mProphet: A general and flexible data model and algorithm for automated SRM data processing and statistical error estimation

Lukas Reiter[1,2,3,4,5], Oliver Rinner[1,4,5], Paola Picotti[4], Ruth Hüttenhain[4,7], Martin Beck[4], Mi-Youn Brusniak[6], Michael O. Hengartner[2,3], Ruedi Aebersold[4,7,8]

1 contributed equally

2 Institute of Molecular Biology, University of Zurich, Zurich, Switzerland

3 PhD Program in Molecular Life Sciences Zurich, Zurich, Switzerland

4 Institute of Molecular Systems Biology, Department of Biology, ETH Zurich, Zurich, Switzerland

5 Biognosys AG, Zurich, Switzerland

6 Institute for Systems Biology, Seattle, WA 98103, USA

7 Competence Center for Systems Physiology and Metabolic Diseases, Zurich, Switzerland

8 Faculty of Science, University of Zurich, Zurich, Switzerland

Corresponding Author:
Prof. Ruedi Aebersold
Institute of Molecular Systems Biology
Wolfgang-Pauli-Str. 16, HPT E 78
ETH Zurich
CH-8093 Zurich
Phone: +41 44 633 31 70
Fax: +41 44 633 10 51
aebersold@imsb.biol.ethz.ch

Keywords: targeted proteomics, selected / multiple reaction monitoring, SRM / MRM, peak group picking, automated processing, false discovery rate

Running Title: Reiter, Rinner et al. *mProphet*

ABSTRACT

Selected reaction monitoring (SRM[1]) is a targeted mass spectrometric method that is increasingly used in proteomics for the detection and quantification of sets of pre-selected proteins at high sensitivity, reproducibility and accuracy. Currently data from SRM measurements are mostly evaluated subjectively by manual inspection based on ad hoc criteria, precluding the consistent analysis of different datasets and an objective assessment of their error rates. Here we present *mProphet*, a fully automated system that computes accurate error rates of the identification of the signals correctly identifying the targeted peptide in SRM data sets and maximizes specificity and sensitivity by combining relevant features in the data into a statistical model. The presented method and software tool will be of critical importance for the full exploitation of the unique potential of SRM measurements in quantitative proteomics, a prerequisite for the meaningful comparison of data across studies and laboratories.

[1] Also referred to as multiple reaction monitoring or MRM

INTRODUCTION

In recent years novel strategies to detect and quantify selected proteins in complex samples by targeted mass spectrometry have been suggested and implemented(Aebersold 2003; Kuster, Schirle et al. 2005; Domon and Aebersold 2006; Jaffe, Keshishian et al. 2008; Schmidt, Claassen et al. 2009). A particularly sensitive targeted mass spectrometry method is selected reaction monitoring (SRM) (also referred to as multiple reaction monitoring or MRM) on triple quadrupole instruments(Anderson and Hunter 2006; Wolf-Yadlin, Hautaniemi et al. 2007; Lange, Picotti et al. 2008; Picotti, Bodenmiller et al. 2009). In contrast to the more widely used shotgun proteomics method where peptides are selected for identification from the pool of sample peptides via a simple heuristics, in SRM sets of pre-determined peptides are detected and quantified selectively in complex samples(Domon and Aebersold). This is accomplished by the selective acquisition of fragment ion signals that are unique for the targeted peptide. A pair of a precursor ion signal of the targeted peptide (detected in Q1) and a diagnostic fragment ion signal (detected in Q3) is referred to as a transition, and the precursor ion, along with several transitions constitutes a definitive SRM assay for the detection of the respective peptide in a complex sample. The ability of the SRM technique to generate highly reproducible and quantitatively accurate data sets matches it in an ideal way with hypothesis driven research and projects that require the consistent analysis of a set of proteins under various conditions(Jovanovic, Reiter et al.), as is the case e.g. in biomarker studies(Anderson and Hunter 2006; Keshishian, Addona et al. 2007; Whiteaker, Zhang et al. 2007; Addona, Abbatiello et al. 2009; Keshishian, Addona et al. 2009; Oberg and Vitek 2009). Besides the actual data acquisition, an SRM experiment involves two major steps. The first is the design of the assay that unambiguously identifies the targeted peptide in a sample and the second is the analysis of the acquired data. Significant advances have been realized to speed up and automate the design of the SRM assays(MacLean, Tomazela et al.; Picotti, Rinner et al.; Sherwood, Eastham et al. 2009). Empirical proteomic data deposited in databases and prediction tools support the selection of the best proteotypic peptides (PTPs)(Martens, Hermjakob et al. 2005; Ahrens, Brunner et al. 2007; Brunner, Ahrens et al. 2007; Mallick, Schirle et al. 2007; Baerenfaller, Grossmann et al. 2008; Deutsch, Lam et al. 2008; Vogel and Marcotte 2008; Fusaro, Mani et al. 2009; Schrimpf, Weiss et al. 2009) for targeted proteins and are powerful resources for SRM assay design(Picotti, Lam et al. 2008; Prakash, Tomazela et al. 2009; Sherwood, Eastham et al. 2009). Libraries of crude synthetic peptides representing the selected PTP's have been

introduced to generate SRM assays on a proteome wide scale at high throughput(Picotti, Rinner et al.).

In contrast to assay development, the downstream processing of SRM data is still in its infancy and represents a bottleneck of the technology. The analysis of SRM data involves the detection, qualification and quantification of the relevant peaks in the raw data. The process of detection and qualification is currently carried out essentially manually using subjective decision criteria. A comparison with the history of shotgun proteomics shows that the development of sophisticated algorithms for the automated generation of peptide-spectrum matches(Eng, McCormack et al. 1994) and for the statistical evaluation of their quality(Keller, Nesvizhskii et al. 2002; Moore, Young et al. 2002; Elias and Gygi 2007) has been critically important for the robust implementation of the technology(Aebersold 2009).

A typical SRM experiment starts with the selection of transitions that are most sensitive and unique for a given peptide(Lange, Malmstrom et al. 2008; Prakash, Tomazela et al. 2009; Sherwood, Eastham et al. 2009). Usually between 3 and 5 transitions per peptide are chosen. If the chromatographic elution time of the targeted peptide is known, it can be used to schedule the SRM measurement, i.e. a transition group is only measured within a defined retention time window, thus increasing the number of peptides that can be measured in a LC-MS run(Stahl-Zeng, Lange et al. 2007). The transitions monitored for each peptide result in extracted ion currents over time for each Q1/Q3 pair (transition). Often, these signals cannot unambiguously be assigned to the targeted peptide, i.e. the experimenter is faced with the decision which, if any, of the detected signals arose from the target peptide. The challenge of assigning transition signals to the targeted peptide sequence is compounded for signals of low signal to noise ratio.

So far, mainly ad hoc criteria have been used to solve these cases. Prakash et al.(MacLean, Tomazela et al.; Prakash, Tomazela et al. 2009) have recently shown how a single score (relative intensity similarity between MS2 spectra and SRM data) can be used for a more systematic discrimination of true and false peak groups. However, this only works if the relative intensities are known and, especially for low abundant signals, random correlations can occur that are difficult to assess with a single score. Abatiello et al.(Abbatiello, Mani et al.) use internal reference peptides and technical replicates to find interferences in transitions and to assign a score to the signals. However, currently no general applicable strategy exists to process any type of SRM data.

Furthermore, some very basic questions have not been systematically analyzed so far. It is, for instance, not quantified how much internal reference peptides support an accurate scoring, nor is it clear which of the applied criteria for peak group verification are really useful, i.e. have a high discriminatory power.

The major impact of the SRM technique in quantitative proteomics will come from its application to large cohorts of samples such as those derived from different patients in clinical studies, or from systematic perturbations or time or dosage series in systems biology studies. With more than 1000 transitions that can be measured in a single run(Stahl-Zeng, Lange et al. 2007), the number of data points acquired in such experiments is no longer amenable to manual evaluation, and the absence of probabilistic scoring will significantly lower the value of such large datasets.

To address this urgent need we developed *mProphet* a system that integrates multiple dimensions of information available in SRM data in a probabilistic scoring model for the automated, objective, flexible and consistent scoring of SRM data sets. Using a novel "decoy-transition" approach, *mProphet* automatically adapts the error model for each dataset and assigns a confidence measure to each peak group for quality control. The signal intensities in the thus identified peak groups are then used for the subsequent quantification of the targeted analytes. Therefore, the work presented in this manuscript is an essential step towards the full exploitation of the potential of SRM based targeted proteomics for quantitative biology.

RESULTS

Definition of data structure and *mProphet* algorithm

The *mProphet* algorithm and software are based on a data structure that describes the basic properties of SRM data. In the following we describe the data structure as well as the algorithm that operates on the data.

To make use of the full information content of SRM recordings we structure the data as follows: we name the assembly of transitions for one peptide a *transition group*. Further, we name the whole data record of such a measurement a *transition group record*. One transition group record consists of the *traces* (retention time vs. signal intensity) of each measured transition. In one transition group record, a number of *peak groups* can be identified, but at most one peak group represents the peptide of interest (**Fig. 1a**).

While complete transition groups with high signal to noise (*S/N)* ratio are usually easy to correctly identify, transition groups with lower *S/N* and/or those containing signals from interfering analytes can lead to ambiguous assignment of the peak group to the target peptide(Sherman, McKay et al. 2009). Since SRM measurements are typically carried out in a complex sample background it is likely that concurrently present multiple precursors generate interfering traces, complicating the identification of the true peak group. Examples of ambiguous peak groups are shown in **Supplementary Fig. 1**.

To separate true from false peak groups and to resolve ambiguous peak groups we developed a scoring system that integrates sub-scores representing different features in the data structure into an overall discriminant score. The features include the temporal concurrence and the shape of the peaks in a peak group, and, if applicable, the global similarity between the peak groups of endogenous and reference peptides (**Fig. 1b**). Furthermore, we use the relative intensity of the transition traces, defined in a prior assay development step(Picotti, Rinner et al.; Prakash, Tomazela et al. 2009). Specifically, we used the following scores to describe peak group features:

1) *Intensity score:* Represents the summed intensity for all peaks of one peak group.
2) *Intensity correlation:* Expresses the conservation of the relative intensities of fragment ion signals derived from the targeted peptide precursor(Picotti, Rinner et al.; Prakash, Tomazela et al. 2009). In order to use this feature the expected relative intensities must be known. Because prediction of fragment ion intensities is very difficult this score uses

prior information from specific experiments such as public libraries of SRM transitions(Picotti, Lam et al. 2008) or public repositories of spectral libraries(Craig, Cortens et al. 2006; Frewen, Merrihew et al. 2006; Deutsch, Lam et al. 2008; Stein and Rudnick 2009).

3) *Co elution score*: Expresses the alignment of the peak apexes in a peak group along the chromatographic time axis.
4) *Shape score*: Expresses the peak shape similarity for all peaks in a peak group (see **Supplementary Fig. 2**).
5) *Reference correlatio*n: Expresses the conservation of the signal intensity ratios for corresponding light and heavy transitions in case stable isotope labeled reference peptides are present. The calculation is done as in 2), except that instead of the expected values, the measured values for the heavy transitions are used.
6) *Retention time deviation*: Indicates the absolute value of the difference between measured and expected retention time of a peptide. The expected retention times were derived from SRM measurements of the corresponding synthetic reference peptide.
7) *Reference co elution score*: Expresses the co elution of light and corresponding heavy transition signals in case stable isotope labeled reference peptides are present.
8) *Reference shape score*: Expresses the shape similarity of light and corresponding heavy transition signals in case stable isotope labeled reference peptides are present.

The above described scores can be used, either by themselves or in combination, to discriminate between true and false peak groups. Ideally, the scores would be combined such that the separation between true and false peak groups is maximized. Because the scores depend on the specifics of a particular experiment such as the inclusion of reference peptides, the availability of trace intensity ratios and instrument settings, it is not beneficial to combine the sub-scores with a fixed set of weights. Therefore, we developed the concept of decoy transitions to optimally combine the scores in each experiment. Decoy transitions are transitions for peptide species that do not exist in biological samples and can be regarded as negative controls (see **Methods, Generation of decoy transitions**). The principle of decoy transitions is conceptually related to the decoy database approach used for the database searches of shotgun proteomic data(Moore, Young et al. 2002; Elias and Gygi 2007). However, while the decoy database approach operates on the database search level, we introduce the decoy transitions at the measurement level. The decoy transition data is used in each experiment to optimize a linear combination of the available sub-scores and to derive statistical error rate estimates by parameterizing a null distribution.

To find an optimal combination of sub-scores, the data set is split into training and test dataset as a first step in the analysis of an experiment. Sets of true (the highest scoring) and false (decoy) peak groups from the training data set are used to refine the model iteratively. In this semi-supervised learning process(Kall, Canterbury et al. 2007; Choi and Nesvizhskii 2008) the weights of the sub-scores converge such that discrimination between true and false peak groups is optimized. Since each transition group record can be associated to several peak groups but generally at most one can be true, the peak groups are re-ranked within each transition group record and only the top ranked peak groups are considered in each iteration. The data derived from the decoy transitions is essential in the training step because they can be fixed as false by default. The normalized discriminant score of the top ranked peak groups resulting from the final iteration is called *mProphet* score and is used to filter the SRM data of a particular experiment. For more information on the semi-supervised learning algorithm see **Methods, Semi-supervised learning algorithm and false discovery rate estimation** and **Supplementary Methods 2**.

To estimate false discovery rates (FDR) depending on the discriminant score (*mProphet* score) cutoff we parameterize a null distribution based on the decoy transition data in the test data set. The null distribution is assumed to be normal based on data in **Supplementary Fig. 3**. The FDR is used, in turn, to filter the final data set according to a user defined quality. For more information on how the FDR is estimated see **Methods, Semi-supervised learning algorithm and false discovery rate estimation** and **Supplementary Methods 3**.

***mProphet* software working steps.** *mProphet* fully automates SRM data processing by implementing the following workflow: First, *mProphet* converts raw machine data into an annotated xml file format which assigns transition traces into transition group records. This annotation defines attributes such as expected retention time, relative intensities, whether a transition is associated with a reference peptide or an endogenous peptide, and whether a transition group record is a target or decoy record. These attributes are essential for the later scoring of the peak groups (see **Methods, Scoring of peak groups**). Second, from this annotated data peak groups are detected and associated with the corresponding targeted peptide using criteria described in the methods section. Third, all extracted peak groups are statistically analyzed. At most one peak group per transition group record is assumed to be correct. *mProphet* maximizes the separation of true and false peak groups and assigns a confidence score (*mProphet* score) to the final top ranked peak group based on the null distribution derived from the decoy transitions. Fourth, the top scoring peak groups for each

transition group record, representing the peptides identified with a certain level of confidence, can be filtered by a user defined quality (FDR). Quantification information such as height and area of the transitions is included in the output as well as the summed height and area for a certain precursor. Also included are the exact coordinates of the peak groups (transition group record i.e. LC-MS run and precursor id, retention time) which allows to combine the quantification of other tools (e.g. a Skyline) with the identification of *mProphet*.

A typical processing time for a data set of 16 LC-MS runs with a 2-fold cross validation repeated 50 times on a single node and including the conversion of the mzXML format to a MRM specific format is 6 minutes.

Generation and analysis of a gold standard data set

To test the performance of the *mProphet* under realistic conditions, to assess the significance of each sub-score and to optimize a composite discriminate score we used a "gold standard" reference sample mixture to acquire a dataset consisting of confidently assigned true and false peak groups. The assignment is based on synthetic isotopically light and heavy labeled peptides spiked into a sample applying a dilution series. To generate the reference sample we chemically synthesized 100 peptides on a micro scale using the SPOT-synthesis technology(Wenschuh, Volkmer-Engert et al. 2000; Hilpert, Winkler et al. 2007) and developed SRM assays for each peptide(Picotti, Rinner et al.; Picotti, Bodenmiller et al. 2009). The peptides were synthesized in an isotopically light and in a heavy form to simulate experiments with quantification by internal standard. The peptides were added in three different concentrations into three different background matrices of increasing complexity, consisting of trypsinized protein extracts from *L. interrogans*, *C. elegans, or H. sapiens* cells (**Table 1**). The samples were subjected to SRM analyses in which 10 transitions (5 light and 5 heavy, **Supplementary Table 1**) were measured per synthetic peptide in the respective background matrices (**Table 1**).

After peak extraction, peak groups for 792 target transition group records were detected. To distinguish between true or false peak groups we took advantage of the dilution series in this data set by comparing peak groups across the diluted samples (**Fig. 2a,b**). Only peak groups which showed the expected intensity ratios between dilutions were accepted as true (**Fig. 2b**). 169 transition group records did not contain any true peak groups after stringent filtering and were discarded. 41 transition group records contained

two true peak groups and were discarded. 567 transition group records contained one true peak group and were retained. This data set of true transition groups was complemented with decoy transition groups representing 702 transition group records after data processing. For a more detailed description of the data processing procedure see **Methods, Assignment of true peak groups in gold standard data set.**

A combined data set, consisting of true peak groups derived from the dilution experiments and false peak groups resulting from the measurement of decoy transitions that act as internal negative controls (see **Methods, Generation of decoy transitions**), was then used to investigate the ability of individual features (**Fig. 1b**) to separate true from false peak groups. The data in (**Fig. 2c**) indicate that the "*reference shape"* and "*intensity correlation"* sub-scores with an area under the curve (AUC) of 0.985 and 0.975 had the highest separation power. This reflects the stability of fragmentation patterns over the course of chromatographic elution and the reproducibility of fragment ion intensities, respectively. The "*reference correlation"*, "*reference co elution"* and "*reference shape"* scores are all based on the presence of a reference peptide, indicating the significance of reference peptides in SRM measurements. The "*intensity score"* showed an interesting characteristic: Above a certain signal intensity the probability of random signals was very low, while the separation power for signals of low abundance was weak. Thus, the intense true signals can be recovered with virtually no false positives. However, if one would use this feature only, the sensitivity of reliable SRM measurements would be limited to abundant peptides. The "*retention time deviation"* score showed the weakest separation power of the properties tested (AUC 0.850), reflecting the variability in chromatography and the fact that retention time was already constrained to a retention time window for scheduled SRM. However, this feature is the only one that is independent of peak group intensity. The "*co elution"* and "*shape score"* are scores derived from only the light isotopic channel. They could be used even if no labeled reference peptide would be measured. Especially the "*shape score"* showed a high separation power. In general, none of the features by themselves fully separated all true and false peak groups and they displayed different separation potential, suggesting that a suitably weighted combination of sub-scores would add to the overall performance of the scoring scheme.

mProphet validation: Development and benchmarking of a composite discriminant score

Using the data derived from the decoy transitions *mProphet* linearly combines the sub-scores into the *mProphet* score each time a data set is analyzed (for a description see **Methods** part **Semi-supervised learning algorithm and false discovery rate estimation** and **Supplementary Methods 2**). To assess the performance of the *mProphet* score we tested *mProphet* on the gold standard data set and compared the separation power of the *mProphet* score to that of the single sub-scores. Further, we validated the accuracy of the statistical false discovery rate estimate using the gold standard data set. We also investigated the influence of the number of transitions on the most commonly used sub-score, the "*intensity correlation*". Finally we investigated the separation power of the *mProphet* score as a function of signal intensity and the fraction of decoy transitions used in the analysis.

***mProphet* scores for gold standard data set.** In order to evaluate the separation power of the *mProphet* score we applied *mProphet* to the gold standard data set. The gold standard data set only consists of transition group records that contain one correct peak group and decoy transition group records which are by definition false. Usually, however, depending on the efficiency of the experiment, an arbitrary fraction of (non decoy) transition group records can be false, i.e. the top scoring peak group of this transition group record is not derived from the targeted peptide. In order to simulate real data and to challenge *mProphet* we treated part of the decoy data as if it was data resulting from target transitions. 243 decoy transition group records (30% of final target transition group records) were treated as target yielding 810 target transition group records in total (see **Supplementary Fig. 4** for results when the number of decoy signals treated as target was varied over a wide range). Although, the *mProphet* algorithm was not aware of the true labels of the target data it separated true (blue) from false (green) very well using its implemented semi-supervised learning algorithm (**Fig. 3a**). As expected in the test data set, the decoy peak groups (red) showed a similar distribution as the decoy peak groups that were treated as target data. These data demonstrate that *mProphet* can process data with an unknown fraction of false target peak groups and separate true from false target peak groups.

Combining scores for better separation power. In order to assess the improvement of the *mProphet* score compared to the single sub-scores we compared the corresponding Receiver-Operating Characteristic (ROC) curves (**Fig. 3b**). The same data set as in **Fig. 3a** was processed with *mProphet* to generate the ROC curves. Only the test data set is shown to avoid overfitting. The *mProphet* score which is a weighted combination of the eight sub-scores, showed the best separation power. The following weights were assigned to the sub-

scores by *mProphet*: "*Intensity score"* 0.0006, "*intensity correlation*" 1.2271, *"co elution score*" -0.1761, *"shape score*" 3.0502, *"reference correlation*" 0.8011, *"retention time deviation*" -0.00971, *"reference co elution score*" -0.1613, *"reference shape score*" 4.6528. In this data set, the second best separation could be achieved with the "*reference shape score"* and the third best with the "*intensity correlation"*. Weights with opposite signs are caused by scores with opposite directionality, as e.g. for the *co elution* and *retention time deviation scores* (**Fig. 2c**). Despite the negative weight signs, these scores also contribute to an enhanced discrimination of the *mProphet* score.

Benchmarking of false discovery rate estimation. *mProphet* estimates a false discovery rate (FDR), which can be used to filter the data according to a user defined quality, analogously to the filtering of tandem mass spectrometry identifications(Keller, Nesvizhskii et al. 2002; Moore, Young et al. 2002; Elias and Gygi 2007; Kim, Gupta et al. 2008; Reiter, Claassen et al. 2009). To determine how well the estimated FDR corresponded to the real FDR in our gold standard data set we processed the same data set as in **Fig. 3a** with *mProphet*. Because the actual classes of the target peak groups are known from the generation of the data set, we could compare the estimated FDR and sensitivity to the real FDR and sensitivity (**Fig. 3c**). The data indicate that the estimated values are very similar to the real values.

Sensitivity. One of the major goals of *mProphet* is to recover signals with low signal to noise ratio. We therefore plotted the signal to noise ratios (S/N) of the highest trace of a peak group as function of the *mProphet* score (**Fig 3d**) using the same data as shown in **Fig 3a**. The data indicate that for peaks groups exceeding a S/N ratio of approximately 10, the true and false peak groups were completely separated. Between a S/N of 5 and 10, 11 out of 16 true peak groups could still be completely separated. We therefore conclude that the *mProphet* is very sensitive and that signals close to the noise level can be recovered.

Dependency of *intensity correlation* score on number of transitions. The number of recorded transitions per targeted peptide is directly related to the information content in the data and SRM measurement time needed. We tested how the number of recorded transitions affected the separation power of the *intensity correlation* (**Fig. 3e**). For this purpose we extracted 3-5 transitions (6-10 including heavy transitions) from the raw data and ran *mProphet* on each of these data sets (data set generated analogously to **Fig. 3a**). As expected the separation power increased with an increasing number of transitions recorded.

Number of decoy transitions needed. We next determined how the performance of *mProphet* is affected by the number of decoy transitions recorded. We varied the number of

decoy transition groups as percentage of total transition group records used for the analysis and determined ROC curves resulting from these analyses (**Fig. 3f**). The data show that in general, an increasing number of decoy transitions improved sensitivity/selectivity. However, the observed gain of separation was very small if more than roughly 90 decoy transition groups were added. Since these numbers are independent of the total number of transitions measured in an experiment, the measurement overhead for decoy transitions is typically around 10% for a larger data set of e.g. 20 LC-MS runs.

In principle *mProphet* works with as few as 20 decoy and 20 target transition group records. However, for a decent lower bound we think of 50 decoy and 50 target transition group records. Generally there should be some true target transition group records that can be used for the positive training data set.

Application of *mProphet* to measurements in a total human lysate

We next assessed the performance of *mProphet* for the automated detection of endogenous peptides by large scale SRM measurements in a complex sample background. The sample consisted of a tryptic digest of an extract of the human cell line u2os that had been isotopically labeled to equilibrium with SILAC medium(Ong, Blagoev et al. 2002) and to which 591 isotopically light synthetic peptides corresponding to 265 proteins had been added. The synthetic peptides and their endogenous counterparts were targeted by SRM, whereby 9648 transitions were acquired for both isotopic forms (**Supplementary Table 2**). The targeted peptides spanned a wide range of cellular abundance as judged by their spectral count number (1-171 counts) in an extensive shotgun sequencing dataset (**Fig. 4b**). Absolute abundances were determined for 4 peptides matching to the proteins CBX5 (90’000 copies/cell), SEH1L (10’000), HS71L (10’000), and MD2L1 (35’000) by using calibrated heavy labeled synthetic peptides in an independent experiment with unlabeled u2os cells(Gerber, Rush et al. 2003) (**Supplementary Table 3**).

In the data analysis, peak groups were considered only if they contained the traces of the spiked in reference peptide which in most cases could be unambiguously identified. In order to facilitate fair decoy scoring the same reference was associated with the decoy transition groups, i.e. a decoy group was scored against the same corresponding reference peptide as an endogenous peptide. The rationale behind this is to provide an intense peak group from the spiked-in peptide as reference for the decoy peak groups (which otherwise

would hardly be observed for the decoys). The *mProphet* scores resulting from the measurement of target (spike-in and endogenous) and decoy transitions are shown in **Fig. 4c**. On average 460 transitions including decoy transitions were recorded per run in scheduled mode in altogether 21 injections.

The results indicated in **Fig. 4d** show that 457 out of 591 targeted peptides and 238 of 265 targeted proteins could be identified with a FDR of 1% (**Supplementary Table 2**). The FDR never rises above 10% because the total number of failed measurements is estimated to 10%. Furthermore, the data indicate that *mProphet* automated scoring confidently identified signals to S/N below 1.5 (**Fig. 4c-e**). Consequently, *mProphet* can be applied to large scale experiments for automated verification of SRM based protein detection and quantification in complex matrices with a large number of transitions measured.

We wanted to investigate the influence of the number of transitions recorded on the separation power of *mProphet*. For this purpose we repeated the analysis using only part of the data corresponding to two to six most intense fragment ions. **Fig. 4f** shows the ROC curves of the *mProphet* score derived from the analysis with two to six transitions (4 to 12 transitions including the reference peptides). In this data set four processed transitions resulted in the highest separation power of the *mProphet* score. The separation power corresponding to two processed transitions was unexpectedly high. A possible explanation is that the presence of a reference peptide, that can be very easily identified based on the high intensity, contributes to a large part to the separation power. This hypothesis is supported by the fact that the reference shape score is assigned to a high weight by *mProphet*.

In order to study the influence of the reference peptides we performed the same analysis as in **Fig. 4f** but ignoring the reference peptide information **Fig. 4g**. Generally, the separation power is lower if reference peptides are missing. The separation power is increasing with the number of transitions recorded. The relative change, when increasing the number of transitions, is stronger when compared to the case including references. This is in agreement with the fact that the intensity correlation is assigned with a high weight by *mProphet* if the reference peptide is missing. The separation power of six transitions without reference is lower than that of two with a reference which underlines the additional power gained by the reference spiked into the sample.

Application of *mProphet* to measurements in human plasma

In order to test the influence of interferences in transitions on quantification we monitored a number of peptides in human plasma in two technical replicates. 67 peptides corresponding to 58 proteins were selected based on the availability of isotopically heavy labeled peptides in the lab. The sample consisted of peptides that are N-glycosylated in the intact protein. The peptides were enriched from human plasma in their de-glycosylated form as described in Zhang et al.(Zhang, Li et al. 2003). We analyzed the data with mProphet and were able to detect approximately 55% percent of the targeted proteins spanning a wide range of concentrations from an estimated 0.92 to 330000 ng/ml (**Fig. 5a**). The concentration estimates are based on a method using a spectral counting approach developed by Terry Farrah et al. (manuscript in preparation). For most of the proteins that were not detected there is no concentration prediction available, indicating that these are probably very low abundant proteins. Using *mProphet* we were able to achieve a sensitivity of 95% using a false discovery rate cutoff of 5% (**Fig.5b,c**). Based on these identifications we investigated the influence of interferences on quantification. For this purpose we also ran AuDIT(Abbatiello, Mani et al.) on the data and identified the transitions with the highest level of interferences for each transition group record. We made the quantification three times, once with all transitions, once with one transition removed based on the AuDIT p-value score and once based on the AuDIT cv score (**Fig. 5d-f**). The increasing correlation when removing transitions with interferences indicates, as expected, that a careful selection of the transitions for quantification as implemented in AuDIT will improve the precision of quantification. We also tested the discriminating power of the AuDIT scores for identification peak groups in our data set and found a significantly smaller separation compared to the strongest sub score of *mProphet*, the *reference shape score* and also to the *mProphet* score (**Supplementary Fig. 5**). The reason is that signals derived from decoy transitions are reproducible because they don't represent noise but interfering signals from real peptide species with similar Q1/Q3 combinations as the targets (**Supplementary Fig. 6**)(Sherman, McKay et al. 2009).

Characterization of decoy transitions

To investigate whether the decoy transitions are a good representation for false identifications of targeted peptides we recorded a number of different decoy transition types in the human plasma sample (see above). For each human peptide targeted we generated two different types of decoys: One was based on reversing the peptide sequence,

decreasing the precursor m/z by 10Th and selecting fragment ions with m/z values that are similar to those of the targeted peptides, the other was based on adding random integers to the precursor m/z and fragment ion m/z (see **Methods, Generation of decoy transitions**). After *mProphet* analysis we compared the discriminant score distributions derived from the two different types of decoys with a quantile quantile plot (Q-Q plot) (**Fig. 5g**). The observed strong linearity indicates that the two different types of decoys produce very similar signals (for Q-Q plots of not only the discriminant score see **Supplementary Fig. 7**).

Further we targeted a random number of yeast peptides in this human sample and again generated the two different types of decoy transitions for these peptides (see **Methods, Generation of decoy transitions**). Since no yeast peptides are expected in the human sample these identifications should be a good approximation of false identifications derived from transitions of real peptides. Again we compared the signals derived from the yeast transitions to the signals derived from the two types of decoy transitions using Q-Q plots. **Fig. 5h,i** show Q-Q plots of the *reference shape score* when comparing the yeast transitions to each type of decoy transitions. The strong linearity indicates that the decoy transitions are good representations for false identifications of targeted peptides (for Q-Q plots of all other scores see **Supplementary Fig. 8**).

DISCUSSION

SRM has been established as an important tool to detect and quantify proteins in a complex sample in a targeted manner(Lange, Malmstrom et al. 2008; Picotti, Bodenmiller et al. 2009). In spite of its favorable performance profile the method has not been widely used in proteomics for two major reasons: 1) The development of validated SRM assays is time consuming and 2) Data analysis is often semi-automated and involves manual verification of raw data. Whereas the throughput of assay development(MacLean, Tomazela et al.; Mallick, Schirle et al. 2007; Deutsch, Lam et al. 2008; Lange, Malmstrom et al. 2008; Lange, Picotti et al. 2008; Picotti, Lam et al. 2008; Fusaro, Mani et al. 2009; Sherwood, Eastham et al. 2009) has recently been improved, the challenge of assigning transition traces to the target peptide has hardly been addressed(Abbatiello, Mani et al.; Prakash, Tomazela et al. 2009). This is somewhat analogous to the situation in early shotgun proteomics, where the tools to control the quality of the identifications were developed long after the principal measurement methods(Nesvizhskii, Keller et al. 2003; Elias and Gygi 2007; Kall, Canterbury et al. 2007; Reiter, Claassen et al. 2009) and where the development of such tools was critical for the assessment of the quality of proteomics data.

Manual verification of SRM data is tedious but more importantly, subjective and not reproducible. In a manual process usually only a small number of features (e.g. co elution, intensity) are considered with no defined quality criteria and weight. In applying SRM we encountered numerous examples similar to those shown in **Supplementary Fig. 1** where several candidate peak groups are present in a single transition group record. Likewise, in the decoy data one can find many examples where several transitions co elute. These problems are compounded for the lower S/N range. It is therefore difficult to objectively identify the true peak groups associated with the targeted peptides using ad hoc, manually selected criteria. We also observed that single scores have the potential to generate erroneous results, even if the data look convincing. E.g. in some peak groups that scored high on the co elution and shape scores could only be identified as false positives when the relative transition peak intensities were invoked. This source of error is particularly frequent if a highly abundant peptide leaks into the precursor mass window of a target peptide and shows residual intensity in many Q3 channels. We also observed that a high correlation of observed and predicted fragment intensity ratios is not sufficient by itself for maximum separation of true and false signals. Correlations can become very high for pure noise peaks

by chance, especially if the spectrum is of low information content with one intense peak among low intensity peaks.

We have shown that similarity of shape and co-elution of peaks within one peak group and between endogenous and reference peptides are powerful scores. We have introduced a scoring metrics based on cross-correlation of peak shapes, which captures global similarity also in cases of low intensity signals where determination of peak apexes is difficult (**Supplementary Fig. 2**). This score can also be used to find interferences in transitions and in principle to automatically filter out transitions with interferences similar to as described in Abbatiello et al. (Abbatiello, Mani et al.) (see also **Fig. 5d-f**).

The idea of *mProphet* is to separate identification from quantification, which is analogous to the practice commonly used in shotgun proteomics. Specifically, the signal intensities of the correctly identified peak groups are an excellent basis for accurate quantification, using one of several available quantification tools. In SRM as in shotgun proteomics, interferences of unrelated peptides can give rise to mixed MS2 spectra which complicates identification and at the same time overlapping features on the MS1 level require special care for quantification. Once a set of signals has been accepted, effects like interferences in transitions and outliers can be considered for a statistical analysis of regulation.

mProphet uses a combination of peak group features for a highly sensitive recovery of true peak groups. We could show that the combined score is more sensitive than any single peak group sub-scores. Moreover, the weight of sub-scores can vary based on the experimental set-up. We observed that the relative signal intensity is a strong score when intensities derived from the same instrument are used, but less so if intensities from a different instrument were used (see **Supplementary Fig. 9**). Based on the availability of decoy data, *mProphet* always combines sub-scores in a way that maximizes discrimination power. Thereby, it is not limited to a certain number of transitions, the presence of a synthetic reference peptide (**Fig. 4g**) or the availability of relative intensities e.g. derived from an MS2 spectrum.

We introduce the concept of decoy transitions to model randomly occurring signals in SRM data. *mProphet* uses decoy transitions for its semi-supervised learning approach, similar to Kall et al. and Choi et al.(Kall, Canterbury et al. 2007; Choi and Nesvizhskii 2008). Our decoy transition concept is related to decoy database searching(Moore, Young et al. 2002; Elias and Gygi 2007). The main difference is that in the context of SRM the decoy

element is not introduced in the search step but as negative controls in the actual measurement. As a consequence, the FDR cannot be determined by simple counting of decoy hits as in the shotgun-proteomics case. *mProphet* estimated the FDR based on cross validation and a model of the null distribution which is essentially based on scores of the decoy peak groups (**Supplementary Methods 3**). The measurement time that needs to be allocated to decoy transitions is in the order of one run at most in studies involving the analysis of multiple samples or of a large number of target peptides. In a hypothetical example where 300 proteins are measured with 3 peptides each, this is less than 10% of the total measurement time. Ideally the decoy transitions are distributed equally among all LC-MS runs of a study to account for differences in machine sensitivity and sample differences. Selection of the transition groups that are used to generate the decoy transition groups can be done random by sampling without replacement. For very similar experiments it might in the future also be possible to reuse the classifier of *mProphet* which would make the inclusion of decoy transitions unnecessary.

In very complex sample matrices we experienced that peak groups can occur where a number of peaks perfectly co elute and by change match the expected relative intensities. The occurrence of such events depends on the sample and is difficult to quantify without using negative controls in the measurement. We use the decoy transitions to quantify these signals. Decoy transitions are different from classical background noise estimation. This is because the best peak group derived from the decoy transitions is processed therefore ensuring a fair competition between target and decoy peak groups. If one would try to accomplish the same by extracting random signals from a target measurement (e.g. the second peak group) they could be confused with signals derived from peptide isoforms which behave differently in the chromatography and therefore have multiple elution maxima.

We showed that the sensitivity of an SRM measurement is often limited by the ability to distinguish the true from false peak groups and not the technical sensitivity of the instrument. We provide a strategy/software that allows retrieving significantly increased amounts of information from a dataset compared to conventional evaluation methods. The gain is particularly strong for low S/N peak groups. The workflow is completely automated, very flexible and provides a foundation for large scale application of SRM in targeted proteomics. We tested *mProphet* in a total human cell lysate sample measuring peptides of high and low abundance as judged by their spectral counts in a shotgun LC-MS/MS experiment of OGE fractionated samples. Unlabeled synthetic peptides spiked into the heavy

labeled, unfractionated SILAC lysate served as reference. This inverted labeling experimental design allowed us to use inexpensive synthetic un-labeled peptides as reference. To our knowledge this is the first time that SRM has been applied to unfractionated human cell lysates with a large number of targeted peptides.

Evaluation of such datasets from complex matrices with high background of noise signals and hundreds of peptides measured in a large number of samples proved to be unreliable and prohibitively tedious before. Such experiments, however, are the most powerful application for the SRM technique. By providing a foundation for data analysis *mProphet* will help SRM to fulfill its true potential.

METHODS

Gold standard data set peptide selection and generation. We selected a set of 100 tryptic peptides mapping to randomly chosen yeast proteins and with a length between 6 and 20 amino acids. Peptides with a high likelihood of being observed in a mass spectrometer were chosen, based on their number of observations in the *S. cerevisiae* build of the proteomics data repository PeptideAtlas(Deutsch, Lam et al. 2008) (www.peptideatlas.org, ISB, Seattle) or on bioinformatic prediction(Mallick, Schirle et al. 2007). Peptide sequences existing in any of the protein databases of the three organisms (*L. interrogans, C. elegans, H. sapiens*) that were used as a background matrix were not included in the peptide list. The final peptide set was synthesized on a micro-scale in an unpurified form using the SPOT-synthesis technology and lyophilized in a 96-well plate format (~50 nmol of each peptide/well, JPT Peptide Technology, Berlin, Germany). Peptides were synthesized in an isotopically light and a heavy form. In each heavy peptide the C-terminal K or R residue was substituted with the corresponding heavy version with a mass shift of +8 Da or +10 Da, respectively. Peptides were re suspended in 20% acetonitrile, 1% formic acid, vortexed for 20 minutes and sonicated for 15 minutes in the 96-well plate. Two mixtures containing the 100 peptides in only light or in both light and heavy version were prepared, evaporated on a vacuum centrifuge to dryness and re solubilized in 0.1% formic acid.

Sample generation for gold standard data set. For the synthetic data set ten different samples were prepared (**Table 1**). One sample contained the 100 synthetic peptides in light and heavy form alone. For the other nine samples the 100 synthetic peptides (light and heavy) were mixed with three different backgrounds and in three different dilutions. The three different backgrounds were protein digests in similar concentrations prepared from *L. interrogans*, *C. elegans* and *H. sapiens* u2os total cell extracts.

The unpurified synthetic peptides were spiked into these backgrounds in three different dilutions as a dilution series. The amount of synthetic peptides for the highest concentration was determined empirically. The concentration maximizing the peak height of the most intense components, but still below the saturation limit of both the MS detector and chromatography was chosen. The other two dilutions were at 4-fold respectively 64-fold lower concentration of the synthetic peptides. The highest dilution was chosen at 64-fold to simulate very low intense signals, as low as 100 counts per second.

Generation of decoy transitions. For each target transition group a corresponding decoy transition group was generated such that they resembled the target transitions but did not interfere with them. Two methods of generating the decoy transitions were tested. The first method is based on reversing the peptide sequence, the second on adding random integers to Q1 and Q3.

For the first method all potential fragment ions were calculated (b- and y-ion series, z=1 or z=1-2 for triply charged peptides) and the transition that was closest to the original transition in terms of m/z was selected. Decoy transitions, with a similar Q3 to other target or decoy transitions from the same group, were removed (|Q31 – Q32| >= 1.5 Th). Decoy transitions with a similar Q3 to Q1 were removed (|Q1 – Q3| >= 5 Th). For each decoy transition 10 Th was added to Q1.

For the second method a random integer between 3 and 10 was substracted from Q1. For Q3 a random integer between -5 and 5 was added such that the difference between Q1 and Q3 was bigger than 5 Th.

By default all other attributes, like e.g. the retention time or dwell time, are copied from the corresponding target transition groups. The decoy transitions were spread equally among the LC-SRM methods, i.e. a target transition group was always followed by a decoy transition group in the method definition file.

If not otherwise mentioned the first method was used to generate the decoy transitions.

Design of SRM assays. For each peptide one precursor/fragment ion transition was calculated for each of the two main charge states (doubly and triply charged), corresponding to the first fragment ion of the y-series with m/z greater than [$m/z_{precursor}$ + 20 Th]. The precursor/fragment ion transitions were used to detect by SRM the peptides of interest in the peptide mixtures and to trigger acquisition of the corresponding full fragment ion spectra. In detail, the light peptide mixture was analyzed on a hybrid triple quadrupole/ion trap mass spectrometer (4000QTrap, ABI/MDS-Sciex, Toronto) equipped with a nanoelectrospray ion source. Chromatographic separations of peptides were performed on a Tempo nano LC system (Applied Biosystems) coupled to a 15 cm fused silica emitter, 75 μm diameter, packed with a Magic C18 AQ 5 μm resin (Michrom BioResources, Auburn, CA, USA). Peptides were loaded on the column from a cooled (4°C) Tempo autosampler and separated with a linear gradient of acetonitrile/water, containing 0.1% formic acid, at a flow

rate of 300 nl/min. A gradient from 5 to 30% acetonitrile in 30 or 45 minutes was used. The mass spectrometer was operated in SRM mode, triggering acquisition of a full MS2 spectrum upon detection of an SRM trace (threshold 300 ion counts). SRM acquisition was performed with Q1 and Q3 operated at unit resolution (0.7 m/z half maximum peak width) with 200 or 300 transitions (dwell time 10 or 7 ms/transition, respectively) per run. MS2 spectra were acquired in enhanced product ion (EPI) mode for the highest SRM transitions, using dynamic fill time, Q1 resolution low, scan speed 4000 amu/s, m/z range 300-1400, 2 scans summed. Collision energies used for both SRM and MS2 analyses were calculated according to the formulas: CE = 0.044 * m/z + 5.5 and CE = 0.051 * m/z + 0.55 (CE, collision energy, m/z, mass-to-charge ratio of the precursor ion) for doubly and triply charge precursor ions, respectively.

Fragment ion spectra collected in the QQQ MS were used to validate peptide identities and to extract optimal fragment ions for SRM analysis. MS2 data were searched with Mascot (MatrixScience, Boston, MA) against the yeast SGD database (version dated 01/26/2007). A decoy database was generated from this sub set by reverting amino acid sequences in between tryptic cleavage sites, and appended to the target database. Precursor mass tolerance was set to 2.0 Da and fragment mass tolerance to 0.8 Da. Data were searched allowing only fully tryptic termini, and no missed cleavages. The search results were validated and assigned probabilities using a cut-off for the Mascot ion score where the cut-off was defined by the proportion of assignments to decoy peptides according to Gygi *et al.*(Elias and Gygi 2007).

A spectral library was created from the peptide-spectrum matches. The spectrum with the highest ion score was selected in cases where several spectra matched to a peptide. These fragment ion spectra were used as a reference to derive the optimal coordinates of each SRM assay (e.g. best responding fragments, fragment relative intensities, peptide elution time). Fragments with m/z values close to the precursor ion m/z ($\left| m/z_{Q1} - m/z_{Q3} \right| \leq$ 5 Th) were discarded, as such transitions result in high noise levels. Transitions for one precursor with a similar Q3 were removed ($|Q3_1 - Q3_2| >= 1.5$ Th). Of the remaining transitions the five with the most intense fragment ions in the corresponding tandem mass spectra were selected. Fragments due to neutral loss from precursor were excluded. Collision energies associated to each transition were derived from the formulas given above. Additional features, such as fragment relative intensities and peptide elution times were extracted for each peptide from the corresponding MS/MS data.

Time-constrained SRM acquisition. The optimal transitions derived as previously described for the peptide set were monitored in scheduled SRM (sMRM) mode for each of the samples. A total of 2000 SRM transitions was monitored for each sample (100 peptides x (5 transitions heavy + 5 transitions light + 10 corresponding decoy transitions). The transitions were organized in four LC-MS runs, with 500 transitions (25 peptides) monitored in each run, to mimic a commonly used sMRM setup. Q1 and Q3 were operated at unit resolution and sMRM acquisition was performed using a retention time window of 300 s and a target scan time of 3 s.

Plasma handling. Plasma collection, handling and shipping was carried out by Sera Labratories International Ltd. human blood was collected from 2 healthy individuals, one male and one female, using EDTA K2 as an anticoagulant. Plasma was obtained from each blood sample by centrifuging at 2000 x g for 10 minutes at room temperature. The 2 plasmas were first pooled, and then filtered through a 0.2 µm filter, aliquoted and stored at -80°C.

Glycopeptide enrichment from human plasma. Formerly N-linked glycosylated peptides were isolated using the N-linked glycopeptide capture procedure as described by Zhang et al.(Zhang, Li et al. 2003). Glycoproteins were first oxidized by adding sodium periodate. Thereafter, the sample was conjugated to the hydrazide resin and non-glycoproteins were extensively washed off the resin. Trypsin was added to digest the glycoproteins directly on the solid-phase resin. The trypsin-released peptides were removed by a second washing procedure. N-linked glycopeptides were released using PNGase F.

Design of SRM assays for the human plasma experiment. A mix of heavy isotope labelled peptides for the target proteins was used for the development of the SRM assays. For both the doubly and triply charged species of each peptide target, one precursor-to-fragment ion transition was calculated, corresponding to the first fragment ion in the *y*-series with an *m/z* above the [$m/z_{\text{precursor}}$ + 20 Th]. The calculated transitions were used to detect by SRN the peptides of interest and to trigger acquisition of the corresponding full fragment ion spectra. The peptide mixture was analyzed on a hybrid triple quadrupole-ion trap mass spectrometer (4000QTRAP; ABI/MDS-Sciex, Toronto) equipped with a nanoelectrospray ion source. Chromatographic separations of peptides were performed by a Tempo nano LC system (Applied Biosystems) coupled to a 15 cm fused silica emitter, 75 µm diameter, packed with a Magic C18 AQ 5 µm resin (Michrom BioResources, Auburn, CA, USA). Peptides were loaded on the column from a cooled (4°C) Tempo autosampler and separated in 60 minutes by a linear gradient of acetonitrile (5 – 35%) and water, containing 0.1% formic acid at a flow rate

of 300 nL min^{-1}. The mass spectrometer was operated in SRM mode, triggering acquisition of a full MS2 spectrum upon detection of an SRM trace (threshold 400 ion counts). SRM acquisition was performed with Q1 and Q3 operated at unit resolution (0.7 *m/z* half maximum peak width) with a dwell time of 10 ms for the approximately 150 - 200 transitions per run resulting in a cycle time of 3.5 – 4 s. MS/MS spectra were acquired in enhanced product ion mode for the highest SRM transitions, using dynamic fill time, Q1 resolution low, scan speed 4000 Da s^{-1}, *m/z* range 250-1400. Collision energies (CEs) were calculated according to the formulas: $CE = 0.044 \times m/z_{precursor} + 5.5$ and $CE = 0.051 \times m/z_{precursor} + 0.55$, for doubly and triply charged precursor ions, respectively.

Acquired full MS/MS spectra were used to validate peptide identities and to extract the most intense fragment ions for subsequent SRM analysis. MS/MS data were searched with Mascot (Version 2.3.0, Matrix Science, Boston, MA) against an artificial database of the concatenated peptide sequences plus their decoy sequences. The amino acid sequences between tryptic cleavage sites were reversed to form the decoy sequences, which were appended to the target sequences. Precursor mass tolerance was set to 1.2 Da and fragment mass tolerance to 0.6 Da. Data were searched for fully tryptic termini with no missed cleavages. The search results were filtered for having a significance threshold greater than 0.05 and a Mascot ion score higher than 30. If multiple spectra identified the same peptide, then the spectrum with the highest Mascot ion score was chosen. Peptides not identified in the first run were targeted again by SRM-triggered MS/MS using two precursor-to-fragment ion transitions for the doubly and the triply charged species. These new triggers corresponded to the second fragment ion in the *y*-series with an *m/z* above the [$m/z_{precursor}$ + 20 Th] and the first fragment ion in the *y*-series with an *m/z* below the [$m/z_{precursor}$ - 20 Th]. All the spectrum-peptide matches taken together constituted the spectral library for target peptides.

These fragment ion spectra were used as a reference to derive the optimal coordinates of each SRM assay (e.g. best responding fragments, fragment relative intensities, peptide elution time). Fragments with m/z values close to the precursor ion m/z ($\left| m/z_{Q1} - m/z_{Q3} \right| \leq$ 5 Th) were discarded, as such transitions result in high noise levels. Transitions for one precursor with a similar Q3 were removed ($|Q3_1 - Q3_2| >= 1.5$ Th). Of the remaining transitions the five with the most intense fragment ions in the corresponding tandem mass spectra were selected. Fragments due to neutral loss from precursor were excluded. Collision energies associated to each transition were derived from the formulas given above.

Additional features, such as fragment relative intensities and peptide elution times were extracted for each peptide from the corresponding MS/MS data.

Target detection in plasma. The isotope labeled peptide strategy introduced by Gerber et al.(Gerber, Rush et al. 2003) was used for the quantification of the target proteins. As internal standards at least one absolutely quantified synthetic peptide (AQUA) per protein was ordered by Sigma-Aldrich (Germany) or Thermo Scientific (Germany). The heavy peptides were spiked into the N-glycosite enriched plasma sample. For each peptide the best 5 transitions for the internal standard as well as the endogenous peptide were monitored in a scheduled fashion with a retention time window of 4 min and a cycle time fixed to 2.75 s. Retention time peptides for calibrating to the adjusted gradient and decoy transitions were also monitored within each run. The total transition number per run was between 320 and 360. SRM analyses was performed on the hybrid triple quadrupole-ion trap mass spectrometer (4000QTRAP; ABI/MDS-Sciex, Toronto) equipped with a nanoelectrospray ion source. Chromatographic separations of peptides were performed by a Tempo nano LC system (Applied Biosystems) coupled to a 15 cm fused silica emitter, 75 µm diameter, packed with a Magic C18 AQ 5 µm resin (Michrom BioResources, Auburn, CA, USA). Peptides were loaded on the column from a cooled (4°C) Tempo autosampler and separated in 35 minutes by a linear gradient of acetonitrile (5 – 35%) and water, containing 0.1% formic acid at a flow rate of 300 nL min^{-1}. SRM acquisition was performed with Q1 and Q3 operated at unit resolution (0.7 *m/z* half maximum peak width).

Assignment of true peak groups in gold standard data set. In order to have a well defined data set of true peak groups we made use of the dilution series. This was necessary for the following reasons: 1) Some of the synthetic peptide signals shifted out of the retention time window in the beginning of the chromatographic gradient. 2) We wanted to include peak groups in our data set that are of very low intensity. The unambiguous assignment of those peak groups without the dilution series would be difficult. 3) Some of the synthetic peptides showed multiple signals only differing by their intensity and retention time. This is most likely an artifact of the synthesis where peptides modified with protective groups underwent in-source removal of the respective group, leading to apparent masses identical to the unmodified peptide. With the stringent dilution alignment strategy we could exclude most of these artifacts. Seven such peak groups were removed manually.

The dilution alignment and filtering for the target peak groups was done as follows: We extracted the two most intense peak groups for each transition group record including

the corresponding peak group features. After that we made use of the three dilutions by comparing each peak group to the peak groups of the two other dilutions within a certain retention time (Tr) window ($|Tr_1 - Tr_2| \leq 30$ s). If in both other dilutions there was at least one peak group with the expected signal intensity (XIC_{tot}) ratio ($|\text{delta } \log_2 XIC_{tot}| \leq 1.5$) the peak group was kept. Finally, only transition group records, where only one peak group fulfilled the prior criteria, were kept (see also **Supplementary Method 1**).

Peak group identification and feature extraction. Raw data from the ABI QTRAP was converted to mzXML(Pedrioli, Eng et al. 2004) by the program mzWiff which is part of the Trans-Proteomic Pipeline distribution(Keller, Eng et al. 2005). Ion traces for all transitions were reconstructed and mapped to an annotated transition list using the Q1 and Q3 m/z values that are reported in the mzXML file format. Annotation contained the peptide sequence, charge state of precursor and transition, ion-type (b- or y-ion and position), expected relative intensity, the expected elution time, isotope form (light or heavy), and target or decoy (yes or no).

All transition ion traces for one transition group record were binned to provide equal spacing and allow for later scoring with cross-correlation measures. Traces were smoothened by a sliding average with a window size of 3.

Transitions were grouped into transition group records based on their annotation, and peak detection was run to detect up to 5 peaks on each trace. Light and heavy labeled versions of a peptide were treated separately but assigned to the same transition group record. Decoy transitions constituted separate decoy transition group records and had no logical connection to any target transition group record. Peak groups were constructed by matching co eluting peaks from all transitions for each isotopic form (light and heavy) within one transition group by the following algorithm:

The most intense peak (seed peak) within all transitions of a transition group record (seed peak) is selected. For all other transitions matching peaks (difference of apex < 5 s) are grouped with the seed peak giving priority to the most intense peaks in ambiguous cases. Peaks that were assigned to a peak-group are then taken out. In the following round the process starts with the remaining most intense peak as seed peak and proceeds as before until all peaks are assigned to peak groups or a maximum of 5 peak groups is reached.

A peak group is not required to contain peaks from all transitions. We accepted peak groups if they contained peaks in 3 out of 5 transitions. In cases where a transition was not

represented in the peak group the trace within the boundaries of the peak group was assigned and the maximum intensity within that section was considered as apex. The rationale for that is to avoid missing data by failure of peak extraction Absence of a detectable signal is not treated as missing data but as the noise line being the maximum possible intensity present.

Peak groups derived from isotopically light and heavy labeled peptides within one transition group record were grouped based on overall similarity of elution time (difference of elution time < 5s) giving rise to pairs of peak groups (peak group pair) which represent the basis for quantification. In cases where no corresponding heavy peak group could be assigned to a light peak group a mock peak group was constructed consisting of traces of the heavy transitions within the elution time boundaries of the light peak group. Again, this serves to allow for full scoring even in cases where peak detection might have failed.

Scoring of peak groups. Peak group pairs were associated with different scores according to the described criteria of co elution, peak shape similarity, intensity and correlation of relative intensities between peak group and assay (expected intensities) or light peak group vs. heavy peak group:

Intensity score: total ion current of all peaks constituting a light peak group

Intensity correlation: product moment correlation between apex intensities of peaks in light peak group and expected intensities derived from the assay.

Co elution score: a measure of co elution based on the average shift in the cross-correlation function between each pair of peak traces. Each peak in a light peak group is cross-correlated to each other peak and the apex of the resulting cross-correlation function is taken as elution time difference. This a more global measure of elution characteristics compared to a simple comparison of apex positions, especially in cases where noisy or jagged data is compared. The pair wise comparisons result in a matrix of elution time differences. The mean plus one standard deviation is the reported score.

Shape score: a measure of global shape similarity based on the same cross-correlation matrix like for the *co elution score*. We used the property of a normalized cross-correlation that an apex intensity of 1 is an indication for perfect peak shape similarity and smaller values indicate non-corresponding peak shapes. The average of the apex intensity in the cross-correlation functions constitutes the *shape score*.

Reference correlation: Product moment correlation between apex intensities of corresponding light and heavy peaks of a peak group pair.

Retention time deviation: difference of expected and measured retention time in seconds.

Reference shape score: A measure of global shape similarity. A cross-correlation matrix is generated as described above, but only for corresponding light and heavy transitions. Average cross-correlation function apex intensities is reported as score.

Reference co elution score: Average shift of cross-correlation functions for corresponding light and heavy transitions.

Semi-supervised learning algorithm and false discovery rate estimation. The goal of *mProphet* is to maximize the number of correctly identified peak groups for a user defined quality (false discovery rate). Furthermore, *mProphet* is applicable to any type of SRM experiment (e.g. independent of the number of transitions recorded for a peptide). In order to employ an optimal classifier for each experiment, *mProphet* implements the following functionality:

1. As input the algorithm uses a vector of peak groups assigned to transition group records. Each peak group was processed and assigned with a number of sub-scores describing features of the peak group. The transition group records are either derived from measurements of target or decoy transition groups. The target and decoy transition group records are split into train and test data set in equal parts. A classifier is trained on the train data set using an iterative semi-supervised learning strategy and linear discriminant analysis. In each iteration step the decoy peak groups are used as negative set, the best target peak groups as positive set for the training (**Supplementary Fig. 10**). After an initialization step each selection of the positive set is based on the classifier of the last round. In each iteration step the peak groups for one transition group record are ranked using the current discriminant score. In each step only the top scoring peak groups are used. After a user defined number of iterations (convergence is usually reached after less ten iterations) the final classifier is applied to the complete data set.
2. The top ranked peak groups of the test data set are then used for the false discovery rate estimation. Using the data points of the decoy transitions a null distribution is parameterized. The weight of the null distribution is estimated using part of the data points of the target transitions (Storey and Tibshirani 2003) (for a comparison to a more

advanced method using the R package qvalue from Storey see **Supplementary Fig. 11**). Finally the false discovery rate and the sensitivity can be estimated using the null distribution. For a more detailed description of the semi-supervised learning and false discovery rate estimation see **Supplementary Methods 2/3**. The semi-supervised learning algorithm and false discovery rate estimation is implemented in R(2008) (Version 2.8.0) and uses the library MASS (Built 2.8.1).

Separation power in dependence of number of transitions. In order to study the influence of the number of recorded transitions on the separation power of the *intensity correlation*, we extracted only part of the signals from the gold standard data set. The highest intense three to five transitions per transition group were extracted (light and heavy version). The intensity ranking was based on the data from the assay development. For these four data sets we derived Receiver-Operating Characteristic (ROC) curves for the *intensity correlation* from only the target peak groups of the test data set and only for the top scoring peak groups of each transition group record.

Separation power in dependence of number of decoy transitions. To study the influence of the amount of recorded decoy transition on the separation power, we randomly selected decoy transition group records making up 36% (459), 20% (202), 10% (90), 5% (43) and 2% (17) of the resulting amount of data (all numbers refer to transition group records) from the gold standard data set. On these five data sets we ran *mProphet* with a 50 times 2-fold cross validation (30% false target peak groups among all target peak groups). Receiver-Operating Characteristic (ROC) curves were derived from only the target peak groups of the test data set and only for the top scoring peak groups of each transition group record.

Sample generation for human total lysate data set. Human u2os cells were grown to ~70% confluence in dishes containing Dulbecco's modified Eagle, heavy SILAC medium supplemented with 10% FBS, penicillin G (100 U/ml) and streptomycin (100 µg/ml). The cells were harvested by trypsinization and washed twice in PBS. 250 µl of cell suspension was mixed with the same volume of 2x lysis buffer (50 mM Ammonium Bicarbonate, 10 M urea, 0.1% Rapigest) and sonicated for 5 min. The resulting lysate was centrifuged in a micro centrifuge for 10 min at maximum speed. The proteins contained in the supernatant were reduced with 5 mM DTT for 45 min at 37°C and alkylated with 25 mM iodoacetamide for 45 min in the dark before diluting the sample with water to a final urea concentration below 1.5 M. Proteins were digested by incubation with trypsin (1/100, w/w) for at least 6 hours at

37°C. The peptides were cleaned up by C18 reversed-phase spin columns according to the manufacturer's instructions (Harvard Apparatus).

Absolute quantification of proteins. Human u2os cells were grown and peptide samples were prepared as described above in non-labeled medium. Cell number was determined by counting in a Neubauer Chamber. An equivalent of 60'000 cells was injected with 50 fmol of each peptide (LTWHAYPEDAENK, VQIFEYNENTR, LAEGEQILSGGVFNK, WEESGPQFITNSEEVR) spiked in. Peptides were ordered from Thermo Scientific (Germany) as absolute quantified synthetic peptides (AQUA). Absolute copy numbers were determined by the ratio of the most intense transitions of the light endogenous and heavy synthetic peptide.

ACKNOWLEDGEMENTS

We thank Johan Malmström and Marko Jovanovic for providing the samples that were used as background matrix. Furthermore, we thank Marko Jovanovic for careful reading of the manuscript. We acknowledge Manfred Claassen for discussions on the subject of machine learning. We thank Holger Wenschuh (JPT Peptide Technologies, Berlin, Germany). We'd like to thank Adam Srebniak for his help in generating a software package. This work was supported by grants from the Forschungskredit of the University of Zurich, University of Zurich Research Priority Program in Systems Biology and Functional Genomics, GEBERT-RÜF Stiftung, Swiss National Science Foundation (SNF) under Grant No. 31000-10767 and with Federal (US) funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, under contract No. N01-HV-28179 and by SystemsX.ch, the Swiss initiative for systems biology. MOH is Ernst Hadorn Endowed Professor of Molecular Biology.

TABLES AND FIGURE LEGENDS

Table 1 Samples for assay development and synthetic data set

dilution of synthetic peptides relative to highest concentration (heavy and light)	background matrix	LC-MS runs	target transitions	decoy transitions
1X	L. interrogans	4	250	250
4X	L. interrogans	4	250	250
64X	L. interrogans	4	250	250
1X	C. elegans	4	250	250
4X	C. elegans	4	250	250
64X	C. elegans	4	250	250
1X	H. sapiens	4	250	250
4X	H. sapiens	4	250	250
64X	H. sapiens	4	250	250

Samples used for the generation of the synthetic data set. Nine samples were generated as follows: We spiked in 100 synthetic peptides (in a isotopically light and heavy form) in three different concentrations into three different background matrices derived from different organisms. For each peptide five light and five heavy transitions were measured as well as ten decoy transitions derived from a non existing but similar peptide. The total 2,000 transitions measured for each sample were split up into 4 LC-SRM methods.

Figure 1. Structure of SRM data and definition of terms. (**a**) Representation of the SRM measurement of one peptide. We name the precursor (Q1) to fragment ion (Q3) transitions, used to measure one targeted peptide, a *transition group*. The data resulting from the measurement of one transition / transition group is named *trace / transition group record.* In one transition group record, a number of *peak groups* can be identified that potentially represent the peptide of interest. (**b**) Peak group features that can be used to identify a true peak group. Red indicates an unexpected behavior for true peak groups. If the peak group is derived from the targeted peptide, the peaks tend to be very similar in terms of retention time profile and shape. Further the relative intensities of the fragment ions must correspond to previously measured intensity ratios (e.g. from a consensus spectrum). If a reference peptide was in the sample, the relative intensities for all corresponding traces as well as peak shape and elution time should be similar for intrinsic peptide and reference.

Figure 2. Generation of a gold standard data set with assigned true peak groups. 100 synthetic peptides in an isotopically light and heavy form were added at three different concentrations to three different background matrices of increasing complexity. They consisted of trypsinized protein extracts from *L. interrogans*, *C. elegans* and *H. sapiens o2u2 cells, respectively.* (**a**) Dilution series of a synthetic peptide mixed into a background matrix. The peptide was measured using SRM with five transitions in three different samples and at three different concentrations. The signal intensities (square root of counts per second) in dependence of the retention time (tR) is shown for one transition group at three different dilutions. The square root is used to visualize the full intensity range. The true peptide signal at 34 minutes behaves, as expected, proportional to the peptide concentration, whereas a second signal* is constant among all three samples and thus designated a wrong peak group and neglected. (**b**) Systematic discrimination between true and false peak groups. Every peak group was compared to the peak groups of the other two dilutions (tR tolerance of 30 s) as shown here for the peak groups of the 64-fold dilution. The comparisons in black fulfill the stringent filtering regarding the expected intensity ratio patterns (delta $\log_2$ ratio tolerance of 1.5). Only transition group records with one peak group fulfilling the criteria were accepted. (**c**) Histograms of sub-score distributions of true and false peak groups detected in dilution experiments; inserts show the corresponding receiver-operator characteristic (ROC) plots and area under the curves (AUC).

Figure 3. Combining features improves the separation of true and false peak groups. (**a**) Separation of true and false target peak groups in the test data set by *mProphet* after training of a classifier with a semi-supervised learning strategy. (**b**) Receiver-Operator characteristics (ROC) plots for all the single sub-scores compared to the composite *mProphet* score. The *mProphet* score showed the best separation power. In this data set the *"reference shape score" showed* the second best separation after the *mProphet* score. (**c**) Comparison of *mProphe*t computed and true sensitivity and FDR in the test data set. (**d**) Signal to noise ratio of peak groups versus the *mProphet* score in the test data set. Signals above a signal to noise of approximately 10 were completely separated from false peak groups. 11 out of 16 true peak groups between a signal to noise ratio of 5 to 10 could be completely recovered. (**e**) Dependency of the true/false intensity correlation score separation power on the number of transitions. ROC curves are shown for the data set using three to five transitions recorded (six to ten including the heavy transitions). (**f**) Dependency of the *mProphet* separation power on the fraction of recorded decoy. ROC curves are shown for 36%, 20%,

10%, 5% and 2% decoy transition group records relative to the total amount of recorded data.

Figure 4. Separation of true from false peak group signals in a total human cell lysate using decoy transitions and *mProphet* scoring. (**a**) Decoy transition groups were designed as pairs of decoy transitions for the endogenous iso-form and the true transitions for the reference form. Both, decoy and target transition groups were thereby scored against the same reference (the spiked peptide). (**b**) Cumulated spectral counts in a shotgun MS experiment of an OGE fractionated u2os total human cell lysate of the peptides that were selected for targeting with SRM. (**c**) *mProphet* score distribution for target and decoy peak groups. Most of the target signals were separated from the decoy distribution. (**d**) Sensitivity and FDR as function of the *mProphet* score cut-off. Most peptides could be detected with a high confidence (FDR<1%). (**e**) High confidence discrimination of low S/N signals by *mProphet*. (**f**) Dependency of the *mProphet* separation on the number of transitions. ROC curves are shown using two to six transitions recorded (six to ten including the heavy transitions). (**g**) Dependency of the *mProphet* separation on the number of transitions when completely neglecting the reference peptide data. ROC curves are shown for the data set using two to six transitions recorded.

Figure 5. Measurements in a human plasma N-glycopeptide enriched sample using isotopically heavy labeled synthetic peptides as internal standard. (**a**) Measurement of 58 proteins (67 peptides) in a human N-glycopeptide enriched sample and classification according to estimated concentrations in blood. We were able to detect 55% of the targeted proteins (blue) when applying a false discovery rate cutoff of 5%. (**b**) Separation of true target signals from decoy signals using *mProphet*. (**c**) Sensitivity and false discovery rate as a function of the *mProphet* discriminant score cutoff. (**d**) Reproducibility of quantification between the two technical replicates. Identifications were filtered with a 5% false discovery rate. The ratios were calculated as endogenous signal versus internal standard signal. (**e**) Increased reproducibility of quantification when removing the transitions with the most interferences based on the AuDIT p-value score. (**f**) Increased reproducibility of quantification when removing the transitions with the most interferences based on the AuDIT cv score. (**g**) Qantile quantile plot (Q-Q plot) comparing the *mProphet* discriminant score derived from two different types of decoy transitions. The strong linearity indicates

that the two decoy types result in signals with similar score distributions. (**h**) Q-Q plot of the *reference shape score* derived from measuring yeast transitions in the human plasma sample and decoy transitions based on reversing the peptide sequence. Since there are no yeast proteins expected in the human sample it can be assumed that the yeast transitions will result in false identifications. The strong linearity in the Q-Q plot indicates that signals derived from decoy transitions are a good approximation for false signals derived from target transitions. (**i**) Q-Q plot of the *reference shape score* derived from measuring yeast transitions in the human plasma sample and decoy transitions based on adding random integers to Q1 and Q3. The strong linearity indicates that signals derived from decoy transitions are a good approximation for false signals derived from target transitions.

REFERENCES

(2008). R Development Core Team. R: A Language and Environment for Statistical Computing.

Abbatiello, S. E., D. R. Mani, et al. "Automated detection of inaccurate and imprecise transitions in peptide quantification by multiple reaction monitoring mass spectrometry." Clin Chem **56**(2): 291-305.

Addona, T. A., S. E. Abbatiello, et al. (2009). "Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma." Nat Biotechnol **27**(7): 633-41.

Aebersold, R. (2003). "Constellations in a cellular universe." Nature **422**(6928): 115-6.

Aebersold, R. (2009). "A stress test for mass spectrometry-based proteomics." Nat Methods **6**(6): 411-2.

Anderson, L. and C. L. Hunter (2006). "Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins." Mol Cell Proteomics **5**(4): 573-88.

Choi, H. and A. I. Nesvizhskii (2008). "Semisupervised model-based validation of peptide identifications in mass spectrometry-based proteomics." J Proteome Res **7**(1): 254-65.

Craig, R., J. C. Cortens, et al. (2006). "Using annotated peptide mass spectrum libraries for protein identification." J Proteome Res **5**(8): 1843-9.

Deutsch, E. W., H. Lam, et al. (2008). "PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows." EMBO Rep **9**(5): 429-34.

Domon, B. and R. Aebersold "Options and considerations when selecting a quantitative proteomics strategy." Nat Biotechnol **28**(7): 710-21.

Domon, B. and R. Aebersold (2006). "Mass spectrometry and protein analysis." Science **312**(5771): 212-7.

Elias, J. E. and S. P. Gygi (2007). "Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry." Nat Methods **4**(3): 207-14.

Eng, J. K., A. L. McCormack, et al. (1994). "An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database." J Am Soc Mass Spectrom: 976-989.

Frewen, B. E., G. E. Merrihew, et al. (2006). "Analysis of peptide MS/MS spectra from large-scale proteomics experiments using spectrum libraries." Anal Chem **78**(16): 5678-84.

Fusaro, V. A., D. R. Mani, et al. (2009). "Prediction of high-responding peptides for targeted protein assays by mass spectrometry." Nat Biotechnol **27**(2): 190-8.

Gerber, S. A., J. Rush, et al. (2003). "Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS." Proc Natl Acad Sci U S A **100**(12): 6940-5.

Hilpert, K., D. F. Winkler, et al. (2007). "Peptide arrays on cellulose support: SPOT synthesis, a time and cost efficient method for synthesis of large numbers of peptides in a parallel and addressable fashion." Nat Protoc **2**(6): 1333-49.

Jaffe, J. D., H. Keshishian, et al. (2008). "Accurate inclusion mass screening: a bridge from unbiased discovery to targeted assay development for biomarker verification." Mol Cell Proteomics **7**(10): 1952-62.

Jovanovic, M., L. Reiter, et al. "A quantitative targeted proteomics approach to validate predicted microRNA targets in C. elegans." Nat Methods **7**(10): 837-42.

Kall, L., J. D. Canterbury, et al. (2007). "Semi-supervised learning for peptide identification from shotgun proteomics datasets." Nat Methods **4**(11): 923-5.

Keller, A., J. Eng, et al. (2005). "A uniform proteomics MS/MS analysis platform utilizing open XML file formats." Mol Syst Biol **1**: 2005 0017.

Keller, A., A. I. Nesvizhskii, et al. (2002). "Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search." Anal Chem **74**(20): 5383-92.

Keshishian, H., T. Addona, et al. (2007). "Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution." Mol Cell Proteomics **6**(12): 2212-29.

Keshishian, H., T. Addona, et al. (2009). "Quantification of cardiovascular biomarkers in patient plasma by targeted mass spectrometry and stable isotope dilution." Mol Cell Proteomics **8**(10): 2339-49.

Kim, S., N. Gupta, et al. (2008). "Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases." J Proteome Res **7**(8): 3354-63.

Kuster, B., M. Schirle, et al. (2005). "Scoring proteomes with proteotypic peptide probes." Nat Rev Mol Cell Biol **6**(7): 577-83.

Lange, V., J. A. Malmstrom, et al. (2008). "Targeted quantitative analysis of Streptococcus pyogenes virulence factors by multiple reaction monitoring." Mol Cell Proteomics **7**(8): 1489-500.

Lange, V., P. Picotti, et al. (2008). "Selected reaction monitoring for quantitative proteomics: a tutorial." Mol Syst Biol **4**: 222.

MacLean, B., D. M. Tomazela, et al. "Skyline: an open source document editor for creating and analyzing targeted proteomics experiments." Bioinformatics **26**(7): 966-8.

Mallick, P., M. Schirle, et al. (2007). "Computational prediction of proteotypic peptides for quantitative proteomics." Nat Biotechnol **25**(1): 125-31.

Moore, R. E., M. K. Young, et al. (2002). "Qscore: an algorithm for evaluating SEQUEST database search results." J Am Soc Mass Spectrom **13**(4): 378-86.

Nesvizhskii, A. I., A. Keller, et al. (2003). "A statistical model for identifying proteins by tandem mass spectrometry." Anal Chem **75**(17): 4646-58.

Oberg, A. L. and O. Vitek (2009). "Statistical design of quantitative mass spectrometry-based proteomic experiments." J Proteome Res **8**(5): 2144-56.

Ong, S. E., B. Blagoev, et al. (2002). "Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics." Mol Cell Proteomics **1**(5): 376-86.

Pedrioli, P. G., J. K. Eng, et al. (2004). "A common open representation of mass spectrometry data and its application to proteomics research." Nat Biotechnol **22**(11): 1459-66.

Picotti, P., B. Bodenmiller, et al. (2009). "Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics." Cell **138**(4): 795-806.

Picotti, P., H. Lam, et al. (2008). "A database of mass spectrometric assays for the yeast proteome." Nat Methods **5**(11): 913-4.

Picotti, P., O. Rinner, et al. "High-throughput generation of selected reaction-monitoring assays for proteins and proteomes." Nat Methods **7**(1): 43-6.

Prakash, A., D. M. Tomazela, et al. (2009). "Expediting the Development of Targeted SRM Assays: Using Data from Shotgun Proteomics to Automate Method Development." J Proteome Res **8**(6): 2733-9.

Reiter, L., M. Claassen, et al. (2009). "Protein identification false discovery rates for very large proteomics data sets generated by tandem mass spectrometry." Mol Cell Proteomics **8**(11): 2405-17.

Schmidt, A., M. Claassen, et al. (2009). "Directed mass spectrometry: towards hypothesis-driven proteomics." Curr Opin Chem Biol **13**(5-6): 510-7.

Sherman, J., M. J. McKay, et al. (2009). "How specific is my SRM?: The issue of precursor and product ion redundancy." Proteomics **9**(5): 1120-3.

Sherwood, C., A. Eastham, et al. (2009). "MaRiMba: a Software Application for Spectral Library-Based MRM Transition List Assembly." J Proteome Res.

Stahl-Zeng, J., V. Lange, et al. (2007). "High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites." Mol Cell Proteomics **6**(10): 1809-17.

Stein, S. E. and P. A. Rudnick (2009). NIST Peptide Mass Spectral Libraries. National Institute of Standards and Technology, Gaithersburg, MD, 20899.

Storey, J. D. and R. Tibshirani (2003). "Statistical significance for genomewide studies." Proc Natl Acad Sci U S A **100**(16): 9440-5.

Wenschuh, H., R. Volkmer-Engert, et al. (2000). "Coherent membrane supports for parallel microsynthesis and screening of bioactive peptides." Biopolymers **55**(3): 188-206.

Whiteaker, J. R., H. Zhang, et al. (2007). "Integrated pipeline for mass spectrometry-based discovery and confirmation of biomarkers demonstrated in a mouse model of breast cancer." J Proteome Res **6**(10): 3962-75.

Wolf-Yadlin, A., S. Hautaniemi, et al. (2007). "Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks." Proc Natl Acad Sci U S A **104**(14): 5860-5.

Zhang, H., X. J. Li, et al. (2003). "Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry." Nat Biotechnol **21**(6): 660-6.