean of the data, and the model effects reflect deviations from the global mean attributable to each effect.) Second, by writing a single model reflecting the complete data and conducting an analysis on that data, we use the data more efficiently than an analysis that constructs ratios one at a time, as is the case with vendor software. Finally, but most important from a practical standpoint, the re-parameterization facilitates estimation by least squares, the method used to obtain estimates of the model parameters. Fortunately, all standard statistical software packages perform the described re-parameterization automatically, and the process is transparent to the analyst. As a final note, we acknowledge a slight abuse of notation in translating from Model 3 to 4. By way of example, in Model 3, p_i represents the contribution of protein i to the response, while in Model 4, p_i is the effect of protein i relative to a reference category. Nonetheless, despite this ambiguity in notation, we feel that to introduce new notation in Model 4 would further complicate unnecessarily an already complicated problem.

Model 4 is an example of a *linear* model, that is a model of the form $Y = \text{mean}(\mathbf{x}) + \text{error}$, where Y is a quantitative (i.e. measured on a continuous scale) response (log reporter ion peak area here), and the mean relates a collection of predictor variables, $\mathbf{x}$ (e.g. protein, condition, etc.),

to the response through a linear function of unknown model parameters (u , $\{p_i\}$, etc. in Model 4). It is common to assume the error follows a zero-mean normal distribution with constant variance. This implies the response, Y , is likewise normally distributed with mean = mean(x) and variance equivalent to that of the error distribution. The average response, mean(x), is non-random but variability about that mean occurs due to random fluctuations in the data induced by the error.

When the predictors are qualitative variables (e.g. condition = treated or control), the linear model is classified as an *analysis of variance* (ANOVA) model. These qualitative predictor variables are commonly called *factors* and the associated categories are the factor *levels*. Inference based on the simplest ANOVA model with only a single two-level factor reduces to the familiar Student's *t*-test. A thorough and accessible overview of ANOVA is presented in Kleinbaum et al.¹² For a more rigorous presentation, see Neter et al.¹³

Model Fitting and Ratio Estimation

The process of estimating model parameters is referred to as *model fitting*. For ANOVA Model 4, the parameters are estimated using the method of least squares. The least squares estimator of the model parameter vector is a matrix equation, the presentation and interpretation of which is beyond the scope of this manuscript. However, the interested reader can consult previously noted references for the derivation of this solution.^{12, 13} As mentioned previously, the usual assumption for the model errors is to assume they follow a zero-mean normal distribution. However, it is notable that least squares estimates are unbiased whether or not the errors are normally distributed, and interval estimates [i.e. 95% confidence intervals (CIs)] are only slightly affected by departure from normality.¹⁴ That is, parameter estimates are unaffected by deviation from this assumption, while statistical hypothesis testing is affected only slightly.

One issue that may impact estimation, however, pertains to proteins identified by a single peptide. In such cases, the protein (p_i) and peptide ($f_{j(i)}$) effects are confounded; that is to say, the model can not distinguish between the protein and the peptide contributions. Similarly, the protein by condition ($r_{i,c}$) and peptide by condition ($g_{j(i),c}$) effects are confounded. One solution to this problem removes the so called 'one-hit wonders' from the data prior to analysis, a practice we use in the example that follows.

Ratio estimation follows directly from the fitted model parameter estimates. However, the structure of the protein ratio estimator depends both on the experimental design and on the terms included in the ANOVA model. Let θ_i be the expression ratio for protein i comparing the treated relative to the control condition. On the log scale, estimation of the ratio θ_i equates to constructing the difference in average response between the treated ($c = 2$) and control ($c = 1$) conditions for the i th protein based on the fitted ANOVA model. That is,

$$\log \hat{\theta}_i = \hat{y}_{i,j(i),c=2,q,s,\ell} - \hat{y}_{i,j(i),c=1,q,s,\ell}.$$

The circumflex (or *hat* notation) indicates use of parameter estimates obtained from the least squares fit of Model 4. Model parameters common to both conditions will cancel in the difference, while those that interact with or are nested within condition are averaged. For a design that assigns conditions to all possible tags across iTRAQ experiments, the i th log ratio is given by

$$\log \hat{\theta}_i = \left(\hat{r}_{i,2} + \hat{r}_2 + \frac{1}{J_i} \sum_{j=1}^{J_i} \hat{g}_{j(i),2} \right) - \left(\hat{r}_{i,1} + \hat{r}_1 + \frac{1}{J_i} \sum_{j=1}^{J_i} \hat{g}_{j(i),1} \right), \quad (5)$$

where J_i is the number of peptides associated with the i th protein. Exponentiation of $\log \hat{\theta}_i$ provides a point estimate of θ_i .

Because an interval estimate provides a range of plausible values for the true parameter consistent with the observed data, we consider a 95% CI of θ_i to be more informative as a measure of uncertainty than the estimated standard error of $\hat{\theta}_i$. We obtain this interval estimate by first estimating the endpoints of a 95% CI for $\log \theta_i$ based on the approximate normality of the sampling distribution of $\log \hat{\theta}_i$. Exponentiation of the endpoints of that interval yield an approximate 95% CI for θ_i , given by

$$\exp \left[\log \hat{\theta}_i \pm 1.96 \times \widehat{\text{SE}}(\log \hat{\theta}_i) \right]. \quad (6)$$

We advocate interval estimation of θ_i based on Equation (6) as opposed to the more familiar $\hat{\theta}_i \pm 1.96 \times \widehat{\text{SE}}(\hat{\theta}_i)$, which produces an unrealistically symmetric interval, suffers from the possibility of yielding a negative lower bound, and is based on the unlikely assumption that the sampling distribution of $\hat{\theta}_i$ is approximately normal. The approach to interval estimation described by Equation (6) for parameters with skewed sampling distributions (e.g. odds ratio estimates based on a fitted logistic regression model, or hazard ratio estimates derived from the fit of a proportional hazard model) is common in statistical applications where the logarithm of the estimated parameter more closely follows a normal distribution than the estimated parameter itself.^{15, 16}

ABRF-like Experiment

ABRF-like Sample Preparation

Two protein mixtures (mixture A and mixture B) were prepared in house to mimic the round robin samples provided by the 2006 Proteomics Research Group (PRG) of the Association of Biomolecular Resource Facilities (ABRF).⁹ Mixtures A and B contained the same eight proteins in the following proportions (mixture A to mixture B): beta casein (bovine milk), 4:1; catalase (bovine liver), 5:1; glycogen phosphorylase (rabbit), 1:76; carbonic anhydrase I (human), 1:3; and each of peroxidase P (horseradish), ribonuclease A (bovine), albumin (bovine serum), and lactoperoxidase (bovine milk) in a 1:1 ratio. All proteins were purchased from Sigma, and used without further purification. Stock solutions of each protein were prepared in HPLC grade water, and mixed to obtain the desired amounts in mixtures A and B. Each mixture was aliquoted, dried, and stored at -20°C until further use.

One aliquot each of mixtures A and B was dissolved in 60 μL of a buffer containing 1 M urea, 20 mM Hepes, and 0.1% CHAPS at pH 8.0.

Modified iTRAQ Experiment

Four aliquots containing 14 μg of total protein (two from each of mixtures A and B) were labeled with the isobaric tag reagents for absolute and relative quantitation (iTRAQTM) [Applied Biosystems (ABI)- Framingham, MA] using the manufacturer's protocol with the following modifications in order to mimic the procedure typically used in our laboratory: all four aliquots were brought up to 43 μL volume using the sample preparation buffer, and 20 μL of 1 M tetraethyl ammonium bicarbonate (TEAB) (Fluka- Switzerland) were added to each. The aliquots were denatured, reduced, and alkylated using the chemicals supplied with the iTRAQ kit, but adjusting the volumes added to maintain the final concentrations similar to those in the standard ABI protocol. Two vials of trypsin (ABI) were diluted, and 10 μL of the combined trypsin solution were added to each aliquot. The trypsin digestion was carried out

overnight at 37°C. Two tubes of the 114, 115, 116 and 117 iTRAQ reagents were dissolved each in 70 μ L of ethanol. Both tubes of each label were combined and added to the corresponding aliquot. Aliquots from mixture A were labeled with the 114 and 116 reagents, while aliquots from mixture B were labeled with the 115 and 117 reagents. The labeled aliquots were incubated at room temperature for one hour, combined into one sample, dried down, and subjected to off-line strong cationic exchange fractionation.

Strong Cationic-Exchange Chromatography (SCX) Fractionation

The combined sample was subjected to SCX fractionation on a Waters 600-MS HPLC system connected to a Waters 484-MS UV detector. A PolySULFOETHYL A™ column (200 $\times$ 2.1 mm I. D., 5 μ m, 200 Å) (PolyLC Inc., Columbia, MD) was used. Solvent A was 10 mM KH₂PO₄, 25% acetonitrile (ACN), pH 2.7–3.0; solvent B was similar to solvent A but with the addition of 0.5 M KCl. A 40 minute gradient from 10% to 50% solvent B, followed by 40 minutes at 50% solvent B provided acceptable separation of the peptides. The flow rate was 250 μ L/minute, and the elution of peptides was monitored by UV at 220 nm. Fractions were collected at five minute intervals. The fractions were completely dried in a SpeedVac Concentrator, and stored at –20°C until further fractionation by reversed phase HPLC (RP-HPLC).

Reversed-Phase Chromatography Fractionation

Thirteen SCX fractions were selected based on the UV trace for further fractionation by RP-HPLC on an Ultimate-Switchos- Probot system (LC Packings, Sunnyvale, CA). The peptides were first loaded using the Switchos system on a C₁₈ pre-column cartridge [5 mm $\times$ 300 μ m I.D. packed with PepMap 100, 5 μ m, 100 Å - (LC Packings)] using 2% ACN, 0.1% trifluoroacetic acid (TFA) in water at 40 μ L/minute. After 20 minutes of desalting, the peptides were eluted from the pre-column onto a C₁₈ (150 mm $\times$ 100 μ m I.D., 3 μ m, 300 Å) column (Micro-Tech Scientific, Vista, CA) using the Ultimate system at 600 nL/minute. Solvent A was 2% ACN, 0.1% TFA in water; solvent B was 85% ACN, 5% 2-propanol, 0.1% TFA. A 50 minute gradient from 12% B to 41% solvent B at 600 nL/minute was used. Peptide elution was monitored at 214 nm.

MALDI Fraction Collection

The eluent from the reversed-phase HPLC separation was mixed at a 1:2 (eluent:matrix) ratio with an 8 mg/ml solution of α -cyano-4-hydroxycinnamic acid (Bruker Daltonics, Germany) in 70% ACN, 0.1% TFA, and 0.15 mg/ml ammonium citrate being continuously delivered from the syringe pump of the Probot system. The mixture was spotted on stainless steel MALDI plates (ABI) in a 20 $\times$ 20 pattern, every ten seconds during the peptide elution phase of the RP-HPLC. Six mass calibration spots were manually pipetted on the perimeter of the plate, and two mass accuracy verification spots were manually placed on the top center and bottom center of each plate. The plates were kept in the dark until MS and MS/MS were acquired.

Mass Spectrometry Analysis

MS and MS/MS analyses were performed on a 4700 Proteomics Analyzer matrix assisted laser desorption ionization (MALDI) time of flight (TOF)-TOF mass spectrometer equipped with the 4000 Series Explorer (Version 3.0) data acquisition software (Applied Biosystems). After plate calibration, alignment, and default calibration update for each MALDI plate, an MS spectrum was acquired from all 400 spots on each plate. Subsequently, the mass spectra were subjected to an interpretation method to select the 15 most intense precursors from each spot. MS/MS spectra were acquired for the selected precursors using collision induced dissociation (CID) gas.

Peptide and Protein Identification

All MS/MS spectra were searched using GPS Version 3.6 software (ABI) and MASCOT Version 2.1 (Matrix Science-London) using a protein sequence database created in house containing a total of 160,763 entries from the NCBI nr protein database for bovine, rabbit, human, and horseradish. iTRAQ labeled lysine, iTRAQ N-terminal labeling, and MMTS alkylation of cysteines were used as fixed modifications; oxidation of methionine and iTRAQ labeling of tyrosine were used as variable modifications. A minimum confidence of 95% was used for peptide identification as defined by the GPS software. The identified peptides were grouped according to their parent proteins by the GPS software.

Results

ANOVA Analysis of ABRF-like Data

Data Description and Univariate Analyses—We removed from the ABRF-like data all observations associated with the known contaminants trypsin and keratin. The remaining data contained 22 unique accession numbers. We assumed accession numbers linked to homologous proteins identified the same putative protein. (We use the word *putative* to refer to any condition specified *a priori* in the ABRF-like sample preparation protocol.) For example, we grouped peptides associated with distinct accession numbers identifying homologous forms of glycogen phosphorylase (human liver, human brain, and rabbit muscle) into a single collection identifying the same protein. [For complex proteome analyses, the GPS protein assignments (equivalently, accession numbers) would be used to group peptides, unless a grouping program or manual grouping was implemented.] This resulted in the identification of 18 unique proteins, the eight putative proteins contained in the ABRF-like sample, and an additional ten contaminants. We then removed all proteins identified by a single peptide, reducing the data further to observations associated with the eight putative proteins and only two contaminants. This data comprised 414 total (168 unique) peptide identifications, yielding $414 \times 4 = 1656$ reporter ion peak areas. Four additional observations with reporter ion peak areas equal to zero were eliminated (the logarithm of zero is undefined), resulting in 1652 reporter ion peak areas available for analysis. These zero areas corresponded to glycogen phosphorylase peptides present at a theoretical 1:76 ratio. The smaller amount of protein in mixture A produced too little signal to provide a positive reporter ion peak area after correction for isotopic overlap.

Figure 1 shows boxplots of log reporter ion peak areas for the protein by mixture (panel A), peptide by mixture (panel B), and tagging (panel C) effects. The observed distributions for the protein by mixture effects are consistent with the putative mixture A to mixture B ratios for each protein. To illustrate the peptide by mixture effect, we provide in panel B boxplots of log reporter ion peak areas for the five most abundant peptides of liver catalase classified by mixture. (Despite the data's simplicity relative to a complex biological sample, there are still too many unique peptides to present them all in Figure 1, panel B. We therefore limit presentation to a subset of the total number of unique peptides identified in the analysis.) Note the distribution of log reporter ion peak areas varies somewhat across peptides, but the mixture A to mixture B relationship within each peptide appears relatively uniform. Finally, we note there is little difference in the distribution of the response across the levels of the iTRAQ tags (panel C), indicating negligible bias attributable to differences in sample loading, tagging efficiencies or pipetting errors.

ANOVA Model Fit and Ratio Estimation—We fit the following modified version of Model 4 to the ABRF-like data:

$$y_{i,j(i),c,s,\ell(c)} = u + p_i + r_{i,c} + r_c + f_{j(i)} + v_{\ell(c)} + h_{i,j(i),c,s,\ell(c)}. \quad (7)$$

Because our data was derived from a single iTRAQ experiment, we dropped the experiment-level subscript, q , from all relevant model effects, and eliminated the experiment effect (b_q) as it is subsequently confounded with the intercept. We also dropped the peptide by mixture effect ($g_{j(i,c)}$) since, given the nature of the study (a spike-in experiment), there is no biological reason to expect the effect of mixture on log reporter ion peak area to change depending on the peptide under consideration. This view is supported empirically with the boxplots depicted in Figure 1, panel B. Additionally, we note that the iTRAQ labels are nested within mixtures [hence the subscript $\ell(c)$] due to the design of this single-experiment study; the mixture A samples were labeled only with the 114 and 116 tags, while the mixture B samples were labeled only with the 115 and 117 tags.

We constructed log A:B protein expression ratios using the estimator

$$\log \hat{\theta}_i = \left(\hat{r}_{i,A} + \hat{r}_A + \frac{1}{2} (\hat{v}_{114(A)} + \hat{v}_{116(A)}) \right) - \left(\hat{r}_{i,B} + \hat{r}_B + \frac{1}{2} (\hat{v}_{115(B)} + \hat{v}_{117(B)}) \right). \quad (8)$$

We note that Equations (5) and (8) differ somewhat due to the nesting of iTRAQ labels within condition. As mentioned previously, all condition-nested effects are averaged in the protein ratio estimates. All statistical analyses were conducted using SAS version 9.1 (Cary, NC).

The R^2 for Model 7 fit to the ABRF-like data was 0.55, indicating that 55% of the observed variation in log reporter ion peak areas is explained by the model. Table 1 summarizes the ratio estimates based on Equation (8) for the eight putative proteins in the ABRF-like sample. We also obtained ratio estimates for the two contaminant proteins, but neither ratio differed significantly from one (results not shown). With the exception of the most extreme ratio (1:76) corresponding to glycogen phosphorylase, all ratio estimates are reasonably close to their putative values. In addition, with the exception of glycogen phosphorylase, all 95% CIs contain the putative ratios.

Ratio Estimation Using GPS Software Formulas: Our ratio estimates are not directly comparable with ratios generated by GPS software, since the latter constructs ratios from peptides linked to the same accession number. Recall we elected to group peptides associated with homologous proteins with different accession numbers. We therefore constructed ratios and associated standard deviations from our peptide groupings based on the formulas stated in the GPS software users manual (Applied Biosystems). Specifically, let $X_1, \dots, X_{J_i}$ be the J_i peptide ratios associated with the i th protein. The GPS software calculates the average ratio, GM_i , for protein i as

$$GM_i = \exp \left[\sum_{j=1}^{J_i} \log(X_j) / J_i \right], \quad (9)$$

where $j = 1, \dots, J_i$ and GM_i is the geometric mean of the peptide ratios. The standard deviation of the J_i peptide ratios, SD_i , for protein i is calculated by the GPS software as

$$SD_i = \exp [\log(GM_i) + \log(sd_i)] - \exp [\log(GM_i)], \quad (10)$$

where sd_i is the standard deviation of the log peptide ratios, $\{\log(X_j)\}$.

Protein quantitation using the GPS software results in output similar to that shown in Table 1. Specifically, a denominator is selected by the user from among the four tags, peptide ratios

subsequently are constructed for each tag pair (three in all), and the corresponding protein ratios and standard deviations are calculated from these peptide ratios based on Equations (9) and (10). Typically, GPS software normalizes peptide ratios prior to protein quantitation based on the assumption that the majority of ratios are equal to one. Importantly, we did *not* utilize GPS software normalization of peptide ratios prior to producing the protein ratios presented in Table 1 (“Analysis using GPS software formulas”) since this assumption is not satisfied for the ABRF-like data.

Table 1 shows the ratios and corresponding standard deviations for the eight ABRF proteins obtained using the GPS formulas [Equations (9) and (10)] with either the 115- or 117-tagged reporter ion peak area as the denominator. We note the ANOVA ratio estimates are similar in value to the GPS ratio estimates. Quantitation based on the GPS reported measures is complicated by the fact that each of two iTRAQ tags was used for the two conditions. Assuming A:B (rather than B:A) ratios are of interest, denominator selection is completely arbitrary (mixture B samples were tagged with both 115 and 117), and there is therefore no clear choice of which pair of ratios to consider (114/115 and 116/115, or 114/117 and 116/117). Ideally, we prefer a single A:B ratio to address the question of scientific interest (“Which proteins are differentially expressed in mixtures A and B?”), but for designs such as the one presented here, it is unclear how to combine the collection of GPS-reported ratios to generate such a single summary ratio. For example, we might average the reported 114/115 and 116/115 ratios to generate a single A:B ratio, but this ignores information provided by the 117 reporter ion peak areas.

Discussion

In this paper, we derive a statistical model for the analysis of iTRAQ reporter ion peak areas that incorporates both biological and experimental sources of variability. We demonstrate this model fits within the broader class of ANOVA models, which facilitates the use of established statistical methods of estimation and inference. Our step-by-step model derivation allows us to be explicit about the sources of variation from the start, so that one is fully aware of those sources of variation being ignored or possibly grouped in other analyses of the same type of data.

An ANOVA analytic approach to iTRAQ data incorporates normalization (bias removal), relative quantitation (fold-change estimation), and uncertainty estimation (CI estimation) into a single model. As pointed out by Kerr et al.³, ANOVA integrates normalization and differential expression into a unified analysis, which is a more efficient use of the data than constructing protein ratios one at a time. Furthermore, this unified approach naturally accounts for the uncertainty associated with the normalization process, is based on testable assumptions, and correctly accounts for the loss in degrees of freedom attributable to normalization.³

A limitation of the model-based normalization presented here is that the estimation of global normalization parameters (e.g. experiment and tagging effects) is based only on reporter ion peak areas from spectra yielding peptide identification. In the case of analysis of data from a single iTRAQ experiment, one could separate normalization from ratio estimation and elect to use all the data, including that for unidentified peptides, in the normalization step. For example, the quantile matching normalization technique presented in Keshamouni et al.⁷ and Jagtap et al.⁸, or the approach implemented in the GPS Explorer software (normalizing to the median ratio constructed from data derived only from identified peptides) easily could be adapted to use all available data. However, care must be taken when adopting these approaches for data arising from multiple experiments. Wang et al.¹⁷ show how differences in overall peptide abundance across samples (as may result from variation in sample loading or protein degradation over time) can result in biased quantile estimates due to the missingness of features

across experiments for ions below the minimum instrument detection level. They use a missing-data model to estimate the probability that a peptide actually present in a sample fails to produce a detectable signal, and use this probability to impute a value for that peptide's abundance. Thus, while percentile-based normalization is internally valid for a single iTRAQ experiment, naively generalizing this approach across experiments does not work, as demonstrated by Wang et al.¹⁷ Oberg et al.¹⁰ demonstrate how our ANOVA model can be properly used for normalization of multiple iTRAQ experiments.

Keshamouni et al.⁷ propose an alternative ANOVA model for the analysis of data from a single iTRAQ experiment comparing a normal and treated condition. Prior to analysis, they perform quantile normalization to remove global tag effects, and construct peptide ratios from the average of duplicate measures across channels. While averaging replicates to produce a single measure is commonplace, this simple preprocessing step results in a loss of information about variability. For each protein, they then model the logarithm of associated protein-specific peptide ratios as a function of a protein main effect (equal to the intercept), a peptide-specific random effect, and an error term. In contrast, the response variable for our model is a simple transformation of the raw data, that is, the logarithm of reporter ion peak areas, thereby eliminating the need to average duplicates prior to analysis while preserving and utilizing variability information.

Additionally, our model differs from that proposed by Keshamouni et al.⁷ by treating peptide as a fixed rather than a random effect. Factors for which model-based conclusions pertain only to those levels included in the study (e.g. condition = treated or control) are called *fixed* effects. Alternatively, when factor levels can be thought of as a random sample from some underlying distribution, such factors are called *random* effects. In the case of iTRAQ experiments, it is reasonable to consider the observed peptides as arising from a population of peptides: peptides are not specified *a priori*, and their identification depends on the search engine used as well as instrument-level detection.¹⁷ We fit our model with peptide included as a random rather than fixed effect, but found no differences in protein ratio point or interval estimates (results not shown).

In fitting separate models for each protein, Keshamouni et al.⁷ assume both the variance of the peptide random effect and the error variance are protein-specific rather than global parameters. In the microarray literature, there are arguments for and against using gene-specific error estimates. For example, Jain et al.¹⁸ state that extremely large outlying expression values can result in poorly powered tests of differential expression, while chance occurrences of genes with similar expression patterns across replicates can lead to increased false positive rates. To mitigate these potential problems¹⁰, they propose the use of a locally pooled error estimate for their gene-specific differential expression analyses, while Tusher et al.¹⁹ add a constant to all error estimates in their popular 'Statistical Analysis of Microarrays' method. For proteomics applications, analysis of proteins one at a time may alleviate some of the computational burden, in particular when combining data across iTRAQ experiments. For example, Oberg et al.¹⁰ use the hybrid approach of Wolfinger et al.²⁰ that performs normalization using an ANOVA model fit to the complete data estimating the global normalization effects, and one-at-a-time protein-level analysis on the residuals (observed minus fitted values) from that model to estimate differential expression effects. We graphically assessed the distribution of residuals across proteins from our fitted model, and found little evidence to support protein-specific error estimation for our data.

A significant advantage of an ANOVA-based analysis is the ease with which changes in experimental design are incorporated into the model, while leaving unchanged the machinery for relative protein quantitation and associated inference. A well-constructed ANOVA model will always be able to address questions of scientific interest. In contrast, standard iTRAQ

protein quantitation software will only provide 'tag' ratios which may or may not coincide with the ratios of interest. Even for simple experimental designs, the available software does not combine information across iTRAQ experiments to produce a single protein-level expression ratio. As it is not possible to accommodate all potential experimental designs in one package, current software understandably provides limited protein quantitation. Oberg et al.¹⁰ discuss the application of ANOVA models for more complicated experimental designs, particularly, studies that utilize multiple experiments.

That an ANOVA model can accommodate data from multiple experiments is a substantial analytic improvement for iTRAQ-based studies, resulting in increased power and subsequently enhancing the researcher's ability to identify potential biomarkers. By analyzing the complete data within a single model, ANOVA produces one summary ratio and one interval estimate per protein. In contrast, current quantitation software forces the researcher to construct multiple ratios and standard deviations for the same protein identified in multiple iTRAQ experiments within the same study, but the manner in which to combine these ratios to obtain a single summary and uncertainty measure is unclear.

In Table 1, we intentionally report only ratio estimates and 95% CIs for each putative protein. It is also possible to conduct formal inference within the ANOVA framework and obtain a p-value for each protein corresponding to the hypothesis that the associated ratio differs significantly from one. We chose, however, to emphasize point estimates and associated uncertainty measures since we find these to be the more biologically meaningful quantities.

With respect to outlying observations, we make no formal recommendation as to their treatment, but feel their identification should be based on model residuals rather than raw data: a raw data value deemed outlying may in fact be well-predicted by the fitted model. We note that logarithmic transformation of the data can reduce the effects of large outlying values, lending further support for the use of a logarithmic scale for analysis of iTRAQ data. Furthermore, outliers can provide clues as to important sources of variability that have been overlooked, thereby guiding model-development as well as indicating factors to control in the experimental process (see Oberg et al.¹⁰ Figure 2A). It is our position that outliers should be discarded only if their values are truly erroneous such as might arise from machine malfunction or technician error. Otherwise, the impact of the outlying values on protein ratio estimation can be assessed directly by fitting the model with and without the offending values. We direct the reader to Kutner et al.¹³ for a thorough discussion of outlier analysis in the context of ANOVA models.

A related point pertains to the treatment of peak intensities that reach saturation. Although we make no attempt to identify such values in our data, in general, ignoring their influence potentially results in an increased false negative rate. That is to say, peak intensities that reach saturation will contribute to protein ratio estimates attenuated toward the null value of 1. To deal with the small number of spectra showing saturation of the reporter ion areas, additional data could be obtained modifying the acquisition parameters to avoid saturation. It is our experience however that the opposite problem poses greater concern, specifically, obtaining sufficient information to achieve both relative quantitation and identification.

As with any model-based approach, one is limited by the nearly universal fact that no model can incorporate every source of variability. For example, our model does not explicitly include terms for fraction-specific variation, variations across batches of reagent, instrument-specific variation, differences in labeling efficiency for different peptides, differences in digestion efficiency across samples, or variations in peak intensities due to protein degradation, to name a few. To illustrate the influence of fraction-specific conditions on observed reporter ion peak intensities, Figure 2 shows MS spectra and corresponding reporter ion peak areas for the same

precursor ion identifying the same peptide from two different RP HPLC fractions. The reporter ion peak intensities corresponding to the precursor shown in panel (A) were between 800,000 and 1,000,000, while those associated with the precursor shown in panel (B) were all approximately equal to 45,000. Differences in fraction composition induce variability in a peptide's propensity for ionization. Furthermore, differences in a peptide's concentration across fractions contribute to variability in MS precursor ion intensity measures and subsequent reporter ion peak areas. Figure 2 shows striking differences in intensities for the same precursor ion and differences in ionized species between the two spectra. When we fit our model including a fraction effect, we observed negligible change in protein ratio point estimates, narrower interval estimates, but an increase in the model R^2 from 0.55 to 0.84. Variation in the observed reporter ion peak areas due to these unmodeled sources of variability contributes uncertainty to estimates of the parameters of interest (protein and peptide-level by condition effects). These limitations, however, do not invalidate the model.

We note that our analysis is limited to a simple mixture of eight proteins while most iTRAQ analyses will be conducted on complex biological samples. Given the didactic nature of the present manuscript, we felt the aims of the paper justified the relative simplicity of the data. Furthermore, by knowing in advance the pre-specified protein ratios, we were able to assess the ability of the model to capture the 'truth', although what protein ratios were specified in the protocol and what actually happened in sample preparation are possibly not the same. At the Medical University of South Carolina, we have successfully used ANOVA to analyze iTRAQ data collected from complex samples and from studies comprising multiple iTRAQ experiments. We elected not to include such an analysis in the current manuscript as our focus was on model development and proof of principle in an application where 'truth' is known. We direct the interested reader to the paper by Oberg et al.¹⁰ for an application highlighting the use of ANOVA models to analyze data from a case study composed of multiple iTRAQ experiments.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

EGH was partially supported by NIH grant number NIDCR K25 DE016863. JHS, SCW and KLS were partially supported by NIH grant number NHLBI NO1-HV-28181. EHS gratefully acknowledges partial support from the MUSC Proteomics Center, NSF grant DMS-0604666, the NIH/NCRR COBRE project P20 RR017696, and NIH grant R01DE016353. ALO and TMT greatly appreciate funding from the David Woods Kemper Memorial Foundation. ALO further acknowledges the University of Minnesota Biomedical Informatics and Computational Biology Program for their generous support.

References

1. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlett-Jones M, He F, Jacobson A, Pappin DJ. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* 2004;3(12):1154–69. [PubMed: 15385600]
2. Aggarwal K, Choe LH, Lee KH. Quantitative analysis of protein expression using amine-specific isobaric tags in *Escherichia coli* cells expressing *rhsA* elements. *Proteomics* 2005;5(9):2297–308. [PubMed: 15887184]
3. Kerr MK, Martin M, Churchill GA. Analysis of variance for gene expression microarray data. *J Comput Biol* 2000;7(6):819–37. [PubMed: 11382364]
4. Seshi B. An integrated approach to mapping the proteome of the human bone marrow stromal cell. *Proteomics* 2006;6(19):5169–82. [PubMed: 16947122]

5. Salim K, Kehoe L, Minkoff MS, Bilsland JG, Munoz-Sanjuan I, Guest PC. Identification of differentiating neural progenitor cell markers using shotgun isobaric tagging mass spectrometry. *Stem Cells Dev* 2006;15(3):461–70. [PubMed: 16846381]
6. Unwin RD, Pierce A, Watson RB, Sternberg DW, Whetton AD. Quantitative proteomic analysis using isobaric protein tags enables rapid comparison of changes in transcript and protein levels in transformed cells. *Mol Cell Proteomics* 2005;4(7):924–35. [PubMed: 15849271]
7. Keshamouni VG, Michailidis G, Grasso CS, Anthwal S, Strahler JR, Walker A, Arenberg DA, Reddy RC, Akulapalli S, Thannickal VJ, Standiford TJ, Andrews PC, Omenn GS. Differential protein expression profiling by iTRAQ-2DLC-MS/MS of lung cancer cells undergoing epithelial-mesenchymal transition reveals a migratory/invasive phenotype. *J Proteome Res* 2006;5(5):1143–54. [PubMed: 16674103]
8. Jagtap P, Michailidis G, Zielke R, Walker AK, Patel N, Strahler JR, Driks A, Andrews PC, Maddock JR. Early events of *Bacillus anthracis* germination identified by time-course quantitative proteomics. *Proteomics* 2006;6(19):5199–211. [PubMed: 16927434]
9. Turck CW, Falick AM, Kowalak JA, Lane WS, Lilley KS, Phinney BS, Weintraub ST, Witkowska HE, Yates NA. The Association of Biomolecular Resource Facilities Proteomics Research Group 2006 Study: Relative Protein Quantitation. *Mol Cell Proteomics* 2007;6:1291–1298. [PubMed: 17513294]
10. Oberg AL, Mahoney DW, Eckel-Passow JE, Malone CJ, Wolfinger RD, Hill EG, Cooper LT, Onuma OK, Spiro C, Therneau TM, Bergen Iii HR. Statistical analysis of relative labeled mass spectrometry data from complex samples using ANOVA. *J Proteome Res* 2008;7(1):225–33. [PubMed: 18173221]
11. Gan CS, Chong PK, Pham TK, Wright PC. Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ). *J Proteome Res* 2007;6(2):821–7. [PubMed: 17269738]
12. Kleinbaum, D.; Kupper, LL.; Muller, KE.; Nizam, A. *Applied Regression Analysis and Multivariable Methods*. Vol. Third ed.. Duxbury Press; Pacific Grove: 1998.
13. Kutner, MH.; Nachtsheim, CJ.; Neter, J.; Li, W. *Applied Linear Statistical Models*. Vol. Fifth ed.. McGraw-Hill; Boston: 2005.
14. Kutner, MH.; Nachtsheim, CJ.; Neter, J.; Li, W. In *Applied Linear Statistical Models*. McGraw-Hill; Boston: 2005. p. 793-794.
15. Hosmer, DW.; Lemeshow, S. *Applied Logistic Regression*. John Wiley & Sons; New York: 1989. p. 44
16. Collett, D. *Modelling Survival Data in Medical Research*. Chapman & Hall; London: 1994. p. 68
17. Wang P, Tang H, Zhang H, Whiteaker J, Paulovich AG, McIntosh M. Normalization regarding non-random missing values in high-throughput mass spectrometry data. *Pac Symp Biocomput* 2006;315–26. [PubMed: 17094249]
18. Jain N, Thatte J, Braciale T, Ley K, O'Connell M, Lee JK. Local-pooled error test for identifying differentially expressed genes with a small number of replicated microarrays. *Bioinformatics* 2003;19(15):1945–51. [PubMed: 14555628]
19. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 2001;98(9):5116–21. [PubMed: 11309499]
20. Wolfinger RD, Gibson G, Wolfinger ED, Bennett L, Hamadeh H, Bushel P, Afshari C, Paules RS. Assessing gene significance from cDNA microarray expression data via mixed models. *J Comput Biol* 2001;8(6):625–37. [PubMed: 11747616]

Figure 1.

Graphical summaries of ABRF-like data. Panels A through C contain boxplots of log reporter ion peak areas across the levels of: (A) the cross-classification of the eight putative proteins by mixture (A versus B); (B) the cross-classification of the five most abundant liver catalase peptides by mixture; and (C) iTRAQ tags 114 – 117. The middle line in each box is the median data value, and the left and right box extremes are the 25th and 75th percentiles, respectively. The boxplot ‘whiskers’ extend to extreme data values not considered outliers, and asterisks indicate potential outlying values.

Figure 2. MS spectra for the same precursor ion (1302.824 Da) from two different RP HPLC fractions, illustrating the effect of fraction on variability in observed reporter ion peak areas. The inset in each panel shows the reporter ion peak areas obtained from the corresponding tandem mass spectra. Certain features of the spectra are enhanced to facilitate visualization.

Table 1

Comparison of protein quantitation for ABRF-like iTRAQ data based on ANOVA model versus GPS software formulas.

Protein	No. unique peptides (total no. peptides)	Putative ^a Ratio (A:B)	ANOVA analysis Est. Ratio (A:B) (95% CI)	Analysis using GPS software formulas	
				114/115 mean (SD)	114/117 mean (SD) 116/117 mean (SD)
Beta casein	3 (7)	4 (4:1)	3.19 (1.44, 7.09)	3.34 (0.62) 3.24 (0.86)	3.15 (0.88) 3.05 (1.03)
Liver catalase	31 (100)	5 (5:1)	4.18 (3.38, 5.16)	4.80 (3.51) 4.04 (2.85)	4.32 (2.00) 3.63 (2.33)
Glycogen phosphorylase	43 (76)	0.013 (1:76)	0.074 (0.058, 0.095)	0.067 (0.13) 0.068 (0.14)	0.072 (0.12) 0.079 (0.11)
Carbonic anhydrase	18 (41)	0.33 (1:3)	0.43 (0.31, 0.60)	0.49 (0.18) 0.40 (0.17)	0.46 (0.13) 0.37 (0.17)
Horseradish peroxidase	6 (23)	1 (1:1)	1.43 (0.92, 2.22)	1.66 (0.46) 1.46 (0.23)	1.39 (0.35) 1.23 (0.22)
Ribonuclease A	3 (5)	1 (1:1)	0.78 (0.30, 2.01)	0.72 (0.25) 0.77 (0.13)	0.80 (0.23) 0.86 (0.13)
Bovine serum albumin	57 (151)	1 (1:1)	1.12 (0.94, 1.33)	1.27 (0.49) 1.04 (0.23)	1.21 (0.28) 0.99 (0.37)
Lactoperoxidase	3 (7)	1 (1:1)	1.00 (0.45, 2.22)	0.95 (0.15) 1.03 (0.16)	0.96 (0.10) 1.05 (0.18)

^aRatio specified by mixing protocol.

Table 2

Summary of Model 4 terms and their interpretations. Indices represent: protein (i); condition (c); peptide (j); experiment (q); tag (l); and spectrum (s).

Model terms	Interpretation
$y_{ij(i),c,q,s,l}$	Log reporter ion peak area
u	Model intercept
p_i	Protein effect
r_c	Condition effect
$r_{i,c}$	Protein by condition effect
$f_{j(i)}$	Peptide effect
$g_{j(i),c}$	Peptide by condition effect
b_q	Experiment effect
$v_{q,l}$	Experiment by tag effect
$h_{i,j(i),c,q,s,l}$	Biologic and experimental error