MSstatsPTM statistical relative quantification of post-translational modifications in global proteomics experiments

Supplementary Information

Devon Kohler[1], Tsung-Heng Tsai[2], Ting Huang[1], Erik Verschueren[4], Trent Hinkle[3], Meena Choi*[3], and Olga Vitek*[1]

[1]Khoury College of Computer Science, Northeastern University, Boston, MA, USA
[2]Kent State University, Kent, OH, USA
[3]MPL, Genentech, South San Francisco, CA, USA
[4]ULUA BV, Arendstraat 29, 2018 Antwerp, Belgium
*Corresponding Authors

March 5, 2022

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 and analysis 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 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.[1]

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

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

Adjustment for changes in the unmodified peptides is performed by taking the difference of log estimates, as follows

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

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 in a simple experimental design: 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   Design of PTM experiments

The proposed approach applied to experiments with complex designs still allows for sample size calculation and power analysis. To make the calculation the standard deviation and degrees of freedom are needed. The estimation of these quantities is dependent on the experimental design. In Figure S2 different experimental designs that the proposed approach can be applied to, along with their estimation of standard error and degrees of freedom is shown.

The standard error and degrees of freedom are estimated separately for both the PTM and unmodified protein models. Once this is done we can use these these values to calculate the sample size and perform power analysis. Given q, the desired false discovery rate, β, the average Type II error rate, $m_0/(m_0+m_1)$, the fraction of truly differentially modified PTM sites in the comparison. The minimal number of replicates is determined by the variance of the estimated log-fold change $\mathrm{SE}^2(\hat{\Delta})$ as

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

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. The degrees of freedom are estimated for the appropriate model using restricted maximum likelihood.

		Model	Estimated Log-fold change	Theoretical variance	Estimated variance	Degrees of freedom
Label-free (Y_{cs} is $\log_2$ intensity in Condition c and Subject s)	**Group comparison**	$Y_{cs} = \mu + Condition_c + \varepsilon_{cs}$ $\sum_{c=1}^{c} Condition_c = 1$ $\varepsilon_{cs} \sim iid\ N(0,\sigma^2)$	$\bar{Y}_{c.} - \bar{Y}_{c'.}$	$\dfrac{2\sigma^2}{S}$	$\dfrac{2S\sum_{c=1}^{C}(y_{cs} - \bar{y}_{c.})^2}{S(CS - C)}$	$CS - C$
	Time course	Y_{cs} $= \mu + Condition_c + Subject_s$ $+ \varepsilon_{cs}$ $\sum_{c=1}^{c} Condition_c = 1$ $Subject_s \sim iid\ N(0,\sigma_S^2)$ $\varepsilon_{cs} \sim iid\ N(0,\sigma^2)$	$\bar{Y}_{c.} - \bar{Y}_{c'.}$	$\dfrac{2\sigma^2}{S}$	$\dfrac{2\sum_{c=1}^{C}\sum_{s=1}^{S}(y_{cs} - \bar{y}_{s.} - \bar{y}_{c.} + \bar{y}_{..})^2}{S(C-1)(S-1)}$	$(C-1)(S-1)$
TMT (Y_{mcs} is $\log_2$ intensity in Condition c and Subject s from Mixture m)	**Group comparison**	Y_{mcs} $= \mu + Condition_c + Mixture_m$ $+ \varepsilon_{mcs}$ $\sum_{c=1}^{C} Condition_c = 1$ $Mixture_m \sim iid\ N(0,\sigma_M^2)$ $\varepsilon_{mcs} \sim iid\ N(0,\sigma^2)$	$\bar{Y}_{.c.} - \bar{Y}_{.c'.}$	$\dfrac{2\sigma^2}{MS}$	$\dfrac{2S\sum_{c=1}^{C}\sum_{m=1}^{M}(y_{mcs} - \bar{y}_{m..} - \bar{y}_{.c.} + \bar{y}_{..})^2}{MS(MCS - C - M + 1)}$	$MCS - C - M + 1$
	Time course	Y_{mcs} $= \mu + Condition_c + Subject_{sm}$ $+ \varepsilon_{mcs}$ $\sum_{c=1}^{C} Condition_c = 1$ $Subject_{sm} \sim iid\ N(0,\sigma_S^2)$ $\varepsilon_{mcs} \sim iid\ N(0,\sigma^2)$	$\bar{Y}_{.c.} - \bar{Y}_{.c'.}$	$\dfrac{2\sigma^2}{MS}$	$\dfrac{2S\sum_{c=1}^{C}\sum_{m=1}^{M}(y_{mcs} - \bar{y}_{ms.} - \bar{y}_{.c.} + \bar{y}_{...})^2}{MS(C-1)(MS-1)}$	$(C-1)(MS-1)$

Figure S2: Different models that are fit depending on the experimental design. Models are fit for Label-free and TMT acquisition methods, as well as group comparison and time course experimental designs. The estimation of standard error and degrees of freedom for each model is shown.

4   Experimental datasets in this manuscript

4.1   Experiments with known ground truth

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

4.1.1   Computer simulation

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 six different approaches as follows: 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. While multiple conditions were generated, all comparisons were still made with the same fold change between conditions (0.75). This ensured that all peptides were differentially abundant by the same amount and the only difference in TP/FP/TN/FN calculations was using more conditions in the modeling.

4.1.2   Dataset 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 unmodified protein: 10

- Missing data: no missing value

The results are summarized from Figure S3 to Figure S4. These results include the false discovery rate with and without adjusting for changes in unmodified protein abundance, 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 as true positives with a differential fold change between conditions, while the other 500 were not differential and treated as true negatives. In terms of the true negatives, 250 were simulated such that any abundance changes were entirely driven by changes in the unmodified peptide. We would expect any model without performing protein adjustment to treat these as false positives. The other 250 were simulated with no significant change in either PTM or overall protein abundance.

4.1.3   Dataset 2 : Computer simulation 2 - Label-free missing values and low features

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 in a biological experiment and provides a more realistic expectation of model performance. The number of PTM features and missing data percentage where determined by looking at the features and missing data in the biological experiments in this paper.

- 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 unmodified protein: 10

- Missing data: 20% of the observations for PTMs and Proteins were masked with NA at random

The results are summarized from Figure S6 to Figure S7.

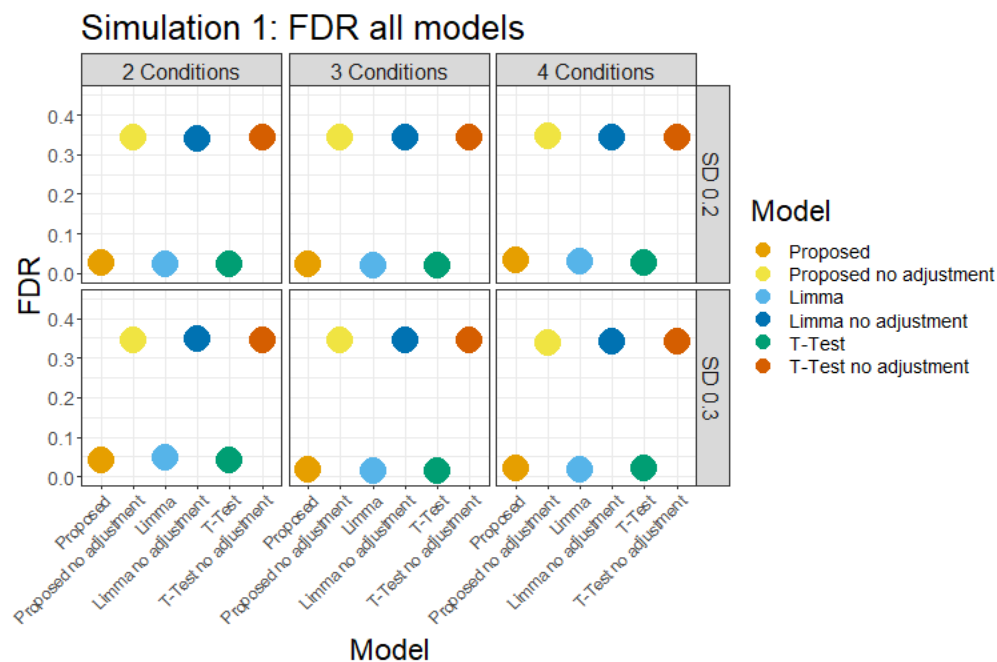The portion of significant peptides were done in the same was as described in simulation 1.

Figure S3.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. The performance of the models without adjustment was much lower than those with adjustment, thus only models with adjustment are compared going forward.
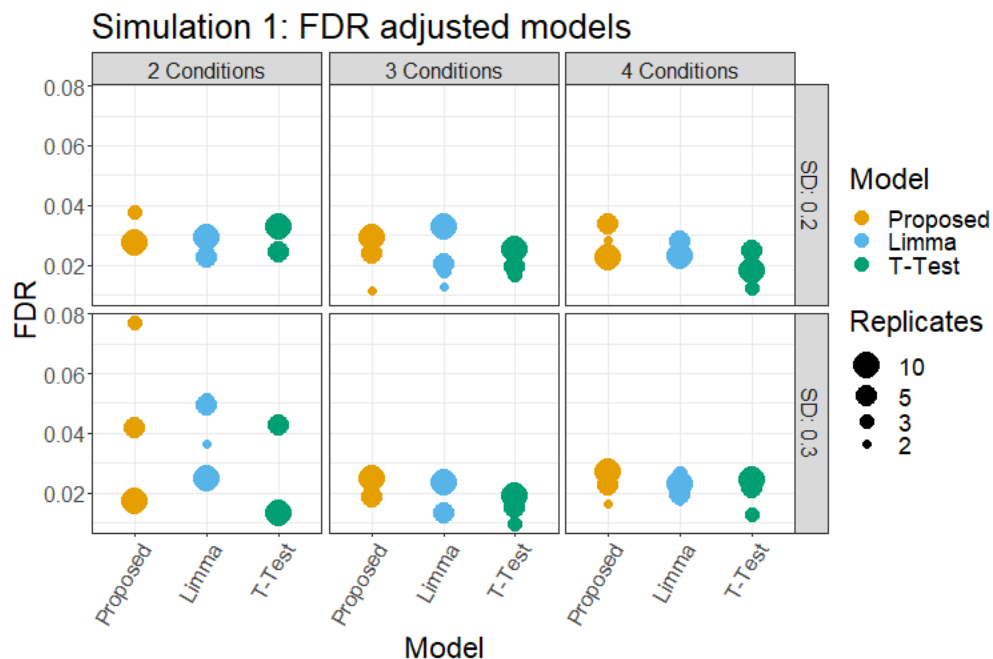

Figure S3.b: The considered methods with protein adjustment are compared in detail. All three methods with adjustment generally performed similarly in terms of FDR.

Figure S3: FDR of Simulation 1.
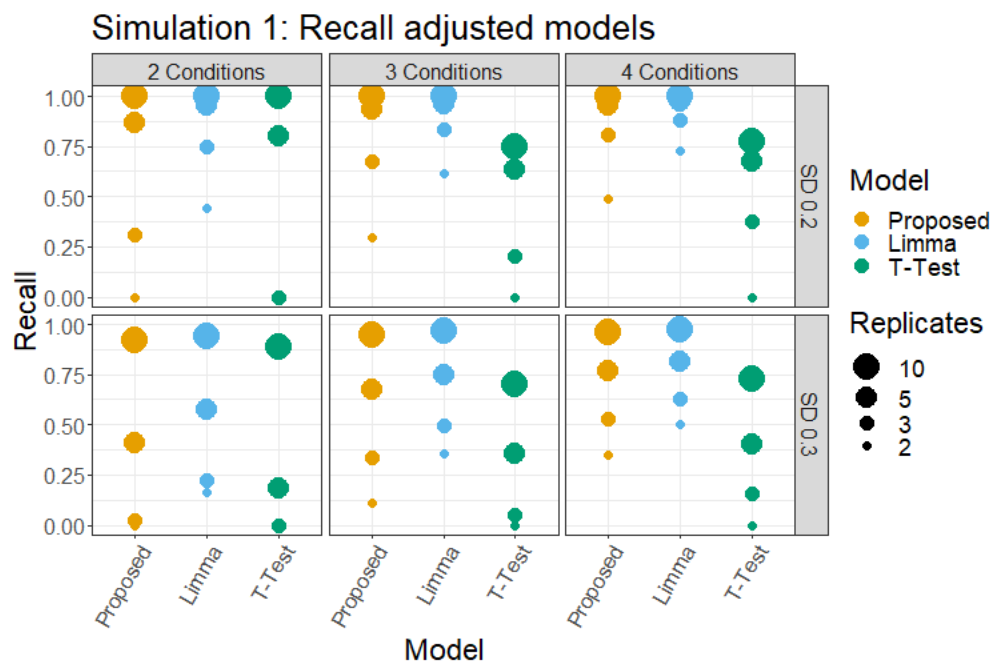
Figure S4.a: The recall for simulation 1 methods with adjustment were compared. It is clear that Limma performs the strongest here when the number of replicates were low. At higher replicates the performance of the proposed methods and Limma are comparable. *t*-test clearly performs worse across all methods.
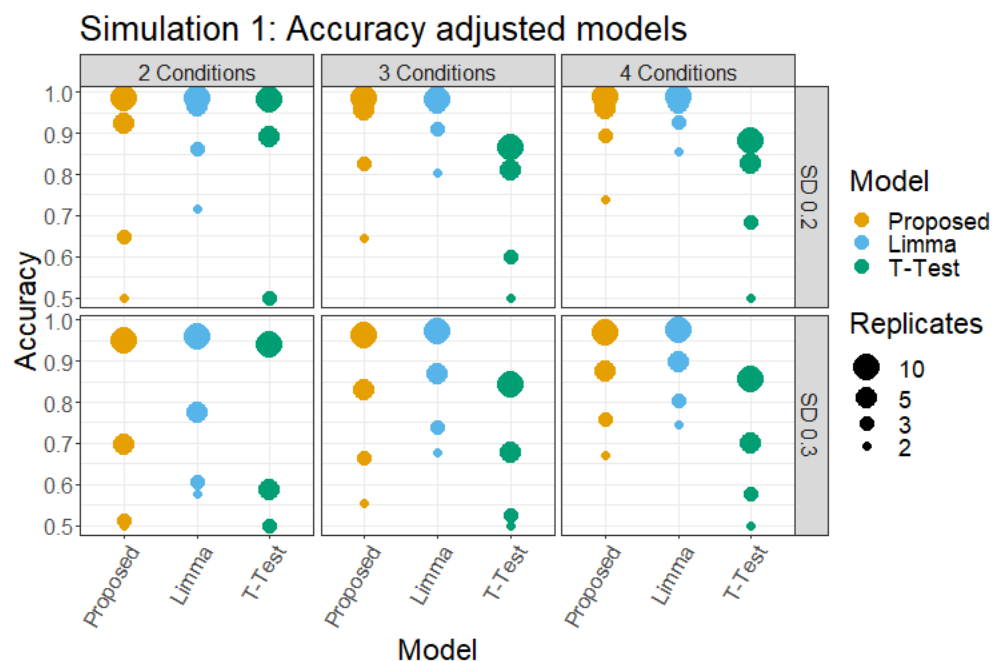

Figure S4.b: The overall accuracy plot mimics the observations in the recall plot. Limma performs stronger than the proposed method at lower replicates, while at higher they are comparable.

Figure S4: Recall and Accuracy results of Simulation 1.
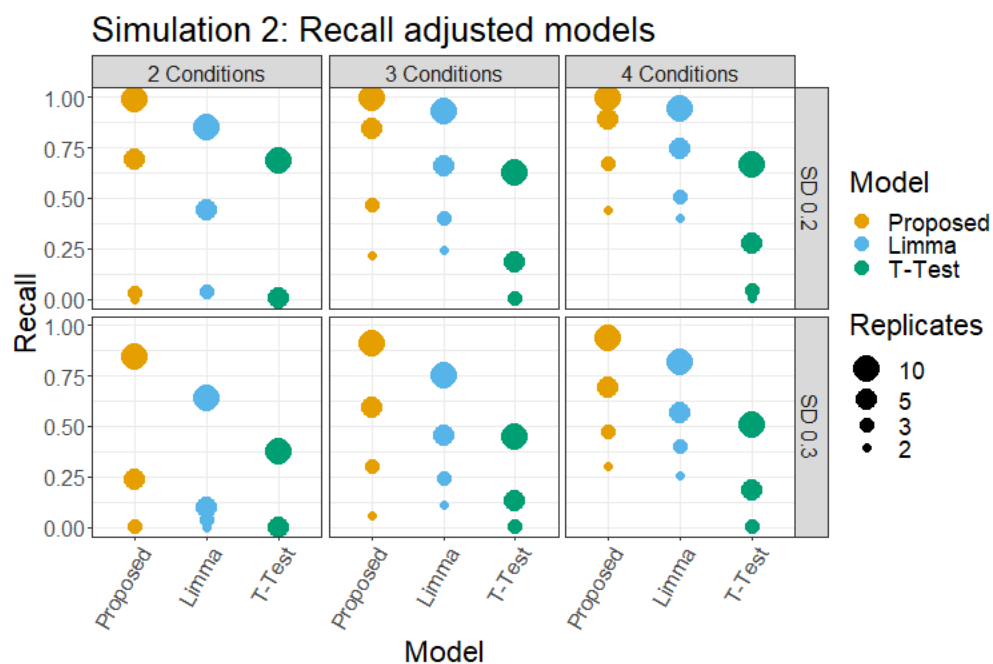
Figure S5

Figure S6: Recall results of Simulation 2. 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.
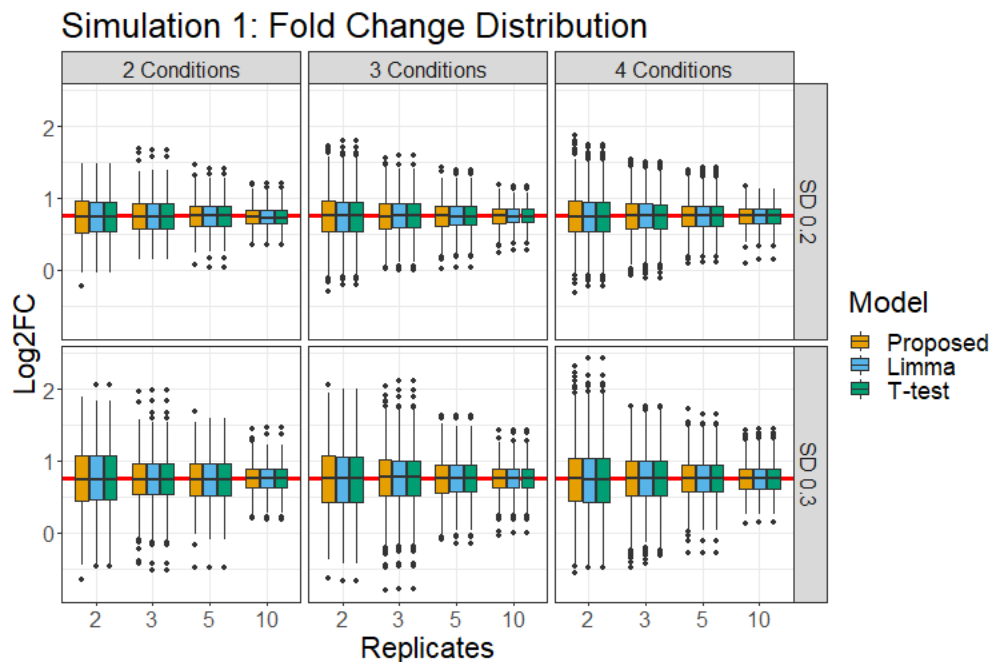
Figure S7.a: In simulation 1 all considered methods correctly estimated the fold change between conditions, with a median fold change estimation of .75. The distributions around the median were also consistent across all methods. Predictably the distribution quantiles decrease as the number of replicates increases.
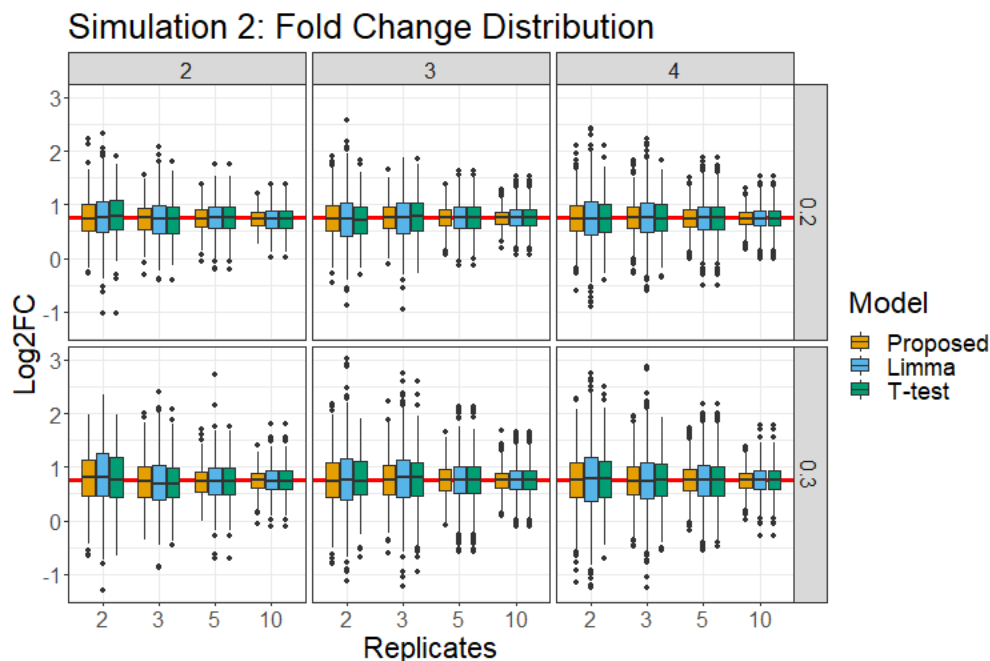

Figure S7.b: In simulation 2 all methods correctly estimated the fold change with a median log change of .75. The proposed method in this simulation had a visibly tighter distribution around the median. Both Limma and t-test showed a wider range around the fold change. The inner quartile range of the Proposed method was on average 10.4% smaller than t-test and 21.8% smaller than Limma. In this case the proposed method showed a stronger performance in correctly estimating the fold change for all peptides.

Figure S7: Fold change distribution comparison between Simulations 1 and 2.

4.1.4   Dataset 3 : SpikeIn benchmark - Ubiquitination - Label-free

A custom designed experiment with labeling was used to assess the performance of the proposed method in a real experimental setting. Heavy-labeled 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 have a known fold change between conditions and are expected to be differential in most cases. Additionally, background E coli lysate was used to normalize total protein levels prior to the enrichment and global protein profiling. The background lysate was treated as controls and are not expected to be differential in any comparison.
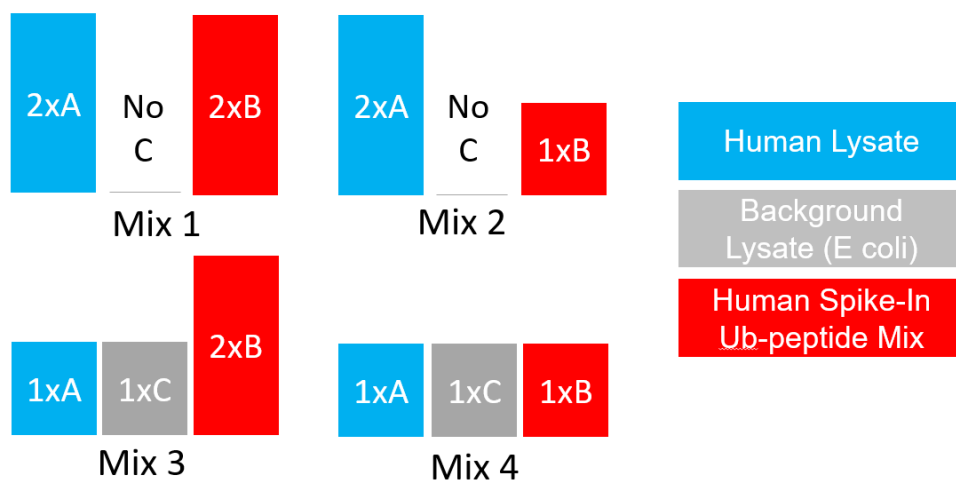

Figure S8: An overview of the spike-in benchmark experimental design. The corresponding levels of spike-in peptide, human lysate, and background lysate for each mixture are shown. The spike-in peptide levels are adjusted for the human lysate levels. For example, the true spike-in abundance of mixture 1 is entirely accounted for by human lysate, whereas mix 2 has double the amount human lysate compared to spike-in abundance. When comparing mix 1 vs mix 2 we would expect to see a log fold change of -1. In this way we know the expected log fold change between mixture comparisons.

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 S9 to Figure S12, including volcano plots, model summary statistics, and fold change analysis.
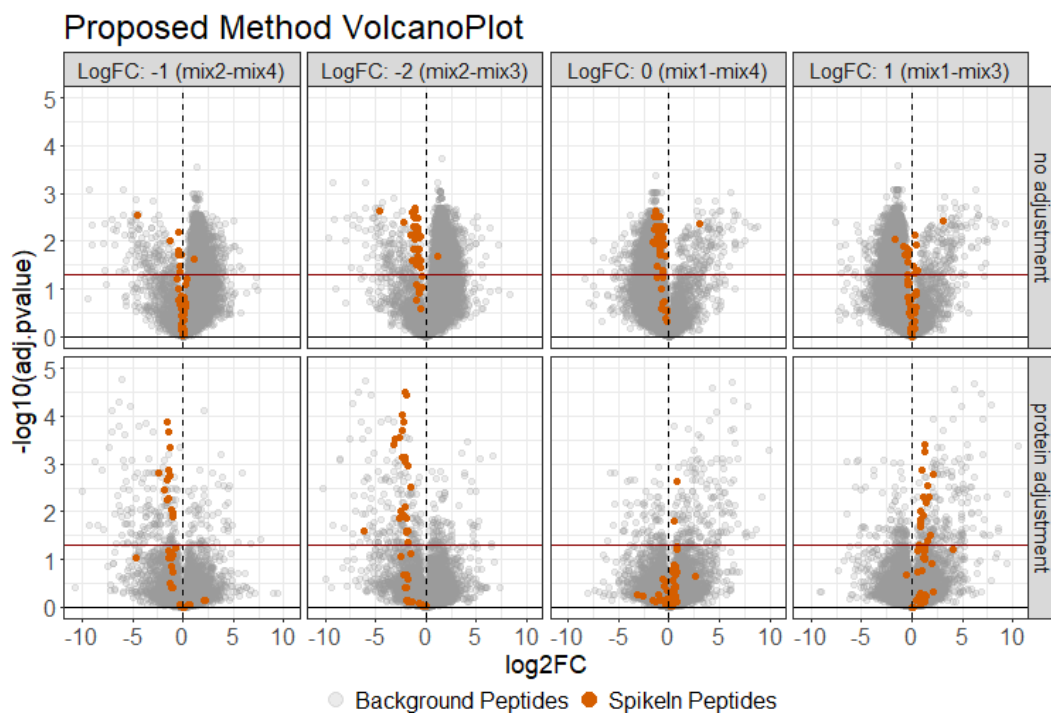
Figure S9: The modeling results of the proposed method both before and after adjustment. The spike-in peptides are colored red and the background peptides are colored grey. All grey peptides are expected to be insignificant. Using the proposed method to model the benchmark experiment, the spike-in peptides (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 these false positives were decreased considerably.
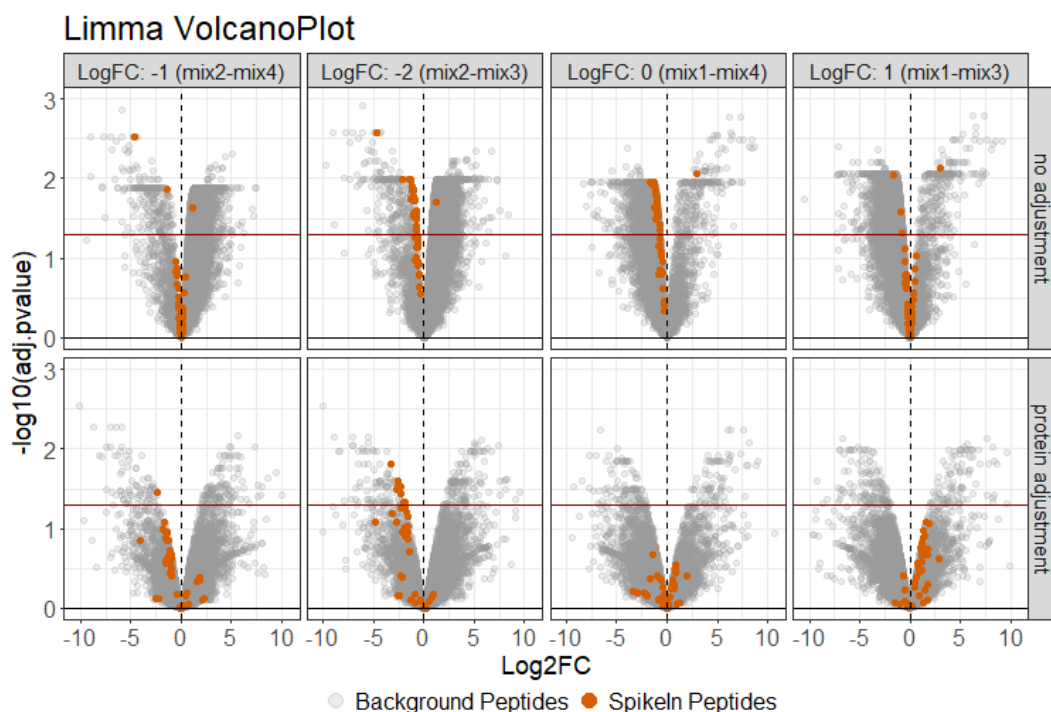
Figure S10: When modeling the experiment with the Limma method, the spike-in peptides again follow the expected log fold change better after adjusting for changes in protein level. However, while the fold change is much more accurate, the majority of spike-in peptides do not have a significant adjusted pvalue. In this case, the known differential 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.
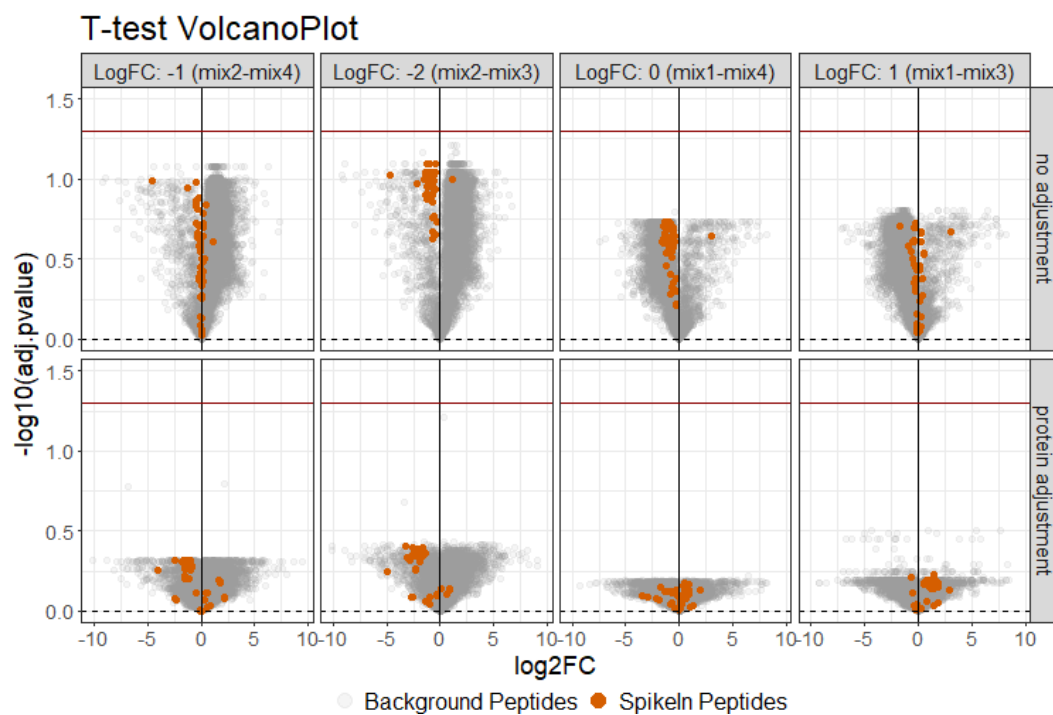
Figure S11: Using the two sample *t*-test, none of the comparisons either before or after adjustment show any significant peptides. With that being said, the fold change of the spike-in peptides is much closer to expectation after adjusting for global protein abundance.
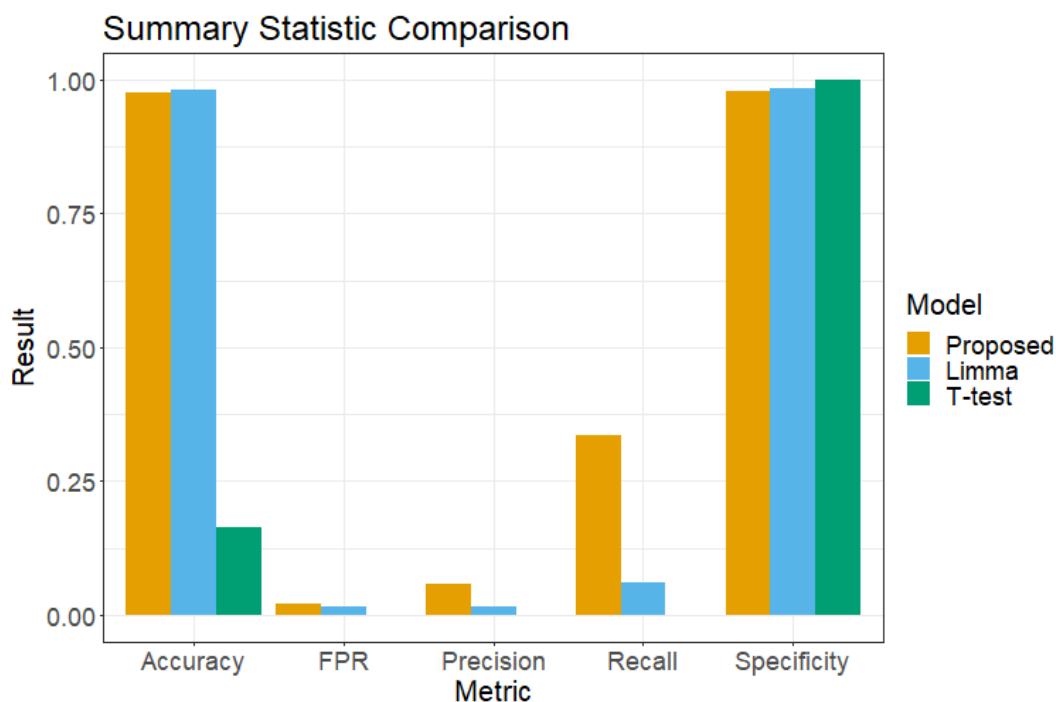
Figure S12: Comparing the summary statistics between methods it is clear that the proposed method performs the strongest. In terms of accuracy and specificity the three methods are close, with Limma and *t*-test showing slightly higher values. Accuracy and specificity are dominated by the large number 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.

4.2   Biological investigations

4.2.1   Dataset 4 : Human - Ubiquitination - 1mix-TMT

In this experiment Shigella ubiquitin ligase IpaH7.8 was shown to inhibit the protein gasdermin D (GSDMD). [2] 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 unbalanced with two biological replicates per condition for all conditions except for infected 1 hour. The experiment was modeled as a group comparison, with different replicates measured at each time point. 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 Dataset 4

A model was fit for the total protein and ubiquitination separately. The model formula can be seen below.

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

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

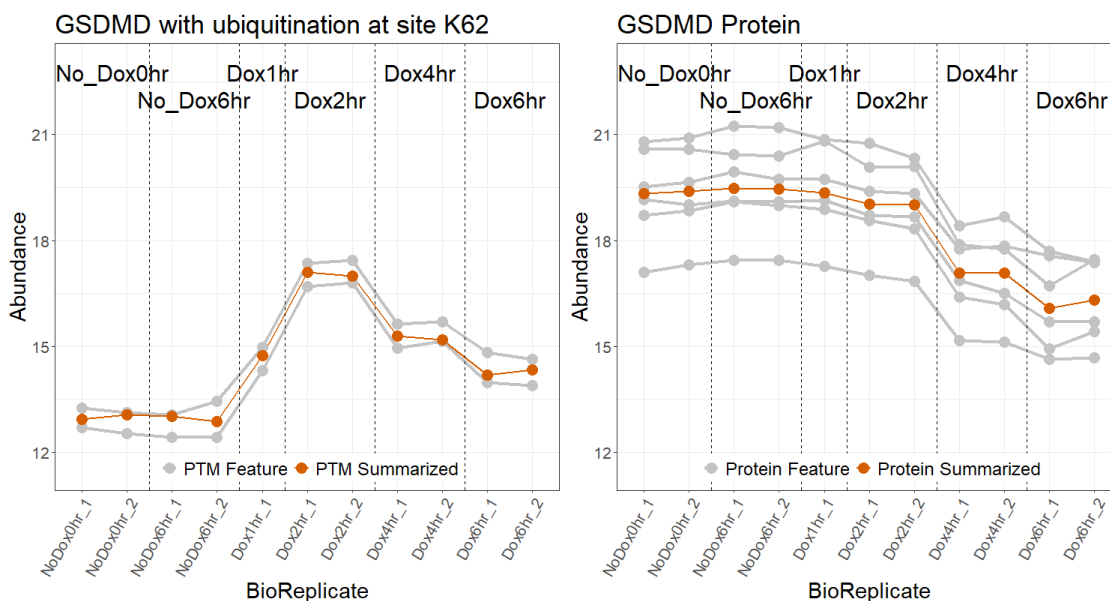The results of the proposed method to this experiment can be seen in Figure S13.

Figure S13.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.
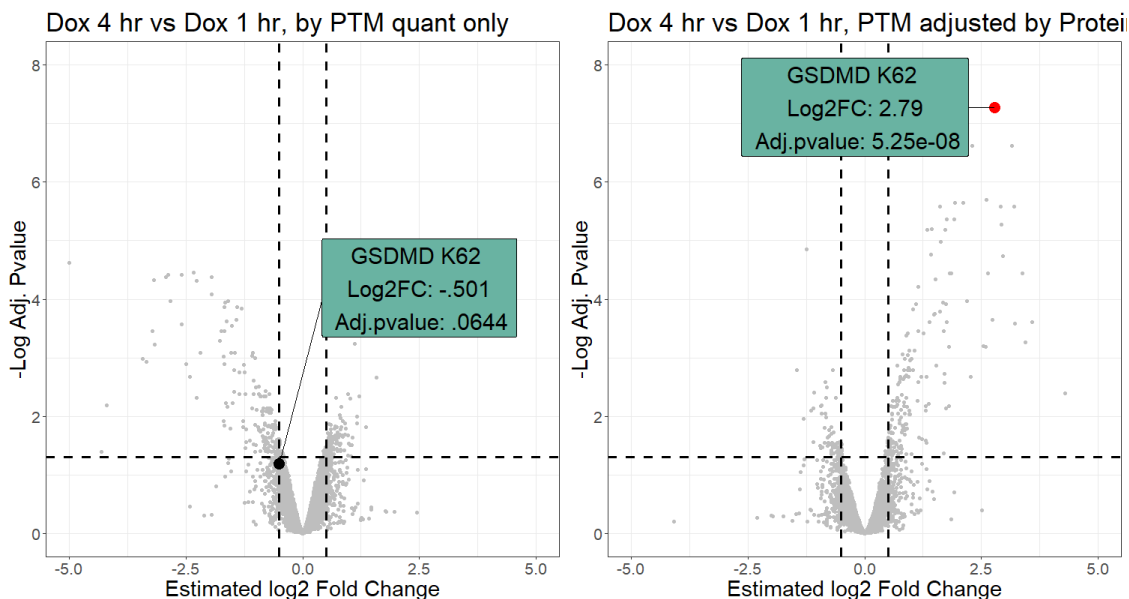

Figure S13.b: Volcano plots of Dox4hr vs Dox1hr both before and after protein adjustment. The $GSDMD\_HUMAN|P57764\_K62$ modification is highlighted. Before adjustment the modification has a small fold change and insignificant adjusted pvalue. After adjustment the fold change is much larger and the adjusted 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 S13: Summary plots for modification of protein GSDMD at site K62.

19

4.2.2   Dataset 5 : Mouse - Phosphorylation - 2mix-TMT time series

In this study, the correlation between the gene Atg16L1 and killing of Shigella flexneri (S.flexneri) was assessed. [3] 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). The experiment was modeled as a time series, with the same biological replicate being measured at each time point. 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 Dataset 5

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.

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

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

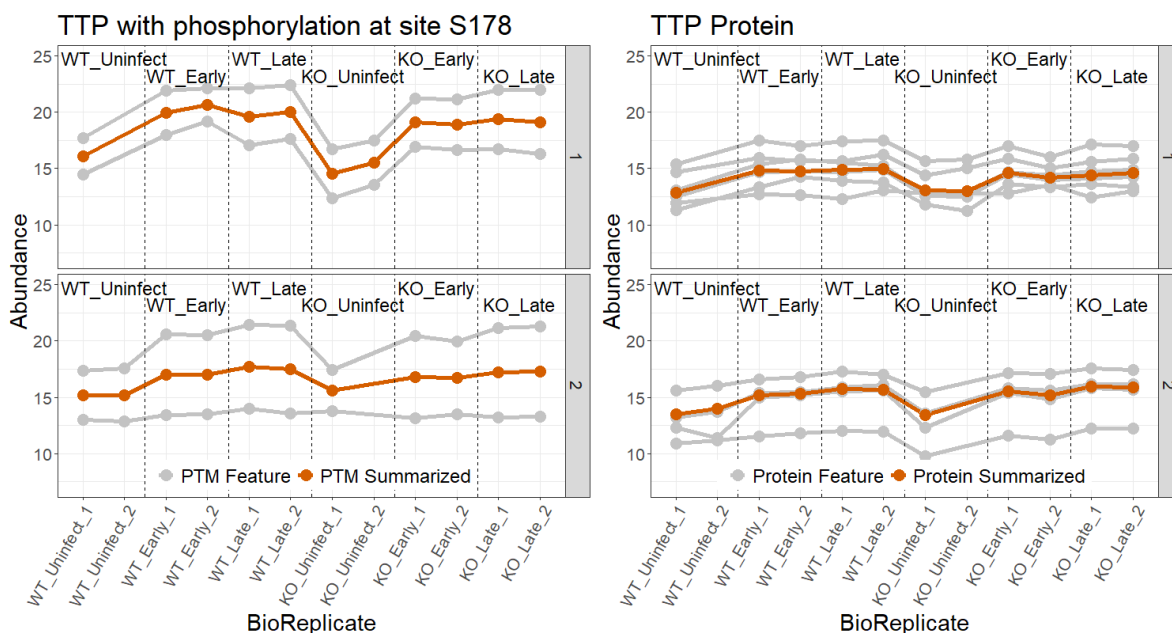The results of the proposed method to this experiment can be seen in Figure S14 and Figure S15.

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.
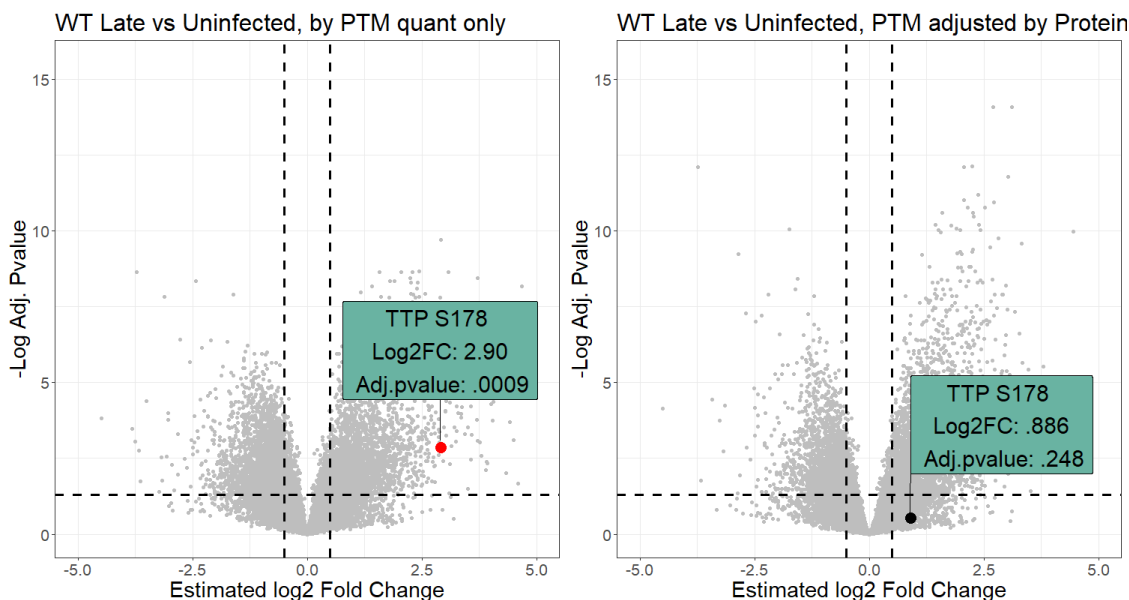

Figure S14.b: Volcano plots of WT_Late vs WT_Uninfect both before and after protein adjustment. The $TTP\_MOUSE|P22893\_S178$ modification is highlighted. Before adjustment the modification has a large fold change and significant adjusted pvalue. After adjustment the fold change is much smaller and the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

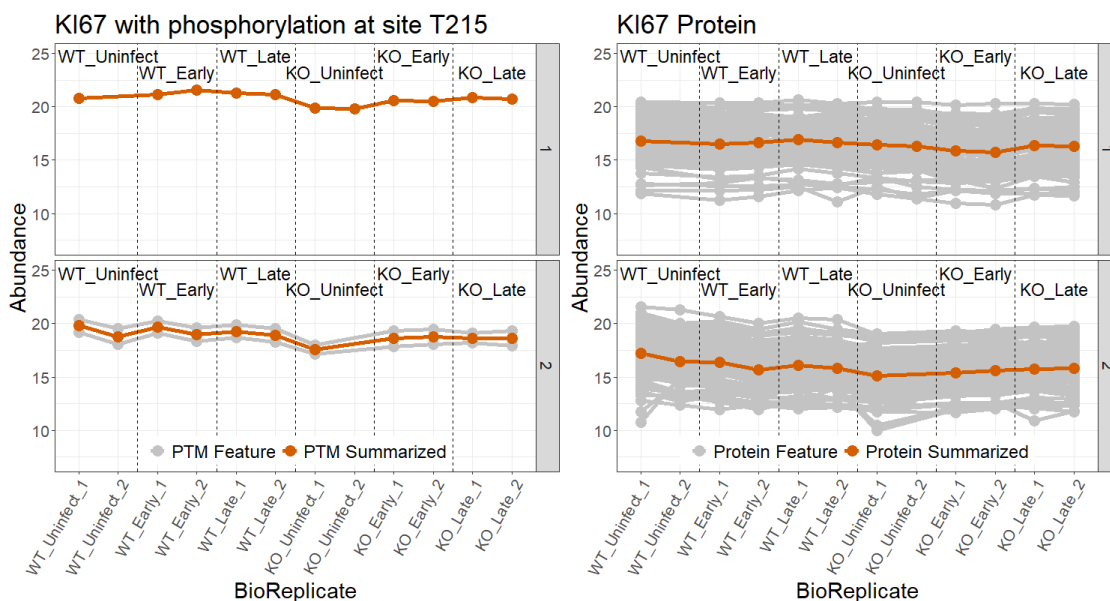Figure S14: Summary plots for modification of protein TTP at site S178.

Figure S15.a: Comparing the global profiling of protein $KI67\_MOUSE|E9PVX6$ with the modification at site $T215$. In this case the modification and global protein appear to show very small to no differences between conditions, however after adjusting for change in global protein abundance, the modification is statistically significant. Additionally, this profile plot shows the large difference in available features between modified peptides and global proteins.
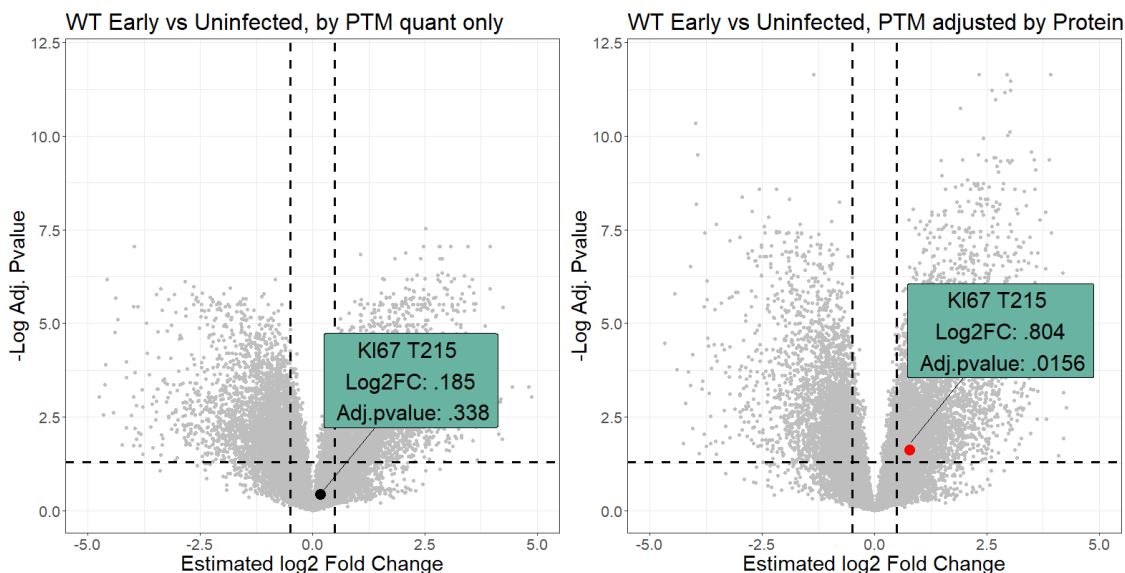

Figure S15.b: The volcano plot of the WT_Uninfect and WT_Early comparison show the specifics of the adjustment. The modification looks to be flat, with a log fold change of .185, while the global profiling showed a small negative fold change of $-.619$. While both exhibit small changes, when combined we can see a log fold change of .804 and adjusted p-value of .0156. In this case we can see the strength of the proposed method, showing noteworthy modifications which would be missed otherwise.

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

4.2.3   Dataset 6 : Human - Ubiquitination - Label-free 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. The experiment was modeled as a group comparison.

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 and low feature counts for the unmodified protein model.

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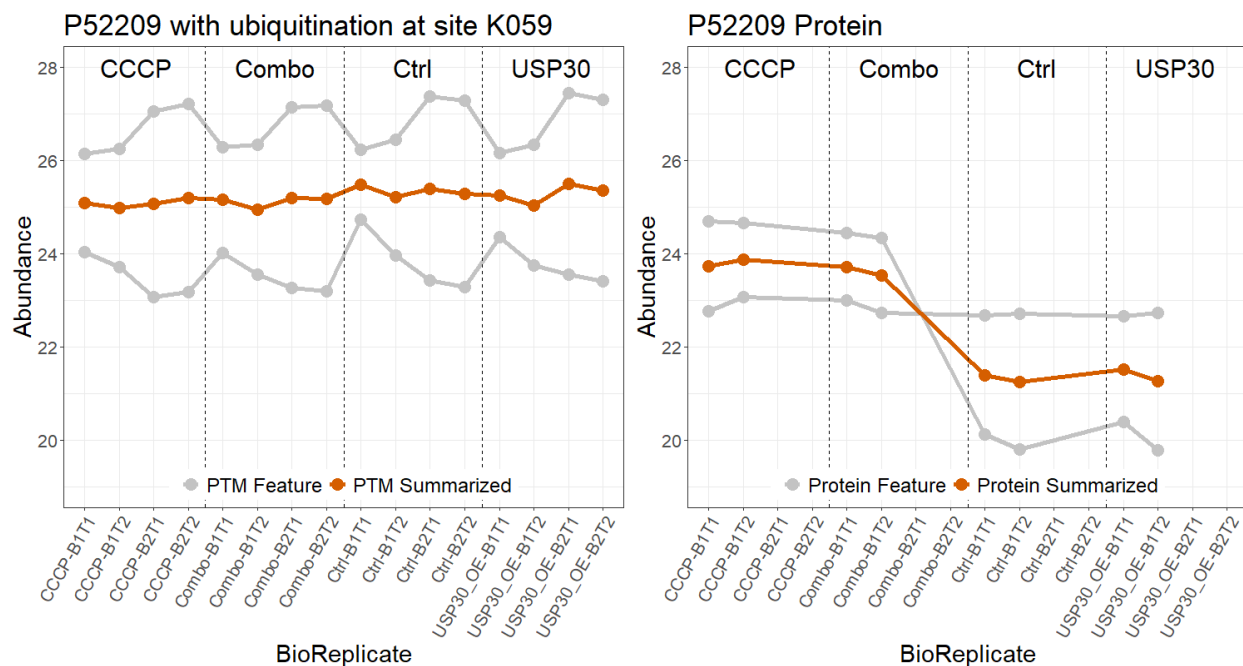

Figure S16: Comparing the global profiling of protein $P52209$ with the modification of the protein at site $K059$. The modification appears generally unchanged between all conditions, whereas the global profiling run shows the CCCP and Combo conditions have a higher relative abundance compared to the Control and USP30_OE. This indicates that the modification actually had an effect when comparing CCCP and Combo to Control and USP30_OE. However it is not entirely clear as one unmodified peptide feature appears to be changed, while the other does not. This uncertainty is another result of not running a separate global profiling run. With a global profiling run, many unmodified features are generally quantified, removing the uncertainty that comes with low feature counts.
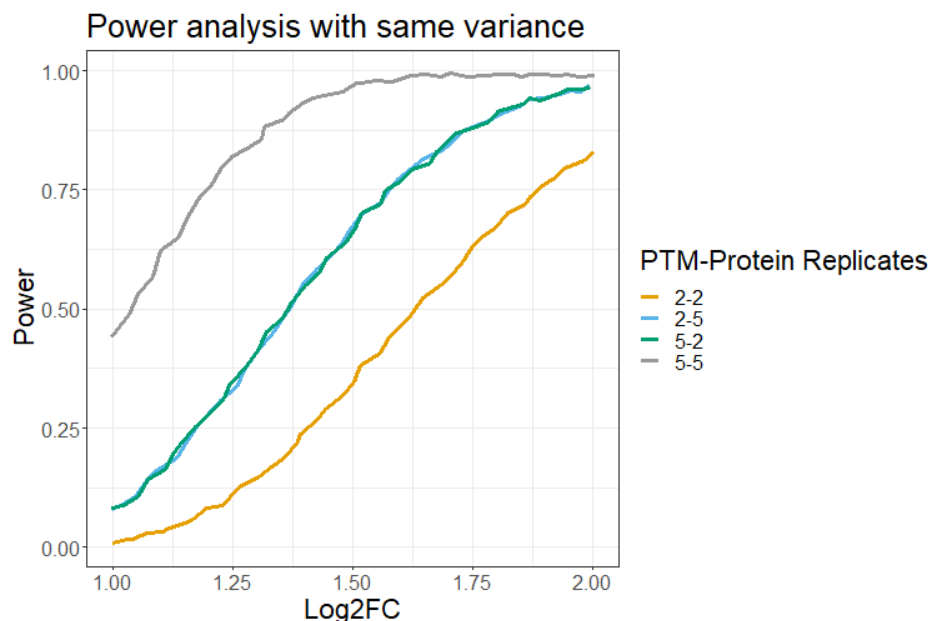
5   Sample size calculation and power analysis

Figure S17.a: The power of an experiment targeting PTMs with the same variance, .15, for the modified and unmodified peptides. Predictably when the replicates are high for both modified and unmodified peptides the power is much higher. Conversely at low replicates for each the power is much lower. The interesting part of this chart is when the replicates are different between runs. With equal variance, it does not matter if the PTM replicates or protein replicates are higher.
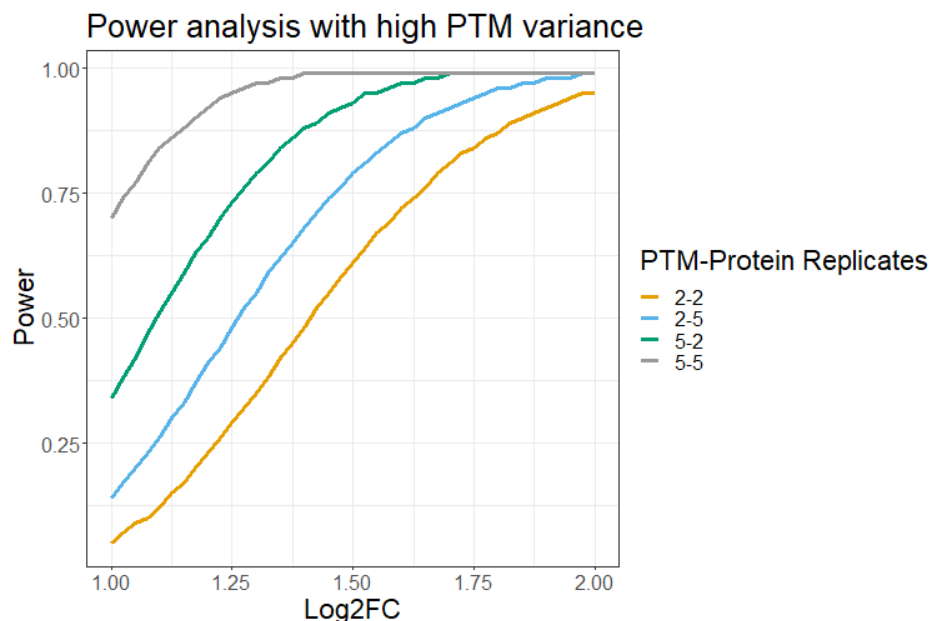

Figure S17.b: In this chart the variance for the PTM is higher than the unmodified protein. The PTM variance is .2, while the unmodified protein variance is .1. With equal replicates the results are the same as above, more replicates equals more power. When the replicates are not the same we can clearly see that having more replicates for the PTM leads to higher power.

Figure S17: Power analysis of experiments with differing variances.

References

[1] 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`.

[2] 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`.

[3] 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`.