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

# Supplementary Information

Devon Kohler[1], Tsung-Heng Tsai[2], Erik Verschueren[4], Ting Huang[1], Trent Hinkle[3], Lilian Phu[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

# Contents

# 1 Proposed approach

## 1.1 Design of PTM experiments - Extension to complex designs

| | | Model | Estimated Log-fold change | Theoretical variance | Estimated variance | Degrees of freedom |
|---|---|---|---|---|---|---|
| **Label-free** ($y_{ij}$ is $\log_2$ intensity in Condition $i$ and BioReplicate $j$) | **Group comparison** | $y_{ij} = \mu_i + \varepsilon_{ij}$ $\sum_{i=1}^{I} \mu_i = 0$ $\varepsilon_{ij} \sim iid\ N(0,\sigma^2)$ | $\bar{Y}_{i.} - \bar{Y}_{i'.}$ | $\frac{2\sigma^2}{I}$ | $\frac{2J\sum_{i=1}^{I}(y_{ij} - \bar{y}_{i.})^2}{J(IJ-I)}$ | $IJ - I$ |
| | **Time course** | $y_{ij} = \mu_i + BioReplicate_j + \varepsilon_{ij}$ $\sum_{i=1}^{I} \mu_i = 0$ $BioReplicate_j \sim iid\ N(0,\sigma_J^2)$ $\varepsilon_{ij} \sim iid\ N(0,\sigma^2)$ | $\bar{Y}_{i.} - \bar{Y}_{i'.}$ | $\frac{2\sigma^2}{I}$ | $\frac{2\sum_{i=1}^{I}\sum_{i=1}^{J}(y_{ij} - \bar{y}_{i.} - \bar{y}_{j.} + \bar{y}_{..})^2}{J(I-1)(J-1)}$ | $(I-1)(J-1)$ |
| **TMT** ($y_{mij}$ is $\log_2$ intensity in Condition $i$ and BioReplicate $j$ from Mixture $m$) | **Group comparison** | $y_{mij} = \mu_i + Mixture_m + \varepsilon_{mij}$ $\sum_{i=1}^{I} \mu_i = 0$ $Mixture_m \sim iid\ N(0,\sigma_M^2)$ $\varepsilon_{mij} \sim iid\ N(0,\sigma^2)$ | $\bar{Y}_{i.} - \bar{Y}_{i'.}$ | $\frac{2\sigma^2}{MJ}$ | $\frac{2J\sum_{i=1}^{I}\sum_{m=1}^{M}(y_{mij} - \bar{y}_{m..} - \bar{y}_{.i.} + \bar{y}_{...})^2}{MJ(MIJ-I-M+1)}$ | $MIJ - I - M + 1$ |
| | **Time course** | $y_{mij} = \mu_i + BioReplicate_{jm} + \varepsilon_{mij}$ $\sum_{i=1}^{I} \mu_i = 0$ $BioReplicate_{jm} \sim iid\ N(0,\sigma_J^2)$ $\varepsilon_{mij} \sim iid\ N(0,\sigma^2)$ | $\bar{Y}_{i.} - \bar{Y}_{i'.}$ | $\frac{2\sigma^2}{MJ}$ | $\frac{2J\sum_{j=1}^{I}\sum_{m=1}^{M}(y_{mij} - \bar{y}_{mj.} - \bar{y}_{.i.} + \bar{y}_{...})^2}{MJ(I-1)(MJ-1)}$ | $(I-1)(MJ-1)$ |

Figure S1: 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. Only one side of the PTM and unmodified protein is shown, variance must be multiplied by two to combine.

# 2 Experimental datasets

## 2.1 Experiments with known ground truth

[TODO: Could you check that we have the same method names and fonts for the models here as in the main text?] We compared six different approaches as follows: a) proposed approach, b) proposed approach without adjusting for unmodified peptides c) ANOVA [1] (with adjustment), d) ANOVA (no adjustment), e) Limma [2] (with adjustment), and f) Limma (no adjustment).

### 2.1.1 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.

- 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, 1., 2., 3.
- Difference in protein abundance between conditions: 0, 1., 2., 3.
- 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

### 2.1.2 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.

- 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, 1., 2., 3.
- Difference in protein abundance between conditions: 0, 1., 2., 3.
- 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

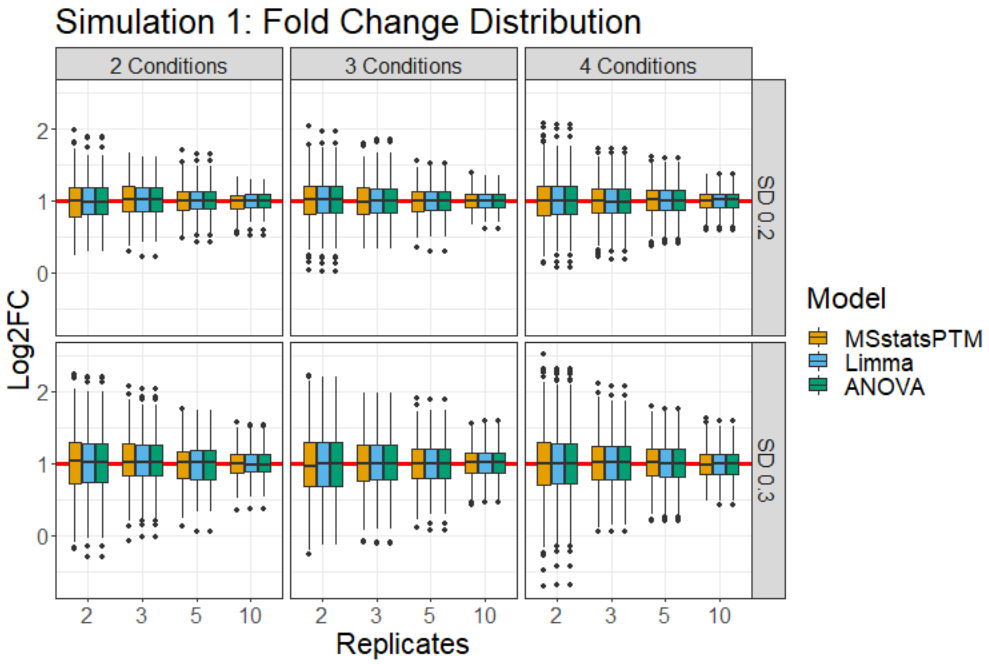


Figure S2.a: [TODO: not sure about the a and b labeling style. Could yo label the subfigures as (a) and (b), and combine the captions into a single caption, as in the main text? Use the same code names for the datasets as in the main text] In simulation 1 all considered methods correctly estimated the fold change between conditions, with a median fold change estimation of 1. The distributions around the median were also consistent across all methods.

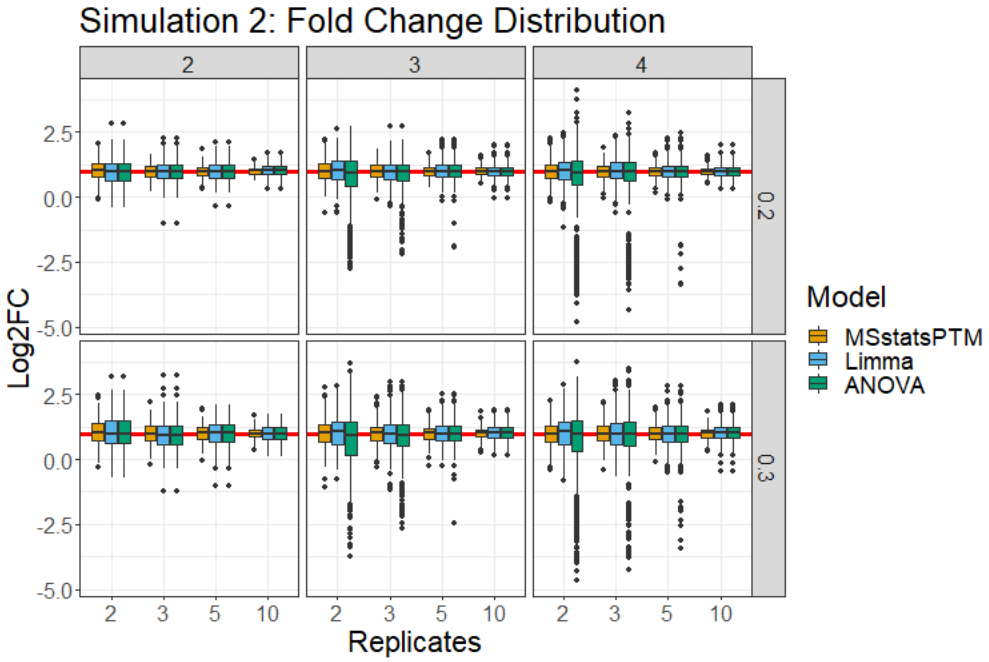


Figure S2.b: In simulation 2 all methods correctly estimated the fold change with a median log change of 1. The proposed method in this simulation had a visibly tighter distribution around the median. Both Limma and ANOVA showed a wider range around the fold change.

Figure S2: Fold change distribution comparison between Simulations 1 and 2.

### 2.1.3 Dataset 3 : SpikeIn benchmark - Ubiquitination - Label-free

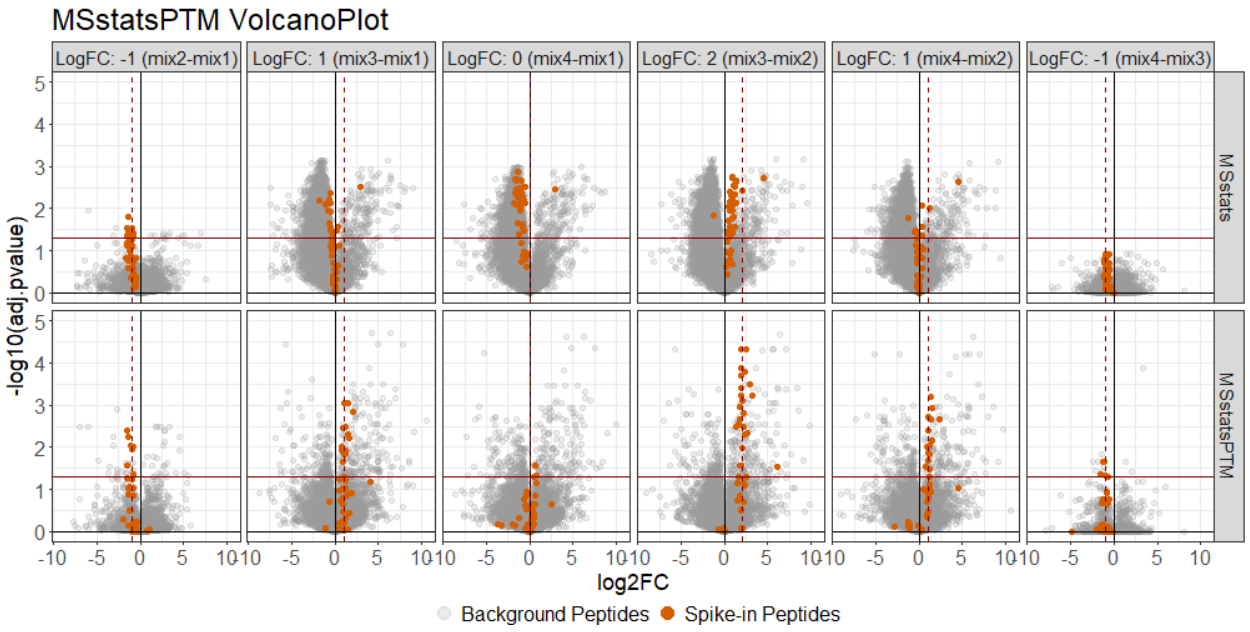


Figure S3: The modeling results of the proposed method both before and after adjustment. The spike-in peptides are colored red and the background peptides are colored grey. All grey peptides are expected to be insignificant. Using the proposed method to model the benchmark experiment, the spike-in peptides (colored red) did not follow the expected log fold change before adjustment. After adjusting for changes in overall protein abundance the spike-in peptides were more in line with expectation. Additionally the background grey colored peptides showed many false positives before adjustment. After adjustment these false positives were decreased considerably.

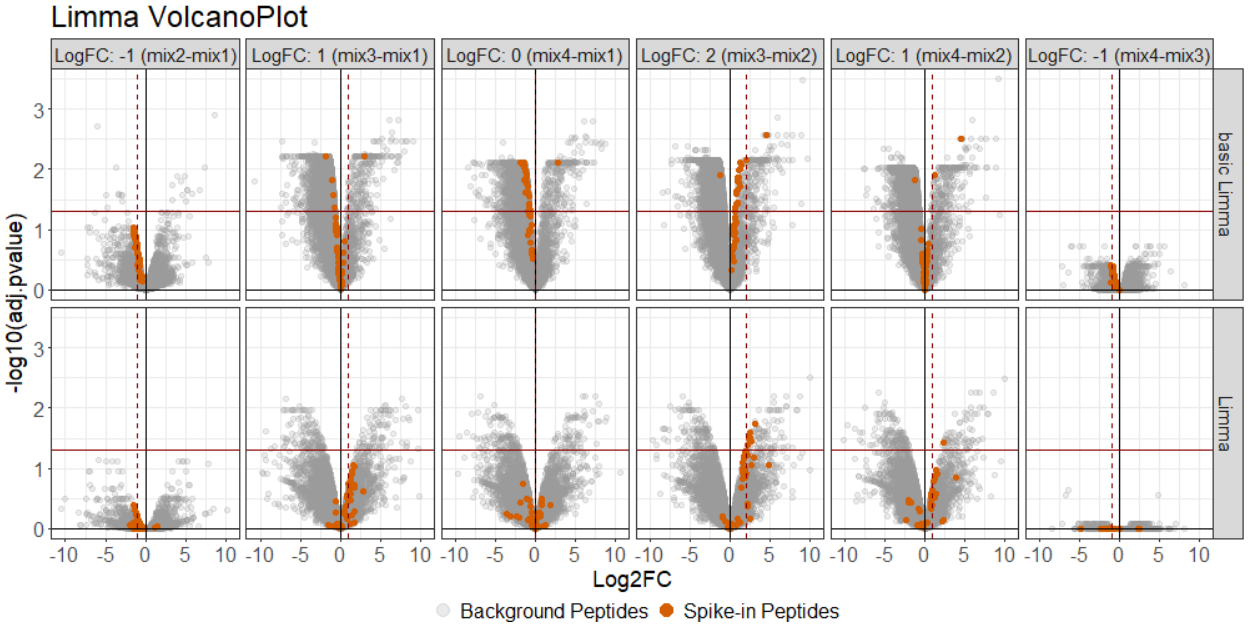


Figure S4: When modeling the experiment with the Limma method, the spike-in peptides again follow the expected log fold change better after adjusting for changes in protein level. However, while the fold change was more accurate, the majority of spike-in peptides did not have a significant adjusted pvalue. In this case, the known differential peptides were missed by the model. In terms of false positives, the results were very similar to the proposed method, with many false positives before adjustment and fewer after.

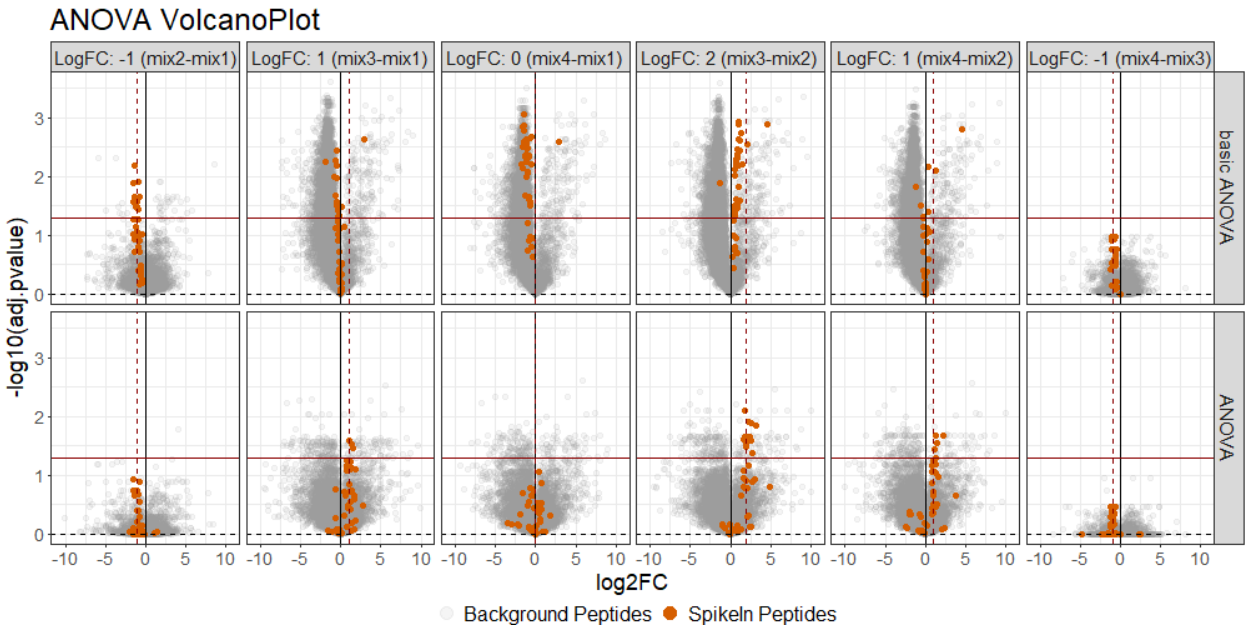


Figure S5: Using ANOVA the fold change of the spike-in peptides was much closer to expectation after adjusting for global protein abundance. The log FC estimation was the same as Limma, however the pvalues were different. We actually see a few more True Positives using Anova compared to Limma.

## 2.2 Biological investigations

### 2.2.1 Dataset 4 : Human - Ubiquitination - 1mix-TMT [3]

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.

$$Y_{mij} = \mu_i + \epsilon_{mij}$$

$$\sum_{i=1}^{I} \mu_i = 0,\ \epsilon_{mij} \ \sim N(0, \sigma^2)$$

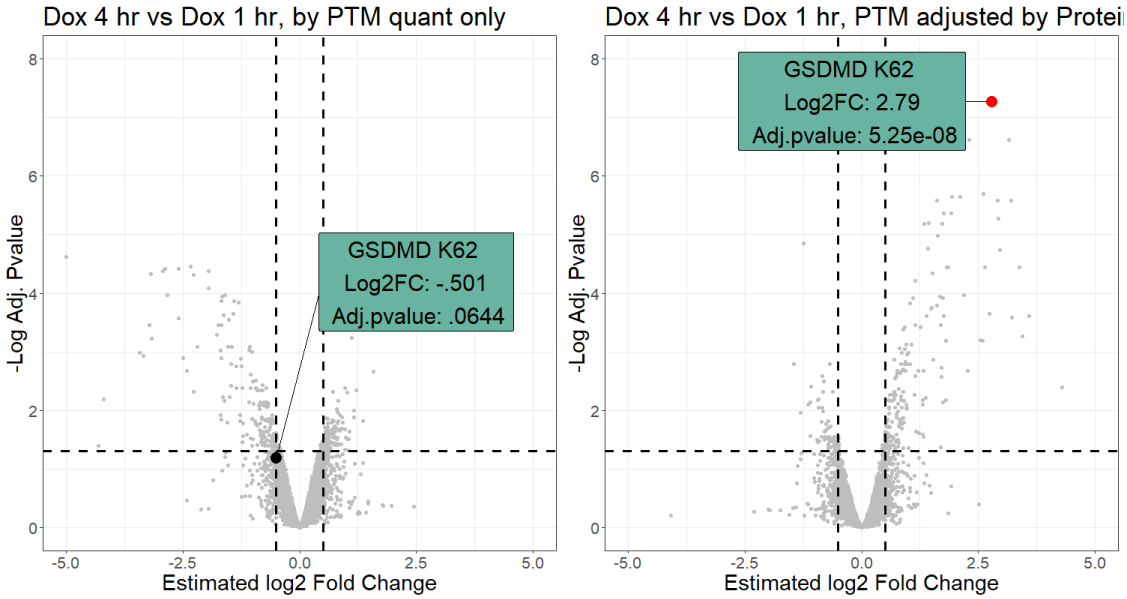


Figure S6: Volcano plots of Dox4hr vs Dox1hr both before and after protein adjustment. The $GSDMD\_HUMAN|P57764\_K62$ modification is highlighted. Before adjustment the modification had a small fold change and insignificant adjusted pvalue. After adjustment the fold change was much larger and the adjusted pvalue was significant. In this case the proposed method allowed us to identify a differential modified peptide that could have otherwise been missed.

### 2.2.2 Dataset 5 : Mouse - Phosphorylation - 2mix-TMT [4]

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, and phosphorylation. The model formula can be seen below.

$$Y_{mij} = \mu_i + Mixture_m + \epsilon_{mij}$$

$$Mixture_m \sim N(0, \sigma_M^2), \sum_{i=1}^{I} \mu_i = 0, \ \epsilon_{mij} \sim N(0, \sigma^2)$$

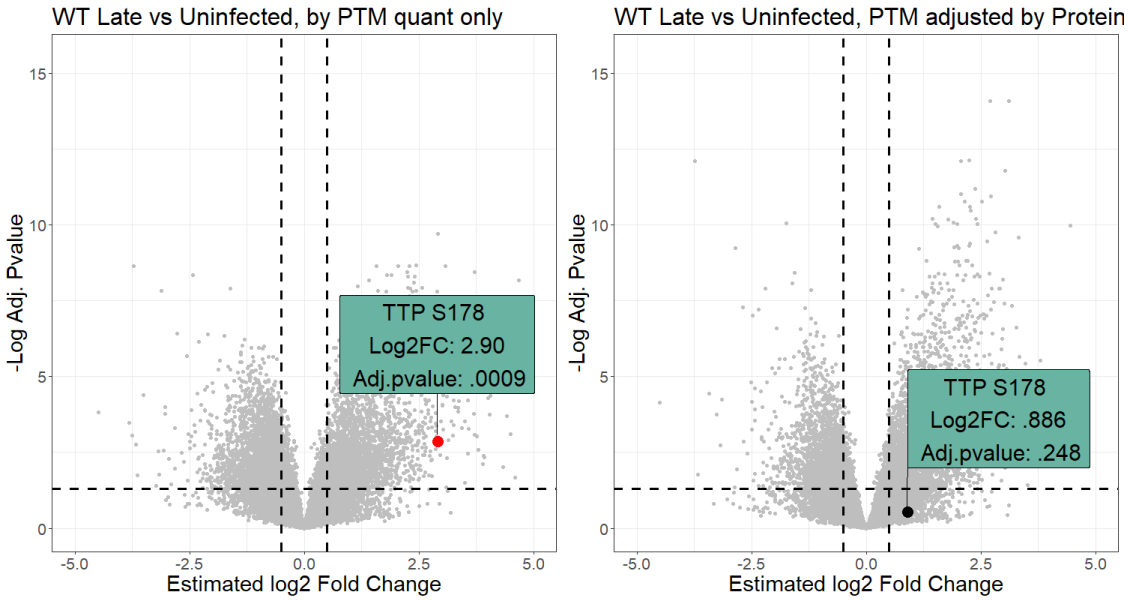


Figure S7: Volcano plots of WT_Late vs WT_Uninfect both before and after protein adjustment. The $TTP\_MOUSE|P22893\_S178$ modification is highlighted. Before adjustment the modification had a large fold change and significant adjusted pvalue. After adjustment the fold change was much smaller and the adjusted pvalue was insignificant.

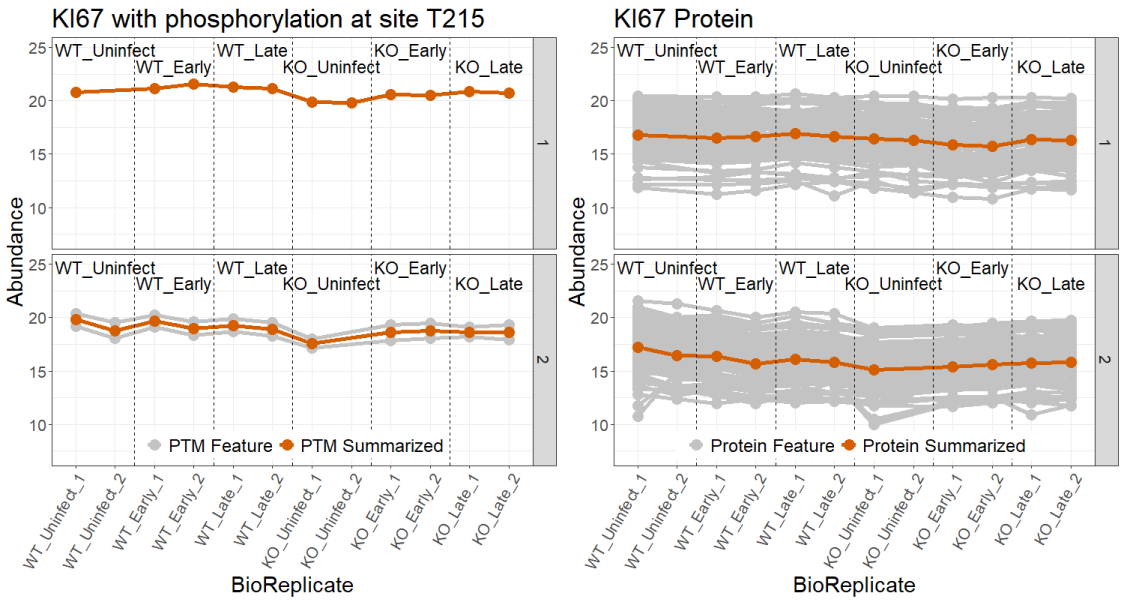


Figure S8.a: The global profiling of protein $KI67\_MOUSE|E9PVX6$ with the modification at site $T215$ was compared. In this case the modification and global protein appeared to show small or no difference between conditions, however after adjusting for change in global protein abundance, the modification was statistically significant. Additionally, this profile plot showed the large difference in available features between modified peptides and global proteins.

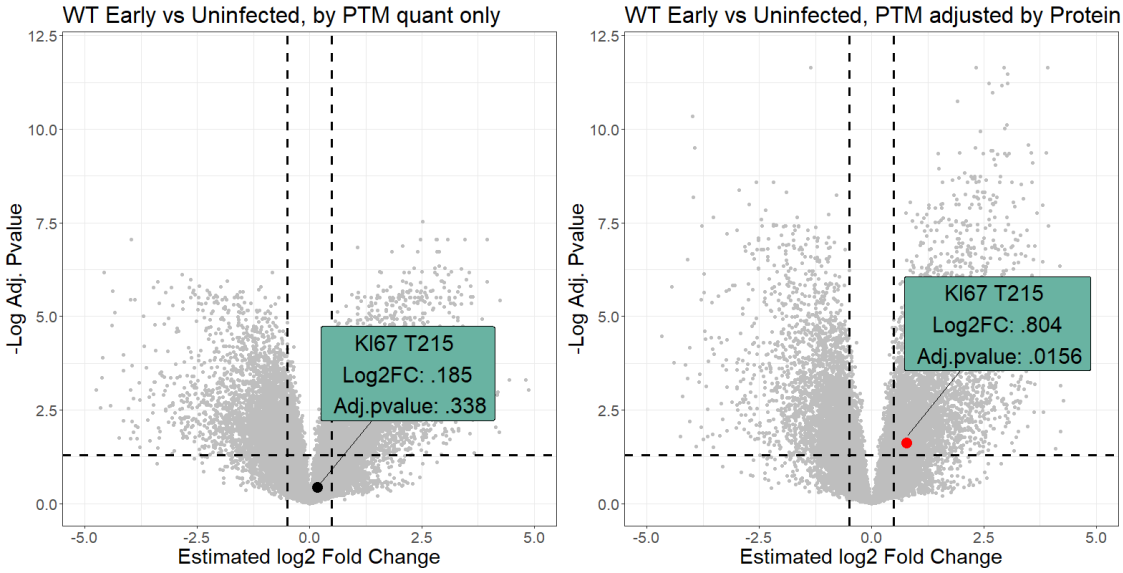


Figure S8.b: The volcano plot of the WT_Uninfect and WT_Early comparison showed the specifics of the adjustment. The modification looked to be flat, with a log fold change of .185, while the global profiling showed a small negative fold change of $-.619$. While both exhibit small changes, when combined we saw a log fold change of .804 and adjusted p-value of .0156.

Figure S8: Summary plots for modification of protein KI67 at site T215.

### 2.2.3 Dataset 6 : Human - Ubiquitination - Label-free no global profiling run [5]

An example profile plot for this experiment can be seen in Figure S9.

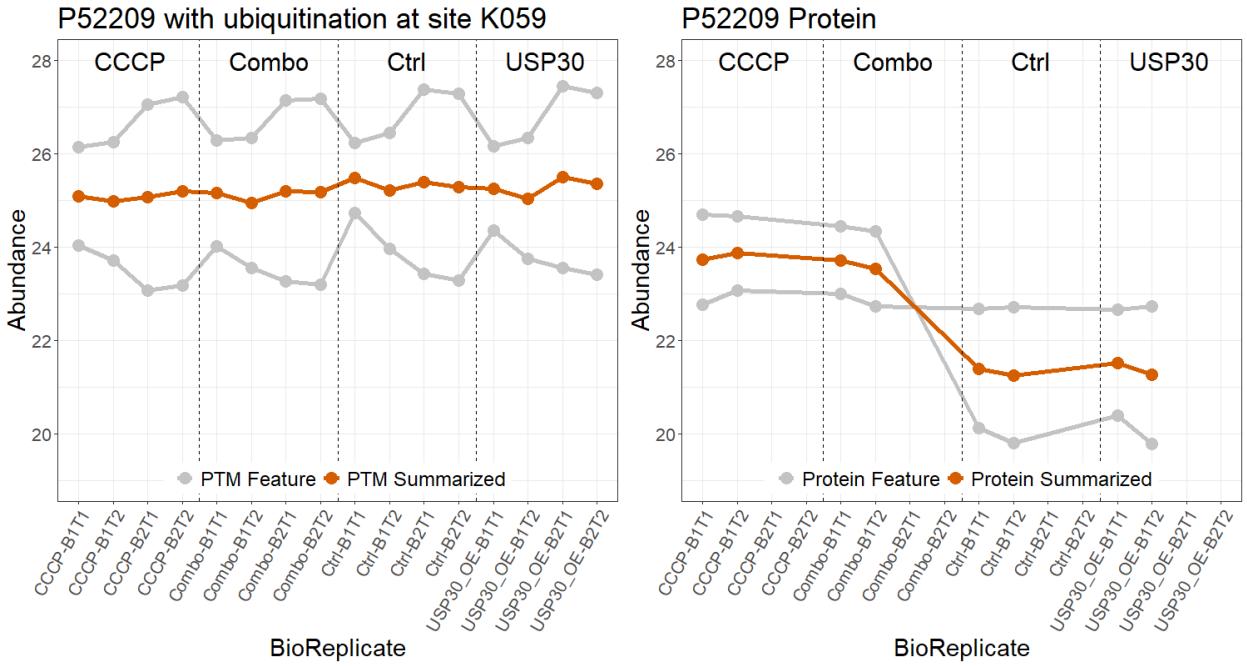


Figure S9: The global profiling of protein $P52209$ with the modification of the protein at site $K059$ was compared. The modification appeared generally unchanged between all conditions, whereas the global profiling run showed the CCCP and Combo conditions had a higher relative abundance compared to the Control and USP30_OE. This indicated that the modification actually had an effect when comparing CCCP and Combo to Control and USP30_OE. However it was not entirely clear as one unmodified peptide feature appeared to be changed, while the other did not. This uncertainty was 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.

# 3 Sample size calculation and power analysis

### Noisy PTM measurements benefited from additional biological replicates

Here we analyzed the sample size needed to achieve a desired statistical power. We compared the statistical power in experiments with differing numbers of replicates, variance, and fold change for both the modified and unmodified runs. In terms of the number of replicates, we tested scenarios with equal replicates in both the modified and unmodified runs, as well as scenarios where the replicates differed between runs. We used the biological experiments to determine what variance values to test. In datasets 4 and 5 the variance of the PTM was higher than the global protein. In dataset 6 the variance of the PTM and Protein were generally the same. We mimicked these scenarios and analyzed the power of experiments when the PTM variance was higher than the protein and when they were equal. When the PTM and protein were the same we chose a variance of .15, whereas when the PTM was higher than the protein we chose a PTM variance of .2 and a protein variance of .1. The results of the power and sample size analysis can be seen in Figure S10.

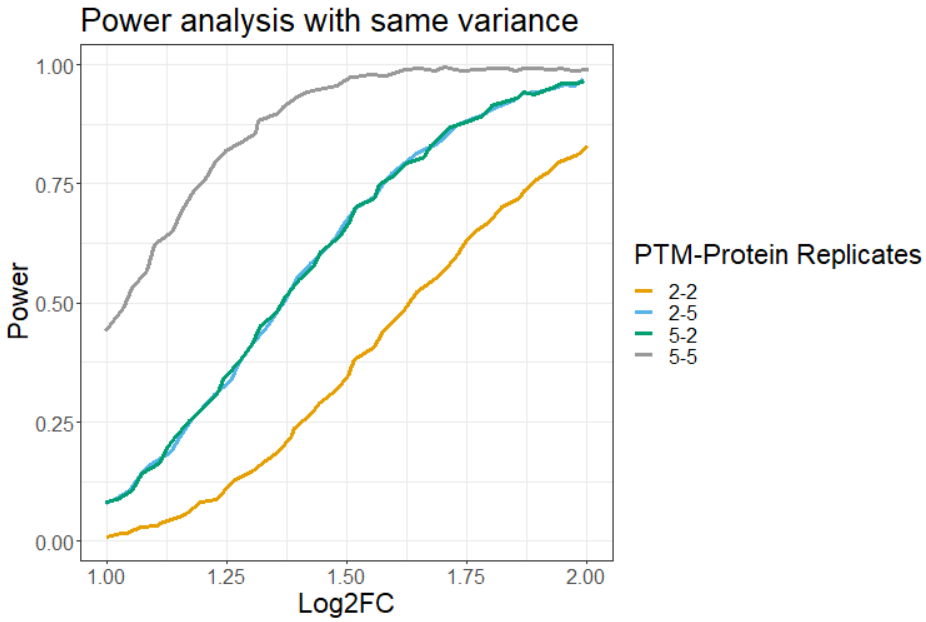


Figure S10.a: The power of an experiment targeting PTMs with the same variance, .15, for the modified and unmodified peptides. Predictably when the replicates were high for both modified and unmodified peptides the power was much higher. Conversely at low replicates for each the power was much lower. With equal variance, it did not matter if the PTM replicates or protein replicates were higher.

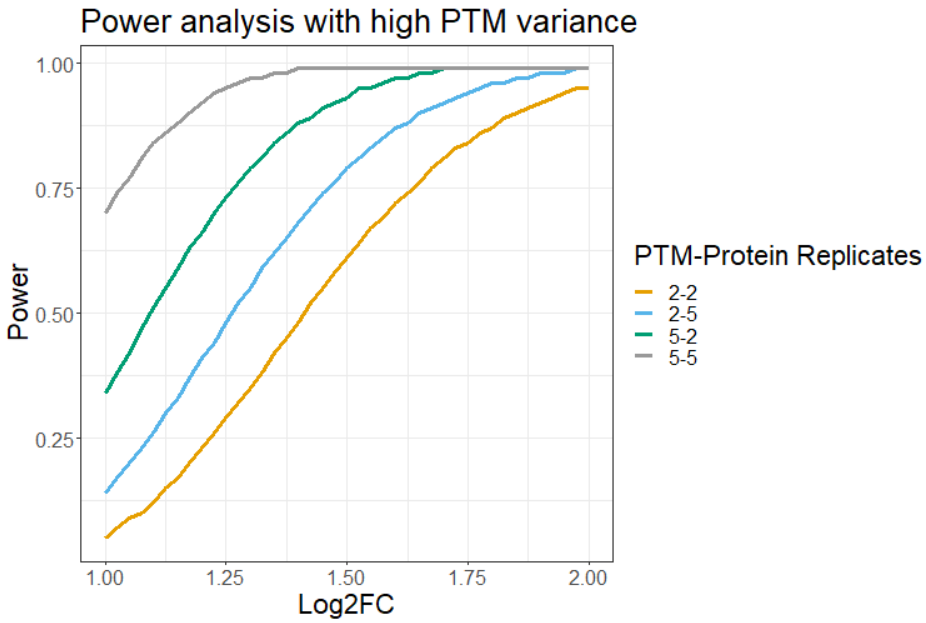


Figure S10.b: In this chart the variance for the PTM was higher than the unmodified protein. The PTM variance was .2, while the unmodified protein variance was .1. With equal replicates the results were the same as above. When the replicates were not equal, more replicates allocated to the PTM runs lead to higher power.

Figure S10: Power analysis of experiments with differing variances.

# References

[1] Michael H. Kutner et al. *Applied linear statistical models*. 5th ed. McGraw-Hill/Irwin, 2004.

[2] G. K. Smyth. "Linear models and empirical bayes methods for assessing differential expression in microarray experiments". In: *Statistical Applications in Genetics and Molecular Biology* 3.1 (2003), Article 3.

[3] G. Luchetti et al. "Shigella ubiquitin ligase IpaH7.8 targets gasdermin D for degradation to prevent pyroptosis and enable infection". In: *Cell Host & Microbe* 29.10 (2021), pp. 1521–1530.

[4] T. Maculins et al. "Multiplexed proteomics of autophagy-deficient murine macrophages reveals enhanced antimicrobial immunity via the oxidative stress response". In: *eLife* 10 (2021), e62320.

[5] C. Cunningham et al. "USP30 and Parkin homeostatically regulate atypical ubiquitin chains on mitochondria". In: *Nature Cell Biology* 17.2 (2015), pp. 160–169.