At the centre of the reviewers' cogent critique of our manuscript was the questioned statistical validity of our approach. Succinctly, the issue at question is whether or not the R package `edgeR` is an appropriate tool for analysis of protein mass spectrometry data.

Statistical inference in `edgeR` is built on a negative binomial (NB) generalized linear model (GLM) framework. Therefore, the data are assumed to be adequately described by a NB distribution parameterized by a dispersion parameter, ϕ. [1]

Our previous approach used a customized workflow [2] to preprocess and normalize the data prior to performing statistical testing using `edgeR's` flexible GLM framework.

Our prior decision to use `edgeR` was motivated by numerous conceptual and practical considerations. However, we failed to thoughougly consider its overall adequacy. Here we reconsider its appropriatness for our TMT proteomics dataset.

We considered `MSstatsTMT` as an alternative analytical tool. `MSstatsTMT` utilizes a linear mixed-model framework. The strength of linear mixed models is thier ability to account for complex sources of variation in an experimental design. We evaluated the overall adequacy of the linear models fit by `edgeR` and `MSstatsTMT` by plotting the residual deviance of all proteins against their theoretical, normal quantiles in a quantile-quantile plot. **Figure 1** illustrates the overall lack of fit for the model fit by `edgeR`. In contrast, the data seem to be well fit by the linear mixed-models used by `MSstatsTMT`, providing justification for this approach.

MSstatsTMT fits following linear mixed effects model to the protein-level data:

$$Y_{mtcb} = \mu + Mixture_m + TechRep(Mixture)_{t(m)} + Condition_c + Subject_{mtcb} + \epsilon_{mtcb} \quad (1)$$

The response, protein Abundance, is as a function of the mixed effects of Mixture Where $Mixture_m \sim N(0, \sigma^2_M)$, $TechRep(Mixture)_{t(m)} \sim N(0, \sigma^2_T)$, and $Subject_{mcb} \sim N(0, \sigma^2_S)$ are the mixed-effects of Mixture, TechRep(Mixture), and Subject. ϵ is a mixed-effect $\epsilon_{mtcb} \sim N(0, \sigma^2)$ and quantifies any remaining variation not explained by Mixture, TechRepMixture or Subject.
A TMT proteomics experiment consists of one or more contatenatino of TMT-label samples or Mixtures. Each TMT channel is dedicated to analysis of an indivual Condition from several Subject's. A Mixture may be fractionated into multiple MS runs in order to increase analytical depth or analyzed in technical replicate. An experiment consists of M mixtures, T

[1] The dispersion parameter can take several forms. `edgeR` supports three dispersion models: 'common', 'trended', and 'tagwise'. However, when using `edgeR's` robust quasi-likelihood test methods, only global (i.e. 'common' or 'trended') dispersion metrics are appropriate (see `?edgeR::glmQLFit`).

[2] The most important step in our normalization approach is IRS normalization. IRS normalization scales protein measurements using an internal reference standard to normalize protein measurments between TMT MS runs. This is essential to account for the stochasticisity of peptide quantification in MS experiments. Phillip Wilmarth's GitHub offers an excellent exploration of IRS normalization.

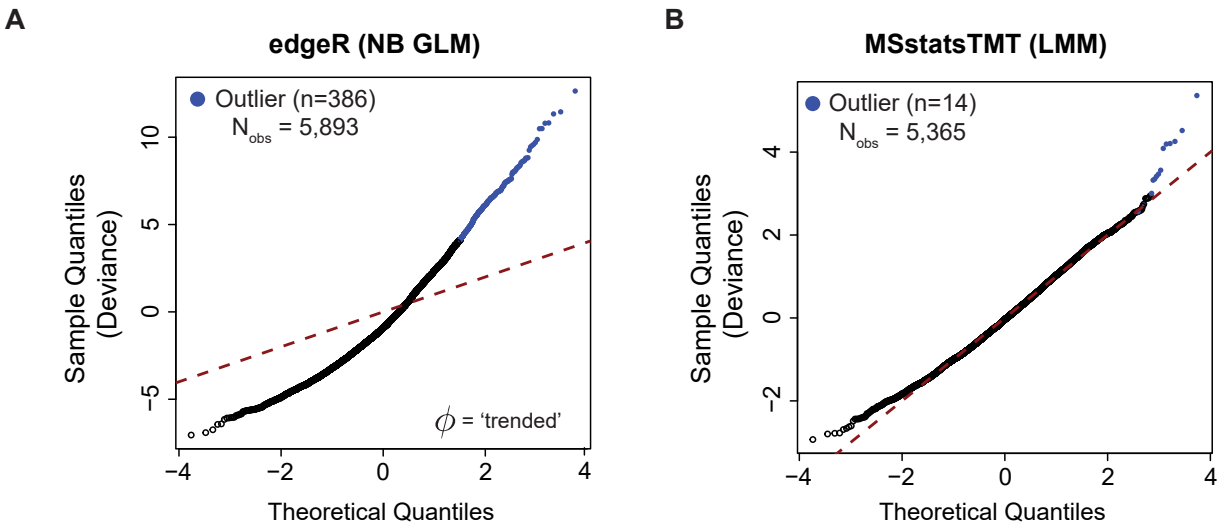

Figure 1: **Goodness-of-fit of `edgeR` (A), and `MSstats` (B) statistical approaches.** The overall adequacy of the linear models fit to the data were assessed by plotting the residual deviance for all proteins as a quantile-quantile plot (McCarthy *et al.*, (2012)). **(A)** The normalized protein data were fit with a NB GLM of the form: `~ Mixture + Condition`. Where `Mixure` is a blocking factor that accounts for sources of variablity between experiments. Protein-wise deviance statistics were transformed to normality and plotted aganis theoretical normal quantiles using `edgeR::gof`. **(B)** The normalized protein data were fit with a linear mixed-effects model (LMM) of the form: `Abundance ~0 + Condition + (1|Mixture)`. Where `Mixture` indicates the random effect of `Mixture`. The residual deviance and degrees of freedom were extracted from the fitted models, z-score normalized, and plotted as in (A). Proteins with significantly poor fit are indicated as outliers in blue (Holm-adjusted P-value < 0.05).

technical replicates of mixture, C conditions, and S biological replicates. TechRepMixture, Subject.