Peptide-level Robust Ridge Regression Improves Estimation, Sensitivity, and Specificity in Data-dependent Quantitative Label-free Shotgun Proteomics

Ludger J. E. Goeminne‡§¶, Kris Gevaert§¶, and Lieven Clement‡

Peptide intensities from mass spectra are increasingly used for relative quantitation of proteins in complex samples. However, numerous issues inherent to the mass spectrometry workflow turn quantitative proteomic data analysis into a crucial challenge. We and others have shown that modeling at the peptide level outperforms classical summarization-based approaches, which typically also discard a lot of proteins at the data preprocessing step. Peptide-based linear regression models, however, still suffer from unbalanced datasets due to missing peptide intensities, outlying peptide intensities and overfitting. Here, we further improve upon peptide-based models by three modular extensions: ridge regression, improved variance estimation by borrowing information across proteins with empirical Bayes and M-estimation with Huber weights. We illustrate our method on the CPTAC spike-in study and on a study comparing wild-type and ArgP knock-out Francisella tularensis proteomes. We show that the fold change estimates of our robust approach are more precise and more accurate than those from state-of-the-art summarization-based methods and peptide-based regression models, which leads to an improved sensitivity and specificity. We also demonstrate that ionization competition effects come already into play at very low spike-in concentrations and confirm that analyses with peptide-based regression methods on peptide intensity values aggregated by charge state and modification status (e.g. MaxQuant’s peptides.txt file) are slightly superior to analyses on raw peptide intensity values (e.g. MaxQuant’s evidence.txt file). 

Molecular & Cellular Proteomics 15: 10.1074/mcp.M115.055897, 657–668, 2016.

High-throughput LC-MS-based proteomic workflows are widely used to quantify differential protein abundance between samples. Relative protein quantification can be achieved by stable isotope labeling workflows such as metabolic (1, 2) and postmetabolic labeling (3–6). These types of experiments generally avoid run-to-run differences in the measured peptide (and thus protein) content by pooling and analyzing differentially labeled samples in a single run. Label-free quantitative (LFQ) workflows become increasingly popular as the often expensive and time-consuming labeling protocols are omitted. Moreover, LFQ proteomics allows for more flexibility in comparing samples and tends to cover a larger area of the proteome at a higher dynamic range (7, 8). Nevertheless, the nature of the LFQ protocol makes shotgun proteomic data analysis a challenging task. Missing values are omnipresent in proteomic data generated by data-dependent acquisition workflows, for instance because of low-abundant peptides that are not always fragmented in complex peptide mixtures and a limited number of modifications and mutations that can be accounted for in the feature search. Moreover, the overall abundance of a peptide is determined by the surroundings of its corresponding cleavage sites as these influence protease cleavage efficiency (9). Similarly, some peptides are more easily ionized than others (10). These issues not only lead to missing peptides, but also increase variability in individual peptide intensities. The discrete nature of MS1 sampling following continuous elution of peptides from the LC column leads to increased variability in peptide quantifications. Finally, competition for ionization and co-elution of other peptides with similar m/z values may cause biased quantifications (11). However, note that in this respect, using data-independent acquisition (DIA), all peptide ions (or all peptide ions within a certain m/z range, depending on the
Peptide-level Robust Regression for Label-free Proteomics

method used) are fragmented simultaneously, resulting in multiplexed MS/MS spectra (12, 13). Hence, issues of missing fragment spectra are less a problem with DIA, however, some of its challenges lie in deconvoluting MS/MS spectra and mapping their features to their corresponding peptides (14).

Standard data analysis pipelines for DDA-LFQ proteomics can be divided into two groups: spectral counting techniques, which are based on counting the number of peptide features as a proxy for protein abundance (15), and intensity-based methods that quantify peptide features by measuring their corresponding spectral intensities or areas under the peaks in either MS or MS/MS spectra. Spectral counting is intuitive and easy to perform, but, the determination of differences in peptide and thus protein levels is not as precise as intensity-based methods, especially when analyzing rather small differences (16). More fundamentally, spectral counting ignores a large part of the information that is available in high-precision mass spectra. Further, dynamic exclusion during LC-MS/MS analysis, meant to increase the overall number of peptides that are analyzed, can worsen the linear dynamic range of these methods (17). Also, any changes in the MS/MS sampling conditions will prevent comparisons between runs. Intensity-based methods are more sensitive than spectral counting (18). Among intensity-based methods, quantification on the MS-level is somewhat more accurate than summarizing the MS/MS-level feature intensities (19). Therefore, we further focus on improving data analysis methods for MS-level quantification.

Typical intensity-based workflows summarize peptide intensities to protein intensities before assessing differences in protein abundances (20). Peptide-based linear regression models estimate protein fold changes directly from peptide intensities and outperform summarization-based methods by reducing bias and generating more correct precision estimates (21, 22). However, peptide-based linear regression models suffer from overfitting due to extreme observations and the unbalanced nature of proteomics data; i.e. different peptides and a different number of peptides are typically identified in each sample.

We illustrate this using the CPTAC spike-in data set where 48 human UPS1 proteins were spiked at five different concentrations in a 60 ng protein/µl yeast lysate. Thus, when comparing different spike-in concentrations, only the human proteins should be flagged as differentially abundant (DA), whereas the yeast proteins should not be flagged as DA (null proteins). Fig. 1 illustrates the structure of missing data in label-free shotgun proteomics experiments using a representative DA UPS1 protein from the CPTAC spike-in study: missing peptides in the lowest spike-in condition tend to have rather low log2 intensity values in higher spike-in conditions compared to peptides that were not missing in both conditions, which supports the fact that the missing value problem in label-free shotgun proteomic data is largely intensity-dependent (23).

Fig. 2 shows the quantile normalized log2 intensity values for the peptides corresponding to the yeast null protein CG121 together with average log2 intensity estimates for each condition based on protein-level MaxLFQ intensities, as well as estimates derived from a peptide-based linear model.

Here, three important remarks can be made:

1) CG121 is a yeast background protein, for which the true concentration is thus equal in all conditions, which appears to be monitored as such by MaxLFQ, except in conditions 6B and 6E (for the latter, no estimate is available). The LM estimate, however, is more reliable but seems to suffer from overfitting.

2) A lot of shotgun proteomic datasets are very sparse, causing a large sample-to-sample variability. Constructing a linear model based on a limited number of observations will thus lead to unstable variance estimates. Intuitively, a small sample drawn from a given population might “accidentally” show a very small variance while another small sample from the same population might display a very large variance just by random chance. This effect is clear from the sizes of the boxes. The interquartile range is twice as large in condition 6E.

Peptide-level Robust Regression for Label-free Proteomics

FIG. 1. Missing peptides are often low abundant. The boxplots show the log2 intensity distributions for each of the 33 identified peptides corresponding to the human UPS1 protein cytoplasmic Histidyl-tRNA synthetase (P12081) from the CPTAC dataset in conditions 6A (spike-in concentration 0.25 fmol UPS1 protein/µl) and 6B (spike-in concentration 0.74 fmol UPS1 protein/µl). Vertical dotted lines indicate peptides present in both conditions. Note, that most peptides that were not detected in condition 6A exhibit low log2 intensity values in condition 6B (colored in red).
Peptide-level Robust Regression for Label-free Proteomics

FIG. 2. Effect of outliers, variability, and sparsity of peptide intensities on abundance estimations. The figure shows log2 transformed quantile normalized peptide intensities for the yeast null protein CG121 from the CPTAC data set for spike-in conditions 6A, 6B, 6C, 6D, 6E. Each color denotes a different condition. Connected crosses: average protein log2 intensity estimates for each condition are provided for a traditional protein level workflow where the mean of the protein-level MaxLFQ values was calculated (MaxLFQ, blue), the estimates of the peptide-based regression model fitted with ordinary least squares (LM, black) and the estimates of the peptide-based regression model fitted with ordinary least squares after omitting the two lowest observations in condition 6B (LM−extremes, orange). In condition 6E there were not enough data points to provide a MaxLFQ protein-level estimate. Boxes denote the interquartile range (IQR) of the log2 transformed quantile normalized peptide intensities in each condition with the median indicated as a thick horizontal line inside each box. Whiskers extend to the most extreme data point that lies no more than 1.5 times the IQR from the box. Points lying beyond the whiskers are generally considered as outliers. Note, that the presence of two low-intensity peptide observations in concentration 6B has a strong effect on the estimates for both MaxLFQ and LM.

compared to condition 6C. This issue leads to false positives since some proteins with very few observations are flagged as DA with very high statistical evidence solely due to their low observed variance (24).

(3) Two observed features at log2 intensities 14.0 and 14.3 in condition 6B have a strong influence on the parameter estimate for this condition. Without these extreme observations, the 6B estimate lies closer to the estimates in the other conditions. As missingness is strongly intensity-dependent, these low intensity values could easily become missing values in subsequent experiments. More generally, a strong influence of only one or two peptides on the average protein level intensity estimate for a condition is an unfavorable property.

These issues illustrate that state-of-the-art analysis methods experience difficulties in coping with peptide imbalances that are inherent to DDA LFQ proteomics data. We here propose three modular improvements to deal with the problems of overfitting, sample-to-sample variability and outliers:

(1) Ridge regression, which penalizes the size of the model parameters. Shrinkage estimators can strongly improve reproducibility and overall performance as they have a lower overall mean squared error compared to ordinary least squares estimators (25–27).

(2) Empirical Bayes variance estimation, which shrinks the individual protein variances toward a common prior variance, hence stabilizing the variance estimation.

(3) M-estimation with Huber weights, which will make the estimators more robust toward outliers (28).

We illustrate our method on the CPTAC Study 6 spike-in data and a published ArgP knock-out Francisella tularensis proteomics experiment and show that our method provides more stable log2 FC estimates and a better DA ranking than competing methods.

EXPERIMENTAL PROCEDURES

CPTAC Spike-in Data Set—The publicly available Study 6 of the Clinical Proteomic Technology Assessment for Cancer (29) is used to evaluate the performance of our method. Raw data can be accessed at https://cptac-data-portal.georgetown.edu/cptac/public?scope=Phase+I. In this study, the Sigma Universal Protein Standard mixture 1 (UPS1, Sigma-Aldrich, St. Louis, MO) containing 48 different human proteins was spiked into a 60 ng protein/μl Saccharomyces cerevisiae strain BY4741 (MATa, leu2Δ1, met15Δ1, ura3Δ1, his3Δ1) lysate in five different concentrations (6A: 0.25 fmol UPS1 proteins/μl; 6B: 0.74 fmol UPS1 proteins/μl; 6C: 2.22 fmol UPS1 proteins/μl; 6D: 6.67 fmol UPS1 proteins/μl; and 6E: 20 fmol UPS1 proteins/μl). These samples were sent to five independent laboratories and analyzed on seven different instruments. For convenience, we limited ourselves to the data originating from the LTQ-Orbitrap at site 86, LTQ-Orbitrap O at site 65 and LTQ-Orbitrap W at site 56. Samples were run three times on each instrument. The used dataset thus features five different samples, each analyzed in threefold on three different instruments. Raw data files were searched using MaxQuant version 1.5.2.8 (30) with the following settings. As variable modifications we allowed acetylation (protein N terminus), methionine oxidation (to methionine-sulfoxide) and N-terminal glutamine to pyroglutamate conversion. As a fixed modification, we selected carbamidomethylation on cysteine residues as all samples were treated with iodoacetamide. We used the enzymatic rule of trypsin/P with a maximum of 2 missed cleavages and allowed MaxQuant to perform matching between runs with a match time window of 0.7 min and an alignment time window of 20 min. The main search peptide tolerance was set to 4.5 ppm and the ion trap MS/MS match tolerance was set to 0.5 Da. Peptide-to-spectrum match level was set at 1% FDR with an additional minimal Andromeda score of 40 for modified peptides as these settings are most commonly used by researchers. Protein FDR was set at 1% and estimated by using the reversed search sequences. We performed label-free quantitation with MaxQuant’s standard settings. The maximal number of modifications per peptide was set to 5. As a search FASTA file we used the 6718 reviewed proteins present in the Saccharomyces cerevisiae (strain ATCC 204508/S288c) proteome downloaded from Uniprot at March 27, 2015 supplemented with the 48 human UPS1 protein sequences. Potential contaminants present in the contaminants.fasta file that comes with MaxQuant were automatically added to the search space by the software. For protein quantification in the proteinGroups.txt file, we used unique and razor
peptides and allowed all modifications as all samples originate in essence from the same yeast lysate and the same UPS1 spike-in sample.

Francisella tularensis Data Set—The data of Ramond et al. (31) is used to illustrate our method on a real biological experiment. Both raw and processed data are publicly available and can be found in the PRIDE repository at http://www.ebi.ac.uk/pride/archive/projects/PXD001584. The authors explored changes in the proteome of the facultative intracellular pathogenic cocccobacillus Francisella tularensis after gene deletion of a newly identified arginine transporter, ArgP. Both wild-type and ArgP mutants were grown in biological triplicate. Each biological replicate was analyzed in technical triplicate via label-free LC-MS/MS. Data were processed with MaxQuant version 1.4.1.2 and potential contaminants and reverse sequences were removed. In addition, only proteins present with at least two peptides in at least 9 out of the 18 replicates were retained. Subsequent data analysis via t-tests on imputed LFQ intensities was performed.

Summarization-based Analysis—MaxLFQ+Persues—This is a standard summarization-based analysis pipeline that is available in the popular MaxQuant-Persues software package (21). Briefly, the MaxQuant ProteinGroups.txt file was loaded into Persues version 1.5.1.6, potential contaminants that did not correspond to any UPS1 protein as well as reversed sequences and proteins that were only identified by site (thus only by a peptide carrying a modified residue) were removed from the data set. MaxLFQ intensities (32) were log2 transformed and pairwise comparisons between conditions were done via t-tests.

MaxLFQ+limma—The MaxQuant ProteinGroups.txt file is used as input for R version 3.1.2 (Pumpkin Helmet) (33). Potential contaminants and reversed sequences (see above) were removed from the data set. The MaxLFQ intensities were log2 transformed and analyzed in limma, an R/Bioconductor package for the analysis of microarray and next-generation sequencing data (34). Limma makes use of posterior variance estimators to stabilize the naive variance estimator by borrowing strength across proteins (see also below).

Peptide-based Model Analysis—Data Preprocessing—MaxQuant’s peptides.txt file was read into R version 3.1.2, the peptide intensities were log2 transformed and quantile normalized (35, 36). Many other normalization approaches do exist, however, comparing them is beyond the scope of this paper (24, 36–38). Reversed sequences and potential contaminants were removed from the data. For the CPTAC dataset, we only removed potential contaminants that did not map to any UPS1 protein. MaxQuant assigns proteins to protein groups using an Occam’s razor approach. However, to avoid the added complexity of proteins mapped to multiple protein groups, we discarded peptides belonging to protein groups that contained one or more proteins that were also present in a smaller protein group. Next, peptides were grouped per protein group in a data frame. Finally, values belonging to peptide sequences that appeared only once were removed as the model parameter for the peptide effect for these sequences is undistinguishable. For notational convenience, a unique protein or protein group is referred to as a protein in the remainder of this article.

Benchmark Peptide-based Model—We start from the peptide-based linear regression models as proposed by Daly et al. (39) Clough et al. (22) and Karpevitch et al. (40), of which we have independently proven their superior performance compared to summarization-based workflows (21). In general, the following model is proposed:

$$\hat{Y}_{ijklmn} = \hat{\beta}_{i}^{\text{treat}} + \hat{\beta}_{k}^{\text{pep}} + \hat{\beta}_{im}^{\text{biorep}} + \hat{\beta}_{m}^{\text{techrep}} + e_{ijklmn}$$  \hspace{1cm} (Eq. 1)

with $Y_{ijklmn}$ the $i^{th}$ log$_{2}$-transformed normalized feature intensity for the $j^{th}$ protein under the $k^{th}$ treatment (treat), the $k^{th}$ peptide sequence (pep), the $k^{th}$ biological repeat (biorep) and the $m^{th}$ technical repeat (techrep) and $e_{ijklmn}$ a normally distributed error term with mean zero and protein specific variance $\sigma_{i}^{2}$. The $\hat{\beta}$‘s denote the effect sizes for treat, pep, biorep and techrep for the $i^{th}$ protein.

Robust Ridge Model—Our novel approach improves the estimation of the model parameters in (Eq. 1) via three extensions: (1) ridge regression, which leads to shrunken yet more stable log$_{2}$ fold change (FC) estimates, (2) Empirical Bayes estimation of the variance, which further stabilizes variance estimators, and (3) M-estimation with Huber weights, which reduces the impact of outlying peptide intensities. For the robust ridge model, degrees of freedom are calculated using the trace of the hat matrix.

1. Ridge Regression—The ordinary least squares (OLS) estimates for protein $i$ are defined as the parameter estimates that minimize the following loss function:

$$\sum_{j,m} \sigma_{ijklmn}^{2} = \sum_{j,m} \left( Y_{ijklmn} - \hat{\beta}_{i}^{\text{treat}} - \hat{\beta}_{k}^{\text{pep}} - \hat{\beta}_{im}^{\text{biorep}} - \hat{\beta}_{m}^{\text{techrep}} \right)^{2}$$

(Eq. 2)

With $\hat{\theta}_{ijklmn}$ the residual errors and $X_{ijklmn}$, the row of the design matrix corresponding to observation $Y_{ijklmn}$.

Ridge regression shrinks the regression parameters by imposing a penalty on their magnitude. The ridge regression estimator is obtained by minimizing a penalized least squares loss function:

$$\sum_{j,m} \sigma_{ijklmn}^{2} + \lambda^{\text{treat}} \sum_{j} \beta_{k}^{\text{pep}}^{2} + \lambda^{\text{biorep}} \sum_{k} \beta_{im}^{2} + \lambda^{\text{techrep}} \sum_{m} \beta_{m}^{2}$$

(Eq. 3)

With each $\lambda$ a ridge penalty for the parameter estimator $\hat{\beta}$ corresponding to an effect in Eq. 1. When the $\lambda$s are larger than zero, the ridge estimators for $\hat{\beta}$ will be shrunken toward 0. This introduces some bias but reduces the variability of the parameter estimator, which makes shrinkage estimators theoretically more stable and more accurate (i.e. they tend to have a lower root mean squared error (RMSE) compared to the OLS estimator) (25–27). On the one hand, $\hat{\beta}$’s estimated by only a few observations will experience a strong correction toward 0, protecting against overfitting. On the other hand, $\hat{\beta}$’s that can be estimated based on many observations will exhibit a negligible bias because the ridge penalty will be dominated by the sum of the squared errors, which reflects that these $\hat{\beta}$’s can be estimated more reliably. We choose to tune each $\lambda$ separately because the variability on the peptide effect seems generally much larger than the variability on the other effect terms. $\lambda$ penalties can be tuned via cross-validation, but in this work, we exploit the link between mixed models and ridge regression (41) and estimate the penalties by implementing ridge regression within the lme4 package (42) in R. $\lambda$ then equals $\sigma_{ij}^{2}$ with $\sigma_{ij}$ the estimated residual variance and $\sigma_{ij}$, the estimated variance captured by the treatment effect for peptide intensities from protein $i$ (Chapter 5, 41). The standard errors of the parameter estimators and contrasts of interest are based on the bias-adjusted variance estimator (Chapter 6, 41).

2. Empirical Bayes Variance Estimations—In the introduction we argued that data sparsity can lead to unstable variance estimates. In order to stabilize the residual variance estimation, we shrink the estimated protein specific variances toward a pooled estimate over all proteins using an empirical Bayesian approach (43) implemented in the limma R/Bioconductor-package (44). In that way, the information contained in all proteins is borrowed to stabilize the variance estimates of proteins with few observations. Hence, the small variances will increase, while large variances will decrease. This avoids that proteins that exhibit a small FC and a tiny variance will appear highly
significantly. Also, the number of degrees of freedom for all these so-called moderated t-tests will increase compared to normal t-tests.

3. M-estimation With Huber Weights—As we have shown in the introduction (see Fig. 2), outlying peptides might have a severe impact on the parameter estimates, especially in conditions with few identified features. We therefore propose to adopt M-estimation with Huber weights to diminish the impact of outlying observations. Combining ridge regression with M-estimation leads to the following penalized weighted least squares loss function:

\[ \sum_{j,k,m,n} w_{ijklmn} \beta_{ijklmn}^2 + \lambda_{\text{tech}} \sum_{j} \beta_{ij}^{\text{tech}} + \lambda_{\text{pep}} \sum_{k} \beta_{ik}^{\text{pep}} + \lambda_{\text{cond}} \sum_{l} \beta_{il}^{\text{cond}} + \lambda_{\text{error}} \sum_{m} \beta_{im}^{\text{error}} \]  

(Eq. 4)

Herein, \( w_{ijklmn} \) is a Huber weight that weighs down observations with high residuals.

**False Discovery Rate**—For both peptide-based models and MaxLFQ-limma, \( p \) values are adjusted for multiple testing with the Benjamini-Hochberg FDR procedure (45). Perseus uses a permutation-based FDR that turns out to be very close to the Benjamini-Hochberg procedure. The FDR is controlled at the 5% level in all analyses.

**RESULTS**

We decided to compare our method to three competing methods using data from the CPTAC spike-in benchmark study. Additionally, its advantages are demonstrated in a case study that was originally analyzed with a standard summarization-based approach. For peptide-based models, the use of MaxQuant’s peptides.txt file, which contains summed up peptide level data, slightly increased the discriminative power as compared to analyses using the MaxQuant’s evidence.txt file, which contains individual intensities for each identified feature (supplemental Fig. S7, File S1). Therefore, all peptide-based models are based on the peptides.txt file.

1. **Evaluation Using Data From a Spike-in Benchmark Study**—The performances of the different methods are assessed by comparing different spike-in concentrations in the CPTAC Study 6 data set. This data set consists of identical samples containing a trypsin-digested Saccharomyces cerevisiae proteome spiked with different concentrations (0.25, 0.74, 2.22, 6.67, and 20 fmol/\( \mu \)l) of a trypsin-digested UPS1 mix containing 48 human proteins. As the high spike-in concentrations are known to suffer from ionization competition effects (18, 21), we focus mainly on comparing 6B-6A, 6C-6A, and 6C-6B, which have the lowest spike-in concentrations.

We compared our robust ridge approach to three competing approaches: (1) MaxLFQ-Perseus, (2) MaxLFQ-limma, and (3) a peptide-based linear model (LM). (1) MaxLFQ-Perseus is a standard summarization-based approach where MaxLFQ normalized log\(_2\) protein intensities are compared between conditions by FDR-corrected t-tests. (2) MaxLFQ-limma is a summarization-based approach in which MaxLFQ protein level data are analyzed via limma (34), an R/Bioconductor package that can stabilize variance estimation by borrowing information across proteins via an empirical Bayesian approach. (3) The linear model (LM) is a peptide-based linear regression model (Eq. 1) that is known to outperform summarization-based approaches (21). This model contains a treatment effect, a peptide effect and an instrument effect, and its structure is motivated in supplemental File S1. Note, that our robust ridge method (RR) is an improvement of the LM approach by implementing ridge regression, empirical Bayes variance estimation and M-estimation with Huber weights. We compared the results of the four methods in terms of precision and accuracy of the log\(_2\) fold change estimates, as well as sensitivity and specificity.

**Precision and Accuracy**—Log\(_2\) fold change (FC) estimates using our robust ridge model for yeast null proteins are clearly more precise compared to the other methods; the interquartile range of the log\(_2\) FC estimates is on average three times smaller compared to the LM fit and 4.5 times smaller for comparison 6B-6A; Fig. 3). The accuracy is also increased as most DA estimates for yeast null proteins are estimated very close to zero. This is due to an effect of the ridge penalty, which strongly shrinks the estimates for null proteins identified by a few peptides toward zero. Fig. 3 also shows a general trend for each method: as spike-in concentration differences increase, a negative bias of the log\(_2\) FC estimates appears. This likely reflects ionization suppression effects (18, 21). Note that for our method, the log\(_2\) FC distributions in all comparisons (including 6B versus 6A) in Fig. 3 are skewed toward negative fold changes and that the skewness increases with higher spike-in concentrations, suggesting that ionization suppression effects already occur at very low spike-in concentrations.

When assessing the DA UPS1 proteins (Fig. 4), the log\(_2\) FC estimates for the summarization-based approaches MaxLFQ-Perseus and MaxLFQ-limma are always more biased and more variable compared to the LM and RR methods, except in condition 6B-6A, where MaxLFQ-limma has the lowest bias of all approaches. Median log\(_2\) FC estimates for UPS proteins are very comparable between our RR and the LM method. For eight out of ten comparisons, the RR estimates are even closer to the true log\(_2\) FC. The interquartile range of the log\(_2\) FC estimates of the DA proteins for RR is on average 1.2 times smaller compared to those of the LM model. Thus, shrinkage estimation does not negatively affect estimates for proteins with a strong evidence for DA.

**Sensitivity and Specificity**—The sensitivity and specificity of the different methods are assessed using receiver operator characteristics (ROC) curves (Fig. 5, Tables I and II). For the summarization-based approaches, it turns out that MaxLFQ-limma outperforms MaxLFQ-Perseus. This is not surprising, as it has been shown that limma outperforms standard t-tests, also in proteomics data sets (46). As we have shown before, both peptide-based regression models outperform the summarization-based approaches (21). Our RR method further improves on the LM model in comparison
6B-6A, in which the detection of DA is most challenging since it involves the two lowest spike-in concentrations. For all other comparisons LM and RR have a similar performance (Table II), which was expected because the ROC curves of the LM method for these comparisons are already very steep.

FDR Control—None of the adopted methods are able to control the true FDR at the nominal 5% level in the majority of the comparisons (Table III). MaxLFQ + Perseus can only control the FDR at the 5% level in comparisons 6C-6A and 6C-6B. MaxLFQ + limma only controls the FDR accurately in comparison 6C-6B and both LM and RR can control the FDR only in comparisons 6B-6A and 6C-6B. When comparing RR and LM, RR does a better job in controlling the FDR in comparisons 6B-6A, 6C-6A, 6D-6B and 6D-6C, but not for the other comparisons.

2. Case Study—We further illustrate the performance of our novel method on true biological data in which a single trigger was expected to have an impact on several tightly regulated...
Molecular & Cellular Proteomics 15.2

Peptide-level Robust Regression for Label-free Proteomics

Fig. 5. Comparison of sensitivity and specificity. Receiver operator characteristic (ROC) curves show the superior performance of our robust ridge approach compared to other standard data analysis techniques for comparisons of conditions 6B-6A, 6C-6A, and 6C-6B in the CPTAC spike-in study. Stars denote the cut-offs at an estimated 5% FDR level. Blue (MaxLFQ + Perseus): protein-level analysis consisting of MaxLFQ normalization followed by t-tests in Perseus, yellow (MaxLFQ + limma): protein-level analysis consisting of MaxLFQ normalization followed by limma analysis, black (LM): peptide-based linear regression model containing treatment, peptide and instrument effects, purple (RR): peptide-based ridge regression model containing treatment, peptide and instrument effects with empirical Bayes variance estimator and M-estimation with Huber weights.

TABLE I

| Comparison | MaxLFQ + Perseus | MaxLFQ + limma | LM | RR |
|------------|-----------------|----------------|----|----|
| 6B-6A      | 0.536           | 0.634          | 0.817 | 0.862 |
| 6C-6A      | 0.583           | 0.672          | 0.877 | 0.869 |
| 6D-6A      | 0.583           | 0.680          | 0.880 | 0.878 |
| 6E-6A      | 0.564           | 0.660          | 0.883 | 0.880 |
| 6C-6B      | 0.607           | 0.655          | 0.860 | 0.861 |
| 6D-6B      | 0.618           | 0.657          | 0.858 | 0.859 |
| 6E-6B      | 0.601           | 0.644          | 0.863 | 0.863 |
| 6D-6C      | 0.654           | 0.680          | 0.856 | 0.860 |
| 6E-6C      | 0.663           | 0.678          | 0.864 | 0.864 |
| 6E-6D      | 0.683           | 0.685          | 0.936 | 0.937 |

TABLE II

| Comparison | MaxLFQ + Perseus | MaxLFQ + limma | LM | RR |
|------------|-----------------|----------------|----|----|
| 6B-6A      | 0.061           | 0.075          | 0.083 | 0.091 |
| 6C-6A      | 0.076           | 0.088          | 0.096 | 0.097 |
| 6D-6A      | 0.077           | 0.089          | 0.097 | 0.097 |
| 6E-6A      | 0.076           | 0.089          | 0.097 | 0.097 |
| 6C-6B      | 0.081           | 0.086          | 0.096 | 0.096 |
| 6D-6B      | 0.082           | 0.086          | 0.096 | 0.096 |
| 6E-6B      | 0.081           | 0.087          | 0.096 | 0.096 |
| 6D-6C      | 0.076           | 0.083          | 0.094 | 0.095 |
| 6E-6C      | 0.088           | 0.090          | 0.095 | 0.095 |
| 6E-6D      | 0.090           | 0.092          | 0.092 | 0.093 |

But highly interconnected pathways. Thus, contrary to a spike-in data set where differential abundance typically heads in one direction (either up or down-regulated) and stays limited to the spiked-in proteins, a biological dataset consists of a plethora of both strongly as well as weakly differentially regulated proteins. Moreover, in the CPTAC data set, the same sample was always used to spike in different amounts of UPS1 proteins in the same yeast background instead of isolating a new yeast proteome each time. Hence, variability in biological repeats will be much larger compared to the spike-in data set. Although detection of differential abundance in real biological data is the ultimate goal, there is no known ground truth available for these data sets and our evaluation is based on visual inspection of selected findings.

We made use of the publicly available data published by Ramond et al. (31) in which the authors compared the proteome of Francisella tularensis mutant for the arginine transporter ArgP to wild-type (WT) bacteria in biological triplicate. For each biological repeat, three technical replicates were also available. We compared the authors’ results with the results of our RR method based on the authors-supplied peptides.txt instead of re-searching the raw spectra. In their study, Ramond et al. (31) performed the analysis at the protein level using intensities of at least two different peptides and
keeping those proteins with at least 9 out of 18 valid values, and as such analyzed 842 proteins in total. Within our approach—analysis at the peptide level and filtering out peptides that appeared only once in the data set—we were able to analyze a total of 989 proteins. Both the results of our ranking as well as the ranking from the original Francisella article can be found in supplemental File S2. When we compared the top 100 proteins with the lowest p values in both methods, only 52 proteins overlapped. Ramond et al. (31) found 309 DA proteins at the 5% FDR level, whereas we only found 159 proteins significantly DA at the same FDR level. Thus, our method appears to be more conservative.

We evaluated the differential abundance of the ten proteins that were present in the RR DA list, but not in the original DA list, as well as the ten highest ranked proteins from the original DA list missing in our list. The ten proteins that were only discovered with our method were all lost during the preprocessing procedure of Ramond et al. (31). Among those proteins, six are more highly abundant in the mutant: exodeoxyriboNuclease V subunit gamma and ABC transporter membrane protein, as well as four hypothetical proteins (the membrane protein FTN_0835, an AAA+ superfamily member FTN_0274, an alpha/beta hydrolase FTN_0721 and FTN_1244). Four out of the ten proteins that were only discovered by our method have a lower abundance in the mutant. These proteins are Radical SAM superfamily protein, DNA helicase II, C32 tRNA thiolase and the hypothetical protein FTN_0400. Log2 intensity plots of the individual peptide intensities suggest a differential abundance for most of these proteins (supplemental Figs. S14-S23, File S1).

Eight out of the ten highest-ranked proteins in the original DA list all seem to be highly abundant gauged by the number of peptides identified (supplemental Figs. S24-S33, File S1).

Further inspection reveals that most of these proteins do not appear to bear strong evidence for DA between WT and mutant. Except for the hypothetical protein FTN_1397 and the Mur ligase family protein, all variance estimates are smaller than 0.01, which leads to extreme T statistics. Empirical Bayesian variance estimation does increase these small variances, but not enough to make them insignificant at the 5% significance level. There seems to be a relatively clear effect for hypothetical protein FTN_1397 (supplemental Fig. S31, File S1) in our method just did not pass the 5% FDR level (p value of 9.4*10^-3 and FDR adjusted p value 0.057). The Mur ligase family protein shows no strong visual DA (supplemental Fig. S32, File S1), but combines a moderate DA estimate (~0.34) with a small variance (0.01). The two other proteins only present in the list of Ramond et al. (31) are rather low-abundant. DNA-binding protein HU-beta (supplemental Fig. S30, File S1) shows no visual evidence for DA at all, but hypothetical protein FTN_1199 (supplemental Fig. S33, File S1) might possess weak evidence for DA.

Ramond et al. (31) also noted that several protein modules had an increased enrichment in DA proteins. In fact, all ribosomal proteins and all proteins involved in branched-chain amino acid (BCAA) synthesis were either unchanged (as the p value did not reach the 5% FDR cut-off) or present in lower amounts in the mutant. Based on their data, one could also suspect an up-regulation of several tricarboxylic acid (TCA) cycle proteins (nine out of 12) in the mutant.

**Ribosomal Proteins**—It is known that a global down-regulation of ribosome synthesis is a common response to nutrient starvation in Bacteria and Eukarya (47, 48). One might thus expect a similar drop in abundance in the arginine transporter-mutated *F. tularensis* as its delay in phagosomal escape can be fully restored by supplementation of the medium with excess arginine (31). When considering all 30S and 50S ribosomal proteins (both significant and insignificant in terms of DA), both our method and the original article report log2 FC estimates for 49 ribosomal proteins. As expected, all log2 FC estimates from our method pointed toward down-regulation in the mutant. Contrary, in the analysis of Ramond et al. (31), two proteins, ribosomal protein L7/L12 and 30S ribosomal protein S18 showed, although deemed insignificant, quite large log2 FCs (0.12 resp. 0.54) in favor of up-regulation in the mutant. This again suggest a more stable log2 FC estimation by our method (supplemental Fig. S34, File S1).

**BCAA Synthesis Proteins**—Based on KEGG, we could only identify nine proteins as BCAA synthesis proteins although the authors report on 12 BCAA synthesis proteins. As we were unable to find the remaining 3 proteins in this pathway, we further focus on nine proteins only. Supplemental Fig. S35, File S1 shows the distribution of the log2 FC estimates for both the methodology of Ramond et al. (31) and our method.

| Comparison | MaxLFQ + Perseus | MaxLFQ + limma | LM | RR |
|-----------|-----------------|----------------|----|----|
| 6B-6A     | 0.071           | 0.048          | 0.034 | 0.030 |
| 6C-6A     | 0.040           | 0.083          | 0.095 | 0.050 |
| 6D-6A     | 0.456           | 0.794          | 0.466 | 0.050 |
| 6E-6A     | 0.917           | 0.922          | 0.870 | 0.883 |
| 6C-6B     | 0.037           | 0              | 0    | 0.024 |
| 6D-6B     | 0.640           | 0.879          | 0.528 | 0.494 |
| 6E-6B     | 0.918           | 0.924          | 0.863 | 0.870 |
| 6D-6C     | 0.429           | 0.799          | 0.481 | 0.434 |
| 6E-6C     | 0.886           | 0.906          | 0.842 | 0.848 |
| 6E-6D     | 0.321           | 0.584          | 0.386 | 0.561 |
Here, one insignificant protein, dihydroxy-acid dehydratase shows a log₂ FC of 0.22. Using our method, all log₂ FCs indicate either down-regulation in the mutant (i.e. log₂ FCs smaller than 0) or no change at all (log₂ FCs smaller than 1*10⁻¹⁷). On average, log₂ FC estimates for BCAA synthesis proteins are more negative in our method. Thus, our method provides stronger evidence for down-regulation of some BCAA synthesis proteins in the mutant compared to the method of the original article.

TCA Cycle Proteins—Based on KEGG, we identified 13 TCA cycle proteins, while Ramond et al. (31) reported 12 proteins. When assessing the log₂ FC estimates of these authors for the 13 proteins, 10 TCA cycle proteins appear to be up-regulated in the mutant (supplemental Fig. S36, File S1), of which 8 are of declared significant. Contrary, in our method, only 1 protein of the TCA cycle (2-oxoglutarate dehydrogenase complex, E2 component, dihydrolipoyltranssuccinase) is found significant. Strikingly, 8 out of 13 TCA cycle proteins even show log₂ FC estimates that are in absolute value smaller than 1*10⁻⁹. We therefore zoomed in on the individual log₂ FC estimates of the peptides mapping to these proteins (supplemental Figs. S37-S50, File S1).

2-oxoglutarate dehydrogenase complex, E2 component, dihydrolipoyltranssuccinase (supplemental Fig. S37, File S1) is the only TCA cycle protein that is found at significantly different levels by our analysis. It is also denoted as significant in the original paper’s methodology. Nonetheless, the evidence does not seem to be very strong. 2-oxoglutarate dehydrogenase E1 component, dihydrolipoamide acetyltransferase and succinate dehydrogenase iron-sulfur subunit are also denoted as significant in the original analysis with p values of 5.3*10⁻¹⁰, 8.1*10⁻⁴ and 1.1*10⁻³ respectively, although the spectral evidence also appears quite weak (supplemental Figs. S38, S40, and S41, File S1). In our opinion, any visual evidence for differential abundance for malate dehydrogenase, aconitate hydratase, succinyl-CoA synthetase, alpha subunit and isocitrateg dehydrogenase is negligible. Nonetheless, p values corresponding to DA for these proteins are estimated at 7.4*10⁻⁴, 3.6*10⁻³, 0.02 and 0.02 respectively in the paper of Ramond et al. (31) (supplemental Figs. S42, S43, and S47, File S1).

DISCUSSION

In this work, we introduced three extensions to existing peptide-based linear models that significantly improve stability and precision of fold change estimates. These extensions include minimizing a penalized least squares loss function (ridge regression), weighing down outliers via M-estimation with Huber weights and variance stabilization via empirical Bayes. Our estimation approach is inevitably computationally more complex than the linear regression model, albeit much faster than fully Bayesian approaches that have to be fitted by computationally intensive Markov chain Monte Carlo algorithms (49). In this contribution, we normalized log₂ transformed peptide intensities using quantile normalization. Many other types of transformation and normalization exist and can be adopted prior to applying our method. Our focus however, is on robust estimation procedures for peptide-based linear models. Therefore, a thorough comparison of preprocessing methods is beyond the scope of this paper.

We compared our novel estimation method to three other methods: a standard protein-level summarization followed by t-tests (MaxLFQ+Perseus), a standard protein-level summarization followed by limma (MaxLFQ+limma) and a peptide-based linear regression model. Evaluation was done based on the CPTAC Study 6 data set, where the ground truth is known as well as on a biological data set, where ArgP mutated versus wild-type Francisella tularensis proteomes are compared. For the CPTAC dataset, we found our method to give more precise and more stable log₂ FC estimates for both the yeast null proteins and the spiked-in UPS1 proteins. Thus, our method sufficiently shrinks abundance estimates that are driven by data sparseness in the null proteins, while retaining good DA estimates when sufficient evidence is present (such as for most UPS1 proteins in the CPTAC spike-in study). These findings are supported by theory as ridge regression shrinkage indeed reduces the variability of the estimator and generates overall more precise FC estimates (lower overall RMSE compared to ordinary least squares estimators) (25–27). Furthermore, empirical Bayes variance estimation squeezes the individual residual variances of all models toward a pooled variance which stabilizes the variance estimation. Finally, M-estimation with Huber weights weakens the impact of individual outliers.

The systematic underestimation of the log₂ FC estimates provided by our method, even in comparison 6B-6A (Fig. 3), suggests that ionization competition effects can already come into play at spike-in concentration levels of 0.74 fmol/μl (condition 6B). These effects might partly explain why none of the considered methods performs well in controlling the FDR at the nominal 5% level. When analyzing the Francisella dataset, RR is more conservative than the method used by Ramond et al. (31), which might suggest a better protection against false positives. Indeed, ionization competition effects are also relevant for true biological data sets as an increase in a number of highly abundant proteins in a certain condition might generate a downwards bias in peptide intensities corresponding to non-DA proteins in this condition. Researchers should thus carefully reflect whether a low log₂ fold change is truly biologically relevant, as ionization suppression effects already seem to appear at low differences in concentration. Indeed, even when disregarding these effects, the abundance of a protein typically has to differ by a reasonable amount to be of interest to a researcher. Therefore, we suggest testing against a minimal log₂ FC value that is biologically interesting in a particular experiment; e.g. 0.5 or 1 (50). A similar approach can be easily adopted within our framework, but we have
chosen to test against an FC of 0 to make our results comparable with the analysis of Ramond et al. (31).

In practice, only a handful of true positives will typically be selected for further experimental validation. Therefore, the ranking that is produced by a method is more important than its capability to accurately control the FDR at the 5% level. Here, RR produces superior ranking lists compared to all other methods except in comparisons 6D-6A, 6E-6A and 6E-6B, where large ionization suppression effects are expected due to huge spike-in concentration differences. Hence, it will be very difficult to discern the ionization bias from real DA for these comparisons. Each component of our method contributes to an improvement in performance. ROC curves in supplemental Fig. S4, File S1 show that empirical Bayesian variance estimation slightly but consistently improves the performance of the LM model, while M-estimation seems to cause the largest gain (supplemental Fig. S5, File S1). Ridge regression clearly improves the LM model, EB variance estimation slightly improves the ridge regression model while M-estimation again seems to deliver the largest gain in performance. Corresponding AUC and pAUC values can be found in (supplemental Tables S7-S10, File S1).

In the CPTAC case study, we also confirmed that a peptide-based model starting from summed up intensities over different charge states and modifications but corresponding to identical peptide sequences in the same sample (peptides.txt file) tends to have a higher discriminative ability than a model starting from the individual feature intensities (evidence.txt file, see supplemental Fig. S7, File S1). Others have also shown that analysis on the lowest level of summarization does not automatically lead to the best performance (18).

We also applied our method on a biological dataset and compared our results to the performance of the MaxLFQ summarization-based method used in the original publication (31). Proteins identified as DA by our method that were missed in the original publication were all filtered out because too few peptides were identified. Most of these proteins contain relatively strong evidence for DA based on visual inspection of the individual quantile normalized peptide intensities, illustrating that our method can reliably discover DA in proteins that suffer considerably from missing peptides across samples. Proteins which were denoted as DA in the original article but not by our method were typically identified by a lot of peptides but with a weak visual evidence for DA. Indeed, when a lot of peptide intensities are present, modeling at the peptide level does not contribute much to stabilize the overall log₂ FC estimate. Mostly though, these small fold changes are not biologically relevant. However, when one seeks DA in sparse data, as is the case in most experiments, our method clearly outperforms classical approaches. We indeed identify possibly interesting effects in low abundant proteins without overfitting, omitting the need for extensive a priori data filtering (e.g. dropping proteins with less than two peptides present in at least nine out of 18 samples as done by Ramond et al. (31)). Consequently, our ability to detect DA in low abundant proteins combined with robustness against irrelevant changes in high abundant proteins is a favorable property.

We also showed that our fold change estimates are more stable for both the ribosomal and the BCAA synthesis modules, which are denoted as DA by Ramond et al. (31). For TCA cycle proteins, nine out of 12 were described as significantly more abundant in the mutant by these authors, while in our method, eight out of 13 TCA cycle proteins have log₂ FC estimates that are in absolute value smaller than 1 × 10⁻⁹. Upon visual inspection, most of these proteins indeed contain very limited evidence for DA. Our method thus appears to provide more reliable data for follow-up analysis, thereby aiding researchers in drawing more correct conclusions. The fact that almost 82% of the proteins in our DA list of Fransella proteins indeed contain less peptides in the condition with the lowest abundance of this protein, advocates the inclusion of imputation or censoring approaches, or even the combination of our method with estimates derived from spectral counting, which could be used as a rough but very simple validation technique. For example, hypothetical protein FTN_0400 could be a false positive because more peptides are identified in the mutant, although it appears to be less abundant (supplemental Fig. S18, File S1). This could be an indication of intensity-dependent censoring in the WT. Just like the peptide-based linear regression model, our RR model can handle missing values. The models we presented here assume missingness completely at random, an assumption that is flawed when analyzing shotgun proteomics data. Note, however, that peptide-based models partially correct for this by incorporating peptide-specific effects. Moreover, our strategies to improve robustness of the estimators can be easily plugged into censored regression methods or estimation approaches that adopt advanced imputation techniques for handling missing data (40).

Another interesting outlook is to model all proteins together by incorporating pathway or module level effects in the model in order to make stronger inferences on individual proteins belonging to a certain pathway. Finally, we want to stress the importance of data exploration. Plots of log₂ peptide intensities for proteins that are flagged as differential abundant are very useful for assessing the biological relevance and the degree of belief one can have in the DA proteins that are returned by a method.

We are currently preparing an R/Bioconductor package for our method. Meanwhile, all code and data needed to repeat the data analysis in this manuscript is available in Supplemental File 3.

*This research was supported in part by IAP research network “StU DyS” grant no. P7/06 of the Belgian government (Belgian Science Policy) and the Multidisciplinary Research Partnership “Bioinformatics: from nucleotides to networks” of Ghent University. L.G. is supported by a Ph.D. grant from the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaan-
deren) entitled “Differential proteomics at peptide, protein and module level” (141573).

This article contains supplemental Files S1 to S3.

REFERENCES

1. Oda, Y., Huang, K., Cross, F. R., Cowburn, D., and Chait, B. T. (1999) Accurate quantitation of protein expression and site-specific phosphorylation. Proc. Natl. Acad. Sci. U.S.A. 96, 6591–6596

2. Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol. Cell. Proteomics 1, 376–386

3. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat. Biotechnol. 17, 994–999

4. Hsu, J.-L., Huang, S.-Y., Chow, N.-H., and Chen, S.-H. (2003) Stable-isotope dimethyl labeling for quantitative proteomics. Anal. Chem. 75, 6843–6852

5. Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol. Cell. Proteomics 3, 1154–1169

6. Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., and Hamon, C. (2003) Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal. Chem. 75, 1895–1904

7. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., and Kuster, B. (2007) Quantitative mass spectrometry in proteomics: a critical review. Anal. Bioanal. Chem. 389, 1017–1031

8. Patel, V. J., Thalasinos, K., Stade, S. E., Connolly, J. B., Crombie, A., Murrell, J. C., and Scrivens, J. H. (2009) A comparison of labeling and label-free methods for quantifying human proteins by shotgun proteomics. J. Proteome Res. 8, 3752–3759

9. Rodríguez, J., Gupta, N., Smith, R. D., and Pevzner, P. A. (2008) Does trypsin cut before proline? J. Proteome Res. 7, 300–305

10. Abaye, D. A., Pullen, F. S., and Nielsen, B. V. (2011) Peptide polarity and the position of arginine as sources of selectivity during positive electrospray ionization mass spectrometry. Rapid Commun. Mass Spectrom. 25, 3597–3608

11. Schliekelman, P., and Liu, S. (2013) Quantifying the effect of competition for detection between coeluting peptides on detection probabilities in mass-spectrometry. Mol. Cell. Proteomics 12, 2575–2584

12. Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol. Cell. Proteomics 1, 376–386

13. Clarke, T., Key, M., Ott, I., Ragg, S., Schadow, G., and Vitek, O. (2009) Protein quantification in label-free LC-MS experiments generated by metabolic labeling. Mol. Cell. Proteomics 8, 2227–2242

14. Zhang, Y., Fonslow, B. R., Shan, B., Baek, M.-C., and Yates, J. R. (2013) Protein analysis by shotgun/bottom-up proteomics. Chem. Rev. 113, 2343–2394

15. Goeminne, L. J. E., Argentini, A., Martins, L., and Clement, L. (2015) Peptide-level Robust Regression for Label-free Proteomics. Mol. Cell. Proteomics 15, 2134–2144

16. Old, W. M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K. G., Mendoza, A., Savitski, M., and Kuster, B. (2012) Quantitative mass spectrometry in proteomics: Performance, pitfalls, and data analysis guidelines. J. Proteome Res. 14, 2457–2465

17. Pfeiffer, A., and Mann, M. (2009) Normalization and statistical analysis of quantitative proteomics data present. Mol. Cell. Proteomics 8, 2575–2584

18. Paulovich, A. G., Billheimer, D., Ham, A.-J. L., Vega-Montoto, L., Rudnick, P. A., Tabb, D. L., Wang, P., Blackman, R. K., Bunk, D. M., Cardasis, H. L., Claus, K. R., Kinsinger, C. R., Schilling, B., Tegeler, T. J., Varyiath, A. M., Wang, M., Whiteaker, J. R., Zimmerman, L. J., Fenyo, D., Carr, S. A., Fisher, S. J., Gibson, B. W., Mesi, M., Neubert, T. A., Regnier, F. E., Rodriguez, H., Spiegelman, C., Stein, S. E., Tempst, P., and Liebler, D. C. (2010) Interlaboratory study characterizing a yeast performance standard for benchmarking LC-MS platform performance. Mol. Cell. Proteomics 9, 242–253

19. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. 26, 1367–1372

20. Amaratunga, D., and Cabrera, J. (2001) Analysis of data from viral DNA microarrays. J. Proteome Res. 1, 277–286

21. Goeminne, L. J. E., Argentini, A., Martens, L., and Clement, L. (2015) Peptide-level Robust Regression for Label-free Proteomics. Mol. Cell. Proteomics 15, 2134–2144

22. Callister, S. J., Barry, R. C., Adkins, J. N., Johnson, E. T., Qian, W.-j., Raftery, M. J., and Chvalclochi, R. (2009) Normalization and statistical analysis of quantitative proteomics data present. Mol. Cell. Proteomics 8, 2575–2584

23. Karpievitch, Y. V., Dabney, A. R., and Smith, R. D. (2012) Normalization and missing value imputation for label-free LC-MS analysis. BMC Bioinformatics 13, S5

24. Ting, L., Cowley, M. J., Hoon, S. L., Gillhaus, M., Raftery, M. J., and Chvalclochi, R. (2009) Normalization and statistical analysis of quantitative proteomics data present. Mol. Cell. Proteomics 8, 2227–2242

25. Ahmed, S. E., and Raheim, S. M. E. (2012) Shrinkage and absolute penalty estimation in linear regression models. Computational Stat. 4, 541–553

26. Stein, C. (1956) Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics, pp. 197–206, University of California Press, Berkeley, Calif.

27. Copas, J. B. (1983) Regression, prediction and shrinkage. J. Roy. Statist. Soc. A 145, 311–354

28. Huber, P. J. (1964) Robust estimation of a location parameter. The Annals of Mathematical Statistics 35, 73–101

29. Paulovich, A. G., Billheimer, D., Ham, A.-J. L., Vega-Montoto, L., Rudnick, P. A., Tabb, D. L., Wang, P., Blackman, R. K., Bunk, D. M., Cardasis, H. L., Claus, K. R., Kinsinger, C. R., Schilling, B., Tegeler, T. J., Varyiath, A. M., Wang, M., Whiteaker, J. R., Zimmerman, L. J., Fenyo, D., Carr, S. A., Fisher, S. J., Gibson, B. W., Mesi, M., Neubert, T. A., Regnier, F. E., Rodriguez, H., Spiegelman, C., Stein, S. E., Tempst, P., and Liebler, D. C. (2010) Interlaboratory study characterizing a yeast performance standard for benchmarking LC-MS platform performance. Mol. Cell. Proteomics 9, 242–253

30. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. 26, 1367–1372

31. Raymon, E., Gesbert, G., Guerra, C. I., Chhuon, C., Dupuis, M., Rigard, M., Henry, T., Barel, M., and Charbit, A. (2015) Importance of host cell arginine uptake in Francisella phagocasmal escape and ribosomal protein amounts. Mol. Cell. Proteomics 14, 870–881

32. Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., and Mann, M. (2014) Accurate proteome-wide label-free quantification by delayed normal- and maximal peptide ratio extraction, termed MaxLFQ. Mol. Cell. Proteomics 13, 2531–2526
Improved normalization of systematic biases affecting ion current measurements in label-free proteomics data. Mol. Cell. Proteomics 13, 1341–1351

39. Daly, D. S., Anderson, K. K., Panisko, E. A., Purvine, S. O., Fang, R., Monroe, M. E., and Baker, S. E. (2008) Mixed-effects statistical model for comparative LC-MS proteomics studies. J. Proteome Res. 7, 1209–1217

40. Karpievitch, Y., Stanley, J., Taverner, T., Huang, J., Adkins, J. N., Ansong, C., Heffron, F., Metz, T. O., Qian, W.-J., Yoon, H., Smith, R. D., and Dabney, A. R. (2009) A statistical framework for protein quantitation in bottom-up MS-based proteomics. Bioinformatics 25, 2028–2034

41. Ruppert, D., Wand, M. P., and Carroll, R. J. (2003) Semiparametric Regression, Cambridge University Press, New York

42. Bates D, M. M., Bolker BM, Walker S (2014) lme4: Linear mixed-effects models using Eigen and S4. J. Statistical Software 67, 1–48

43. Lönnstedt, I., and Speed, T. (2002) Replicated microarray data. Statistica Sinica 12, 31–46

44. Smyth, G. K. (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. 3, Article3

45. Benjamini, Y., and Hochberg, Y. (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. Royal Statist. Soc. 57, 289–300

46. Schwämmle, V., León, I. R., and Jensen, O. N. (2013) Assessment and improvement of statistical tools for comparative proteomics analysis of sparse data sets with few experimental replicates. J. Proteome Res. 12, 3874–3883

47. Rudra, D., and Warner, J. R. (2004) What better measure than ribosome synthesis? Genes Dev. 18, 2431–2436

48. Dressaire, C., Redon, E., Gitton, C., Loubiere, P., Monnet, V., and Cocaign-Bousquet, M. (2011). Investigation of the adaptation of Lactococcus lactis to isoleucine starvation integrating dynamic transcriptome and proteome information. Microbial Cell Factories 10, S18

49. Henao, R., Thompson, J. W., Moseley, M. A., Ginsburg, G. S., Carin, L., and Lucas, J. E. (2012) Hierarchical factor modeling of proteomics data. Computational Advances in Bio and Medical Sciences (ICCABS), 2012 IEEE 2nd International Conference on, pp. 1–6

50. McCarthy, D. J., and Smyth, G. K. (2009) Testing significance relative to a fold-change threshold is a TREAT. Bioinformatics 25, 765–771