PART I: LEARNING FROM VARIANCE

Error-Reduced Roll-up Improves Protein Quantification Through Measurement Quality

Protein quantification in mass spectrometry depends on how peptide measurements are combined into protein ratios. Conventional roll-up algorithms, such as MaxLFQ, assume all peptides contribute equally—a flawed assumption in DIA proteomics where measurement quality varies widely. Here we introduce Error-Reduced (ER) Roll-up, Golgi's weighted aggregation strategy that learns from variance, assigning data-driven weights that reflect peptide reliability. Benchmarking and simulation show up to a 40% reduction in variance and improved ratio fidelity, bringing proteomic quantification closer to ground truth and enabling more reproducible biological insight.

INTRODUCTION

Mass spectrometry proteomics experiments estimate protein abundance changes, from measurements made on peptides. This fundamental disconnect makes algorithmic decisions critical: how peptide measurements are translated into protein ratios determines the accuracy of all downstream analysis.

Quantifying proteins from peptide-level data requires a process known as peptide-to-protein roll-up, in which multiple peptide measurements are combined into a single protein abundance measurement. However, this roll-up is complicated by both missing data and variations in measurement quality. Advanced proteomic algorithms, such as MaxLFQ, address the missing data problem and outperform simpler methods for aggregation [1]. However, the MaxLFQ algorithm assumes that all peptides contribute equally and with comparable reliability to the final protein abundance [2].

This is a dangerous assumption in DIA proteomics. With DIA, a variety of parameters impact the quality of peptide quantification relative to a TMT-based approach, including co-eluting precursors, variable interference from complex spectra, and imperfect chromatographic alignment, among others. As a result, while some peptides provide highly precise and accurate quantification, others carry substantial systematic error. Treating all peptides as equal can therefore bias protein-level estimates, inflate variance, and obscure true biological differences.

Using Weights to Account for Variation in Measurement Quality

Not all peptide measurements are created equal. In DIA proteomics, measurement quality varies dramatically due to co-eluting precursors, spectral interference, and chromatographic alignment issues. Some peptides provide highly precise quantification; others carry substantial systematic error. The authors of the MaxLFQ paper acknowledged this limitation [2]:

> *"In principle, weighting factors can be included in the sum for H(N) in order to penalize low-intensity ions. Here we refrained from this in order to keep the parameterization of the model simple"*

A Simple Example

Consider a protein with two peptides measured across two groups (true FC = 2). MaxLFQ will report the median ratio of all peptides. In this case, all measurements are high quality except for one problematic observation which is 10x more variable than the good measurements:

	Group A	Group B	Ratio
Peptide A	100	200	2
Peptide B	100	600 (the bad one)	6

Unweighted median ratio:
Median ratio = 4.0 [100% error]

With inverse-variance weighting:
Weighted ratio = 2.36 [18% error]

The motivating principle is straightforward: weight observations by the inverse of their variance ($w = 1/\sigma^2$). In the theory of linear models this approach is proven to minimize estimation error among all unbiased estimators when measurement quality varies [3]. In our example, the three high-quality measurements (weight = 10/11) appropriately dominate the single low-quality measurement (weight = 1/11), recovering an estimate close to the true fold change.

The Simulation

To demonstrate the importance of incorporating weights into the aggregation process, we simulated 1000 proteins from two experimental groups with a true fold change of 2 ($\log_2$FC = 1). Each protein consisted of three peptides, and the peptide intensities were drawn randomly from log-normal distributions with randomly assigned variances (ranging from 0.5 to 1). Each observation was then assigned a weight equal to 1/variance. This is a scenario that could cause challenges with a typical unweighted median approach.

Since the peptides from group B were drawn from distributions with centers twice as large as those of group A, both aggregation methods produce protein fold changes centered near the expected value of $\log_2$FC = 1 (**Fig. 1**). The difference is in the spread of the distributions. When accounting for the weights, the distribution of estimated fold changes is much tighter than when ignoring the differences in measurement quality. Indeed, the standard deviation drops by 50% (from 1.42 to 0.72).

This simulation shows that when we know the variability of the data generating process, we can greatly improve the precision of the aggregation method by accounting for the differing quality of measurements.

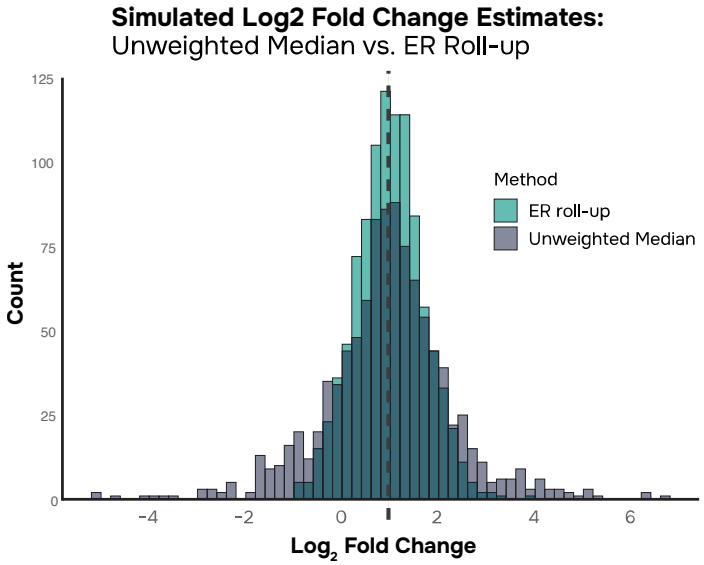

Figure 1. We simulated 1000 proteins, each with three peptides, from two experimental groups with a true FC of 2. First, we calculated the protein intensities using the median peptide ratios (navy). Second, we assigned a weight of w=12 to each observation, where 2 is the variance of the log-normal distribution from which the peptide intensity was randomly drawn. We then aggregated the centered log ratios of the peptide intensities, using ER roll-up (teal). Using the weighted model drops the standard deviation from 1.45 to 0.72, a decrease of 50%.

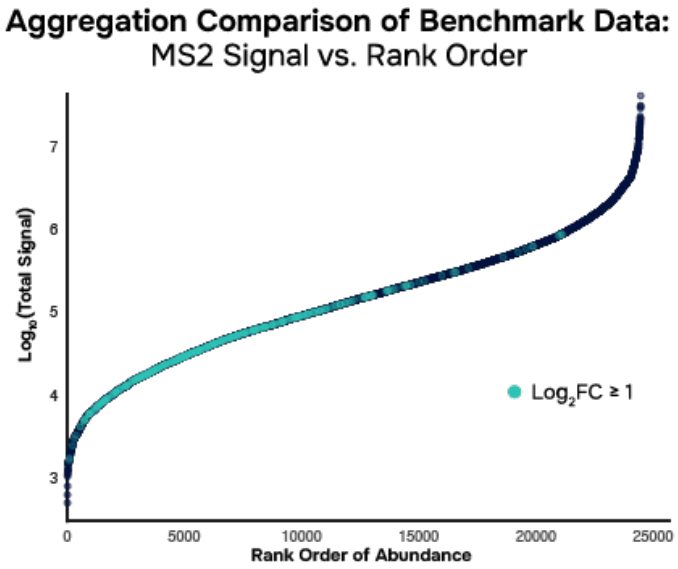

Figure 2. Crude abundance plot showing $\log_{10}$ MS2 total signal versus rank order of abundance for all proteins. Points with a method difference of ≥1 $\log_2$FC are highlighted in teal (n = 488, 2% of total proteins) and all other proteins are in navy (alpha = 0.2). Teal points are plotted last so they overlay navy points and remain visible.

A DIA Benchmarking Experiment

We know we can simulate the variability of data generation - but it is essential to be able to estimate this in real samples. Golgi's approach does exactly that: we leverage heteroskedasticity within the data to predict measurement quality and convert predictions into weights through ER roll-up. For TMT we exploit the relationship between peptide intensity and variance to assign weights equal to 1/ variance. For DIA, the relationship between intensity and variance is more complex due to interference from overlapping fragment ions. We therefore instead combine intensity metrics with additional measures of spectral, identification, and ion-mobility quality metrics to produce more robust weights.

This is Golgi's secret sauce. We specialize in finding patterns in proteomics metadata that are indicative of measurement quality, and using that information to get more value from every experiment.

Benchmarking Data

To evaluate this approach under real experimental conditions, we compared MaxLFQ to ER roll-up using DIA benchmarking data [4]. The dataset consists of a matrix-matched calibration series generated by mixing unlabeled HeLa with SILAC HeLa at multiple dilutions. For this analysis, we used only the 50% and 100% dilution data points from the Astral benchmarking study (expected true FC = 2).

Both aggregation methods produced mean fold changes close to the expected value and yielded similar results for the vast majority of proteins. However, 3,002 proteins (6% of the total) differed in their $\log_2$ fold change estimates by at least 0.5 (**Fig. 2**). While a 6% difference might not seem substantial, the impacted proteins are not randomly distributed—they concentrate in an abundance range where the probability of ratio distortion from poor measurements is substantially higher.

Crude rank-abundance analysis indicates that the majority of the 3,002 diverging proteins occupy a mid-low range of MS signal rather than the extreme low tail. Many of these proteins are known or candidate drug targets, including regulators of the cell cycle (CCNA2, CCNB1, CDK16), components of the autophagy

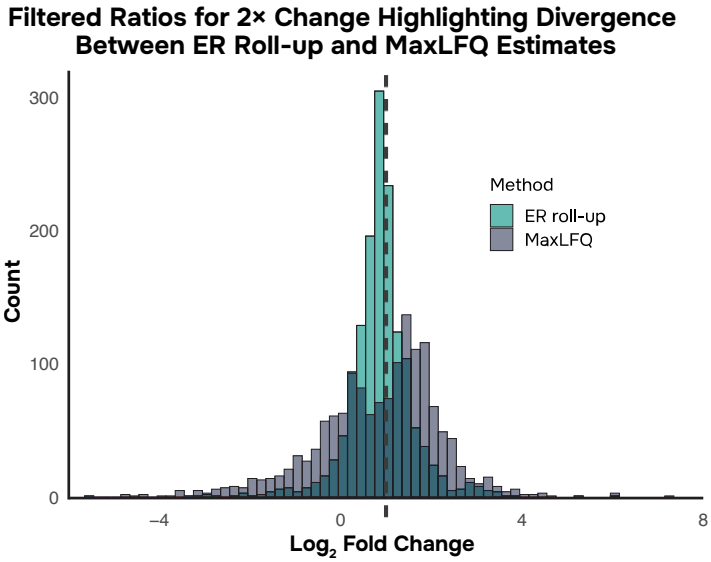

Figure 3. Distribution of $\log_2$ fold-change estimates for 3,002 proteins whose method estimates diverged by ≥0.5 $\log_2$FC in one or more of the three replicates for the 100% and 50% dilution points. ER roll-up is shown in teal and MaxLFQ is shown in navy. The dashed vertical line indicates the expected 2X change ($\log_2$FC = 1). Accounting for peptide-level weights (weighted model) reduces the standard deviation from 1.37 to 0.77 (41% decrease), producing a substantially tighter distribution.

pathway (BECN1, ATG7, ATG2B), and clinically validated therapeutic targets (JAK1, PGH2/COX-2, and CRBN), highlighting the risk of poor reproducibility for proteins in this critical subset.

Within the subset of proteins where measurement quality makes a difference, ER roll-up provides a substantial boost to ratio precision consistent with the simulation results (**Fig. 3**). Accounting for peptide-level weights (Golgi method) produces a much tighter distribution, reflected by a 41% decrease in standard deviation (from 1.37 to 0.77), substantially reducing the risk of inconsistent signatures in this biologically relevant subset.

Variations in measurement quality are fundamental to mass spectrometry proteomics data. We have shown here how understanding and accounting for these variations can improve ratio fidelity. However, this is only one of many downstream tasks where measurement quality proves impactful. Subsequent DIA white papers will expand on these benefits:

Part II: Sensitivity and Specificity

Part III: Eliminating Technical Replicates

Bibliography

1. Yu, F., Haynes, S. E., Teo, G. C., Avtonomov, D. A., Polasky, D. A. & Nesvizhskii, A. I. Analysis of DIA proteomics data using MSFragger-DIA and FragPipe computational platform. Nat Commun 14, 4154 (2023).

2. Cox, Jürgen, et al. "Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ." Molecular & cellular proteomics 13.9 (2014): 2513-2526.

3. Christensen, Ronald. Plane answers to complex questions. Vol. 35. No. 1. New York: Springer, 2002.

4. Heil, Lilian R., et al. "Evaluating the performance of the astral mass analyzer for quantitative proteomics using data-independent acquisition." Journal of proteome research 22.10 (2023): 3290-3300.