Contents

1 Overview of proposed approach and notation

This study addresses three major goals in the characterization of post-translational modifications (PTMs): a) relative PTM quantification, b) PTM significance analysis, i.e., to detect PTM sites that are differentially modified across experimental conditions, and c) statistical design of PTM experiments.

1.1 Data structure of PTM quantification experiments.

A set of fully-cleaved and/or partially-cleaved peptides containing a same PTM (e.g., ubiquitination/phosphorylation) at one site are considered together. There are I conditions and J mass spectrometry runs (technical replicates) per condition in the experiment. The PTM site is represented by K spectral features (peptide ions, distinguished by their cleavage residues and charge states). The log-intensity (base 2) of Feature k, in Run j of Condition i is denoted by y_{ijk}. To account for the underlying protein abundance, features corresponding to the unmodified peptides from the same protein are considered together, except those unmodified peptides containing a modified site to avoid the confounding effect due to the PTM. The log-intensity of Feature l from the unmodified peptides in the same run is denoted by y^*_{ijl}. Figure S1 shows an example data representation of modified peptide ions at one site and unmodified peptide ions of the same protein. Unmodified peptides from the same protein provide additional evidence on the underlying protein abundance, which needs to be integrated for PTM characterization. To address the goals of PTM characterization, statistical analysis needs to summarize values in this table using appropriate statistical models, translate the goal into a model-based quantity of interest, and draw inference (i.e., characterize the uncertainty) about the quantity.

		Condition 1				...	Condition I			
		Run 1	Run 2	...	Run J	...	Run 1	Run 2	...	Run J
Modified	Feature 1	y_{111}	y_{121}	...	y_{1J1}	...	y_{I11}	y_{I21}	...	y_{IJ1}
	Feature 2	y_{112}	$-$	...	y_{1J2}	...	y_{I12}	y_{I22}	...	y_{IJ2}
	...	...	...	...	...	...	...	...	...	...
	Feature K	y_{11K}	y_{12K}	...	y_{1JK}	...	y_{I1K}	y_{I2K}	...	y_{IJK}
Unmodified	Feature 1	y^*_{111}	y^*_{121}	...	y^*_{1J1}	...	y^*_{I11}	y^*_{I21}	...	y^*_{IJ1}
	Feature 2	y^*_{112}	y^*_{122}	...	y^*_{1J2}	...	$-$	$-$	...	$-$
	...	...	...	...	...	...	...	...	...	...
	Feature L	y^*_{11L}	y^*_{12L}	...	y^*_{1JL}	...	y^*_{I1L}	y^*_{I2L}	...	y^*_{IJL}

Figure S1: Representation of the data of modified peptides at one site and unmodified peptides of the same protein, with I conditions and J replicate runs. Abundances of the PTM and protein are quantified by multiple spectral features (peptide ions, K for modified peptides and L for unmodified peptides). Some spectral features can be missing (shown as $-$), either randomly in individual runs or completely in certain conditions. In real practice, the number of runs can vary across conditions.

2 Existing methods

2.1 two-sample t-test

Two-sample t-test is based on the null hypothesis that there is no difference in mean PTM abundance between Conditions i and i'. The abundance in each run is taken as input and is often estimated by sum of peak intensities. The t-test is typically performed based on the log of summarized value. For example, the log-abundance estimate for the PTM in Run j of Condition i is given by

$$\log\left(\sum_{k=1}^{K} 2^{y_{ijk}}\right).$$

For adjustment with respect to unmodified peptides, the estimate of PTM abundance is divided by the protein abundance estimate, and the t-test for the adjusted PTM abundance on log scale takes as input the difference of their log-estimates. The quantity is denoted by d_{ij} and is given by

$$d_{ij} = \log\left(\sum_{k=1}^{K} 2^{y_{ijk}}\right) - \log\left(\sum_{l=1}^{L} 2^{y^*_{ijl}}\right).$$

Alternatively, run-level summary to be described in Section 3.1.1 can also be used. The difference between the means of PTM abundance in Conditions i and i' are estimated as

$$\hat{\Delta} = \frac{1}{J} d_{i+} - \frac{1}{J} d_{i'+},$$

where $d_{i+} = \sum_{j=1}^{J} d_{ij}$ and the test statistic for the t-test is given by $\hat{\Delta}/\mathrm{SE}(\hat{\Delta})$. The statistical significance of the difference is determined by comparing the test statistic against the t distribution, with degrees of freedom $df = 2J - 2$ in balanced designs.

2.2 Limma

Limma uses linear models to test the null hypothesis that there is no difference in mean PTM abundances between conditions. Additionally, it leverages Empirical Bayes moderation to share pooled variance information across individual modification models and moderate the individual residual variances. Using a linear model allows Limma to share variance information across conditions, providing a more accurate estimate.[7]

With respect to PTM analysis, feature level summarization is done in the same way as t-test. The log-abundace estimate for PTM in Run j of Condition i is given by

$$\log\left(\sum_{k=1}^{K} 2^{y_{ijk}}\right).$$

To adjust for unmodified peptides, the unmodified peptide is summarized in the same way and the result is combined with the PTM given by

$$d_{ij} = \log\left(\sum_{k=1}^{K} 2^{y_{ijk}}\right) - \log\left(\sum_{l=1}^{L} 2^{y^*_{ijl}}\right).$$

Alternatively run level summarization can be performed given by

$$\hat{\Delta} = \frac{1}{J} d_{i+} - \frac{1}{J} d_{i'+},$$

After fitting a linear model, the final variance is moderated using Empirical Bayes and the global variance across all modified peptides. This is done as $s_i = P(v_i|v_\Theta)$ where v_i is the resulting variance when modeling peptide i and v_Θ is the global variance.

3 Proposed approach

To characterize the observed feature intensities, different levels of variations are expressed using linear mixed models in consideration of the following factors: modification, condition, run, and feature. As different degrees of variability are present in the feature intensities of modified and unmodified peptides, they are expressed by separate models.

3.1 Statistical modeling and inference

The observed log-intensity of a modified peptide feature is denoted by y_{ijk} and represented as

$$y_{ijk} = \psi + C_i + R_{j(i)} + F_k + (R \times F)_{ijk},$$

where the effects of condition and feature are modeled as fixed effects:

$$\sum_{i=1}^{I} C_i = 0, \quad \sum_{k=1}^{K} F_k = 0,$$

and the effects of run and its interaction with feature are considered as random effects arising from normal distribution with mean 0:

$$R_{j(i)} = \gamma_{j(i)} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2_\gamma), \qquad (R \times F)_{ijk} = \epsilon_{ijk} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2_\epsilon).$$

Similarly, the observed log-intensity of an unmodified peptide feature is denoted by y^*_{ijl} and represented as

$$y^*_{ijl} = \psi^* + C^*_i + R^*_{j(i)} + F^*_l + (R \times F)^*_{ijl},$$

where the effects of condition and feature are modeled as fixed effects:

$$\sum_{i=1}^{I} C^*_i = 0, \quad \sum_{l=1}^{L} F^*_l = 0,$$

and

$$R^*_{j(i)} = \gamma^*_{j(i)} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2_{\gamma^*}), \qquad (R \times F)^*_{ijl} = \epsilon^*_{ijl} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2_{\epsilon^*}).$$

3.1.1 Run-level summarization of feature intensities

Run-level summarization of feature intensities for each PTM site is carried out as in the sub-plot model of MSstats [1], which involves a) imputation of censored missing values, and b) summarization of feature intensities using Tukey's median polish. The run-level summary for the PTM in Run j of Condition i is denoted by $\hat{y}_{ij}$.

3.1.2 Model-based inference of the underlying abundance

The PTM abundance in each run is represented as

$$\hat{y}_{ij} = \psi + C_i + R_{j(i)},$$

where $\sum_{i=1}^{I} C_i = 0$, $R_{j(i)} = \gamma_{j(i)} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2_\gamma)$. Similarly, the protein abundance in each run is expressed as

$$\hat{y}_{ij} = \psi^* + C^*_i + R^*_{j(i)},$$

where $\sum_{i=1}^{I} C^*_i = 0$, $R^*_{j(i)} = \gamma^*_{j(i)} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2_{\gamma^*})$. The expected values of log-abundances of the PTM and protein in Condition i are denoted by μ_i and μ^*_i, respectively, and the values are estimated as:

$$\hat{\mu}_i = \hat{\psi} + \hat{C}_i = \frac{1}{J}\hat{y}_{i+}$$
$$\hat{\mu}^*_i = \hat{\psi}^* + \hat{C}^*_i = \frac{1}{J}\hat{y}^*_{i+},$$

where the standard errors of the estimates are $\text{SE}(\hat{\mu}_i) = (\hat{\sigma}^2_\gamma / J)^{1/2}$ and $\text{SE}(\hat{\mu}^*_i) = (\hat{\sigma}^2_{\gamma^*} / J)^{1/2}$. Based on the estimates $\hat{\mu}_i$ and $\hat{\mu}^*_i$, the adjusted log-abundance of the PTM is given by $(\hat{\mu}_i - \hat{\mu}^*_i)$ and the standard error of the estimate is

$$\left[\frac{1}{J}\left(\hat{\sigma}^2_\gamma + \hat{\sigma}^2_{\gamma^*}\right)\right]^{1/2}.$$

3.2 PTM significance analysis

With protein-level adjustment, the model-based testing is based on the hypothesis that there is no difference in adjusted PTM abundance between Conditions i and i'

$$H_0 : \Delta = (\mu_i - \mu_{i'}) - (\mu^*_i - \mu^*_{i'}) = 0$$
$$H_a : \Delta = (\mu_i - \mu_{i'}) - (\mu^*_i - \mu^*_{i'}) \neq 0$$

The log-fold change in the adjusted PTM abundance, Δ, is estimated by

$$\hat{\Delta} = \left[\frac{1}{J}\left(\hat{y}_{i+} - \hat{y}_{i'+}\right)\right] - \left[\frac{1}{J}\left(\hat{y}^*_{i+} - \hat{y}^*_{i'+}\right)\right],$$

and the standard error of the estimate $\text{SE}(\hat{\Delta})$ is

$$SE(\hat{\Delta}) = \left[\frac{2}{J}\left(\hat{\sigma}^2_\gamma + \hat{\sigma}^2_{\gamma^*}\right)\right]^{1/2}.$$

The test statistic $\hat{\Delta}/\mathrm{SE}(\hat{\Delta})$ is compared against the t distribution, with degrees of freedom approximated by

$$\left(\hat{\sigma}_{\gamma}^2 + \hat{\sigma}_{\gamma^*}^2\right)^2 \Big/ \left(\frac{\hat{\sigma}_{\gamma}^4}{\mathrm{df}(\gamma)} + \frac{\hat{\sigma}_{\gamma^*}^4}{\mathrm{df}(\gamma^*)}\right).$$

A distinctive property of the proposed model-based testing to the two-sample t-test is that even only the PTM abundances in Conditions i and i' are compared, measurements from all conditions are used for the modeling and inference.

3.3 Design of PTM experiments

The proposed statistical framework allows for design of PTM experiments in terms of sample size calculation and power analysis. Sample size calculation takes as input a) q, the desired false discovery rate, b) β, the average Type II error rate, c) Δ, the minimal log-fold change in adjusted PTM abundance that we would like to detect, d) $m_0/(m_0+m_1)$, the fraction of truly differentially modified PTM sites in the comparison, and e) σ_{γ}^2 and $\sigma_{\gamma^*}^2$, the anticipated variances associated to modified and unmodified peptide features, respectively. The variances can be derived based on the dataset being analyzed, assuming similar quantitative properties and variations. With these values and a user-specified number of conditions, the corresponding number of technical replicates per condition can then be derived, as described in [3]. Given the above quantities, the minimal number of replicates J is determined by the variance of the estimated log-fold change $\mathrm{SE}^2(\hat{\Delta})$ as

$$\mathrm{SE}^2(\hat{\Delta}) = \left[\frac{2}{J}\left(\hat{\sigma}_{\gamma}^2 + \hat{\sigma}_{\gamma^*}^2\right)\right] \leq \left(\frac{\Delta}{t_{1-\beta,df} + t_{1-\alpha/2,df}}\right)^2,$$

where

$$\alpha = (1-\beta)\cdot\frac{q}{1+(1-q)\cdot m_0/m_1},$$

and $t_{1-\beta,df}$ and $t_{1-\alpha/2,df}$ are the $100(1-\beta)^{\text{th}}$ and the $100(1-\alpha/2)^{\text{th}}$ percentiles of the t distribution, with $df = I(J-1)$ degrees of freedom in balanced designs. More details can be found in [6].

3.4 TMT experiment

The methods described previously can be extrapolated to Tandem Mass Tag (TMT) experiments targeting PTMs, similarly to how MSstatsTMT targets the global protein.[2] The null hypothesis is still that there is no difference in mean PTM abundance between Conditions i and i' after adjusting for changes in mean unmodified peptide abundance between Conditions i and i'.

$$H_0 : \Delta = (\mu_i - \mu_{i'}) - (\mu_i^* - \mu_{i'}^*) = 0$$
$$H_a : \Delta = (\mu_i - \mu_{i'}) - (\mu_i^* - \mu_{i'}^*) \neq 0$$

For TMT experiments there is an additional source of variation coming from the mixtures used. This increases the model complexity as variation can now come from multiple mixtures, biological replicates, and technical replicates. Going forward we will denote the log-intensity of a modified peptide for Technical Replicate j of Mixture m for Condition i by y_{ijm}, where $m = 1, ..., M$. This is done in the same way for the unmodified peptide y_{ijm}^*.

The model assumes the variability of the mixture is a random effect and is not convoluted with the other variables in the model. The model also assumes that the difference between conditions is the same between mixtures. For a given modified peptide and corresponding unmodified protein a mixed effects model is fit as follows.

$$Y_{mtcb} = \mu + Mixture_m + Techrep(Mixture)_{tm} + Condition_c + Subject_{mcb} + \epsilon_{mtcb}$$
$$Y^*_{mtcb} = \mu^* + Mixture^*_m + Techrep(Mixture)^*_{tm} + Condition^*_c + Subject^*_{mcb} + \epsilon^*_{mtcb}$$

Where $Mixture_m \sim N(0, \sigma^2_m)$, $Techrep(Mixture)_{tm} \sim N(0, \sigma^2_{tm})$, $Subject \sim N(0, \sigma^2_s)$ are random effects, and $\sum_{c=1}^{I} Condition^*_c = 0$ is a fixed effect. Finally the error is expressed as $\epsilon_{mtcb} \sim N(0, \sigma^2)$.

The log-fold change for adjusted PTM abundance is estimated as the average over mixtures and technical replicates

$$\hat{\Delta} = \frac{1}{MJ}\left(\hat{y}_{+,+,i+} - \hat{y}_{+,+,i'+}\right) - \frac{1}{BJ}\left(\hat{y}^*_{+,+,i+} - \hat{y}^*_{+,+,i'+}\right)$$

and the standard error of the estimate is

$$\left[\frac{2}{MJ}\left(\hat{\sigma}^2_\gamma + \hat{\sigma}^2_{\gamma^*}\right)\right]^{1/2}.$$

The test statistic $\hat{\Delta}/\mathrm{SE}(\hat{\Delta})$ is compared against the t distribution, with degrees of freedom approximated by

$$\left(\hat{\sigma}^2_\gamma + \hat{\sigma}^2_{\gamma^*}\right)^2 \Big/ \left(\frac{\hat{\sigma}^4_\gamma}{\mathrm{df}(\gamma)} + \frac{\hat{\sigma}^4_{\gamma^*}}{\mathrm{df}(\gamma^*)}\right).$$

4 Computer simulation

The proposed statistical approach was evaluated and compared to the t-test and Limma using computer simulation. In particular, their statistical properties with respect to protein-level adjustment and real world experimental conditions were evaluated.

4.1 Protein-level adjustment

Differential levels of modified peptides may be due to differential modifications and/or changes in protein abundance. The proposed approach adjusts the PTM abundance with respect to protein abundance as introduced in Section 3. Alternatively, two-sample t-test or Limma taking as input the ratio between feature intensities of modified and unmodified peptides is commonly applied for the same purpose (Section 2). Approaches without considering protein-level adjustment lose track of an important aspect in interpreting observed changes in PTM abundance, which may result in misleading conclusions. To highlight the necessity of the adjustment, we compared the following approaches: a) proposed approach, b) proposed approach without adjusting for unmodified peptides c) t-test (with adjustment), d) t-test (no adjustment), e) Limma (with adjustment), and f) Limma (no adjustment).

In experiments of complex designs, multiple inter-related conditions are often compared together. The proposed approach and Limma leverage measurements in all conditions for the inference of the underlying abundance, whereas t-test uses measurements from the two conditions being compared. To highlight this distinction, multiple conditions of data were generated. Two simulations were generated based on the following parameters:

4.1.1 Computer simulation 1 - label free

In the first simulation an experiment with many features per PTM and unmodified protein was created. Additionally this simulation contained no missing data. These attributes are not representative of a real experiment, but provide a baseline for model performance.

- Mean of log-intensity: 25
- Standard deviations of log-intensities for modified and unmodified peptides: 0.2, 0.3
- Difference in PTM abundance between conditions: 0, 0.75, 1.5, 2.25
- Difference in protein abundance between conditions: 0, 0.75, 1.5, 2.25
- Number of replicates: 2, 3, 5, 10
- Number of conditions: 2, 3, 4
- Number of realizations: 1000
- Number of features per PTM: 10
- Number of features per protein: 10

- Missing data: no missing value

The results are summarized from Figure S2 to Figure S3, including false discovery rate with or without changes in protein abundance by the considered methods, false positive rate, and overall accuracy. Additionally the results are compared to simulation 2 described in the next section.

When simulating the data, 500 of the 1000 peptides were simulated with a differential fold change between conditions, while the other 500 were not. In terms of the peptides that were not differential, 250 were simulated with a change in both PTM and overall protein abundance, such that the change in PTM abundance is entirely due to the change in the overall protein. The other 250 were simulated with no significant change in either PTM or overall protein abundance. In this way there were 500 true positives and 500 true negative PTMs.

4.1.2 Computer simulation 2 - missing values and low replicates

In the second simulation we introduced limited feature observations per PTM as well as masking a portion of the observation to simulate missing values. This is more in line with what we would expect with a real life experiment and provides a more realistic expectation of model performance.

- Mean of log-intensity: 25
- Standard deviations of log-intensities for modified and unmodified peptides: 0.2, 0.3
- Difference in PTM abundance between conditions: 0, 0.75, 1.5, 2.25
- Difference in protein abundance between conditions: 0, 0.75, 1.5, 2.25
- Number of replicates: 2, 3, 5, 10
- Number of conditions: 2, 3, 4
- Number of realizations: 1000
- Number of features per PTM: 2
- Number of features per protein: 10
- Missing data: 20% of the observations for PTMs and Proteins were masked with NA at random

The results are summarized from Figure S4 to Figure S5, including false positive rate with or without changes in protein abundance by the considered methods, false positive rate, and overall accuracy.

The portion of significant peptides were done in the same was as described in simulation 1.

Figure S2.a: All the considered methods in simulation 1 correctly calibrated FDR when adjusting for changes in protein abundance. In comparison, the methods without accounting for the protein-level changes resulted in off-target, high false discovery rates.

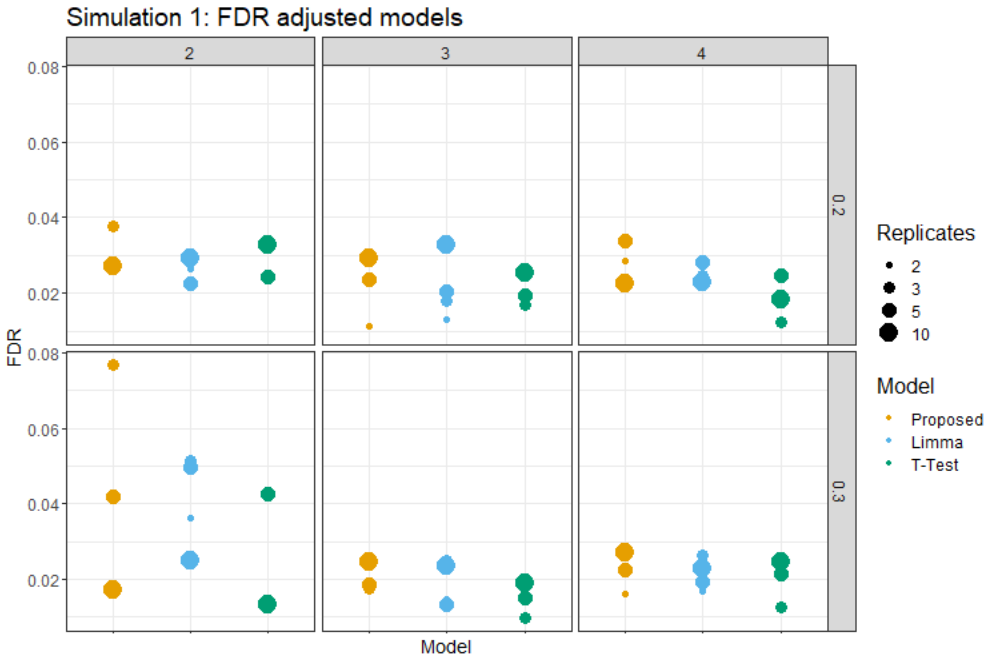

Figure S2.b: The considered methods with protein adjustment are compared in detail. All three methods generally performed similarly in terms of FDR.

Figure S2: FDR of Simulation 1.

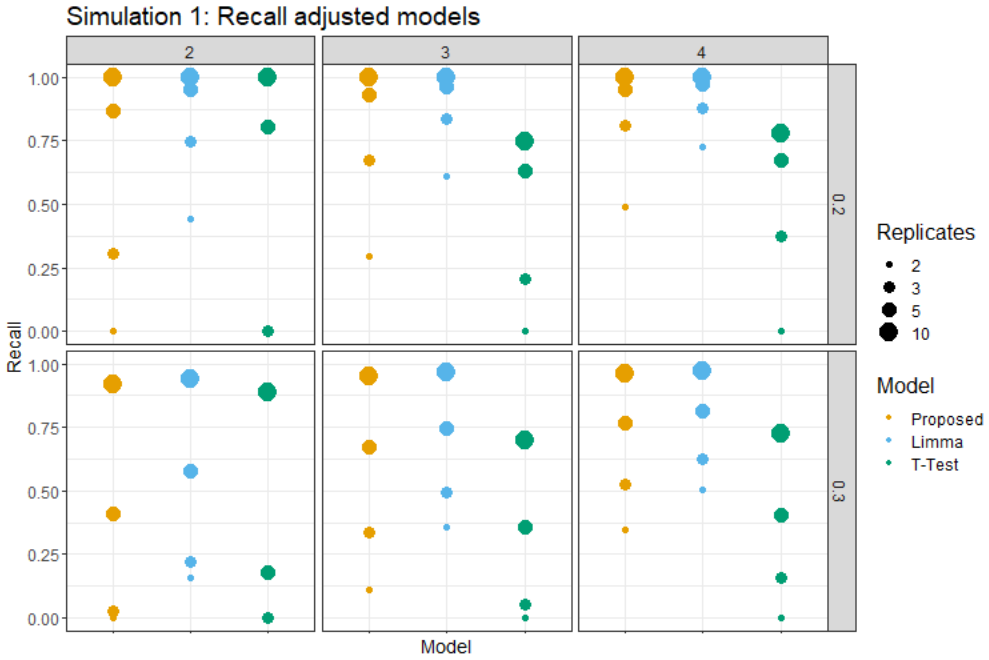

Figure S3.a: The recall for simulation 1 methods with adjustment were compared. It is clear that limma performs the strongest here when the number of replicates were low. At higher replicates the performance of the proposed methods and limma are comparable. *t*-test clearly performs worse across all methods.

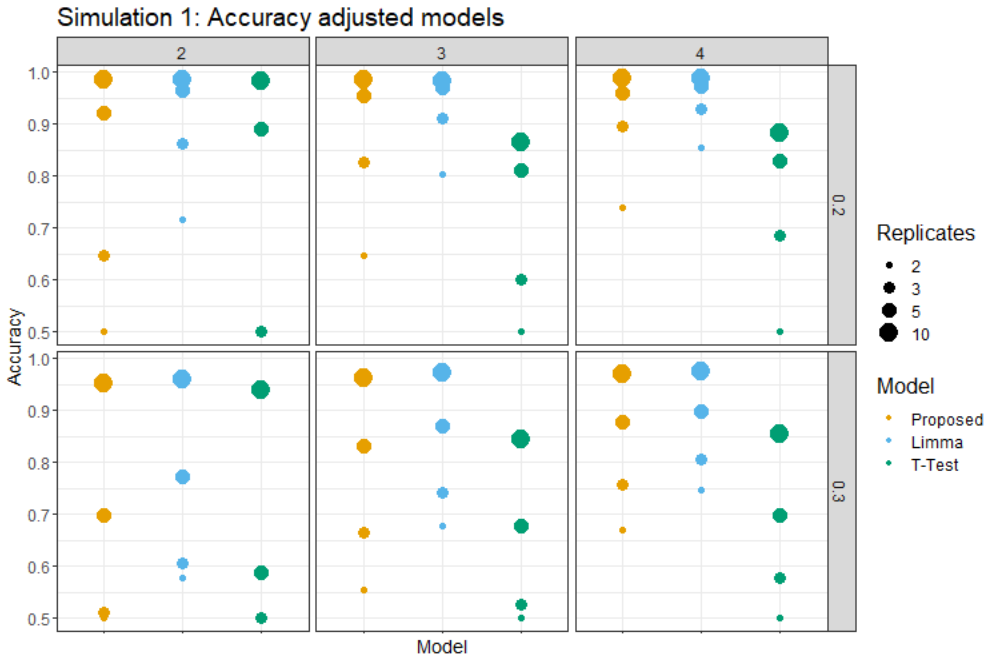

Figure S3.b: The overall accuracy plot mimics the observations in the recall plot. Limma performs stronger than the proposed method at lower replicates, while at higher they are comparable.

Figure S3: Recall and Accuracy results of Simulation 1.

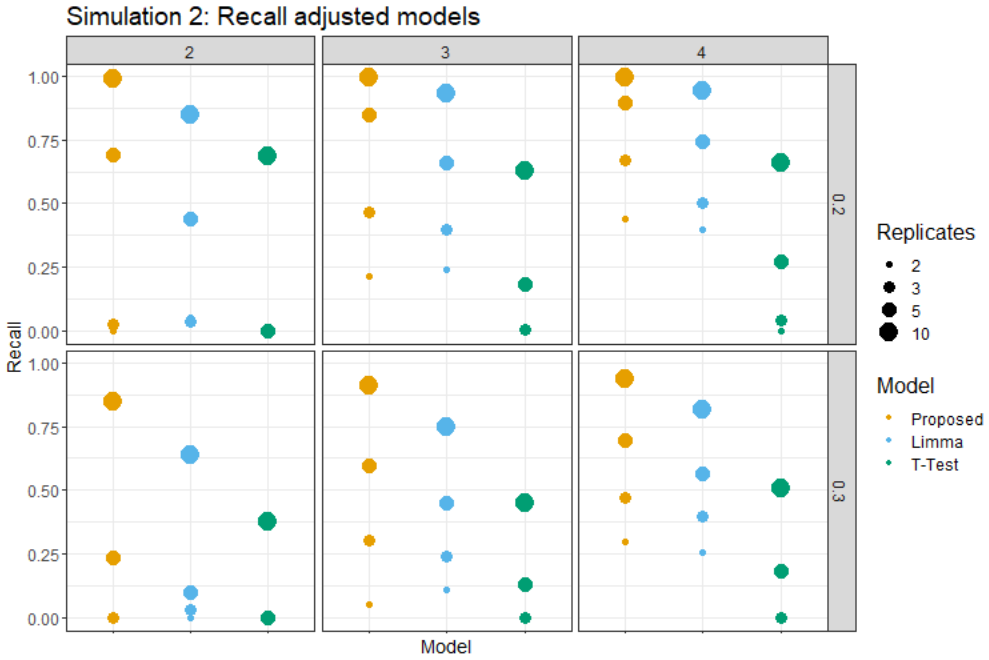

Figure S4.a: The advantage of using the proposed approach is apparent when looking at simulation 2, which includes limited observations and the presence of missing values. In the case of recall the proposed method performs stronger than limma and t-test in nearly every model. Even at lower replicates the proposed method still outperformed limma. Again the lowest performing method was *t*-test.

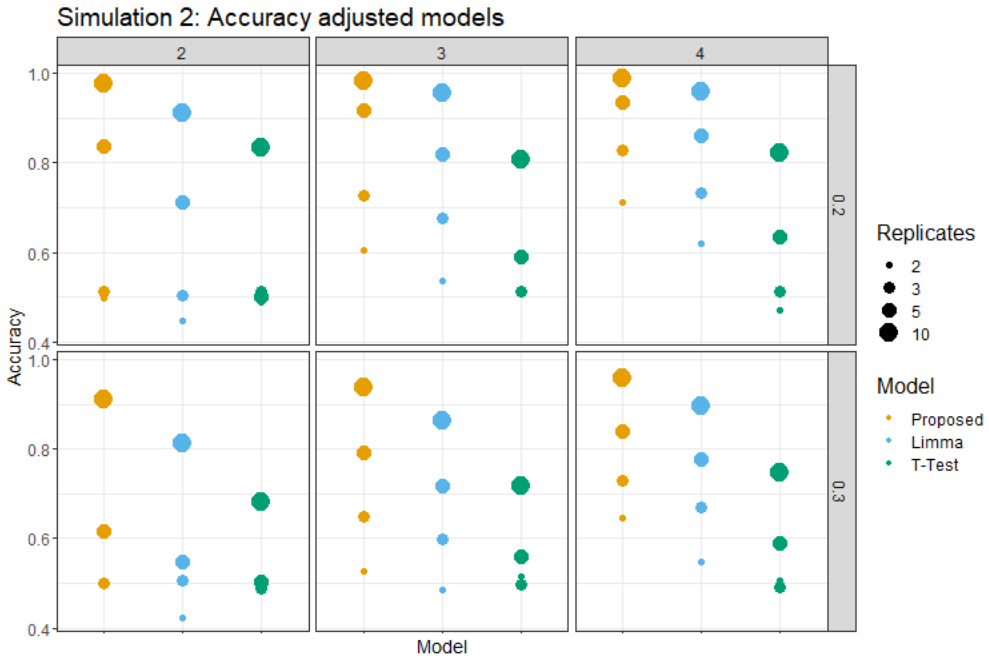

Figure S4.b: The overall accuracy plots show similar results to recall. The proposed method performed the strongest across all methods, even when replicates were low. Limma shows strong performance in a clean experiment, however when real world data problems are introduced it is clear the proposed method is more robust.

Figure S4: Recall and Accuracy results of Simulation 2.

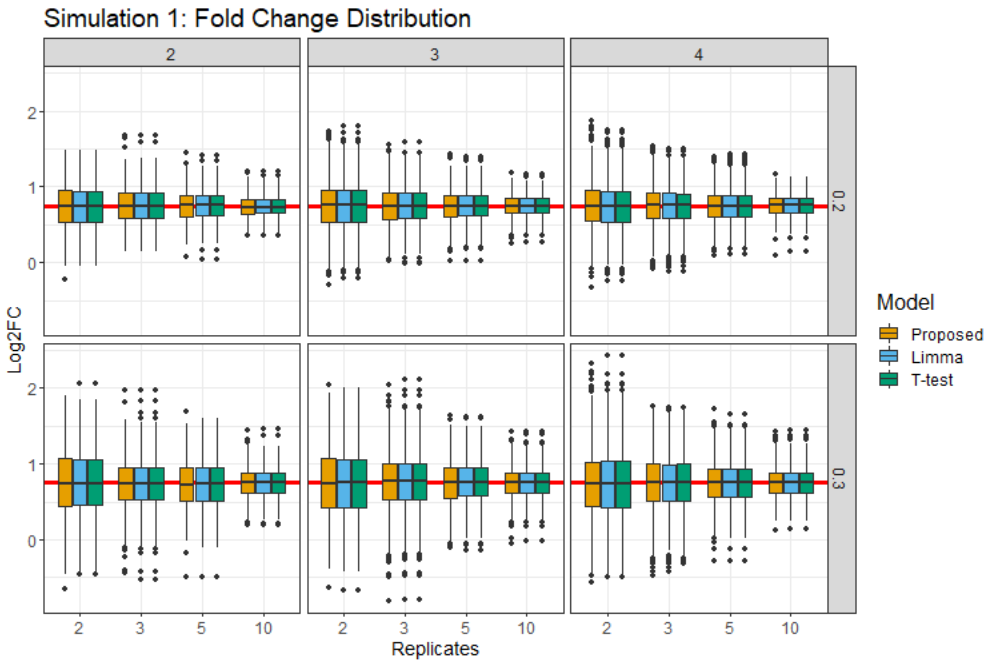

Figure S5.a: In simulation 1 all considered methods correctly estimated the fold change between conditions, with a median fold change estimation of .75. The distributions around the median were also consistent across all methods. Predictably the distribution quantiles decrease as the number of replicates increases.

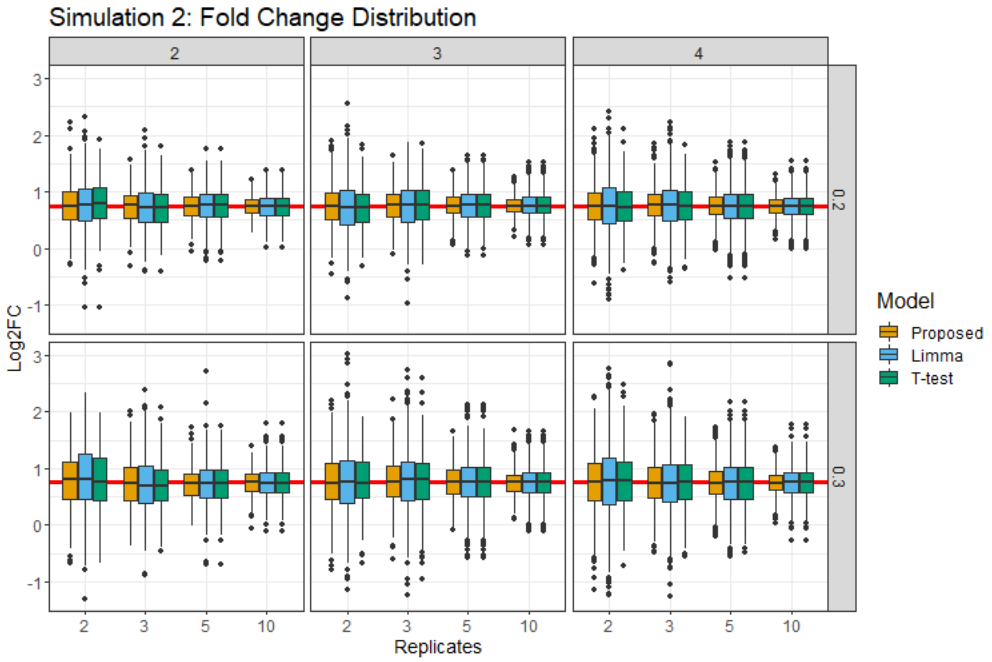

Figure S5.b: In simulation 2 again all methods correctly estimated the fold change with a median log change of .75. The proposed method in this simulation had a visibly tighter distribution around the median. Both limma and t-test showed a wider range around the fold change. In this case the proposed method showed a stronger performance in correctly estimating the fold change for all peptides.

Figure S5: Fold change distribution comparison between Simulations 1 and 2.

4.2 SpikeIn benchmark - KGG enriched

A custom designed experiment with labeling was used to assess the performance of the proposed method in a real experimental setting. Heavy-labeld KGG modified peptides were used as spike-in peptides. The spike-in peptides were mixed with human lysate to create four mixture conditions. Two sets of data were acquired for each mixture: KGG enriched + LC-MS, and LC-MS only. The KGG enriched dataset included the spike-in peptides, as well as modified and unmodified human lysate. The LC-MS dataset included only unmodified peptides. The spike-in peptides where a significant fold change between conditions are expected to be differencial, whereas none of the human lysate peptides in any comparison are expected to be differencial.

Again we consider three different methods and assess their performance: the method proposed in this paper, limma, and two sample t-test. All methods are analyzed after adjusting for changes in overall protein level. The proposed method summarizes feature intensities up to the run level using Tukey's Median Polish, while the other methods use the log of summed feature intensities. The results are summarized from Figure S6 to Figure S10, including volcano plots, model summary statistics, and fold change analysis.

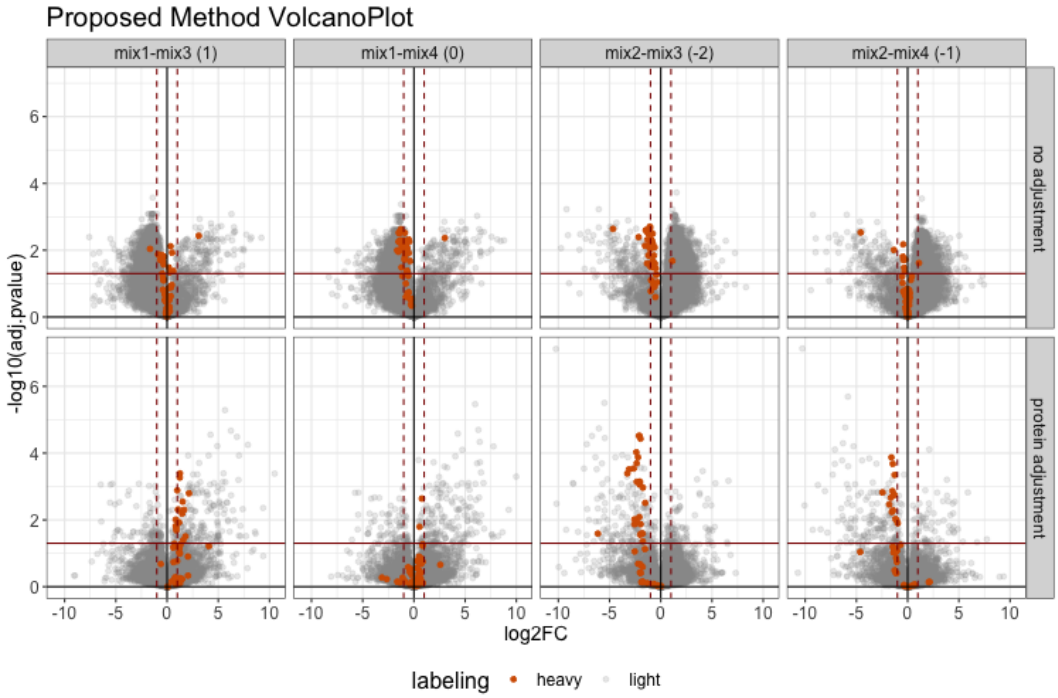

Figure S6: Using the proposed method to model the benchmark experiment, the spike in peptides (colored red) do not follow the expected log fold change before adjustment. After adjusting for changes in overall protein abundance the spike in peptides are more in line with expectation. Additionally the background grey colored peptides show many false positives before adjustment. After adjustment the false positives were decreased considerably.

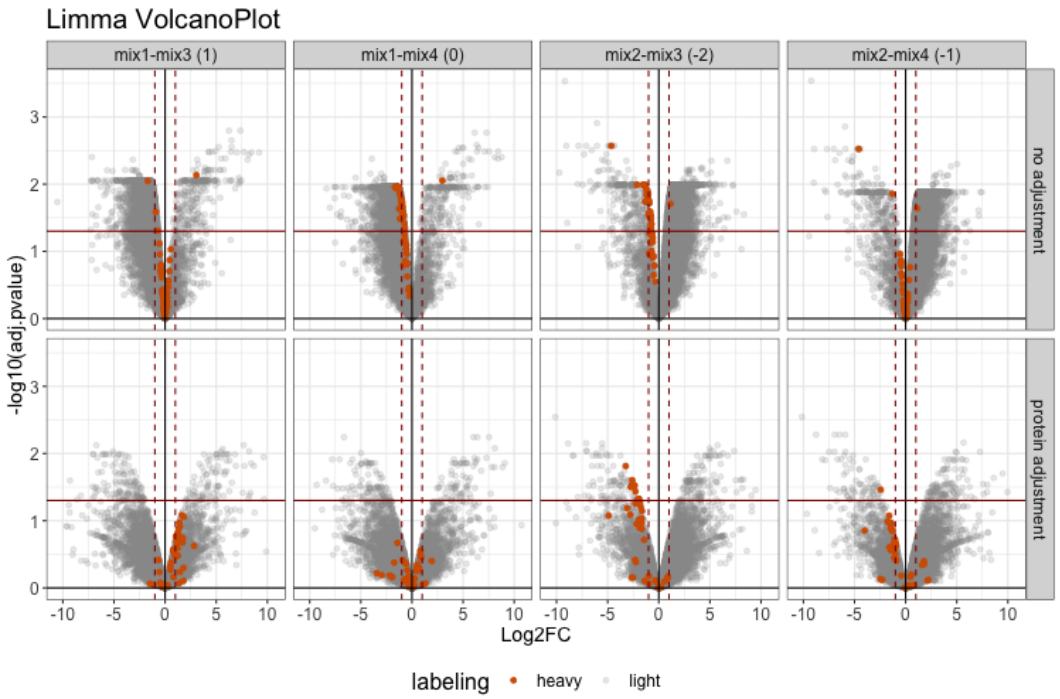

Figure S7: When modeling the experiment with the limma method, the spike in peptides again follow the expected log fold change better after adjusting for changes in protein level. However, while the fold change is much more accurate, the majority of spike in peptides do not have a significant adjusted pvalue. In this case, the known differencial peptides are missed by the model. In terms of false positives, the results are very similar to the proposed method, with many false positives before adjustment and much fewer after.

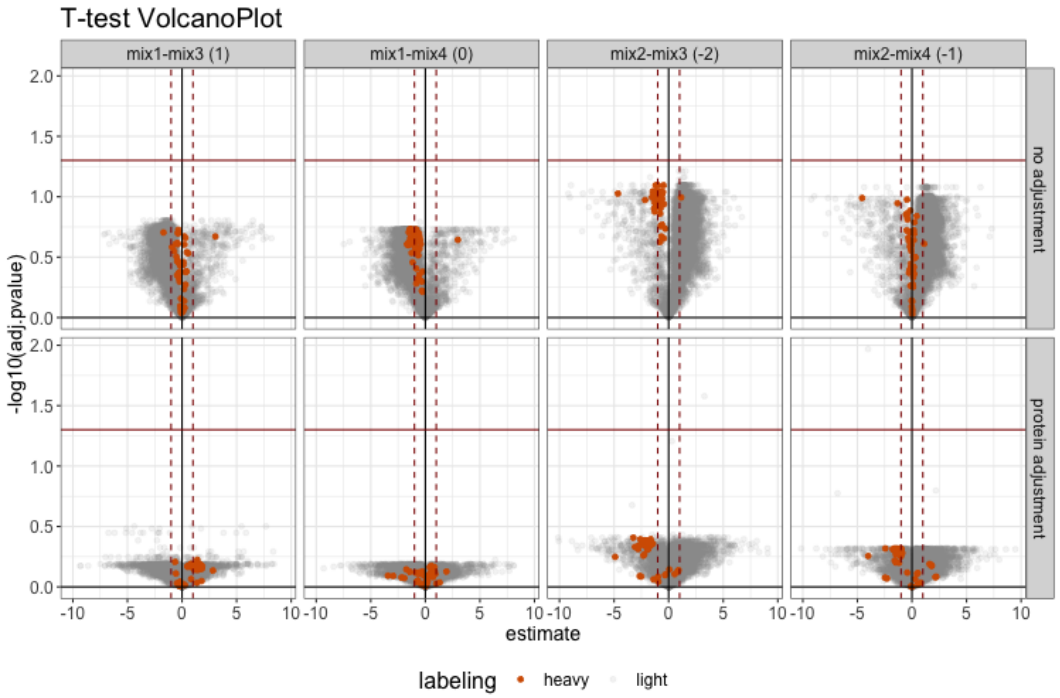

Figure S8: Using the two sample t-test, none of the comparisons either before or after adjustment show any significant peptides. With that being said, the fold change of the spike in peptides is much closer to expectation after adjusting for global protein abundance.

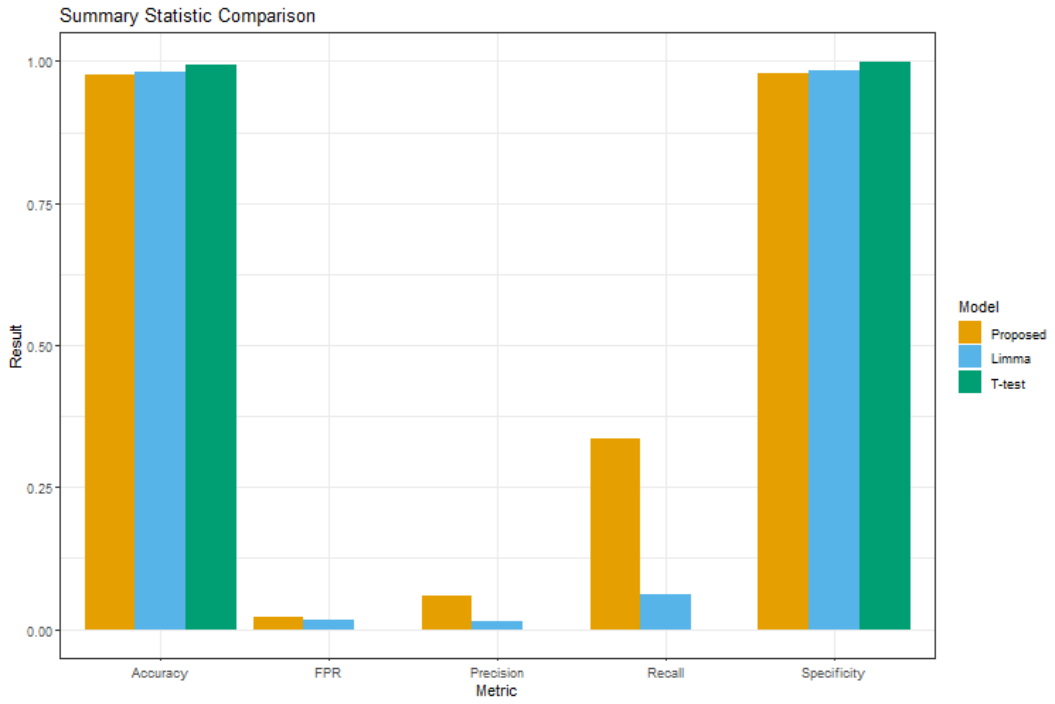

Figure S9: Comparing the summary statistics between methods it is clear that the proposed method performs the strongest. In terms of accuracy and specificity the three methods are close, with limma and t-test showing slightly higher values. Accuracy and specificity are dominated by the large number of true negatives (background peptides) compared to the true positives (spike in peptides). In terms of recall, the proposed approach far out performed the other two methods, showing that it correctly labeled the most spike in peptides.

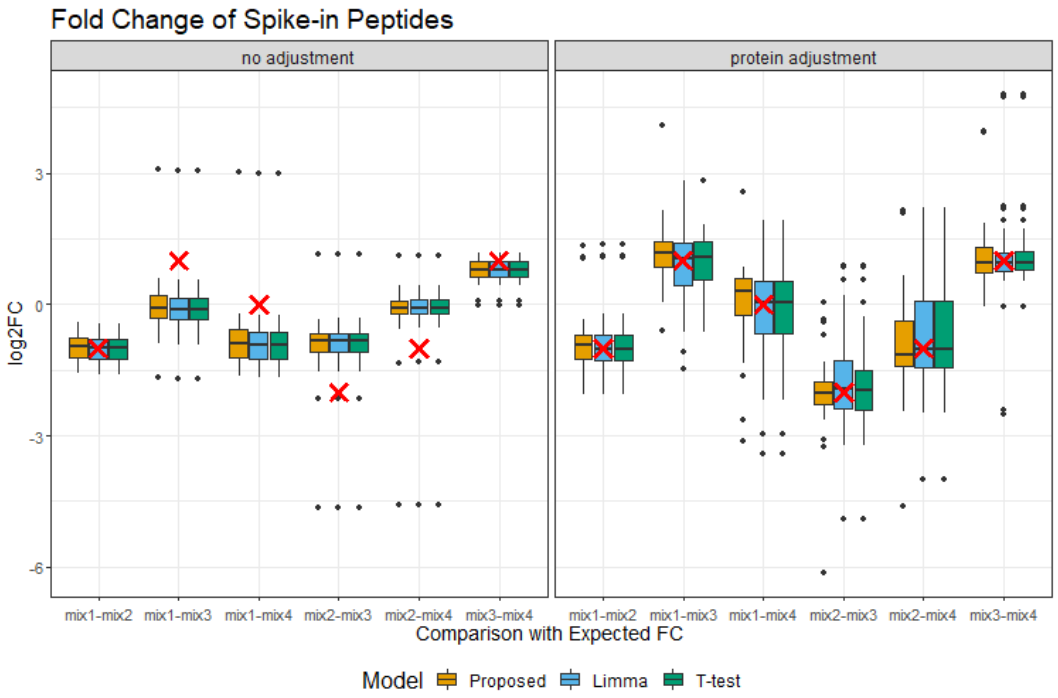

Figure S10: The fold change misalignment before adjustment is illustrated in the boxplots. The true log fold change is indicated by the red 'X' marks. Before adjustment the boxplots are not close to the true fold change. After adjustment all three models generally have a median near the true fold change. In terms of model comparison, the proposed method shows a much tighter distribution around the true fold change, whole both Limma and t-test are much wider.

5 Datasets : Biological investigation

5.1 Human-1mix-TMT - ubiquitination

In this experiment Shigella ubiquitin ligase IpaH7.8 was shown to inhibit the protein gasdermin D (GSDMD). [4] Multiplex proteomics was used to quantify the abundance of total protein, and ubiquitination in human epithelial cells. Cells were either infected or uninfected with IpaH7.8-deficient Shigella flexneri and measurements were taken at different time periods. Uninfected cells were measured at 0 and 6 hours, while infected cells were measured at 1, 2, 4, and 6 hour increments, resulting in six total conditions. The experiment was inbalanced with two bioreplicates per condition for all conditions except for infected 1 hour. The experimental design can be seen in Table S1.

Condition	BioReplicate	Channel
Dox1hr	Dox1hr_1	127C
Dox2hr	Dox2hr_1	128N
Dox2hr	Dox2hr_2	130C
Dox4hr	Dox4hr_1	128C
Dox4hr	Dox4hr_2	131C
Dox6hr	Dox6hr_1	129N
Dox6hr	Dox6hr_2	131N
NoDox0hr	NoDox0hr_1	126C
NoDox0hr	NoDox0hr_2	129C
NoDox6hr	NoDox6hr_1	127N
NoDox6hr	NoDox6hr_2	130N

Table S1: The experimental design of the IpaH7.8 experiment

A model was fit for the total protein, and ubiquitination separately, as described previously for TMT experiments in Section 3.4. The model formula can be seen below.

[TODO: only one mixture so we probably wouldn't include the mixture as a term in the model?]

$$Y_{mcb} = \mu + Condition_c + Subject_{mcb} + \epsilon_{mcb}$$

$$\sum_{c=1}^{C} Condition_c = 0,\ Subject_{mcb} \sim N(0, \sigma_S^2),\ \epsilon_{mcb} \sim N(0, \sigma^2)$$

The results of the proposed method to this experiment can be seen in Figure S11 and Figure S12.

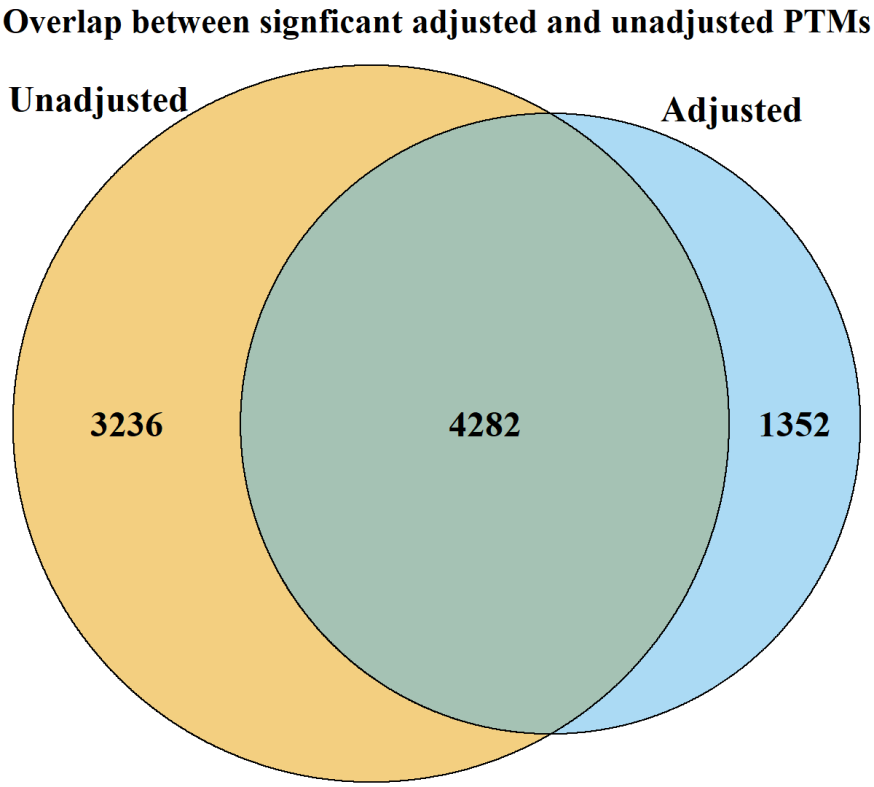

Figure S11: The overlap of differencial modified peptides between the PTM model with and without global protein level adjustment. More PTMs became insignificant after adjustment then became significant. This indicates that for the peptides that became insignificant in the adjusted model, their change in abundance was mainly due to changes in global protein abundance. An additional question that must be addressed is if the decrease in significant peptides is due to the increased variance that comes from adjustment. This was tested by looking for modified peptides who's adjusted log fold change was within 10% of the unadjusted log fold change but became insignificant after adjustment. When this test was applied on this experiment, only one peptide became insignificant due to an increase in variance. Thus we can conclude that the drop off in significant peptides was truly due to changes in global protein abundance.

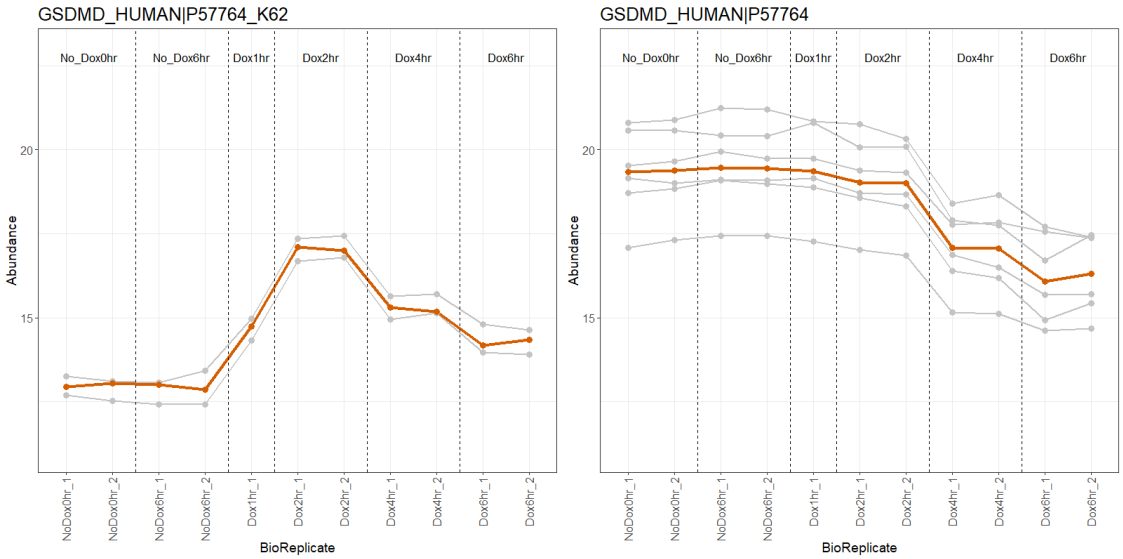

Figure S12.a: Comparing the global profiling of protein $GSDMD\_HUMAN|P57764$ with the ubiquitination of the protein at site $K62$. The individual PSM features are shown in grey, while the feature summarization is shown in red. When looking at the summary of the modification and global protein it is clear the conditions follow different trends. Specifically, there appears to be no change in abundance between Dox1hr and Dox4hr in the modified plot, however there is a large negative change when looking at the unmodified plot. This indicates the modification is confounded with changes in the unmodified protein.

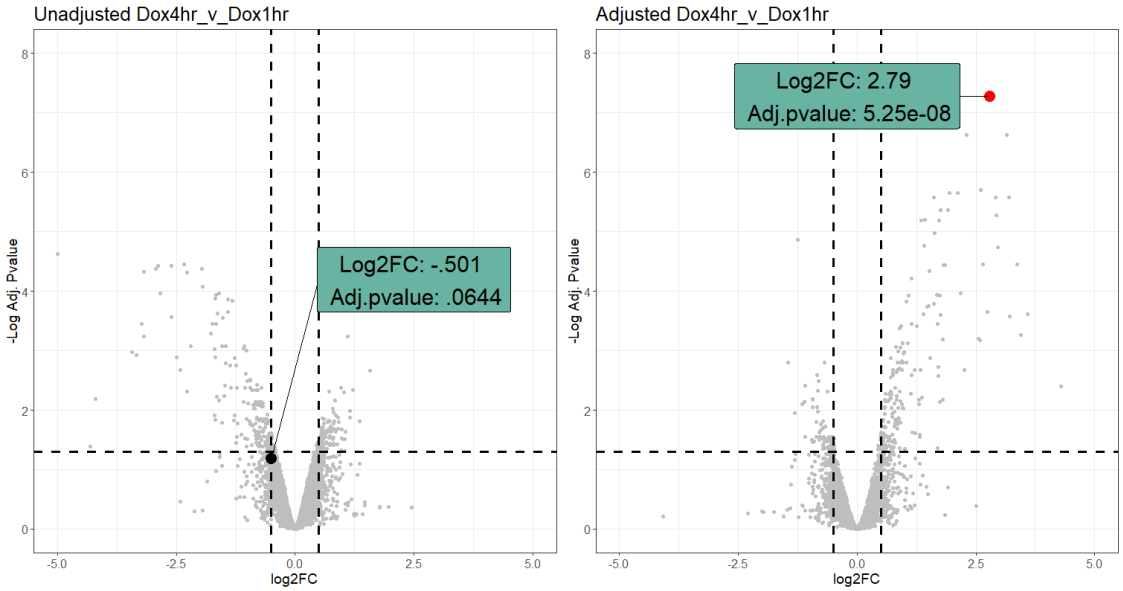

Figure S12.b: Volcano plots of Dox4hr vs Dox1hr both before and after protein adjustment. The $GSDMD\_HUMAN|P57764\_K62$ modification is highlighted. Before adjustment the modification has a small fold change and insignificant adjusted pvalue. After adjustment the fold change is much larger and the adjusted pvalue is very significant. In this case the proposed method allows us to identify a differential modified peptide that could have otherwise been missed.

Figure S12: Summary plots for modification of protein GSDMD at site K62.

5.2 Mouse-2mix-TMT - phosphorylation

In this study, the correlation between the gene Atg16L1 and killing of Shigella flexneri (S.flexneri) was assessed. [5] Multiplex proteomics was used to quantify the abundance of total protein, phosphorylation, and ubiquitination in wild type (WT) and ATG16L1-deficient (cKO) samples, uninfected and uninfected with S.flexneri. The abundance of total protein and post-translation modifications were quantified at three time points, uninfected, early infection (45-60 minutes), and late infection (3-3.5 hours). Quantifying the total protein along with the post-translational modifications allowed us to adjust for changes in total protein and see the true impact of the site specific modifications. Two mixtures using 11-plex were ran over the six conditions. The six conditions were split between 11 channels leading to the experimental design being unbalanced. Each mixture contained two replicates per early and late WT and KO conditions. Mixture one contained one replicate of uninfected WT and two replicates of uninfected KO. Mixture two contained one replicate of uninfected KO and two uninfected WT. The experimental design can be seen in Table S2.

	Mixture 1		Mixture 2		Condition
Uninfected	128C		128C	131C	
Early (1 Hour)	126C	129C	126C	129C	WT
Late (3 Hour)	127C	130C	127C	130C	
Uninfected	129N	131C	129N		
Early (1 Hour)	127N	130N	127N	130N	KO
Late (3 Hour)	128N	131N	128N	131N	

Table S2: The experimental design of the Shigella flexneri experiment

A model was fit for the total protein, phosphorylation, and ubiquitination separately, as described previously for TMT experiments. The model formula can be seen below.

$$Y_{mcb} = \mu + Mixture_m + Condition_c + Subject_{mcb} + \epsilon_{mcb}$$

$$Mixture_m \sim N(0, \sigma_M^2),\ \sum_{c=1}^{C} Condition_c = 0,\ Subject_{mcb} \sim N(0, \sigma_S^2),\ \epsilon_{mcb} \sim N(0, \sigma^2)$$

The results of the proposed method to this experiment can be seen in Figure S13, Figure S14, and Figure S15.

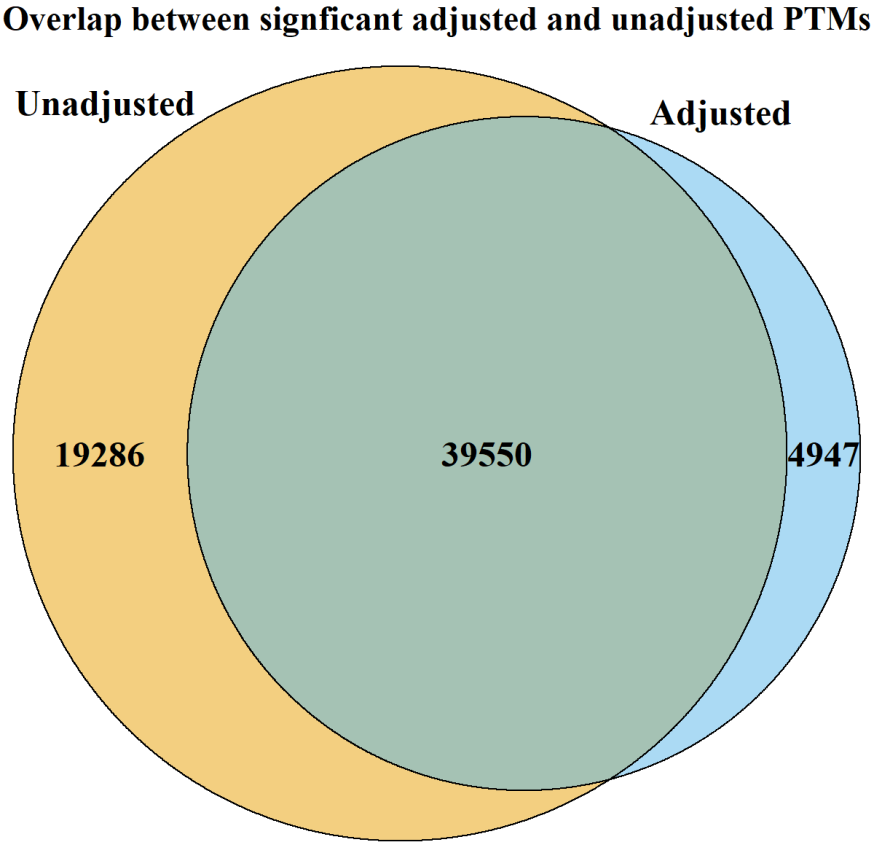

Figure S13: The overlap of differencial modified peptides between the PTM model with and without global protein level adjustment. Much like in Figure S11, more PTMs became insignificant after adjustment then became significant. This indicates that for the peptides that became insignificant in the adjusted model, their change in abundance was mainly due to changes in global protein abundance. Again we tested if the decrease in significant peptides was due to the increased variance that comes from adjustment. This was tested by looking for modified peptides who's adjusted log fold change was within 10% of the unadjusted log fold change but became insignificant after adjustment. When this test was applied on this experiment, 776 peptides became insignificant due to an increase in variance. This is higher than in Figure S11, however it is still a very small portion (4%) of the peptides that became insignificant after adjustment. Thus we can conclude that the drop off in significant peptides was mainly due to changes in global protein abundance.

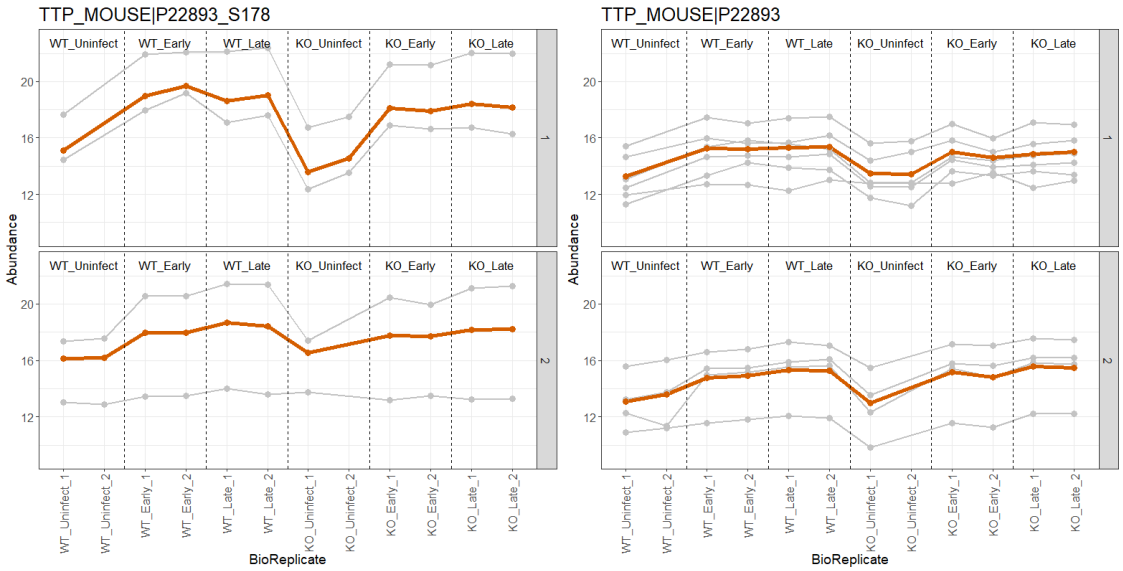

Figure S14.a: Comparing the global profiling of protein $TTP\_MOUSE|P22893$ with the modification of the protein at site $S178$. The individual PSM features are shown in grey, while the feature summarization is shown in red. When looking at the summary of the modification and global protein it is clear the difference between conditions follow the same trend. Specifically, there is a positive adjustment in abundance when comparing WT_Uninfect to WT_Late in both the modification and global profiling run. This indicates the movement is driven by changes in global protein.

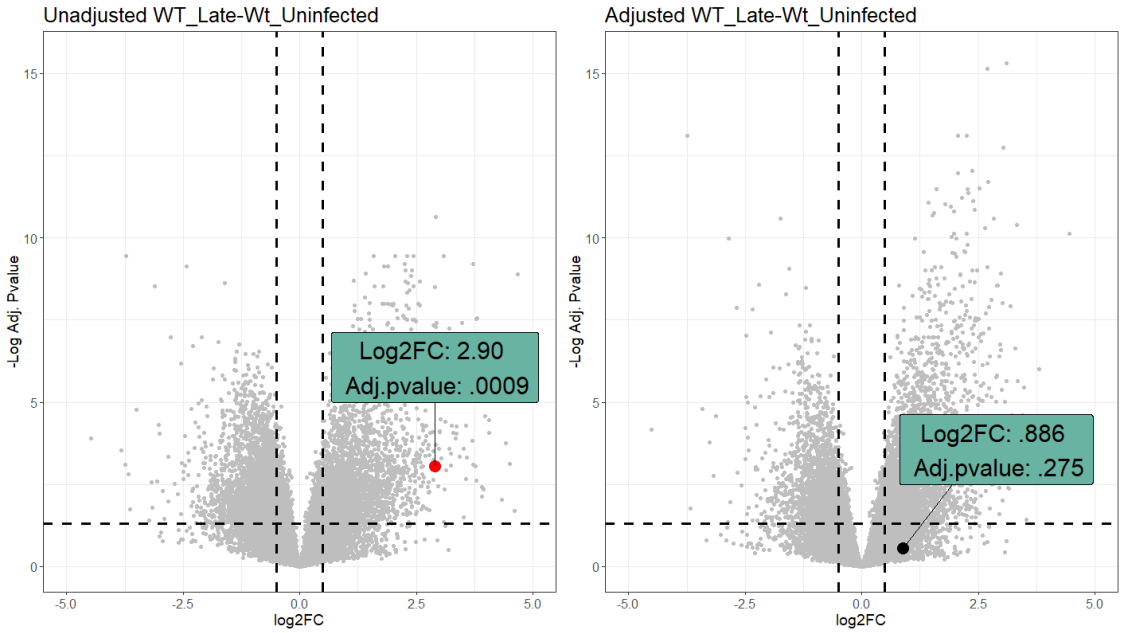

Figure S14.b: Volcano plots of WT_Late vs WT_Uninfect both before and after protein adjustment. The $TTP\_MOUSE|P22893\_S178$ modification is highlighted. Before adjustment the modification has a large fold change and significant adjusted pvalue. After adjustment the fold change is much smaller and the adjusted pvalue is insignificant.

Figure S14: Summary plots for modification of protein TTP at site S178.

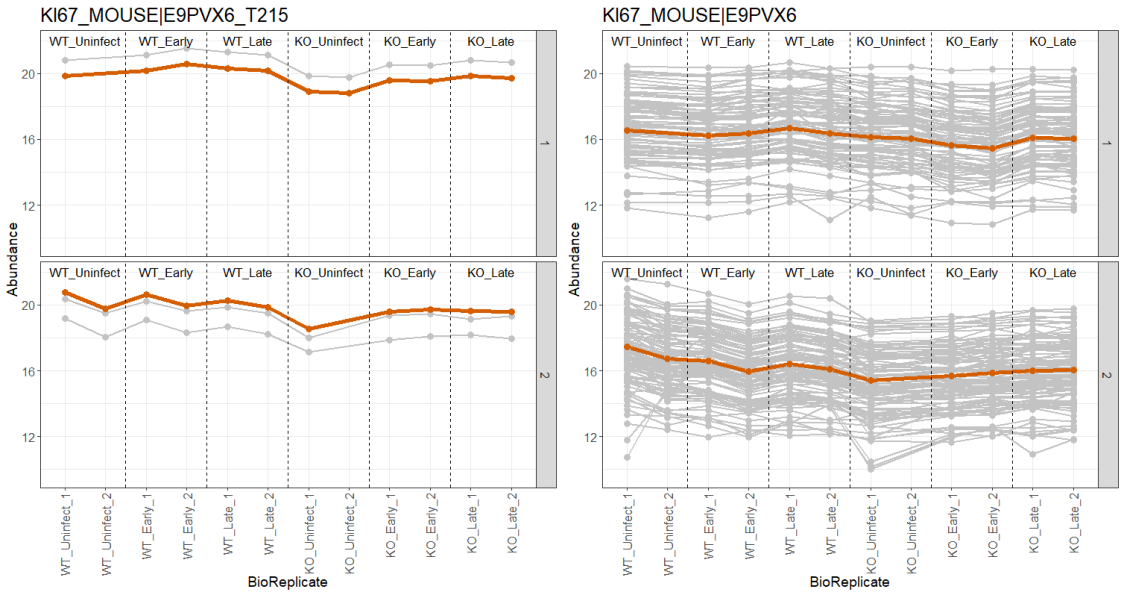

Figure S15.a: Comparing the global profiling of protein $KI67\_MOUSE|E9PVX6$ with the modification of the protein at site $T215$. In this case the modification and global protein trend in different directions. Specifically, comparing WT_Uninfect and WT_Early there is a slightly positive change in abundance, however in the global profiling there was a negative change. In this case the profile plot indicates the effect of the modification is masked by the change in global protein abundance. Additionally this profile plot shows the large difference in available features between modified peptides and global proteins.

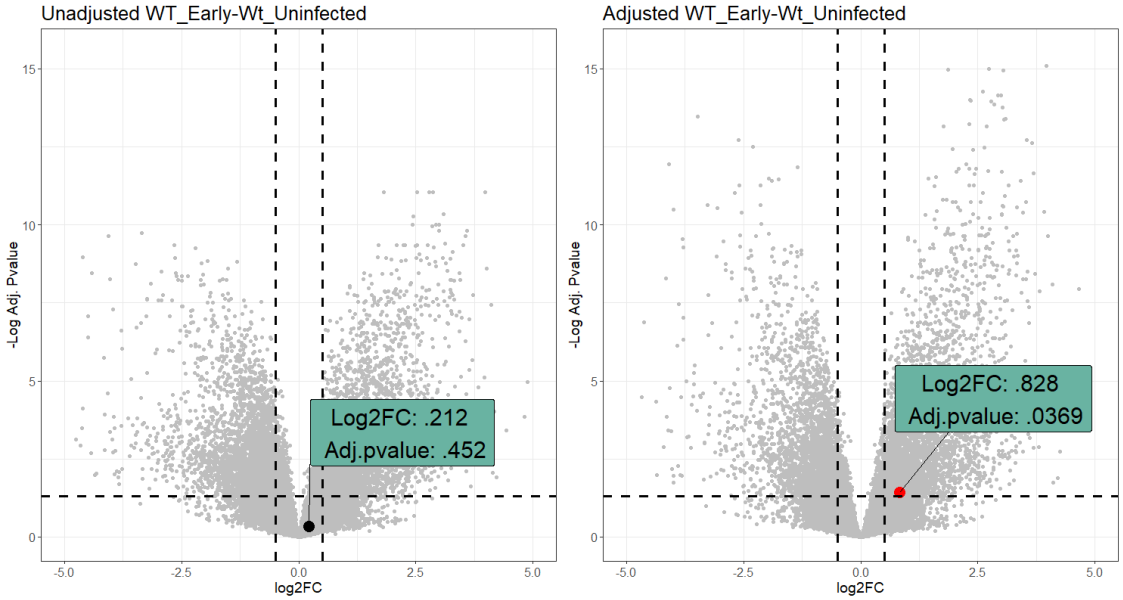

Figure S15.b: Volcano plots of WT_Early vs WT_Uninfect both before and after protein adjustment. The $KI67\_MOUSE|E9PVX6$ modification is highlighted. Before adjustment the modification has a small positive fold change and insignificant adjusted pvalue. After adjustment the fold change increases and the adjusted pvalue is now significant.

Figure S15: Summary plots for modification of protein KI67 at site T215.

5.3 Human label free quantification - no global profiling run

This experiment looked into the relationship between USP30 and protein kinase PINK1, and their association with Parkinson's Disease. Ubiquitination site profiling was performed and the modified site abundance was analyzed. Four conditions were tested with two biological replicates per condition. The conditions were as follows: CCCP, USP30 overexpression (USP30_OE), Combo, and Control. Label-free mass spectrometry quantification was used to quantify the abundance of modified peptides. A corresponding mixed effects model was fit per modification and global protein as described previously in this supplementary.

In contrast to other experiments analyzed in this paper, there was no unmodified global protein profiling run performed in this experiment. Once identification and quantification of the Ubiquitinated profiling was performed, peptides which were unmodified were extracted and used in place of a global profiling run. This resulted in a significant lack of overlap between modified and unmodified peptides. Any modified peptide without a corresponding unmodified protein could not be adjusted. Of the 10,799 modified peptides identified, only 4526 had a corresponding unmodified run and could be adjusted. Not having a separate global profiling run resulted in very low feature counts for the unmodified protein model.

An example profile plot for this experiment can be seen in Figure S16.

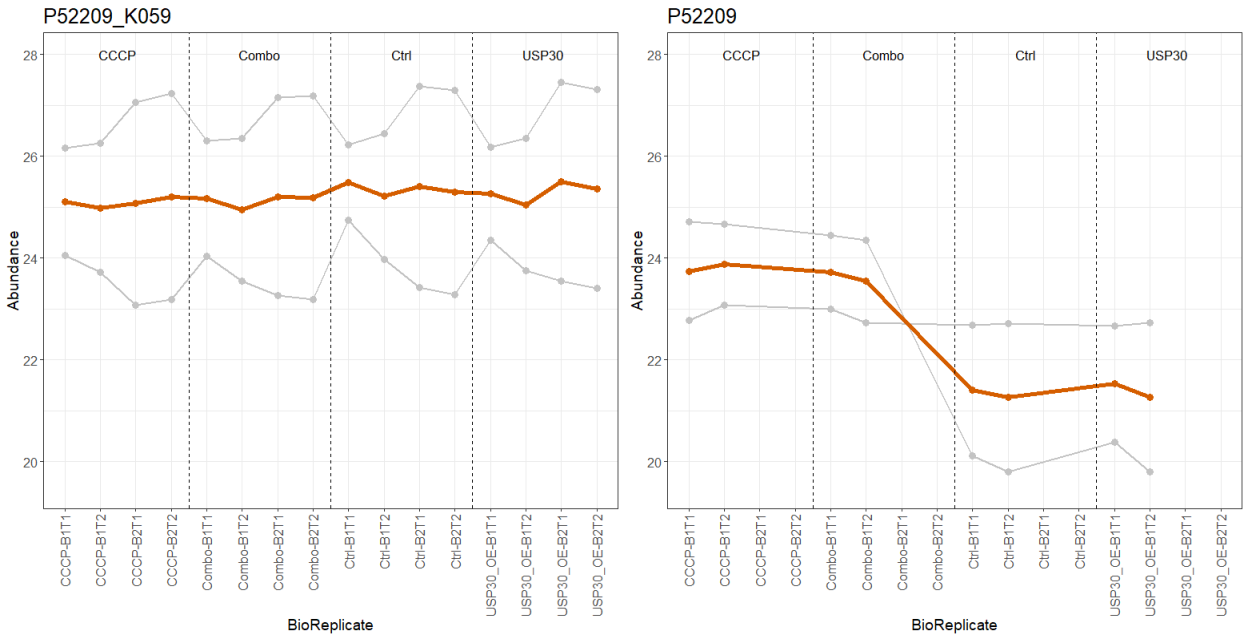

Figure S16: Comparing the global profiling of protein $P52209$ with the modification of the protein at site $K059$. The modification appears generally unchanged between all conditions, whereas the global profiling run shows the CCCP and Combo conditions have a higher relative abundance compared to the Control and USP30_OE. This indicates that the modification actually had a major effect when comparing CCCP and Combo to Control and USP30_OE. This would have been missed without adjusting for global protein changes.

6 Sample size calculation and power analysis

[TODO: Devon: Update this section per Olga's commentsv]

The proposed approach allows us to calculate the sample size needed to achieve a desired statistical power as described in Section 3.3. Given a desired power of .9, varying standard deviations, and an expected fold change, the number of replicates per condition can be seen in Figure S17

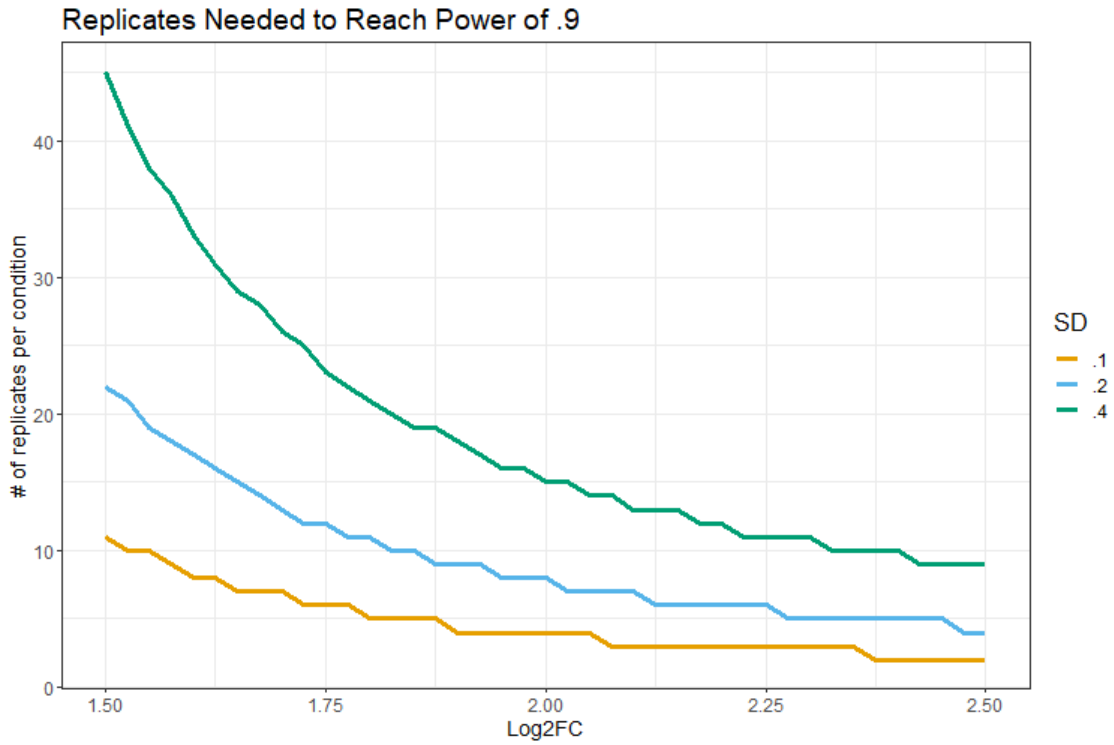

Figure S17: The number of replicates per condition required to reach a power of .9 can be seen in the plot. At a high standard deviation and low expected fold change the number of replicates is unrealistically high. As the expected fold change increases the number of replicates needed decreases dramatically. The standard deviation of .1 expectantly shows the least number of replicates needed regardless of fold change.

As described in Section 3.2 the proposed method adjusts for the underlying protein abundance in the PTM significance analysis, which corrects the confounding factor with a cost of increased variation. When the variation is increased, the experiment requires a larger number of replicates to reach the same power.

We compared the required sample size with varying standard deviations for the modified and unmodified peptide models. Three pairs of standard deviations of log-intensities for modified and unmodified peptides were used: (0.2, 0.1), (0.2, 0.2), and (0.2, 0.4). In the first pair the standard deviation for the unmodified peptide was less than the modified, in the second the standard deviations are the same, and in the third the standard deviation for the unmodified peptide was more than for the modified peptide. Again the desired power was .9 and the expected fold change was

varied. The results can be seen in Figure S18

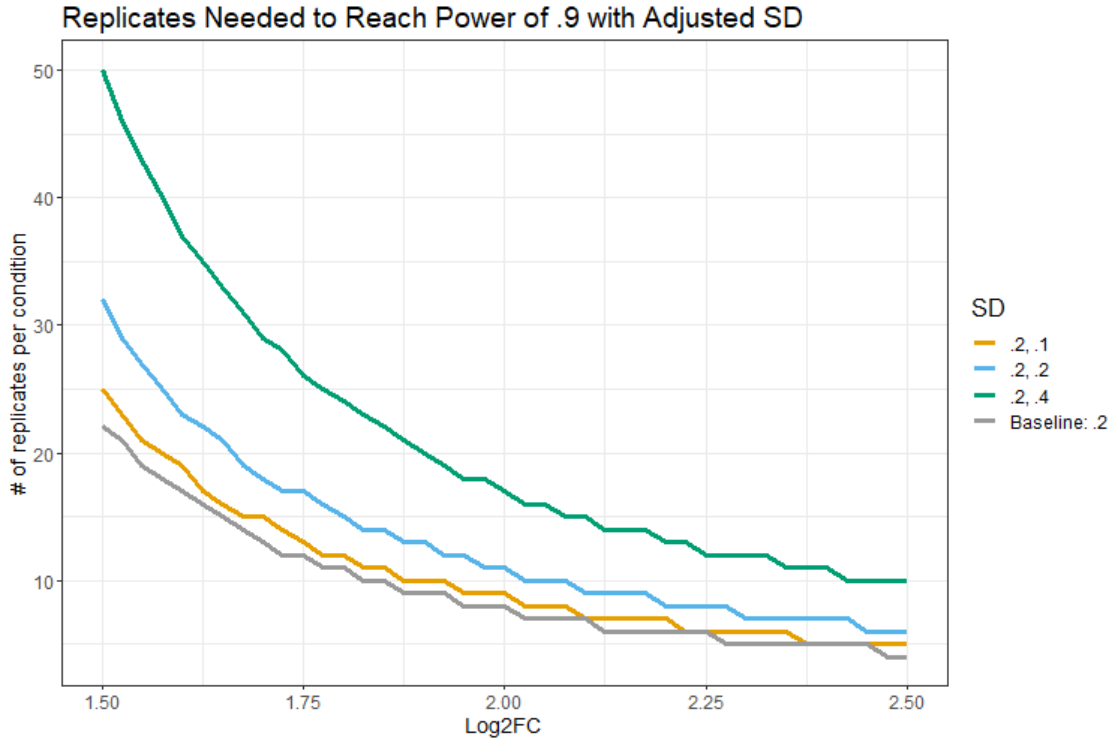

Figure S18: When the unmodified peptide has a low standard deviation the number of replicates is very similar to the baseline. As the standard deviation increases the number of replicates increases greatly. When the unmodified standard deviation is much higher than the modified, we can see that many more replicates are needed to reach the same power.

References

References

[1] Meena Choi et al. "MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments". In: *Bioinformatics* 30.17 (2014), pp. 2524–2526.

[2] T. Huang et al. "MSstatsTMT: Statistical Detection of Differentially Abundant Proteins in Experiments with Isobaric Labeling and Multiple Mixtures". In: *Molecular & Cellular Proteomics* 19 (10 Oct. 2020), pp. 1706–1723.

[3] Michael H. Kutner et al. *Applied Linear Statistical Models*. 5th ed. McGraw-Hill/Irwin, 2004.

[4] Giovanni Luchetti et al. "Shigella ubiquitin ligase IpaH7.8 targets gasdermin D for degradation to prevent pyroptosis and enable infection". In: *Cell Host & Microbe* (2021). ISSN: 1931-3128. DOI: https://doi.org/10.1016/j.chom.2021.08.010. URL: https://www.sciencedirect.com/science/article/pii/S1931312821003863.

[5] Timurs Maculins et al. "Proteomics of autophagy deficient macrophages reveals enhanced antimicrobial immunity via the oxidative stress response". In: *bioRxiv* (2020). DOI: 10.1101/2020.09.10.291344. eprint: https://www.biorxiv.org/content/early/2020/09/12/2020.09.10.291344.full.pdf. URL: https://www.biorxiv.org/content/early/2020/09/12/2020.09.10.291344.

[6] Ann L. Oberg and Olga Vitek. "Statistical design of quantitative mass spectrometry-based proteomic experiments". In: *Journal of Proteome Research* 8.5 (2009), pp. 2144–2156.

[7] Matthew E. Ritchie et al. "limma powers differential expression analyses for RNA-sequencing and microarray studies". In: *Nucleic Acids Research* 43.7 (Jan. 2015), e47–e47. ISSN: 0305-1048. DOI: 10.1093/nar/gkv007. eprint: https://academic.oup.com/nar/article-pdf/43/7/e47/7207289/gkv007.pdf. URL: https://doi.org/10.1093/nar/gkv007.