

# Protein Significance Analysis in Selected Reaction Monitoring (SRM) Measurements\*

Ching-Yun Chang<sup>‡</sup>, Paola Picotti<sup>§</sup>, Ruth Hüttenhain<sup>‡‡</sup>, Viola Heinzlmann-Schwarz<sup>¶</sup>, Marko Jovanovic<sup>\*\*</sup>, Ruedi Aebersold<sup>‡‡§§</sup>, and Olga Vitek<sup>‡¶¶¶</sup>

Selected reaction monitoring (SRM) is a targeted mass spectrometry technique that provides sensitive and accurate protein detection and quantification in complex biological mixtures. Statistical and computational tools are essential for the design and analysis of SRM experiments, particularly in studies with large sample throughput. Currently, most such tools focus on the selection of optimized transitions and on processing signals from SRM assays. Little attention is devoted to protein significance analysis, which combines the quantitative measurements for a protein across isotopic labels, peptides, charge states, transitions, samples, and conditions, and detects proteins that change in abundance between conditions while controlling the false discovery rate. We propose a statistical modeling framework for protein significance analysis. It is based on linear mixed-effects models and is applicable to most experimental designs for both isotope label-based and label-free SRM workflows. We illustrate the utility of the framework in two studies: one with a group comparison experimental design and the other with a time course experimental design. We further verify the accuracy of the framework in two controlled data sets, one from the NCI-CPTAC reproducibility investigation and the other from an in-house spike-in study. The proposed framework is sensitive and specific, produces accurate results in broad experimental circumstances, and helps to optimally design future SRM experiments. The statistical framework is implemented in an open-source R-based software package SRMstats, and can be used by researchers with a limited statistics background as a stand-alone tool or in integration with the existing computational pipelines. *Molecular & Cellular Proteomics* 11: 10.1074/mcp.M111.014662, 1–12, 2012.

From the <sup>‡</sup>Department of Statistics, Purdue University, West Lafayette, Indiana; <sup>§</sup>Department of Biology, Institute of Molecular Systems Biology, ETH Zürich, Switzerland; <sup>¶</sup>Translational Research Group, University Hospital Zürich, Switzerland; <sup>||</sup>Faculty of Medicine, Prince of Wales Clinical School, UNSW 2052, Australia; <sup>\*\*</sup>Institute of Molecular Life Science, University of Zürich, Switzerland; <sup>‡‡</sup>Competence Center for Systems Physiology and Metabolic Diseases, ETH Zürich, Switzerland; <sup>§§</sup>Faculty of Science, University of Zürich, Switzerland; <sup>¶¶</sup>Department of Computer Science, Purdue University, West Lafayette, Indiana

Received October 5, 2011, and in revised form, December 14, 2011

Published, MCP Papers in Press, December 14, 2011, DOI 10.1074/mcp.M111.014662

Selected reaction monitoring (SRM)<sup>1</sup> is a mass spectrometry technique that can accurately and reproducibly quantify proteins in complex biological mixtures (1, 2, 3). It can cover a nearly complete dynamic range of abundance of cellular proteome, with a lower boundary of detection below 50 copies per cell for single cellular organisms (3). Considerable efforts are currently invested into developing high-throughput SRM assays, even for whole proteomes (4, 5). These assays are then used to simultaneously quantify hundreds of proteins with a high degree of reproducibility across multiple samples, and as a result the assays are increasingly used in systems biology and in clinical investigations (3, 6, 7, 8).

SRM experiments quantify *a priori* known protein species. They require knowledge of the peptides of these proteins that are unique to the target proteins and can be observed by a mass spectrometer (9, 10), and of the mass spectrometric characteristics of these peptides such as fragment ion mass, signal intensity distribution, and optimal collision energy (2). Enzymatically digested proteins are subjected to liquid chromatography separation and are monitored in a triple quadrupole mass spectrometer, and the ion signals for an *a priori* selected set of fragment ions are recorded over chromatographic time (11, 12). The intensities of precursor/fragment ion pairs of a peptide, called *transitions*, are then used as measurements of protein abundance. Label-based workflows further enhance the accuracy of the quantification by spiking an isotopically labeled reference version of each target peptide into the samples and then compare the relative intensities of the *endogenous* and the *reference* transitions.

Many SRM experiments aim at *class comparison*, *i.e.* comparing protein abundance across conditions or time points of interests. Statistical and computational tools are essential for this task. In addition to increasing sample throughput, the tools allow us to conduct and interpret the experiment in an objective and reproducible fashion. A typical SRM bioinformatics analysis workflow is overviewed in Fig. 1A. It includes assay development, signal processing, and protein significance analysis. Currently, most available bioinformatics tools focus on the first two steps in Fig. 1A. A variety of stand-alone software packages such as MRMAid, MRM worksheet and others reviewed in (13) help to automatically select transition

<sup>1</sup> The abbreviations used are: SRM, selected reaction monitoring.

lists, schedule retention times, optimize collision energy, etc. Other stand-alone tools, e.g. MultiQuant (Applied Biosystems/MDS Sciex) and Pinpoint (Thermo Scientific), quantify and visualize the acquired signals, and AuDIT (14) and mProphet (15) detect and filter out poor quality transitions. Alternatively, comprehensive pipelines such as Skyline (16) and ATAQS (17) integrate the transition design and the signal processing steps in user-friendly workflows. Therefore, the upstream setup of SRM measurements and the analysis of the raw primary data to quantify the targeted peptides are well supported by the available software tools.

At the same time, there is no generally accepted approach to significance analysis of the identified and quantified proteins. We define *protein significance analysis* as a procedure that appropriately combines the quantitative measurements for a targeted protein across isotopic labels, peptides, charge states, transitions, samples, and conditions, and detects proteins that change in abundance between conditions more systematically than as expected by random chance, while controlling the false discovery rate (18). Significance analysis is challenging, in part, because of the natural biological variation of protein abundance and the experimental variation in sample handling (19). Signal processing also contributes to the variation when individual transitions are missed, misidentified or imprecisely quantified (14). Statistical inference allows us to make objective conclusions in such situations, however the development of statistical methods for protein significance analysis in SRM experiments has not received sufficient attention as of yet.

Many investigations perform significance analysis using simple statistical methods such as the two-sample *t* test, which compares the abundance of all the transitions from one condition to another. Such tests take as input intensities of the individual transitions of the protein (or their averages within a run) in label-free experiments, or ratios of the endogenous and reference transitions in label-based experiments. In this manuscript we argue that more accurate conclusions can be obtained by a more detailed probabilistic modeling. We propose a general and flexible statistical framework for SRM experiments, which is schematically illustrated in Fig. 1B. The framework consists of the following components: (a) definition of the biological populations of interest and of the desired scope of conclusions, (b) exploratory data analysis to control the quality of MS runs, (c) joint representation of the quantitative measurements of the protein using a flexible linear mixed-effects model, and model-based determination of proteins that change in abundance from one condition to another, and (d) statistical design of future follow-up experiments.

The proposed framework is implemented in an open source R-based software package SRMstats, and can be used stand-alone or as a module integrated with comprehensive pipelines such as Skyline and ATAQS. It is implemented for both label-free and label-based SRM workflows.

## EXPERIMENTAL PROCEDURES

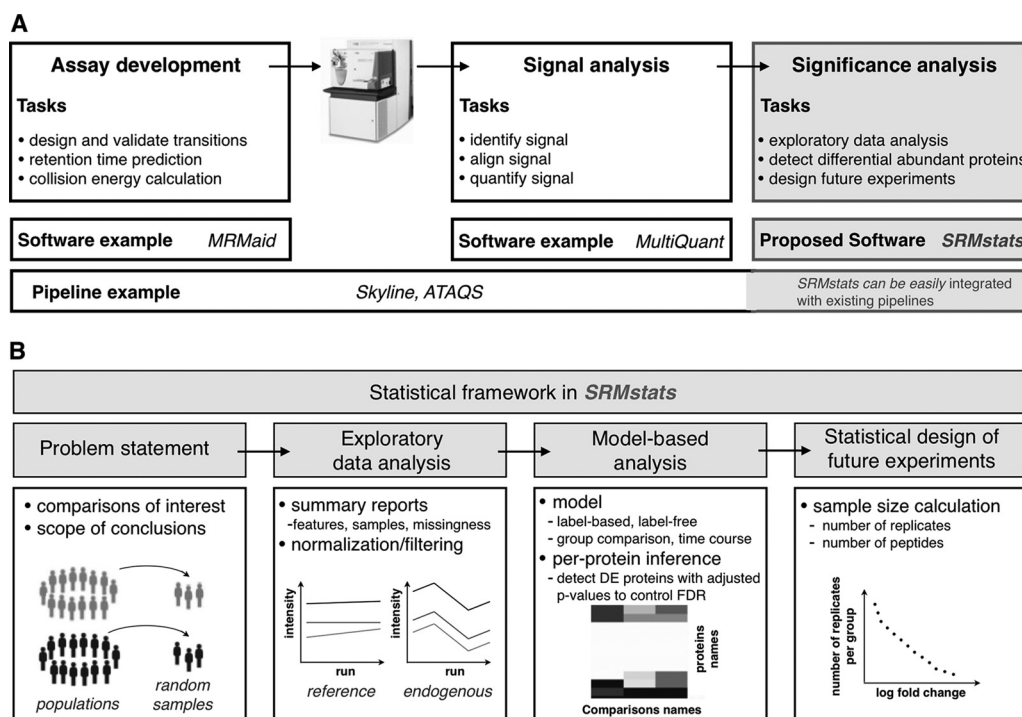
**Controlled Spike-In Experimental Data Sets**—We evaluated the statistical framework using two controlled spike-in experimental data sets. The first was generated by a multilaboratory investigation of the Clinical Proteomic Technology Assessment for Cancer network of the National Cancer Institute (NCI-CPTAC), described in detail in (19). The original goal of this investigation was to assess reproducibility, recovery, linear dynamic range, limits of detection, and quantification of SRM assays. Here we used the data sets to evaluate the ability of significance analysis to detect known changes in protein concentration.

Briefly, the investigation targeted seven proteins, each was represented by up to three peptides and each peptide by up to three transitions. The proteins were spiked into a complex background in a series of nine concentrations, and heavy isotope-labeled synthetic peptides (20) were used as internal standards. The spike-in samples were prepared according to three mixing and digestion protocols (called Study I, II, and III). Study III was subjected to sources of technical variation closest to a real-life experiment. The samples were shipped to eight participating sites, which independently performed SRM analyses to detect and quantify the transitions. We took as input the tabulated transition abundances as quantified by each participating site, and evaluated the ability of significance analysis to detect known fold changes, separately for each study and site.

The second controlled data set was an in-house spike-in experiment (supplemental Material Sec. 1.1). To evaluate the sensitivity of significance analysis, six proteins were spiked into a complex background in varying concentrations according to a Latin Square design (21). The design is advantageous in that it allows us to evaluate the ability to detect a range of fold changes over a series of proteins, baseline concentrations, and replicate runs, while limiting the number of mixtures. To evaluate the specificity of the significance analysis, six additional proteins were spiked into the background in constant concentrations. Each protein was represented by two peptides and each peptide by up to three transitions. Heavy isotope-labeled synthetic peptides (20) were used as references, and each mixture was profiled in two mass spectrometry runs. We evaluated the ability of significance analysis to detect known fold changes, as well as the extent of false positive changes among proteins spiked in constant concentrations.

**Biological Investigations**—We tested the statistical framework using example data sets from two biological investigations. The first studied plasma samples from six patients with epithelial ovarian cancer and ten healthy controls in a group comparison experimental design (supplemental Material Sec. 1.2). Each specimen was analyzed in a single mass spectrometry run. 14 *N*-glycosylated proteins were selected based on prior evidence of differential abundance from the literature and profiled as potential diagnostic ovarian cancer biomarkers. Heavy isotope-labeled reference peptides (20) were spiked into the samples as internal references, and up to three transitions were measured for each endogenous and heavy-labeled peptide.

The second example is the study of central carbon metabolism of *Saccharomyces cerevisiae* described in detail in (3) (supplemental Material Sec. 1.3). The experiment targeted 45 proteins in the glycolysis/gluconeogenesis/TCA cycle/glyoxylate cycle network in yeast, which spans the range of protein abundance from less than 128 to 10E6 copies per cell. Unlike the previous example this study had a time course experimental design. Three biological replicates were analyzed at ten time points (T1-T10), while the cells transitioned through exponential growth in a glucose-rich medium (T1-T4), diauxic shift (T5-T6), post-diauxic phase (T7-T9), and stationary phase (T10). Prior to trypsinization, the samples were mixed with an equal amount of proteins from a common <sup>15</sup>N-labeled yeast sample, which was used as a reference. Each sample was profiled in a single mass spectrom-



**FIG. 1. Overview of the computational approaches.** A, Typical computational steps of design and analysis of SRM experiments. The contribution of this manuscript and the proposed software are colored in gray. B, Details of the proposed statistical framework in A. *Problem statement*: definition of the biological populations of interest, and of the desired scope of conclusions with respect to the biological and technical MS run variation. *Exploratory data analysis*: visualization, normalization and quality control of the intensities of reference and endogenous transitions. *Model-based analysis*: summarization of the raw intensities (*i.e.* not ratios of reference and endogenous transitions) in a linear mixed-effects model, model-based testing for differential abundance, and adjustment of the *p* values to control for the False Discovery Rate (FDR). *Statistical design of future experiments*: comparison of label-based and label-free workflows, and calculation of the number of biological and technical replicates, and of peptides and transition, in a future follow-up experiment.

etry run, where each protein was represented by up to two peptides and each peptide by up to three transitions. Transcriptional activity under the same experimental conditions has been previously investigated in (22). Genes coding for 29 out of the 45 targeted proteins were found differentially expressed between conditions similar to those represented by T7 and T1 in this proteomic study and are used in this manuscript for external validation.

**Synthetic Data Sets**—Additional simulated data sets were generated to demonstrate the performance of the proposed significance analysis in a variety of circumstances, such as in the presence of missing transitions, relatively large between-run variation, and the presence/absence of technical replication of a same biological sample. Pseudocode of the basic algorithm used to generate these synthetic data sets is provided in (supplemental Material Sec. 4.1).

## RESULTS

In the following sections we detail each step of the proposed statistical framework for protein significance analysis in SRM measurements (Fig. 1B). To detect differences in protein abundance with high sensitivity and specificity, the key step is to select an appropriate statistical model, *i.e.* STEP 3 (model-based analysis) of the proposed framework. Prior to the statistical modeling, it is necessary to examine the properties of the data as demonstrated in STEP 1 (problem statement) and STEP 2 (exploratory data analysis). Finally, STEP 4 uses the model and the data to plan the design of future experiments.

**STEP 1 Problem Statement**—The important first step in designing and analyzing an SRM experiment is to determine the comparisons of interest, the experimental design, and the scope of the conclusions of the study, in order to select an appropriate statistical model for the data. In this manuscript we refer to this step as *problem statement*.

**Determine the Comparisons of Interest and the Experimental Design**—First, we define the comparisons of interests, *i.e.* the conditions that will be compared in the study. For example in the ovarian cancer study the only comparison of interest was between the mean protein abundance of subjects with the disease and the controls; in the yeast metabolism study researchers could compare the mean protein abundances in many pairs of time points. In the following we focus on comparing the time points 1 and 7.

Next, we distinguish the experimental designs that are group comparison and time course. Group comparison studies acquire measurements on distinct individuals in each condition. For example the ovarian cancer study had a group comparison design. In contrast, time course studies acquire measurements on each biological source repeatedly at several conditions or time points. The repeated nature of the experiment allows us to profile changes in protein abundance



for each individual over time. The yeast metabolism study had the time course design. Different statistical models will be required for these two experimental designs, and we will discuss this in the following section.

**Determine the Desired Scope of Validity of the Conclusions**—An important aspect of the problem statement is the specification of the desired scope of validity of the conclusions, *i.e.* the scope to which the conclusions from the analysis will be valid, with respect to the biological and technical MS run variation. As an illustration, Fig. 3A shows a hypothetical protein measured with two subjects per condition (Healthy and Disease) and three endogenous transitions per subject in large biological variation scenario. The black curves represent the distribution of protein abundances in the populations of subjects that are of interest to the investigation. The solid dots are the protein abundances of the subjects selected in the study. The black circles are the endogenous transitions measured for this protein and this subject.

Given the observed intensities of the transitions, two comparisons can be potentially of interest. The first compares average protein abundances in the selected subjects (comparing two stars), while taking into account the noise and the variation in the quantified transitions. In other words, we restrict the scope of our conclusions to the selected subjects. The second comparison is between the average protein abundances in the underlying populations (comparing two squares), while taking into account both the variation because of random selection of the subjects and the variation in the transitions. In other words, we expand the scope of our conclusions to the entire population, beyond the selected subjects.

Specification of the scope of the conclusions has two important implications for the results. First, the conclusions of the two comparisons may disagree. As seen in Fig. 3A, when the variation of protein abundance in the population is large, the abundance in the selected subjects can differ between the groups even though the abundance of the underlying populations is the same between groups. The discrepancy is particularly frequent in studies with a small number of subjects, where the subjects do not adequately represent the full diversity of the underlying population. Second, the reduced scope of the conclusions only takes into account the measurement error, while the expanded scope of the conclusions considers both the measurement error and the biological variation. Because the reduced scope accounts for less variation, it is typically easier to find changes in abundance among the selected subjects. Fig. 3B shows that these discrepancies disappear in proteins that are highly regulated and have a small biological variation.

These considerations provide practical guidelines. Because many SRM experiments are exploratory screens with a small sample size, it often helps to restrict the conclusions to the specific biological replicates in the experiment. The restricted scope of the conclusions helps us gain statistical power of the

initial screen and generate initial scientific hypotheses, even though some of these hypotheses may not be confirmed in the populations. On the other hand, expanding the scope of the conclusions to the entire populations is appropriate for confirmatory investigations.

A similar distinction between the restricted and expanded scope of the conclusions applies to the technical variation, *i.e.* variation introduced by the mass spectrometry technology. It may sometimes be meaningful to restrict the scope of the conclusions to the performed mass spectrometry runs. But more frequently we must view the replicate MS runs of a same biological sample as random instances that represent the variability of sample handling and spectral acquisition, and expand our conclusions to fully account for this variation. This distinction is made regardless of whether the experiment acquires technical replicates of a same biological sample.

For the example data sets, both the ovarian cancer study and the yeast metabolism study were initial screening experiments, and therefore we restricted the scope of our conclusions to the selected biological replicates. At the same time, we fully accounted for the technical variation associated with all possible MS runs, and expanded the scope of the conclusions with respect to the technical variation.

The scope of the conclusions should be defined prior to conducting the experiment, according to the purpose of the study. It is reflected in the statistical model in STEP 3. Further discussion of this topic can be found, *e.g.* in (21, 23, 24, 25). The models proposed in this manuscript and implemented in SRMstats enable a flexible representation of both experimental designs, and of the scope of validity of the conclusions.

**STEP 2 Exploratory Data Analysis**—The second step of the proposed statistical framework is to use exploratory analysis plots to visualize the biological and the technical MS run variation in the data. Understanding the properties of the data further helps to select the most appropriate statistical model.

The quantified transitions can be influenced by the technical MS run variability, *e.g.* because of changes in the instrument performance over a large number of MS runs. This technical variation can be partly removed by a global between-run normalization, *e.g.* by constant normalization that equalizes the median peak intensities of all reference transitions between runs. The normalization is based on the assumption that the intensities of the reference transitions are stable across all MS runs.

Exploratory analysis plots in [supplemental Fig. S2](#) help determine whether there is systematic bias because of technical MS run variability and whether this bias should be addressed by global normalization. For example, the plots for the ovarian cancer study ([supplemental Fig. S2b](#)) showed small variability of reference signal intensities across MS runs, and a global normalization had almost no effect. Alternatively, in the in-house spike-in study the intensities of the reference signals varied substantially between runs, and this was primarily because of sample evaporation in the second replicate

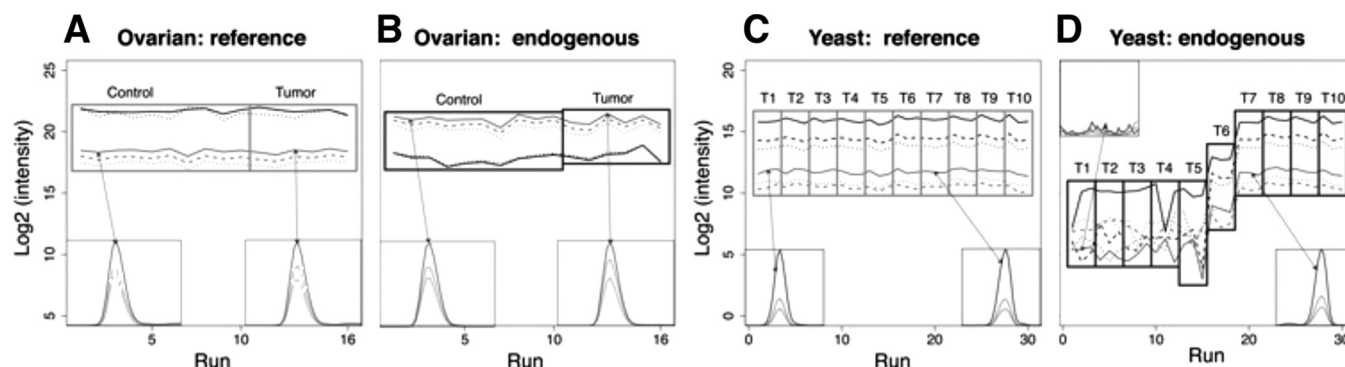


FIG. 2. **Quantitative profiles of reference and endogenous transitions in label-based SRM experiments.** A, B, Group comparison study of ovarian cancer, protein CLU. C, D, Time course study of *S. cerevisiae*, protein IDP2. x-axis: run. y-axis: log-intensities of transitions. Line types indicate transitions, and colors indicate peptides. Insets: selected chromatographic peaks of a peptide. In D, peptides are undetectable at early time points.

set (supplemental Fig. S2a). Global normalization corrected for this systematic difference in signals (supplemental Fig. S2d), and we show below that it lead to meaningful results despite the variation.

Fig. 2 shows quantified transition signals of an example protein from each of the biological investigations, after a global normalization. As can be seen, transitions of the reference peptides had a roughly constant abundance between runs. In contrast, transitions of the endogenous peptides indicated systematic differences between conditions, and these changes are of primary interest. The measurements were subject to biological variation of protein abundance and technical between-run variation. Some transitions were heterogenous, *i.e.* transitions from a same protein produced contradictory evidence of abundance within a run because of, *e.g.* co-eluting interfering peptides and misidentified or misquantified transitions, and this effect is particularly strong for low-abundant proteins. The plots in Fig. 2 can help evaluate the extent of transitions with interferences and the extent of missing values. The plots can also complement computational tools such as AuDIT and mProphet in detecting low quality transitions that should be removed.

For example, protein CLU from the ovarian cancer study did not contain transitions with interferences and missing transitions. In the case of yeast metabolism study, protein IDP2 contained heterogenous transitions particularly in early time points, but no missing transitions. Such properties of the data set are addressed in the following statistical modeling. The plots and the normalization are obtained automatically with SRMstats.

#### STEP 3 Model-based Analysis—

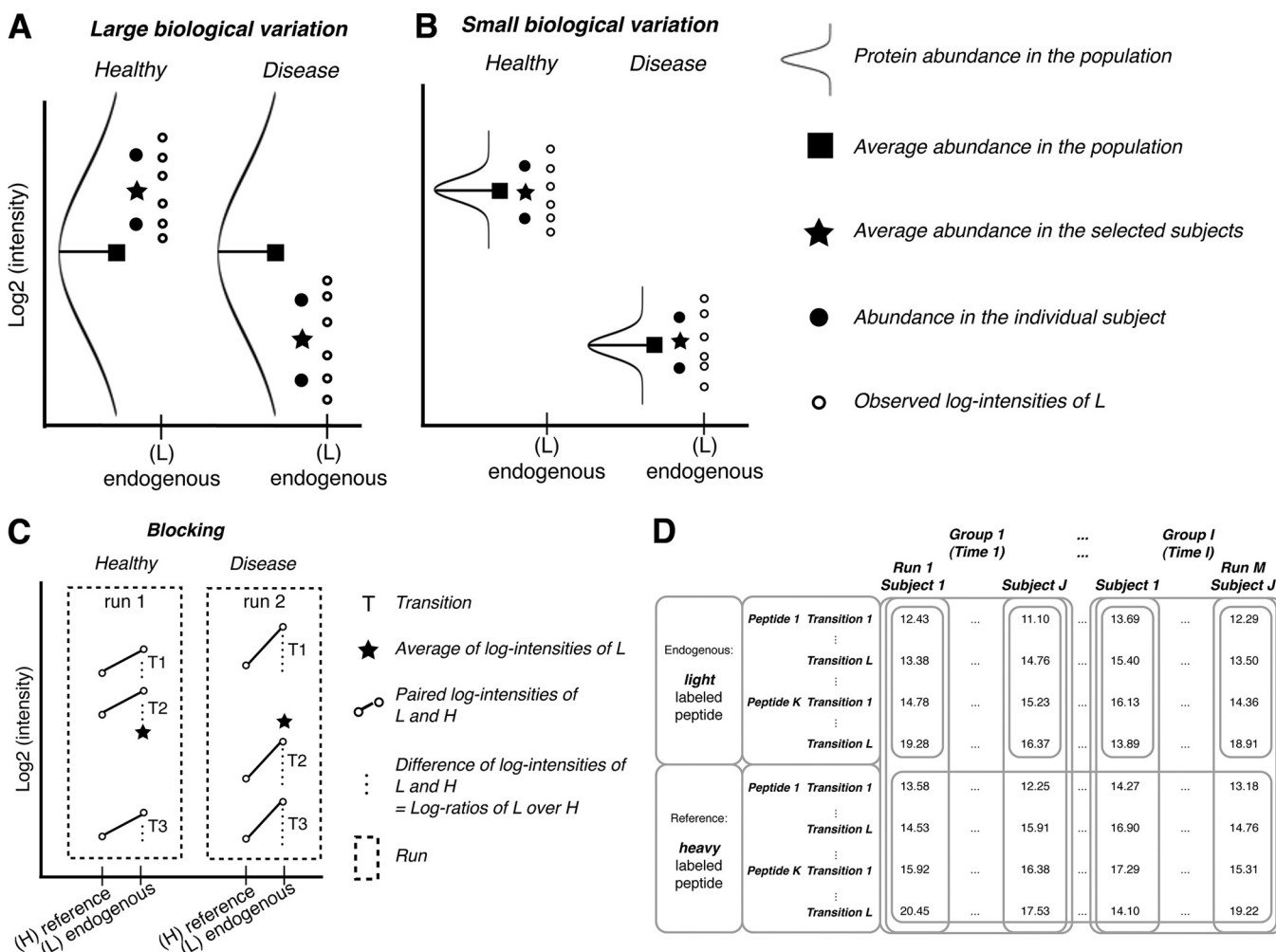
**Linear Models Account for Transition-specific Variation Without Calculating Ratios of Endogenous and Reference Transitions**—We propose a linear mixed-effects model that represents all the quantitative measurements for a targeted protein, in order to detect systematic changes in protein abundance between conditions with high sensitivity and

specificity at a controlled false discovery rate. A distinctive feature of the approach is that, for label-based experiments, we do not calculate the ratios of the endogenous and the reference transitions within a run, but model the individual intensities directly. Although we do not calculate the ratios, the intensity-based approach reflects the pairing between the endogenous and the reference transitions, and accounts for the transition-specific variation that was not removed by global normalization.

To illustrate the advantages of the proposed intensity-based approach, in this subsection we describe the model in a simplified data set, and illustrate the connection between the intensity-based and the ratio-based models in this simple situation. In the next subsection we extend this example to complex real-life data sets, and introduce a family of full intensity-based probability models.

Global normalization cannot remove transition-specific nuisance variation that arises, *e.g.* from variation in peptide ionization in the presence or absence of other peptides in the sample. Fig. 3C illustrates this using a hypothetical protein that was measured with three transitions in samples from two different conditions acquired in two MS runs. In the figure, intensities of the endogenous transitions are affected by both condition-specific variation (of interest) and transition-specific variation (nuisance). In particular, the intensity of the second endogenous transition in the disease group decreased, while the intensities of the other transitions increased. A label-free experiment would compare the averages of the endogenous transitions (stars), and would not distinguish the systematic change from nuisance variation. Label-based workflows consider deviations of the endogenous transitions from their references (dashed vertical lines). These deviations are larger in the disease run than in the control, pointing to a systematic change in protein abundance. Therefore, label-based experiments are more efficient in the presence of transition-specific variation.

Label-based workflows account for this transition-specific variation by spiking reference peptides in constant concen-



**FIG. 3. Data representation in SRM experiments.** A, Illustration of the scope of validity of our conclusions with respect to the biological variation. The reduced scope of conclusions compares the average protein abundance between conditions in the subjects selected for the study (stars). The expanded scope of conclusions compares the average protein abundances in the populations (squares). The two conclusions can disagree when the biological variation in the populations is large. B, As in A, but with small biological variation. There is little difference between the two scopes of conclusions in this case. C, Illustration of blocking. The label-free experiment compares the averages of the endogenous transitions (stars). The label-based workflow considers deviations of the endogenous transitions from their references (dashed vertical lines), and is more sensitive at detecting the change. D, Measurements from a protein for a group comparison or a time course label-based SRM experiment. Entries in the table are log-intensities of transitions, they are grouped according to shared sources of variation.

tration. The variation in the reference transitions was because of the technical factors alone, and we detect the effect of conditions on the protein abundance by studying deviations of the endogenous transitions from their references. In the example of Fig. 3C, these deviations were larger in the disease sample than in the control sample, pointing to a systematic change in abundance. Therefore, the label-based experiment was more efficient in the presence of transition-specific variation.

In statistical terminology, the pairing of endogenous and reference transitions within a run is called *blocking* (24). Blocking aims at controlling the known sources of variation in the data, and is encountered in many areas of statistical experimental design (21) as a tool for improving efficiency. To reflect the beneficial effect of blocking, we model the log-

intensities of the transitions directly instead of the log-ratios, and introduce a dedicated model term *Run*, which accounts for the shared run membership of the endogenous and reference transition pairs in one MS run:

$$\text{LogIntensity}_{ijk} = \text{OverallMean} + \text{Condition}_i + \text{Run}_k + \text{Error}_{ijk}, \text{ where } i = \{\text{Healthy}, \text{Disease}, \text{Reference}\}, j = \{T1, T2, T3\}, k = \{\text{Run1}, \text{Run2}\}, \sum_i \text{Condition}_i = 0, \sum_k \text{Run}_k = 0, \text{Error}_{ijk} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma_{\text{LogIntensity}}^2) \quad (1)$$

In contrast to the proposed approach, currently researchers often perform the *t* test on the log-ratios of the endogenous and the reference transition pairs measured in one MS run. The log-ratios are equivalent to the differences of log-intensities (dashed lines in Fig. 3C). For the simplistic example in Fig. 3C, the model underlying the *t* test is:

$\text{LogRatio}_{ij} = \text{OverallMean} + \text{Condition}_i + \text{Error}_{ij}$ , where

$i = \{\text{Healthy}, \text{Disease}\}, j = \{T1, T2, T3\}$ ,

$$\sum_i \text{Condition}_i = 0, \text{Error}_{ijk} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma_{\text{LogRatio}}^2) \quad (2)$$

In the special case of balanced data sets (*i.e.* data sets where the protein has a same number of quantified measurements across runs), the intensity-based model in Eq. (1) is equivalent to the ratio-based model in Eq. (2) as it yields identical protein-level conclusions. However in other situations the intensity-based model in Eq. (1) has a number of advantages. The next section utilizes this insight, extends the simplified version of the intensity-based model in Eq. (1) to realistic data sets, and discusses the advantages of this approach.

**Linear Models Reflect the Structure of the Data and the Scope of the Conclusions**—To accurately detect proteins that change in abundance between conditions, we need to obtain (a) an accurate estimate of protein abundances in each condition, and (b) an accurate assessment of variation associated with these estimates. A statistical model is essential for these tasks. In this section we introduce the proposed linear mixed-effects statistical modeling framework, and illustrate the choices of the details of the models that best represent the properties of the data.

The potential sources of variation of SRM measurements can be illustrated in a typical data set generated by a label-based experiment, *e.g.* a Table in Fig. 3D produced by detecting, quantifying and normalizing the intensities of transitions. Entries in the table share sources of variation from *Label*, distinguishing endogenous and reference transitions; *Feature*, reflecting the intensities of the signal of each peptide/transition pair; *Group*, reflecting the conditions that may induce systematic changes in protein abundance; and *Run*, reflecting potential changes in signal intensities between MS runs. Finally, *Subject* reflects the natural variation in protein abundance between biological samples. Because the same synthetic peptides are used as references in all runs, they can be viewed as another instance of *Subject*. This is similar to the analysis of experiments with reference designs in cDNA microarrays (24, 26), however, the opportunity for such models in SRM experiments has so far been unrecognized.

To express these sources of variation in the data, we developed a family of linear mixed-effects models shown in Fig. 4 and (supplemental Material Sec. 2.1). The family distinguishes group comparison and time course experiments, and includes the term *Run* that normalizes each endogenous transition with respect to its reference. Statistical interactions represent between-group and between-run heterogeneity of the features. The interaction terms can be omitted when exploratory analysis points toward homogenous transitions to avoid overfitting.

The family encompasses four distinct models, which have the same terms but differ in the scope of the validity of the conclusions attributed to *Subject* and *Run*. The restricted scope of biological interpretation is supported by the model

that views *Subject* as *fixed* effect, and the expanded scope is supported by the model that views *Subject* as *random* effect. The restricted scope of technical MS run variation is supported by the model where *Run* is *fixed* effect, and the expanded scope is supported by *Run* viewed as *random* effect.

To understand the relationship between the proposed models and the models for log-ratios, we derived the models for log-ratios that are special cases of Fig. 4 (supplemental Material Secs. 2.2 to 2.4). We theoretically demonstrated (supplemental Material Sec. 2.5) that in balanced data sets the model in Fig. 4 with *fixed Run* yields the same protein-level conclusions as the model for log-ratios. In other words, the use of the ratios implies the reduced scope of technical MS run variation. Furthermore, the model in Fig. 4 with *fixed Run* and *random Subject* yields the same protein-level conclusions as the *t* test with averaged transition ratios. In other words, the *t* test with averaged transition ratios implies the expanded scope of biological variation.

In comparison to the ratio-based models, the intensity-based approach proposed in Fig. 4 has multiple advantages. First, it can represent a broader class of problem statements. In particular, it can specify the model that simultaneously reduces the scope of biological variation (*fixed Subject*) and expands the scope of technical MS run variation (*random Run*). This is arguably the most interesting specification in exploratory SRM screens with a small sample size, and we show theoretically in (supplemental Material Sec. 1.5) that this specification also yields the highest sensitivity. However, this is impossible to achieve with the ratio-based approaches, because of the assumptions implied by the calculations of the ratios and the averages.

Second, the intensity-based models utilize transitions that miss their endogenous or their reference counterparts, thereby making the most complete use of the data. Finally, the intensity-based models are naturally extended to experiments with more than two isotopic labels, and to label-free experiments. In particular, we show below how the models enable a direct comparison between label-free and label-based experimental designs.

To summarize, there are multiple choices in statistical modeling that need to be made to account for the properties of the data and the desired scope of conclusions. These choices are once again overviewed in Fig. 5. The same statistical model choice is used for all proteins in the study.

For example, in the ovarian cancer study the model was specified to reflect the group comparison design, the restricted scope of biological variation and the expanded scope of technical MS run variation. Unlike protein CLU, which shows the evidence of homogenous transitions (Figs. 2A, 2B), the majority of the proteins in the ovarian cancer study have transitions with some interference. Therefore, the model was specified with the interaction model terms. In the yeast metabolism study the model was specified to reflect the time course design, the restricted scope of biological variation, the



Observed  
log2(int of peak)

=

Overall  
mean

+

Group  
or time

+

Subject+Feature+Run+

Time by  
subject

+

Group by  
feature

+

Run by  
feature

+

Random  
error

(a) Group comparison:

$y_{ijklm}$

=

$\mu$

+

$G_i$

+

$S(G)_{j(i)}$

+

$F_{kl}$

+

$R_m$

+

$(G \times F)_{ikl} + (R \times F)_{klm} + \varepsilon_{ijklm}$

Subject: F; Run: F

$\sum_{i=0}^I G_i = 0$

$\sum_{j=0}^J S(G)_{j(i)} = 0$

$\sum_{k,l=1}^{KL} F_{kl} = 0$

$\sum_{m=1}^M R_m = 0$

$\sum_{i=0}^I \sum_{k,l=1}^{KL} (G \times F)_{ikl} = 0$

$\sum_{m=1}^M \sum_{k,l=1}^{KL} (R \times F)_{klm} = 0$

$\varepsilon_{ijklm} \stackrel{iid}{\sim} N(0, \sigma^2)$

Subject: R; Run: F

$\varepsilon_{ijklm} \stackrel{iid}{\sim} N(0, \sigma_S^2)$

Subject: F; Run: R

$\varepsilon_{ijklm} \stackrel{iid}{\sim} N(0, \sigma_R^2)$

$\varepsilon_{ijklm} \stackrel{iid}{\sim} N(0, \sigma_{RF}^2)$

Subject: R; Run: R

$\varepsilon_{ijklm} \stackrel{iid}{\sim} N(0, \sigma_S^2)$

$\varepsilon_{ijklm} \stackrel{iid}{\sim} N(0, \sigma_R^2)$

$\varepsilon_{ijklm} \stackrel{iid}{\sim} N(0, \sigma_{RF}^2)$

(b) Time course:

$y_{ijklm}$

=

$\mu$

+

$T_i$

+

$S_j$

+

$F_{kl}$

+

$R_m$

+

$(T \times S)_{ij} + (T \times F)_{ikl} + (R \times F)_{klm} + \varepsilon_{ijklm}$

Subject: F; Run: F

$\sum_{i=0}^I T_i = 0$

$\sum_{j=0}^J S_j = 0$

$\sum_{k,l=1}^{KL} F_{kl} = 0$

$\sum_{m=1}^M R_m = 0$

$\sum_{i=0}^I \sum_{j=0}^J (T \times S)_{ij} = 0$

$\sum_{i=0}^I \sum_{k,l=1}^{KL} (T \times F)_{ikl} = 0$

$\sum_{m=1}^M \sum_{k,l=1}^{KL} (R \times F)_{klm} = 0$

$\varepsilon_{ijklm} \stackrel{iid}{\sim} N(0, \sigma^2)$

Subject: R; Run: F

$\varepsilon_{ijklm} \stackrel{iid}{\sim} N(0, \sigma_S^2)$

$\varepsilon_{ijklm} \stackrel{iid}{\sim} N(0, \sigma_{TS}^2)$

Subject: F; Run: R

$\varepsilon_{ijklm} \stackrel{iid}{\sim} N(0, \sigma_R^2)$

$\varepsilon_{ijklm} \stackrel{iid}{\sim} N(0, \sigma_{RF}^2)$

Subject: R; Run: R

$\varepsilon_{ijklm} \stackrel{iid}{\sim} N(0, \sigma_S^2)$

$\varepsilon_{ijklm} \stackrel{iid}{\sim} N(0, \sigma_R^2)$

$\varepsilon_{ijklm} \stackrel{iid}{\sim} N(0, \sigma_{TS}^2)$

$\varepsilon_{ijklm} \stackrel{iid}{\sim} N(0, \sigma_{RF}^2)$

FIG. 4. **Proposed family of intensity-based linear mixed-effects model for a label-based SRM experiment.**  $y_{ijklm}$  denotes the log-intensities of the observed transitions from treatment *Group* or time point  $i = 0, \dots, I$ , biological replicate *Subject*  $j = 0, \dots, J$ , *Peptide*  $k = 1, \dots, K$ , *Transition*  $l = 1, \dots, L$ , *Run*  $m = 1, \dots, M$  of a protein. Index 0 refers to the reference transitions. Model terms are specified as fixed effects subjected to zero-sum constraints. Random terms follow Normal distributions with mean 0 and their corresponding variances. The scope of biological variation can be restricted (*Subject* as fixed effect, denoted *F*), or expanded (*Subject* as random effect, denoted *R*). The scope of technical MS run variation can be restricted (*Run* as fixed effect, denoted *F*), or expanded (*Run* as random effect, denoted *R*). A separate model is specified for each protein. (a) Group comparison experiment. (b) Time course experiment.

FIG. 5. **Summary of modeling choices associated with the proposed intensity-based linear mixed-effects models for SRM experiments.**

Modeling aspects	Design		Aim of research		Technical variation		Quantified transitions	
	Group comparison	Time course	Screening	Validation	Consider performed MS runs	Consider all future MS runs	Frequent interference	Rare interference
Motivation								
Model statement	Time x Subject excluded	Time x Subject included	Subject fixed effect	Subject random effect	Run fixed effect	Run random effect	Group x Feature and/or Run x Feature included	Group x Feature and/or Run x Feature excluded
STEP 1: Problem statement					STEPS 2 and 3			

expanded scope of technical MS run variation, and the presence of interferences in the quantified transitions. Only the proposed model in Fig. 4 allowed the flexibility to address all these circumstances. All the models proposed in Fig. 4 are implemented in SRMstats.

*Linear Models for Protein Significance Analysis are not Equivalent to Linear Models for Reproducibility*—The term “linear mixed-effects models” refers to a very general class of

models that extends well beyond the scope of this manuscript. Different research purposes might lead to different specification of linear mixed-effects models. In this section we discuss how the linear mixed-effects models for protein significance analysis in Fig. 4 differ from the linear mixed-effects models for assessing reproducibility.

Recently Xia *et al.* (27) applied linear mixed-effects modeling to the NCI-CPTAC data set, to assess the reproduc-



ibility of SRM experiments. The goal of reproducibility analysis is to quantify the components of *technical variation*, and find the steps of the experimental protocol that contribute most to the variation. Accordingly, the models in Xia *et al.* partitioned the variation in the data into the contributions of the studies, sites, peptide groups, transitions and their interactions, and expanded the scope of the conclusions with respect to peptides and transitions by treating these terms as *random*. The primary output of this approach were the quantified variances, and their relative magnitude rank. The models were customized to the NCI-CPTAC data set, and different models would be needed for another experiment with a different design.

In contrast, the goal of protein significance analysis in this manuscript is to compare distinct *biological populations* (or specific subjects from these populations) for a pre-defined set of proteins and transitions, in an experiment typically conducted in a single laboratory or site. As the result, the models in Fig. 4 do not contain the terms related to studies and sites, and account for the prior selection of peptides and transitions by treating them as *fixed*. The primary output is a list of proteins that are differentially abundant across conditions, at a controlled false discovery rate. Therefore, although both approaches utilize the same general class of linear mixed-effects models, they are quite different in nature. The flexibility of the experimental design of the NCI-CPTAC data set allowed us to use it to benchmark the performance of protein significance analysis, as illustrated below.

**The Proposed Modeling Framework is Sensitive and Specific**—We compared the sensitivity and the specificity of significance analysis with the proposed approach as implemented in SRMstats to the *t* test in the two controlled investigations, and the two biological investigations. SRMstats used as input log-intensities of individual reference and endogenous transitions. Because both studies were exploratory screens, we restricted the scope of biological variation to the selected samples (*i.e. fixed Subject*), while expanding the scope of technical MS run variation (*i.e. random Run*). The *t* test used as input log-ratios of endogenous and reference transitions. Both SRMstats and the *t* test compared the same two pre-specified conditions, while adjusting the *p* values to control the false discovery rate.

Fig. 6A summarizes the results for the noisiest study (Study III) of the NCI-CPTAC data set. We selected a series of known changes in protein abundance from a low and a high baselines, and in low and high fold changes. Consistent with the goal of protein significance analysis, the results are obtained separately for each study and each site. Figs. 6B, 6C summarize the results for the in-house spike-in data set. In addition to the sensitivity, the experimental design also allowed us to evaluate the specificity of the tests. Overall, SRMstats outperformed the *t* test in sensitivity of detecting true known changes in abundance for all concentration baselines without a large loss of specificity.

In the two biological investigations, SRMstats also strongly outperformed the *t* test in sensitivity. In the investigation of ovarian cancer, the 14 selected proteins had prior evidence of differential abundance from the literature, and SRMstats yielded a maximal sensitivity for these proteins (Fig. 7A). In the investigation of yeast metabolism genes coding for 29 proteins were previously found differentially expressed between conditions similar to time points T7 and T1, and SRMstats produced the highest agreement between the proteomic and the transcriptomic investigations (Fig. 7B). Negative controls were not measured in both studies, which precludes any claim regarding specificity and false positives in these examples. More details on the performance of the models are provided in (Supplemental Material Sec. 3).

**The Intensity-based Linear Models Produce Accurate Results in Most Broad Circumstances**—To further investigate the advantage of the proposed intensity-based models over linear mixed models that utilize log-ratios of endogenous and reference transitions (supplemental Material Sec. 2.4), this section empirically compares these models using synthetic data sets.

First, we empirically showed (supplemental Material Sec. 4.2) that expanding the scope of technical MS run variation to all possible mass spectrometry runs (*i.e. random Run*) gained sensitivity, especially in experiments with low technical MS run variation. This specification is impossible when modeling the log-ratios of endogenous and reference transitions. The reduced scope of biological variation, when it is appropriate, further gained sensitivity. This specification was impossible with the ratio-based *t* test. Second, in presence of missing values, the intensity-based model produced the smallest bias and the smallest variance of estimates of fold changes (Fig. 8A and (supplemental Material Sec. 4.3)).

Third, in label-based workflows, intensity-based models allowed us to distinguish the extent of biological and technical MS run variation even without technical replicated mass spectrometry runs from a same biological sample. Because the same reference peptides were spiked to all samples in constant concentrations, and their transitions were only subject to technical MS run variation, they could be used to separate the extent of technical MS run variation from the biological variation. (Supplemental Material Sec. 4.4) illustrates the accuracy of this estimation using a synthetic data set.

Overall, the intensity-based linear model produced accurate results for the most broad range of data sets, and is therefore recommended for practical applications and implemented in SRMstats.

**STEP 4 Statistical Design of Future Experiments**—While label-based workflows yield accurate quantification, label-free workflows are cheaper and higher in throughput. The proposed statistical framework helps choose between these two workflows when planning a new experiment. We extended the model in Fig. 4 to represent label-free SRM experiments (supplemental Material Sec. 5.1), and this allowed

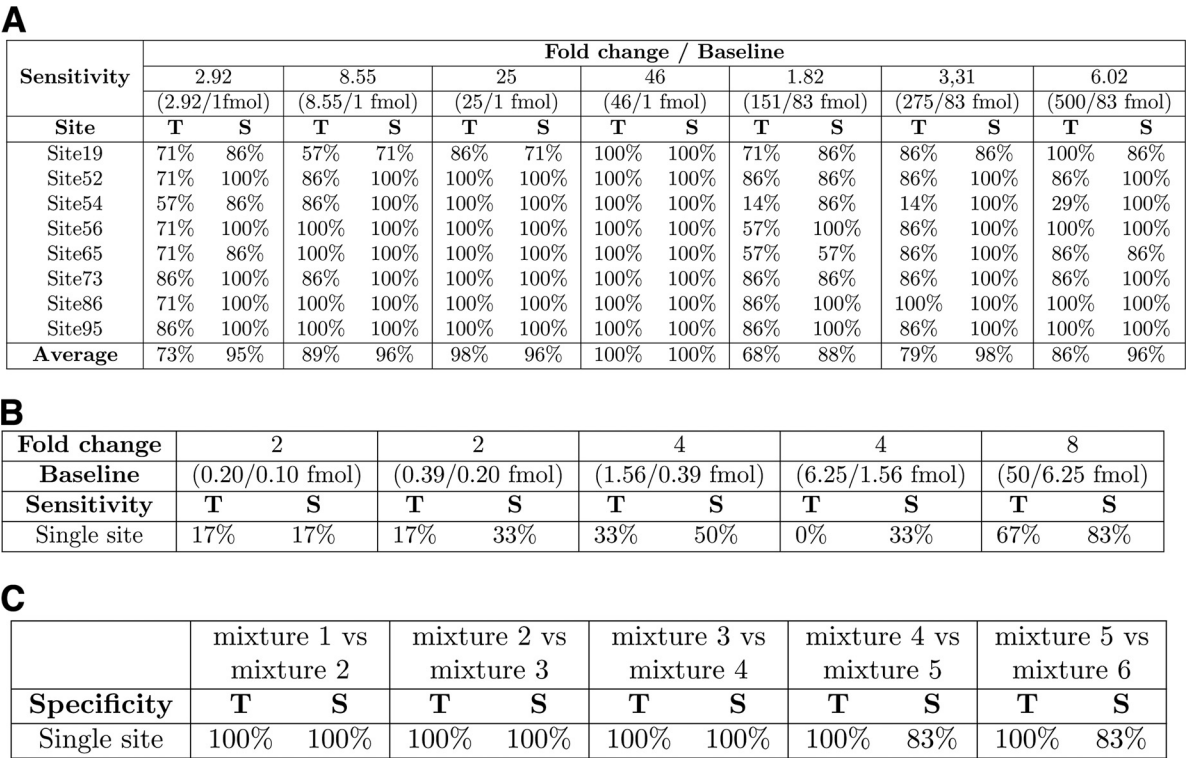


FIG. 6. Sensitivity and specificity of protein significance analysis in two controlled data sets, at the FDR cutoff 0.05. T indicates the *t* test, S indicates the implementation of the intensity-based linear model in SRMstats. A, NCI-CPTAC investigation, Study III (*i.e.* the study with the noisiest background). Each cell represents the sensitivity summarized over seven proteins. B, C, in-house controlled data set. Each cell represents the sensitivity (B) and specificity (C) summarized over six proteins. The in-house investigation has a smaller number of replicates, and a larger between-run variation, as compared with the NCI-CPTAC data set, and as the result the sensitivity for this data set is lower. Overall, SRMstats consistently gains sensitivity as compared with the *t* test, without a large loss of specificity.

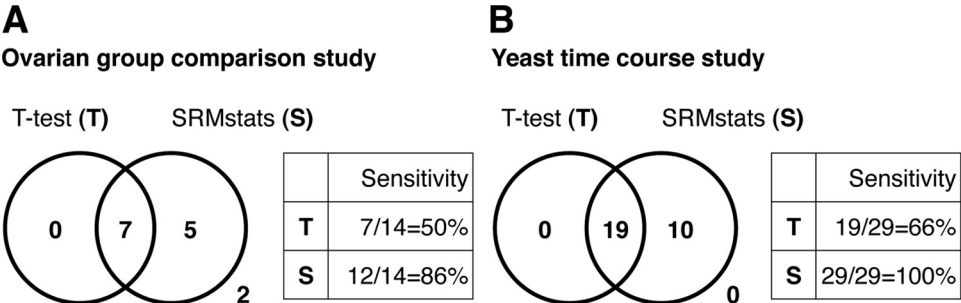
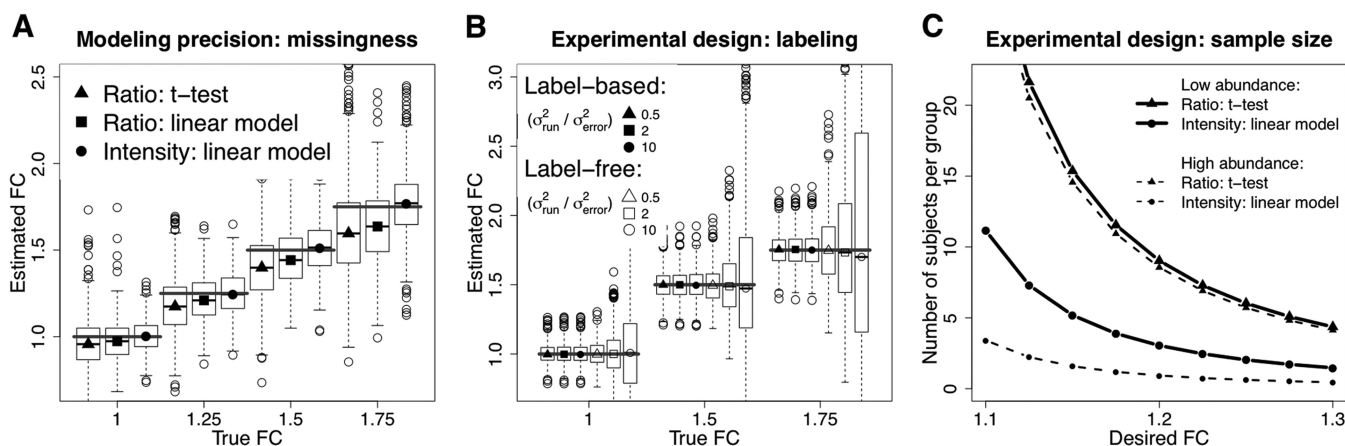


FIG. 7. Number of differentially abundant proteins in the two experimental data sets at the FDR cutoff 0.05. T indicates the *t* test, S indicates the implementation of the intensity-based linear model in SRMstats. Outside the circles are numbers of proteins not detected as differentially abundant by any of the methods. A, Comparison of healthy and disease groups in the investigation of ovarian cancer. B, Comparison of time points T1 and T7 in the investigation of yeast metabolism for the subset of proteins with a prior evidence of differential gene expression. SRMstats has better sensitivity as compared with the *t* test.

us to compare the accuracy of the two workflows using a computer simulation. The results of one such simulation are summarized in Fig. 8B. The label-based workflow consistently yielded accurate estimates of changes in abundance. With relatively small between-run variation both workflows had similar results, and the label-free workflow could be viewed as a viable alternative to the label-based experiments. However, the label-free workflow became less accurate when the between-run variation increased. Such calculations can be per-

formed in any lab, using lab-specific estimates of variation, to evaluate the expected performance of the label-free or label-based workflows.

The proposed framework also allowed us to calculate the required sample size for a future investigation (Fig. 8C and (supplemental Material Sec. 5.2)). The choice of a data analysis strategy had the strongest impact on the sample size, and linear modeling allowed us to substantially reduce the required number of biological replicates as compared with the *t* test. In



**FIG. 8. Simulation-based evaluation of modeling and of experimental designs.** A, Accuracy of the estimation of fold changes (FC) in the presence of heterogeneous transitions and with 25% of peaks missing preferentially at low signal. x-axis: true fold change. y-axis: model-based estimates of fold change. The horizontal line indicates the true fold change. B, Accuracy of the estimation of fold changes in label-free and label-based experiments for three types of between-run variation. The experimental data sets in this manuscript have  $\sigma_{run}^2 / \sigma_{error}^2 \approx 2$ . C, Minimal number of biological replicates per group required to detect a desired fold change for a low- and a high-abundant protein. See [supplementary Section 5.2](#) for details.

addition, larger fold changes and the changes among relatively high-abundant proteins required fewer biological replicates. Similar calculations can determine the number of technical replicates for a same biological sample and the number of peptides and transitions per protein for each experimental system. These calculations are implemented in SRMstats, open-development system R.

**Implementation**—We developed a software package SRMstats, which implements the proposed framework in the open-source (28). The package takes quantified endogenous and reference transitions as inputs, and provides functionalities for data visualization, quality control, model-based protein significance analysis, and experimental planning for both label-based and label-free SRM experiments. The package can be used by researchers with limited statistics background, as a stand-alone tool or in integration with the existing analysis pipelines such as Skyline or ATAQS. We also developed detailed examples of fitting the proposed models in the commercial software system SAS (SAS Institute Inc., Cary, NC), which is frequently used in clinical investigations. The SAS examples can be useful to researchers with statistics expertise. The implementations, examples of their use, and three experimental data sets are available at [www.stat.purdue.edu/~ovitek/Software.html](http://www.stat.purdue.edu/~ovitek/Software.html).

#### DISCUSSION

We proposed a general statistical modeling framework for protein significance analysis for SRM-based quantitative proteomics. It is applicable to a variety of experimental designs and to both label-based and label-free workflows. It strongly outperforms the naïve method underlying the *t* test, and yields accurate results in the most broad type of experimental circumstances, such as in the presence of missing values. The framework allows us to plan subsequent experiments, in particular select between a label-free and a label-based experiment, and

choose the appropriate number of biological and technical replicates. It is implemented in the open-source R-based package SRMstats, which can be used stand-alone or integrated into pipelines, such as Skyline or ATAQS.

The proposed framework can be extended beyond the discussion above. For example, it can represent experimental designs with even more complex structures, such as factorial or cross-over investigations. In addition to testing, the models can be used to quantify the abundance of proteins in individual samples (29) for use with clustering or machine learning algorithms. Straightforward extensions can also be made to specify heterogeneity of biological or technical variance components across features or conditions. Robust estimation (30) can help account for potential outliers, and Empirical Bayes estimation (31) can compensate for a small sample size. Finally, separate customized statistical models can be specified for all proteins in the study. At the same time, the flexibility of the proposed framework should not be misused. In particular, comparisons of interest, the scope of conclusions regarding *Subject* and *Run*, and model terms (illustrated in STEP 1: problem statement of the proposed statistical framework) should be specified prior to collecting the data based on the goals of the experiment as opposed to maximizing the sensitivity of the results.

A potential limitation of the proposed framework is the assumption that all the peptides and transitions are correctly mapped to the underlying protein. This limitation can be addressed by assigning weights to individual transitions according to the confidence in their quantification (14, 15). In the future, this drawback will be partly alleviated by improved experimental optimization of protein-specific SRM assays.

**Acknowledgments**—We thank Michael O. Hengartner (University of Zurich, Switzerland) for the contribution of resources and thank Mrs. Meena Choi for help with the implementation of SRMstats.



\* The project was funded in part by SystemsX.ch, the Swiss initiative for systems biology by the European Research Council (grant #ERC-2008-AdG 233226), by the Swiss National Science Foundation (grant # 31003A-130530), by SNF (grant # 320000-120543), by OncoSuisse, Cancer League Canton Zurich and Cancer Institute NSW (grant #09/CRF/2-02), by the Purdue Research Foundation (grant # 203202-SIO-8000029887) and by the NSF CAREER (grant #DBI-1054826). M.J. was supported by a grant from the Research Foundation of the University of Zurich, the University of Zurich Research Priority Program in Systems Biology/Functional Genomics and by a fellowship from the Roche Research Foundation.

§ This article contains [supplemental Material](#).

||| To whom correspondence should be addressed: 250 N. University Street, West Lafayette, IN 47907. ovitek@stat.purdue.edu.

**AUTHOR CONTRIBUTIONS:** C.-Y.C. developed and implemented the statistical analysis framework, designed the controlled experiment, analyzed the data sets, and wrote the manuscript. P.P. designed and conducted the investigation of the yeast metabolism, and designed the controlled experiment. R.H. designed and conducted the investigation of ovarian cancer. V.H.-S. collected the plasma samples for the investigation of ovarian cancer. M.J. designed and conducted the controlled experiment. R.A. supervised the experimental aspects of the work. O.V. supervised the statistical aspects of the work, and wrote the manuscript.

## REFERENCES

- Kiyonami, R., Schoen, A., Prakash, A., Peterman, S., Zabrouskov, V., Picotti, P., Aebersold, R., Huhmer, A., and Domon, B. (2011) Increased selectivity, analytical precision, and throughput in targeted proteomics. *Mol. Cell. Proteomics* **10**, M110, 002931
- Lange, V., Picotti, P., Domon, B., and Aebersold, R. (2008) Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **4**, 222–235
- Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B., and Aebersold, R. (2009) Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell* **138**, 795–806
- Ahrens, C. H., Brunner, E., Qeli, E., Basler, K., and Aebersold, R. (2010) Generating and navigating proteome maps using mass spectrometry. *Nat. Rev. Mol. Cell Biol.* **27**, 789–801
- Picotti, P., Rinner, O., Stallmach, R., Dautel, F., Farrah, T., Domon, B., Wenschuh, H., and Aebersold, R. (2010) High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nat. Methods* **10**, 43–46
- Cima, I., Schiess, R., Wild, P., Kaelin, M., Schuffler, P., Lange, V., Picotti, P., Ossola, R., Templeton, A., Schubert, O., Fuchs, T., Leippold, T., Wyler, S., Zehetner, J., Jochum, W., Buhmann, J., Cerny, T., Moch, H., Gillissen, S., Aebersold, R., and Krek, W. (2011) Cancer genetics-guided discovery of serum biomarker signatures for diagnosis and prognosis of prostate cancer. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 3342–3347
- Whiteaker, J. R., Lin, C., Kennedy, J., Hou, L., Trute, M., Sokal, I., Yan, P., Schoenherr, R. M., Zhao, L., Voytovich, U. J., Kelly-Spratt, K. S., Krasnoselsky, A., Gafken, P. R., Hogan, J. M., Jones, L. A., Wang, P., Amon, L., Chodosh, L. A., Nelson, P. S., McIntosh, M. W., Kemp, C. J., and Paulovich, A. G. (2011) A targeted proteomics-based pipeline for verification of biomarkers in plasma. *Nat. Biotechnol.* **29**, 625–634
- Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D. A., and White, F. M. (2007) Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5860–5865
- Kuster, B., Schirle, M., Mallick, P., and Aebersold, R. (2005) Scoring proteomes with proteotypic peptide probes. *Nat. Rev. Mol. Cell Biol.* **6**, 577–583
- Mallick, P., Schirle, M., Chen, S. S., Flory, M. R., Lee, H., Martin, D., Ranish, J., Raught, B., Schmitt, R., Werner, T., Kuster, B., and Aebersold, R. (2007) Computational prediction of proteotypic peptides for quantitative proteomics. *Nat. Biotechnol.* **25**, 125–131
- Domon, B., and Aebersold, R. (2006) Mass spectrometry and protein analysis. *Science* **312**, 212–217
- Domon, B., and Aebersold, R. (2010) Options and considerations when selecting a quantitative proteomics strategy. *Nat. Biotechnol.* **28**, 710–721
- Cham, J. A., Bianco, L., and Bessant, C. (2010) Free computational resources for designing selected reaction monitoring transitions. *Proteomics* **10**, 1106–1126
- Abbatiello, S. E., Mani, D. R., Keshishian, H., and Carr, S. A. (2010) Automated detection of inaccurate and imprecise transitions in peptide quantification by multiple reaction monitoring mass spectrometry. *Clin. Chem.* **56**, 291–305
- Reiter, L., Rinner, O., Picotti, P., Hüttenhain, R., Beck, M., Brusniak, M.-Y., Hengartner, M. O., and Aebersold, R. (2011) mProphet: automated data processing and statistical validation for large scale SRM experiments. *Nat. Methods* **8**, 430–435
- MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., and MacCoss, M. J. (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968
- Brusniak, M.-Y. K., Kwok, S.-T., Christiansen, M., Campbell, D., Reiter, L., Picotti, P., Kusebauch, U., Ramos, H., Deutsch, E. W., Chen, J., Moritz, R. L., and Aebersold, R. (2011) Ataq: A computational software tool for high throughput transition optimization and validation for selected reaction monitoring mass spectrometry. *BMC Bioinformatics* **12**, 1–15
- Benjamini, Y., and Hochberg, Y. (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statistical Soc.* **57**, 289–300
- Addona, T. A., Abbatiello, S. E., Schilling, B., Skates, S. J., Mani, D. R., Bunk, D. M., Spiegelman, C. H., Zimmerman, L. J., Ham, A. L., Keshishian, H., Hall, S. C., Allen, S., Blackman, R. K., Borchers, C. H., Buck, C., Cardasis, H. L., Cusack, M. P., Dodder, N. G., Gibson, B. W., Held, J. M., Hiltke, T., Jackson, A., Johansen, E. B., Kinsinger, C. R., Li, J., Mesri, M., Neubert, T. A., Niles, R. K., Pulsipher, T. C., Ransohoff, D. F., Rodriguez, H., Rudnick, P. A., Smith, D., Tabb, D. L., Tegeler, T. J., Variyath, A. M., Vega-Montoto, L. J., Wahlander, A., Waldemarson, S., Wang, M., Whiteaker, J. R., Zhao, L., Anderson, N. L., Fisher, S. J., Liebler, D. C., Paulovich, A. G., Regnier, F. E., Tempst, P., and Carr, S. A. (2009) Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat. Biotechnol.* **27**, 633–641
- Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., and Gygi, S. P. (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6940–6945
- Montgomery, D. (2005). *Design and analysis of experiments*. John Wiley & Sons, Inc., Hoboken, NJ
- DeRisi, J. L., Iyer, V. R., and Brown, P. O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686
- Kreutz, C., and Timmer, J. (2009) Systems biology: experimental design. *FEBS J.* **276**, 923–942
- Oberg, A. L., and Vitek, O. (2009) Statistical design of quantitative mass spectrometry-based proteomic experiments. *J. Proteome Res.* **8**, 2144–2156
- Kutner, M., Nachtsheim, C., Neter, J., and Li, W. (2005). *Applied linear statistical models*. McGraw Hill, New York
- Dobbin, K., and Simon, R. (2002) Comparison of microarray designs for class comparison and class discovery. *Bioinformatics* **18**, 1438–1445
- Xia, J. Q., Sedransk, N., and Feng, X. (2011) Variance component analysis of a multi-site study for the reproducibility of multiple reaction monitoring measurements of peptides in human plasma. *PLoS One* **6**, e14590
- R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>. ISBN 3-900051-07-0
- Clough, T., Key, M., Ott, I., Ragg, S., Schadow, G., and Vitek, O. (2009) Protein quantification in label-free LC-MS experiments. *J. Proteome Res.* **8**, 5275–5284
- Mosteller, F., and Tukey, J. (1977). *Data analysis and regression*. Addison Wesley, New York
- Smyth, G. (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, 1–26



In order to cite this article properly, please include all of the following information: Chang, C.-Y., Picotti, P., Hüttenhain, R., Heinzelmann-Schwarz, V., Jovanovic, M., Aebersold, R., and Vitek, O. (2012) Protein Significance Analysis in Selected Reaction Monitoring (SRM) Measurements. *Mol. Cell. Proteomics* 11(4): M111.014662. DOI: 10.1074/mcp.M111.014662.