

## Benchmarking Quantitative Performance in Label-Free Proteomics

James A. Dowell,\* Logan J. Wright, Eric A. Armstrong, and John M. Denu\*

Cite This: *ACS Omega* 2021, 6, 2494–2504

Read Online

ACCESS |



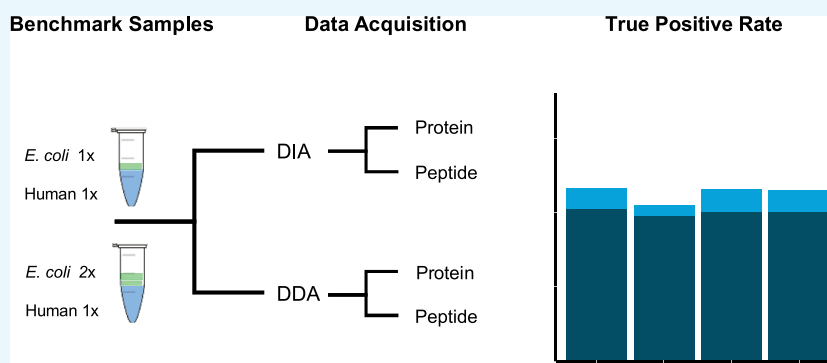
Metrics &amp; More



Article Recommendations



Supporting Information



**ABSTRACT:** Previous benchmarking studies have demonstrated the importance of instrument acquisition methodology and statistical analysis on quantitative performance in label-free proteomics. However, the effects of these parameters in combination with replicate number and false discovery rate (FDR) corrections are not known. Using a benchmarking standard, we systematically evaluated the combined impact of acquisition methodology, replicate number, statistical approach, and FDR corrections. These analyses reveal a complex interaction between these parameters that greatly impacts the quantitative fidelity of protein- and peptide-level quantification. At a high replicate number ( $n = 8$ ), both data-dependent acquisition (DDA) and data-independent acquisition (DIA) methodologies yield accurate protein quantification across statistical approaches. However, at a low replicate number ( $n = 4$ ), only DIA in combination with linear models for microarrays (LIMMA) and reproducibility-optimized test statistic (ROTS) produced a high level of quantitative fidelity. Quantitative accuracy at low replicates is also greatly impacted by FDR corrections, with Benjamini–Hochberg and Storey corrections yielding variable true positive rates for DDA workflows. For peptide quantification, replicate number and acquisition methodology are even more critical. A higher number of replicates in combination with DIA and LIMMA produce high quantitative fidelity, while DDA performs poorly regardless of replicate number or statistical approach. These results underscore the importance of pairing instrument acquisition methodology with the appropriate replicate number and statistical approach for optimal quantification performance.

## INTRODUCTION

Proteomics is a powerful tool to profile global changes in protein expression and protein modification states.<sup>1</sup> For these “discovery” experiments to be useful in guiding follow-up “validation” experiments, they must have a high degree of quantitative accuracy. As with other “omics technologies”, many factors can impact the accuracy of quantitative proteomics, including the type of quantitative workflow, instrument parameters, replicate number, and statistical approach.

In general, there are three different types of quantitative proteomics workflows—label-free, metabolic labeling, and chemical tagging. Each of these techniques has certain advantages and disadvantages. Label-free quantification (LFQ) is the simplest and most straightforward to set up. However, LFQ generally exhibits higher amounts of variance than metabolic labeling or chemical tagging.<sup>2,3</sup> In contrast, metabolic labeling via stable isotope labeling with amino acids

in cell culture (SILAC) yields highly accurate quantification, but is limited to cell culture systems and produces lower proteome coverage than LFQ workflows.<sup>4,5</sup> Chemical tagging via isobaric tandem mass tags (TMTs) also yields highly accurate quantification; however, TMT workflows display quantitative bias unless triple-stage isolation (MS3) or high-field asymmetric-waveform ion mobility (FAIMS) is employed, which requires a specific type of mass spectrometer and an electrospray source.<sup>6–10</sup>

Data-dependent acquisition (DDA) is a highly effective method for identifying large numbers of peptides and

Received: August 20, 2020

Accepted: January 11, 2021

Published: January 20, 2021



proteins.<sup>11–14</sup> However, DDA presents some unique quantitative challenges for LFQ. Specifically, DDA precursor ion selection is based on relative peptide intensities at a given time in the LC–MS run and thus exhibits run-to-run inconsistencies in peptide identification and quantification. This stochastic peptide sampling results in a large number of missing peptide intensities across the samples. These missing values negatively affect downstream statistical analyses, especially for low-abundance peptides/proteins.<sup>2,15,16</sup> While newer, faster instruments provide immense sampling depth, they still suffer from undersampling when near the limit of detection.<sup>2</sup>

To alleviate the issue of missing values in label-free DDA workflows, data imputation has been employed.<sup>17–20</sup> A simple application of data imputation uses the normal distribution to replace the missing values.<sup>21</sup> This methodology has been shown to increase the quantitative fidelity for low-abundance proteins but not for high-abundance proteins.<sup>22</sup> More recently, more sophisticated techniques combining multiple types of data, for example, peptide intensities and spectral counts, have been used to produce more robust quantitative performance from DDA data.<sup>23</sup>

In addition to post hoc data imputation, instrument-centric methods employing data-independent acquisition (DIA) have been developed to reduce missing values in label-free workflows.<sup>24–26</sup> In contrast to DDA workflows that produce tandem MS1/MS2 spectra via isolation and fragmentation of specific peptide precursors, DIA methods cycle through wide mass isolation windows in which all peptides within a window are fragmented. These peptide spectra are then matched to a previously generated peptide library across the LC–MS/MS run. This unbiased isolation and fragmentation produces very few missing values.<sup>15</sup> However, DIA produces complex, multipropeptide fragmentation spectra that are generally not suitable for database searching. To circumvent this issue, DIA workflows perform peptide identification and quantification using an external peptide library.<sup>24,25,27</sup> These libraries are typically generated via parallel DDA experiments and thus require additional instrument time.

A popular alternative to LFQ is the use of isobaric labeling with TMTs.<sup>8,28</sup> In this workflow, tryptic peptides are N-terminal-labeled via amine-reactive isobaric mass tags that contain a sample-specific reporter fragment ion. After labeling, the samples are combined into a single sample and then assayed via standard DDA acquisition. Instead of the quantification being performed on peptide intensities (MS1), the signal intensity from each unique “reporter” fragment is compared against the other “reporter” fragments to yield relative peptide (and protein) quantification. This was traditionally performed on the peptide fragmentation spectra (MS2); however, it was discovered that peptide co-isolation leads to interference in the MS2 channel. To alleviate these issues, triple-stage isolation (MS3) is used, in which the most intense fragment ions of the precursor peptides are isolated and further fragmented to produce peptide-specific, interference-free “reporter” ions.<sup>6–10</sup> This results in almost no missing values because any peptide that undergoes fragmentation yields a complete quantitative profile across all samples.

In addition to instrument acquisition methods, quantification is affected by post-acquisition data handling, including data normalization and statistical analyses.<sup>29</sup> For protein-level analyses, typically the peptide intensities for a specific protein are combined (by sum, mean, or median) to obtain a protein-level intensity value. Statistical analyses are then performed on

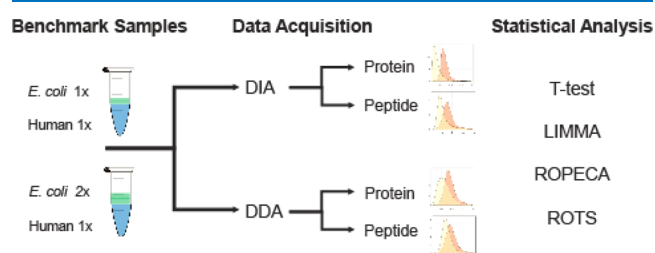
these “rolled-up” protein-level values.<sup>30–32</sup> More recently, peptide-centric statistical approaches have been developed that perform statistical analysis at the peptide level before combining results into protein values. These approaches were shown to have superior quantitative accuracy in comparison with protein-level statistical approaches.<sup>22,23,33–36</sup>

In addition to protein-level experiments, quantitative mass spectrometry is employed to analyze the alterations in peptide modification states, for example, phosphoproteomics and acetylomics.<sup>37–40</sup> The modified peptides are typically enriched and then analyzed via label-free (DDA or DIA) or isobaric labeling (TMT) workflows.<sup>41</sup> In contrast to protein intensities that are a combination of multiple peptide signals, peptide-level analyses rely on a single intensity of an individual peptide analyte. This presents more demands on the quantitative accuracy due to the use of a single intensity for downstream analysis.

While a number of benchmarking studies have evaluated the effects of instrument acquisition, replicate numbers, and statistical approaches independently, a holistic benchmarking approach that examines the synergistic effects of these parameters on quantitative performance has not been undertaken.<sup>2,29,34,42–48</sup> In this study, we have rigorously determined the combined impact of these factors on the quantitative performance in label-free proteomics (DDA and DIA). As a corollary, we also evaluated the impact of statistical approach on a previously published data set employing isobaric labeling via TMTs.

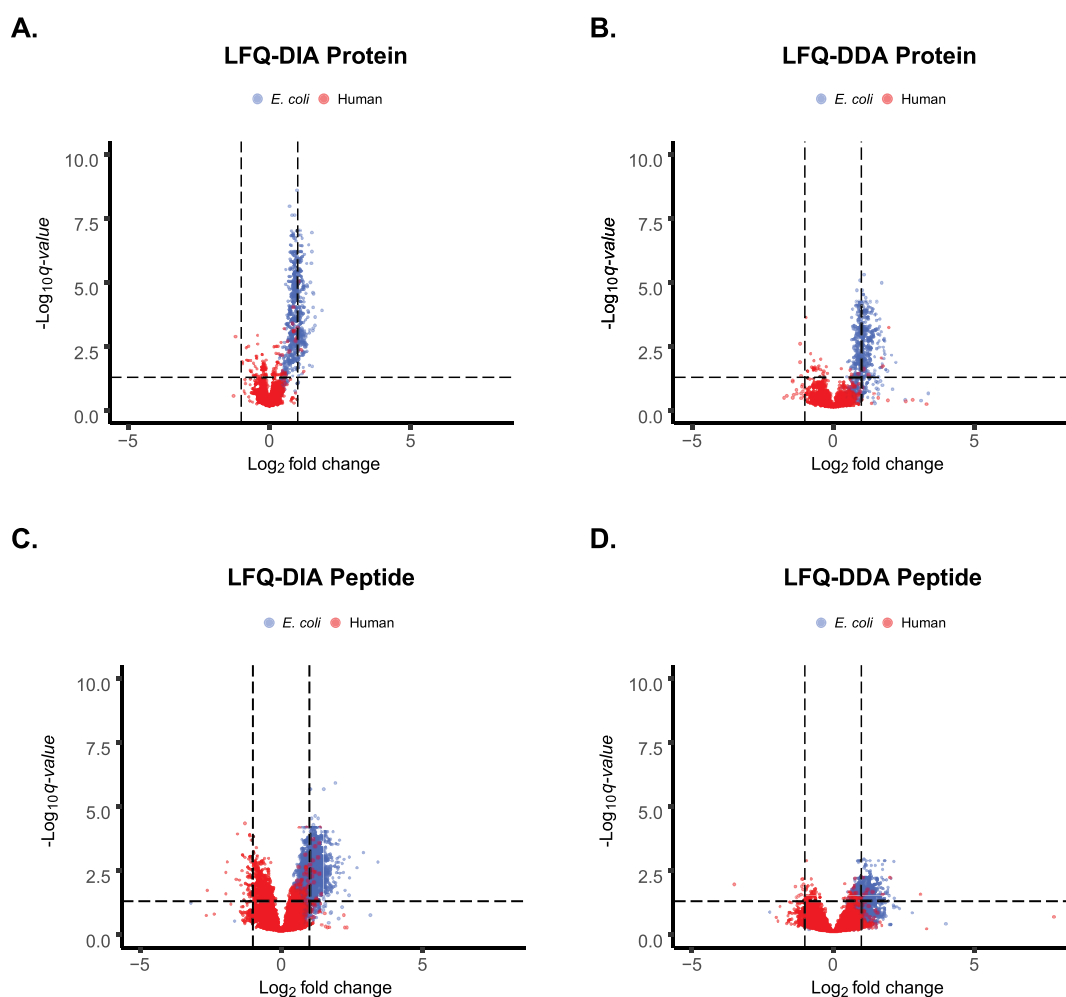
## RESULTS

**Experimental Overview.** To assess the impact of data acquisition, statistical approach, false discovery rate (FDR) calculations, and replicate number on quantitative accuracy, we produced a benchmarking data set by spiking an *Escherichia coli* lysate into a human HEK293 lysate at 1:4 (1x sample) and 1:2 (2x sample), where the 2x sample contains twice as much total *E. coli* protein as the 1x sample, while the HEK293 protein amount is held constant (Figure 1). Specifically, 100



**Figure 1.** Workflow overview. Benchmarking standards were created with an *E. coli* protein digest spiked into a human protein digest (HEK293 cell lysate) at either a 1x or 2x concentration. Benchmark standards were analyzed by LC–MS/MS using two different acquisition modes: DDA or DIA. Statistical analysis was performed with four distinct approaches: *t*-test, LIMMA, ROPECA, and ROTS.

$\mu\text{g}$  (1x sample) or 200  $\mu\text{g}$  (2x sample) of the *E. coli* lysate was spiked into 400  $\mu\text{g}$  of the human lysate (HEK293 cell lysate). Since there is very little protein homology between these species, the expected peptide/protein fold change between the 1x and 2x conditions is known simply based on the species, and thus, human peptides/proteins have an ideal fold change of 1 ( $\log_2 \text{FC} = 0$ ), while bacterial proteins have an ideal fold change of 2 ( $\log_2 \text{FC} = +1$ ). We analyzed our benchmark

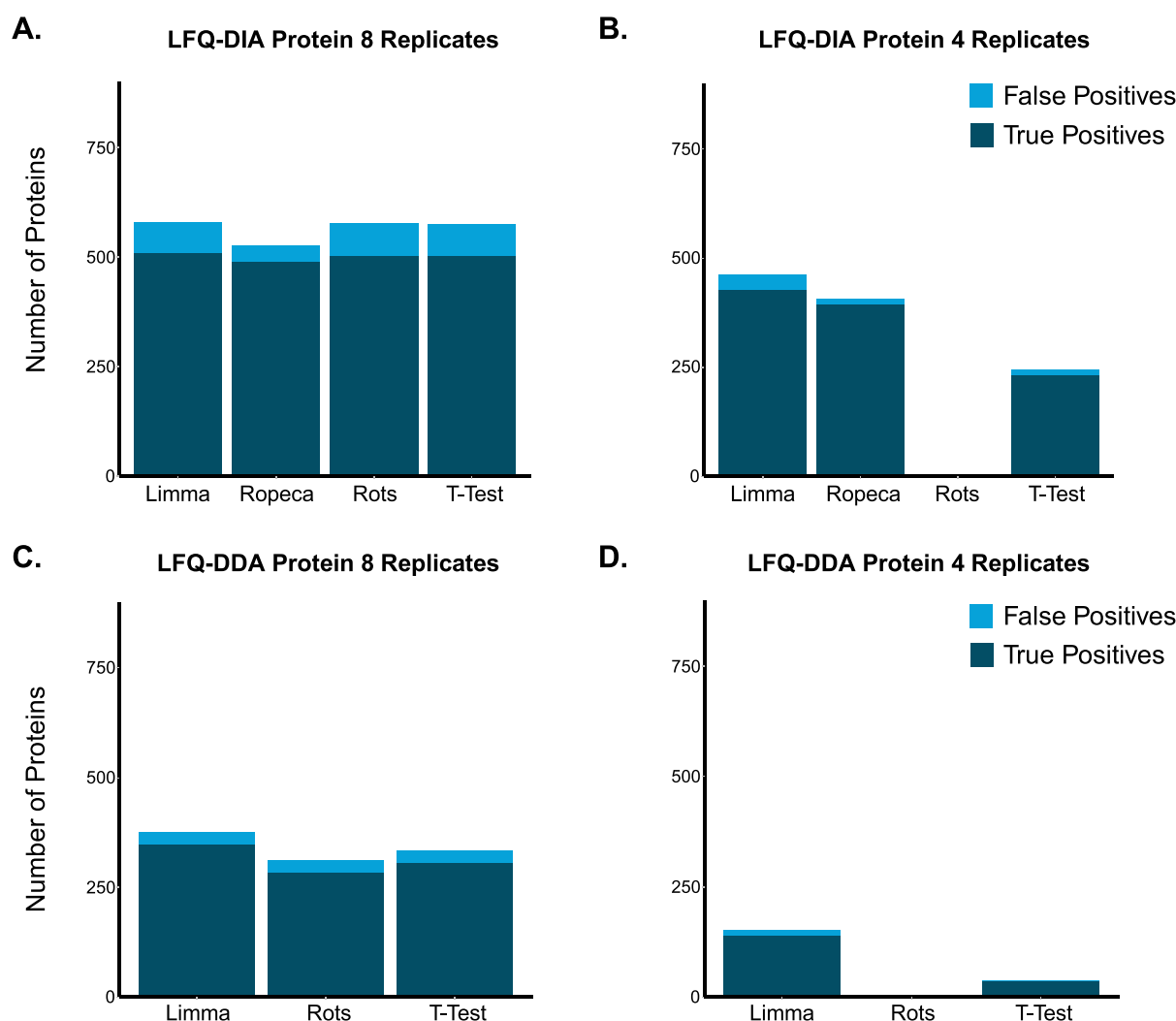


**Figure 2.** Scatter plots. Log<sub>2</sub> fold change plotted against the negative log<sub>10</sub> *q*-value for each protein/peptide. Red dots are human proteins/peptides. Blue dots are *E. coli* proteins/peptides. All proteins above the horizontal dotted line indicate statistical significance ( $n = 8$ , moderated *t*-test with the BH correction,  $q < 0.05$ ). Vertical lines indicate a log<sub>2</sub> fold change of  $\pm 1$ . (A) Proteins analyzed via DIA, (B) proteins analyzed via DDA, (C) peptides analyzed via DIA, and (D) peptides analyzed via DDA.

standard using either label-free DIA (LFQ-DIA) or DDA (LFQ-DDA) in combination with various replicate numbers and statistical approaches. The peptide library for the DIA analysis was generated from the DDA replicates, and thus, the identified peptides and proteins are identical between the two data sets. In this manner, we obviated any identification bias between data sets.

To extend our evaluation beyond label-free quantitative methodologies, we evaluated the impact of statistical approaches on the quantitative performance of a previously published data set that utilized tandem mass tagging (TMT-DDA) and label-free DDA<sup>2</sup>. The inclusion of this data set also served as a laboratory-independent control for our evaluation of statistical approaches. Gygi et al. employed a similar experimental design as the one employed in our study, but instead of using an *E. coli* digest as their spike-in proteome, they used a yeast digest. However, in contrast to our “single-shot” label-free workflow, Gygi et al. prefractionated their benchmarking sample via offline high pH reverse-phase HPLC before chemical tagging and LC–MS/MS analysis. Gygi et al. also employed a Thermo Scientific Lumos mass spectrometer coupled to a nanoflow HPLC, whereas we used a first-generation Thermo Scientific Q-Exactive coupled to a microflow HPLC.

**Acquisition Methodology.** To qualitatively assess the impact of instrument acquisition modes on quantitative performance, we analyzed the *E. coli* benchmarking standard via DDA and DIA acquisition. These analyses were performed using the MS1 signal intensities of eight replicates for both the DDA and DIA acquisition methodologies. We performed moderated *t*-tests with a Benjamini–Hochberg (BH) false discovery correction to identify those peptides and proteins that exhibit significantly different quantities between the 1x and 2x conditions. We plotted log<sub>2</sub> fold change against the negative log<sub>10</sub> *q*-value for each protein or peptide (Figure 2). All probes above the horizontal dotted line are statistically significant ( $q < 0.05$ ). The expected fold change value is designated by the dotted vertical trendline at log<sub>2</sub> fold change = 1. We assessed the quantitative performance via the overall negative log<sub>10</sub> *q*-value of the bacterial proteins/peptides and number of bacterial proteins/peptides that most closely adhere to the expected fold change trendline. For the protein-level analysis, the LFQ-DIA exhibited the highest negative log<sub>10</sub> *q*-values as well as a slightly tighter grouping around the expected log<sub>2</sub> fold change trendlines (Figure 2A,B). For the peptide-level analysis, the LFQ-DIA data set also had the highest negative log<sub>10</sub> *q*-values. However, in contrast to the protein data, the LFQ-DIA and LFQ-DDA peptide-level analysis displayed a



**Figure 3.** Protein true and false positive stacked bar graphs. True positives (dark blue) and false positives (light blue) were plotted as stacked bar graphs for each protein workflow and statistical approach. (A) Eight replicates analyzed via DIA, (B) four replicates analyzed via DIA, (C) eight replicates analyzed via DDA, and (D) four replicates analyzed via DDA.

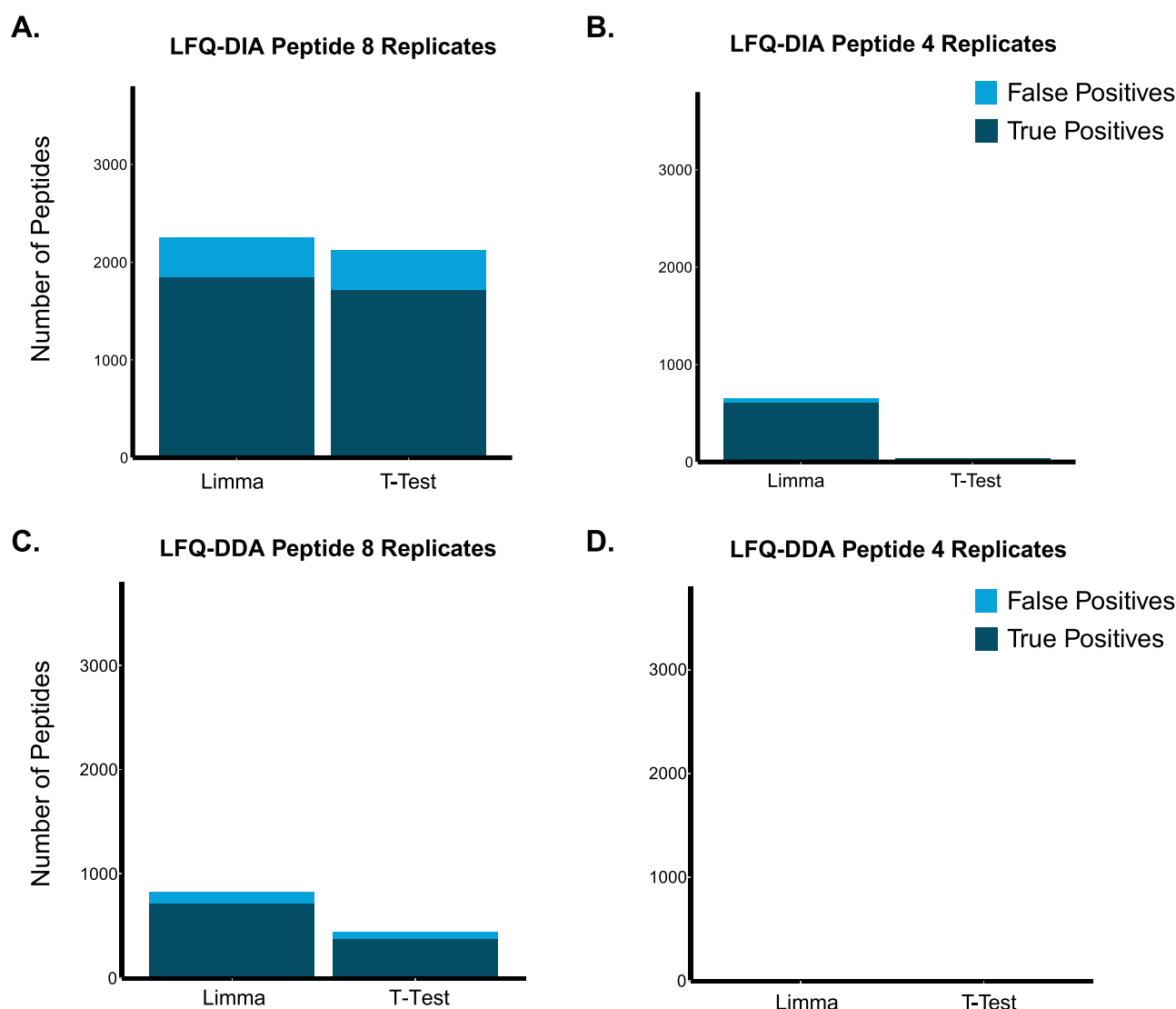
similar data spread and adherence to the expected  $\log_2$  fold change trendlines (Figure 2C,D). We also visualized the data as a standard deviation plot and observed similar trends in the adherence to the expected fold change and data spread as observed in the scatter plots (Figure S1).

As an extension of these observations, we also qualitatively assessed the protein- and peptide-level quantitative performance of the Gygi et al. yeast benchmarking data set using the same metrics of the negative  $\log_{10}$   $p$ -values and adherence to expected fold change values (Figure S2). The yeast TMT-DDA data set exhibited the tightest grouping around the expected  $\log_2$  fold change trendline, while the yeast LFQ-DDA data set exhibited a much larger data spread. Also, the negative  $\log_{10}$   $q$ -values were higher for the yeast TMT-DDA than for the yeast LFQ-DDA data set. For the peptide-level analysis, the yeast TMT-DDA data set also had higher negative  $\log_{10}$   $q$ -values and adhered more closely to the expected  $\log_2$  fold change trendlines than the yeast LFQ-DDA data set.

Interestingly, all the protein data sets exhibited a tighter grouping around the expected trendline than the peptide data sets, indicating a higher quantification fidelity at the protein than at the peptide level. This is likely due to protein-level quantification being a composite of multiple peptide signals,

essentially producing multiple measurements of its intensity value, whereas peptide-level quantification relies on a single intensity value, that of the individual peptide analyte. Thus, peptide quantification is more prone to quantification errors and requires a more robust experimental design than protein-level quantification, a theme we will explore more thoroughly below.

**Statistical Approaches.** While the LFQ-DIA data display slightly less data spread and higher negative  $\log_{10}$   $p$ -values than the LFQ-DDA, these observations are largely qualitative. To quantitatively assess the impact of acquisition methodology, statistical approach, and replicate number on protein-level quantitative performance, three distinct statistical methods were compared— $t$ -tests, linear models for the analysis of microarrays (LIMMAs), and reproducibility-optimized test statistic (ROTS and ROPECA). We used true positive and false positive rates to assess the performance of each workflow. True positive rate (TPR) measures the proportion of actual positives that are identified in an experiment in which the actual positives and negatives are known. In our experimental design, a true positive is any *E. coli* protein/peptide that is identified as being significantly different between the 1x and 2x benchmarking standards. The false positive rate is the



**Figure 4.** Peptide true and false positive stacked bar graphs. True positives (dark blue) and false positives (light blue) were plotted as stacked bar graphs for each peptide workflow and statistical approach. (A) Eight replicates analyzed via DIA, (B) four replicates analyzed via DIA, (C) eight replicates analyzed via DDA, and (D) four replicates analyzed via DDA.

proportion of nonpositives identified as positive. In our experimental design, a false positive is any human protein/peptide identified as being significantly different between the 1x and 2x benchmarking standards.

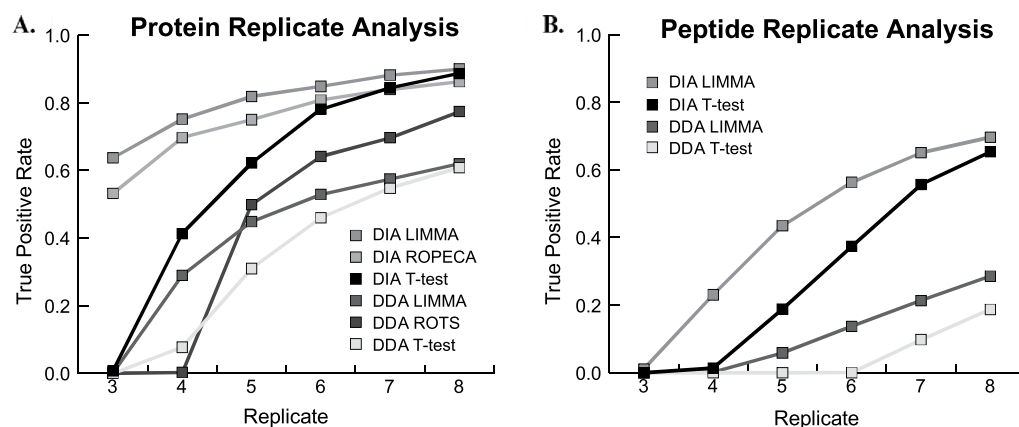
LIMMA fits a linear model to the expression data.<sup>49</sup> ROTS and ROPECA use a version of bootstrapping to optimize the parameters that maximize reproducibility to identify differentially expressed proteins.<sup>29,34</sup> ROTS is specific to protein-level data and ROPECA is the adaptation of ROTS that calculates peptide-level *p*-values and then aggregates expression values of each peptide within a protein into a single value for that protein. In order to control for the multiple testing hypothesis problem, we applied the BH FDR correction with a *q*-value threshold of <0.05.<sup>50</sup>

To compare these various approaches for protein-level analysis, stacked bar graphs of true and false positives (Figure 3), confusion tables (Figure S3), and receiver operating characteristic (ROC) curves (Figure S4) were generated. The LFQ-DIA workflow at high replicate numbers ( $n = 8$ ) generated a high number of true positives and a low number of false positives across all statistical approaches with an

average 90% TPR (~500 out of 568 possible true positives) and a false positive rate of ~3% across all statistical approaches (Figure 3A). In contrast, the quantitative performance of DIA at low replicate numbers ( $n = 4$ ) was highly dependent on the statistical approach with LIMMA and ROPECA performing extremely well—LIMMA had a 75% TPR (427 out of 568 possible) and ROPECA had a 70% true positive rate (396 out of 568 possible)—while the *t*-statistic only had a 40% TPR and ROTS performed even more poorly, identifying a single true positive (Figure 3B).

For all replicate numbers and statistical approaches, the LFQ-DDA workflow had a lower TPR than the LFQ-DIA workflow. Across all statistical conditions, the high replicate analysis ( $n = 8$ ) produced an average 54% TPR (311 out of 573 possible) with LIMMA slightly outperforming the other two statistical approaches (Figure 3C). However, LFQ-DDA exhibited a poor TPR at low replicate numbers, regardless of the statistical approach, with an average 10% TPR (59 out of 573 possible) (Figure 3D). However, there were still large differences in statistical performance at low replicate numbers in the LFQ-DDA data set—LIMMA had a 24% TPR (139 out





**Figure 5.** Replicate analysis. TPR plotted against replicate number for (A) each protein and (B) peptide workflow and statistical approach from the *E. coli* benchmarking data set. The legend indicates the instrument acquisition and statistical approach for each line plot.

of 573 possible), while the *t*-statistic and ROTS had TPRs well below 10%. While the reasons for the difference in quantitative performances between the LFQ-DIA and LFQ-DDA workflows are unclear, one of the major differences between the DDA and DIA data sets is the number of missing values. The LFQ-DDA workflow exhibited a significantly higher number of missing values (12%) versus the LFQ-DIA workflow (0.078%), pointing to the possible importance of missing values on quantitative performance.

To examine if the trends we observed in label-free protein quantification extended to a chemical tagging approach, we compared the performance of the TMT and DDA workflows from a previously published data set.<sup>2</sup> The results from the yeast LFQ-DDA data set from Gygi et al. exhibited very similar trends to the *E. coli* LFQ-DDA data set at low replicates ( $n = 4$ ), with LIMMA strongly outperforming ROTS and the *t*-statistic (Figure S5). The yeast LFQ-DDA analysis yielded a 45% TPR (432 out of 945 possible) and a 3% false positive rate with LIMMA, while ROTS and the *t*-test performed much more poorly. Interestingly, the yeast TMT-DDA workflow at low replicate numbers ( $n = 4$ ) had a very high average TPR of 94% (1222 out of 1304 possible) (Figure S5). These results are in stark contrast to all other workflows at low replicate numbers, which all exhibited highly variable TPRs between statistical approaches. It also has the fewest missing values of any data set (less than 0.15%), once again emphasizing the strong impact of missing values on quantitative fidelity, especially at low replicate numbers. In terms of statistics, LIMMA and ROPECA performed the best across workflows and replicate numbers.<sup>34</sup> In summary, these data demonstrate the complex interplay of replicate number, instrument acquisition, and statistical approach.

For the peptide-level analysis, two statistical approaches were performed—LIMMA and *t*-tests (Figure 4). ROTS and ROPECA can only be performed on protein-level analyses, so these statistical approaches were not included.<sup>33,34</sup> The LFQ-DIA analysis at high replicate numbers ( $n = 8$ ) had a 67% TPR (1786 out of 2,655 possible) (Figure 4A). In contrast, the LFQ-DDA analysis, even at high replicates, performed poorly with a 29% TPR (715 out of 2501 possible) using LIMMA and a 15% TPR using the *t*-statistic (Figure 4C). However, at low replicate numbers even DIA performed poorly, with LIMMA and *t*-tests exhibiting 23% and 1% TPRs, respectively (Figure 4B). The LFQ-DDA workflow at a low replicate number

performed even worse, with less than 10 true positives using LIMMA and no true positives with the *t*-statistic (Figure 4D).

To examine if the trends observed in the peptide quantification via LFQ-DDA and LFQ-DIA workflows extend to a chemical tagging approach, we compared the performance of the Gygi et al. TMT and DDA data sets. The yeast LFQ-DDA data exhibited a 44% TPR for LIMMA and a 2% TPR for *t*-tests (Figure S5). In contrast, the yeast TMT-DDA workflow produced 97% true positives with LIMMA (6434 out of a possible 6610) and a 75% TPR via the *t*-statistic (Figure S5). While not directly comparable to our LFQ-DIA and LFQ-DDA data sets, the statistical performance of LIMMA and the *t*-statistic exhibited highly similar trends across all data sets, with LIMMA outperforming the *t*-statistic regardless of acquisition methodology.

In summary, accurate peptide quantification is much more demanding than protein analyses. The LFQ-DIA and TMT-DDA workflows produced superior results to the LFQ-DDA workflows, with LIMMA producing the highest TPRs of any statistical approach across all workflows.

**Replicate Numbers.** To further examine the impact of replicate numbers on statistical performance, we plotted TPR versus replicate number for each statistical approach in combination with either the LFQ-DDA or LFQ-DIA workflow (Figure 5). As with our previous analyses, the statistical performance was highly variable across the replicate number. For LFQ-DIA protein quantification, both LIMMA and ROPECA exhibited a high, flat TPR across replicates, which is consistent with the previous comparison of low and high replicate data that showed robust performance at both low ( $n = 4$ ) and high ( $n = 8$ ) replicates (Figure 3). However, the *t*-statistic exhibited a steadily increasing curve that approached a plateau near  $n = 8$  for the LFQ-DIA workflow, indicating that the optimal replicate number for the *t*-statistic is closer to  $n = 8$ . For the LFQ-DDA data, both the *t*-statistic and LIMMA analysis steadily increased with more replicates and do not appear to plateau, indicating that the optimal replicate number is above  $n = 8$ . For the peptide LFQ-DIA data, both the LIMMA and *t*-statistic steadily increased to  $n = 8$  without plateauing, once again, indicating that the ideal replicate number is above  $n = 8$ . For the LFQ-DDA peptide data, very few true positives are identified until  $n = 5$  for LIMMA and  $n = 7$  for the *t*-statistic. These data show that the statistical performance is highly dependent on the data acquisition methodology and replicate number.

**FDR Correction.** Controlling for the multiple testing hypothesis problem is essential for the quantitative accuracy of large data sets, including microarrays, RNA-seq, and quantitative proteomics experiments.<sup>50,51</sup> A commonly used FDR correction for type I errors (false positives) is the Benjamini–Hochberg (BH) correction.<sup>50</sup> For all previously displayed data, we applied the BH correction at a  $q$ -value threshold of  $<0.05$  (see the [Experimental Section](#)). To assess the potential impact of FDR corrections on quantitative accuracy, we also applied the Storey correction at the same  $q$ -value cutoff.<sup>51</sup> The Storey correction is slightly less conservative than the BH correction.<sup>51</sup> Overall, the Storey correction produced higher TPRs with a concurrent increase in the false positive rates. However, the choice of FDR correction unexpectedly exhibited variable behavior that depended on replicate number, statistical approach, and acquisition methodology.

The BH correction ( $q$ -value  $<0.05$ ) in combination with the LFQ-DIA protein workflow at a high replicate number ( $n = 8$ ) differed very little from the Storey corrected results, regardless of statistical approach ([Table S4](#)). However, the FDR correction had a slightly larger impact on the high replicate LFQ-DDA workflow with the Storey correction producing a consistently higher TPR. The choice of FDR correction also noticeably affected the results from the low replicate LFQ-DIA workflow. However, the most dramatic effect of altering the FDR correction was manifested in the low replicate ( $n = 4$ ) LFQ-DDA workflow in combination with LIMMA, with the Storey correction producing 37% more true positives than the BH correction ([Table S4](#)).

These results indicate that FDR corrections can greatly impact quantitative performance, especially at low replicates using LFQ-DDA approaches. It is unclear why the FDR correction produces large effects in certain conditions, for example, low replicate LFQ-DDA using LIMMA, and not others, for example, high replicate LFQ-DIA. However, the high replicate workflows clearly exhibit a greatly reduced dependence on the choice of FDR correction, again, highlighting the importance of replicate choice in experimental design.

## ■ DISCUSSION

We have employed a quantitative proteomics benchmarking approach to explore the synergistic impact of instrument acquisition methodology, replicate number, and statistical approach on quantitative performance. We have evaluated two LFQ workflows as well as TMT workflow for protein- and peptide-level quantitative performance. The relative advantages of these quantitative workflows have been evaluated before.<sup>2,29,34,42–48</sup> However, the combined impact of replicate number and statistical approach on these workflows has not been thoroughly explored. Interestingly, we found a complex relationship between quantitative workflow, statistical approach, and replicate number.

A common theme we found across our analyses is those acquisition methodologies with fewer missing values (TMT-DDA and LFQ-DIA) produced superior results at lower replicate numbers. Missing values have been shown to have a detrimental effect on downstream statistical analyses across disparate data types.<sup>2,15,16</sup> The effect of missing values is especially detrimental at low replicate numbers. However, the impact of missing values and/or low replicate numbers on quantitative performance depends on the statistical approach

employed. Specifically, conventional parametric statistical tests, such as the  $t$ -statistic, are especially affected. In contrast, the performance of a linear model fitted to expression values, as implemented in the LIMMA package, was less impacted by low replicate numbers and/or high numbers of missing values. This is not surprising since LIMMA has been optimized for smaller sample sizes compared with the  $t$ -statistic and its improved performance might be due to the elimination of small within-group variance and/or the introduction of fold-change criterion.<sup>49</sup> Similarly, the performance of the peptide-centric, bootstrapping algorithm employed by ROPECA is also relatively unaffected by the low number of replicates.<sup>34</sup> Most likely, this is due to the generation and aggregation of  $p$ -values at the peptide level instead at the protein level. We would expect that other peptide-centric statistical methodologies would also exhibit similar performance, including the algorithm employed in MSstats.<sup>35</sup>

In contrast to protein-level quantification, which is based on multiple peptide measurements, peptide-level quantification relies on a single value. This produces a much more demanding scenario for downstream statistical analysis, with missing values and/or low replicate numbers being especially detrimental to the quantitative performance. Furthermore, the use of a single peptide intensity for quantification precludes the implementation of peptide-centric statistical methods, such as those employed by ROPECA and MSstats, that aggregate multiple peptide  $p$ -values into a single protein-level value. For peptide-level quantification, LIMMA clearly outperforms the  $t$ -statistic. As mentioned above, this is not too surprising given that LIMMA has been optimized for small replicate numbers. However, even LIMMA does not perform well with low replicates in combination with high numbers of missing values, such as those encountered in the LFQ-DDA workflows. Thus, peptide-level quantification requires more robust experimental designs, including the use of data acquisition methodologies with less missing values (TMT-DDA and LFQ-DIA), higher replicate numbers ( $n = 8$ ), and more sophisticated statistical approaches (LIMMA).

Another important consideration is instrument requirements. DDA approaches are the simplest to set up and can be performed on a wide array of instrumentation. A quantitative DDA experiment can be performed on almost any mass spectrometer, whether it is a quadrupole time-of-flight, an ion trap (IT), an Orbitrap, or a hybrid instrument. In contrast to DDA, DIA requires a high-resolution quadrupole instrument to obtain suitable MS2 data for peptide library matching and quantification. Although TMT workflows can theoretically be performed on many proteomics instruments, quantitative accuracy is greatly enhanced when FAIMS or MS3-based quantification is used; however, this requires an IT-based instrument or a FAIMS-capable source. Also, TMT workflows generally require prefractionation, usually via ion exchange or high pH reverse-phase chromatography, which adds to the complexity and level of expertise required to execute this workflow.

Another consideration is the cost of executing these workflows. The TMT-DDA workflow requires the purchase of chemical tagging reagents, which increases the cost per sample. Some LFQ-DIA workflows require the purchase of retention time standards, which also adds to the analysis cost. While LFQ-DDA workflows do not require the purchase of these reagents, they require a higher number of replicates to obtain adequate quantitative performance. This requirement

for higher replicate numbers increases the upstream sample preparation costs as well as the instrument time per experiment.

In summary, excellent quantitative performance in bottom-up proteomics can be achieved using LFQ-DDA, LFQ-DIA, or TMT-DDA workflows. However, each workflow exhibits unique characteristics that require careful attention to experimental design, including the choice of replicate number and downstream statistical approach.

## ■ EXPERIMENTAL SECTION

**Sample Preparation.** HEK293 cells were cultured in Dulbecco's modified Eagle's medium with 10% fetal bovine serum until confluent. The cells were rinsed once with phosphate-buffered saline, lifted with trypsin/ethylenediaminetetraacetic acid, and then pelleted. The pellet was resuspended in 1% sodium dodecyl sulfate in 100 mM Tris-HCl (pH 7.5) and then tip sonicated at 20% amplitude three times for 5 s each. Cysteines were reduced with 25 mM dithiothreitol at 95 °C for 10 min and then alkylated with 50 mM iodoacetamide for 20 min in the dark at room temperature. Proteins were precipitated with 4x volumes of ice-cold acetone for 60 min and then spun at 21,000 g to pellet proteins. Precipitated protein was resuspended in 2 M urea in 25 mM ammonium bicarbonate with tip sonication (as above). Protein concentration was determined via the bicinchoninic acid assay. The protein extract was digested with trypsin at 1:50 overnight at 37 °C. The digest was quenched with 0.5% trifluoroacetyl and protein extracts desalted via Stage-tips.<sup>52</sup> The samples were dried down and resuspended in 0.1% formic acid. *E. coli* were cultured in 2xYT media and then pelleted at 8000 g and processed in the same manner as the HEK293 samples.

We prepared our benchmark sample as previously published by Mann and co-workers.<sup>53</sup> Two separate samples were created by mixing the HEK293 and *E. coli* digests. For the "1x" sample, 100  $\mu$ g of *E. coli* digest was added to 400  $\mu$ g of HEK293 sample and diluted to a total of 600  $\mu$ L. For the "2x" sample, 200  $\mu$ g of *E. coli* digest was added to 400  $\mu$ g of HEK293 sample and diluted to a total of 600  $\mu$ L. iRT peptides (HRM Kit, Biognosys) were spiked into each peptide digest according to the manufacturer's instructions. Eight technical replicates of each sample group (1x and 2x) and each acquisition type (DDA and DIA) were analyzed by LC-MS/MS for a total of 32 runs. For the mapping of sample IDs to raw files, the sequence table was exported into an Excel file (Table S4).

**LC-MS/MS Parameters.** For both DDA and DIA, 20  $\mu$ L of each sample was injected onto a Dionex Ultimate3000 microflow HPLC and separated at a flow rate of 6  $\mu$ L/min using a Thermo Scientific Hypersil Gold C18 column (150  $\times$  0.18 mm, particle size 3  $\mu$ m) at 60 °C coupled to a Thermo Scientific Q-Exactive mass spectrometer with a HESI-II source equipped with a narrow-bore spray needle. HPLC mobile phases consisted of water + 0.1% formic acid and acetonitrile + 0.1% formic acid. Peptides were resolved with a linear gradient of 2–30% ACN over 85 min. HESI parameters were as follows: sheath gas = 5; spray voltage = 3 kV; capillary temperature = 240 °C; S-lens RF = 50; auxiliary gas heater = 30 °C; auxiliary gas flow = OFF; and sweep gas = OFF. The mass spectrometer was operated in the DDA mode with dynamic exclusion enabled (exclusion duration = 15 s), MS1 resolution = 70,000, MS1 automatic gain control target =  $1 \times 10^6$ , MS1 maximum fill time = 100 ms, MS2 resolution =

17,500, MS2 automatic gain control target =  $2 \times 10^5$ , MS2 maximum fill time = 200 ms, and MS2 normalized collision energy = 28. For each cycle, one full MS1 scan range = 400–1000  $m/z$ , followed by MS2 scans with a loop count of 20 and an isolation window size of 2.0  $m/z$ . In DIA, the mass spectrometer was operated with a MS1 scan resolution = 70 000, automatic gain control target =  $1 \times 10^6$ , MS1 maximum fill time = 100 ms, followed by a DIA scan with a loop count of 30 using nonoverlapping windows from 400 to 1000  $m/z$ . The DIA settings were as follows: window size = 20  $m/z$ , resolution = 17,500, automatic gain control target =  $1 \times 1 \times 10^6$ , DIA maximum fill time = AUTO, and normalized collision energy = 30. The total cycle time was 3.1 s.

**Data Processing.** DDA runs were analyzed with MaxQuant (version 1.6.5.0) using default settings. The data were searched against a combined human and *E. coli* Swiss-Prot database (downloaded December 12, 2017). Search criteria included carbamidomethylation of cysteine as a fixed modification, methionine oxidation and N-terminal protein acetylation as variable modifications, and trypsin/P with two missed cleavages as the enzyme. Mass tolerance was set at 4.5 ppm for precursor ions and 20 ppm for fragment ions. Peptide spectrum match and Protein FDR were set at 0.01. For DDA peptide quantification, match between runs was performed with a matching time window of 0.7 min and an alignment window of 20 min. MS1 summed isotope intensities were used. For protein quantification, all peptide MS1 intensities for a given protein were summed to yield the total protein intensity. The MaxQuant output file was exported as an Excel file and is included in the Supporting Information (Table S4). Normalization was performed via the summation of the total intensity of all the identified human peptides. These calculations were performed using an internally generated R-script (see below). The normalized peptide and protein values were exported to an Excel file (Table S4).

The DIA runs were analyzed with Spectronaut Pulsar X (version 12.0.20491.13.20699) using default parameters. The spectral libraries were generated from the MaxQuant results from all 16 DDA runs. The spectral library contained 19,561 precursors, 17,571 modified peptides, and 3092 proteins. For identification, a "mutated" decoy method was used, and the precursor and protein  $q$ -value cutoffs were set at 0.01, and the retention time prediction was set to dynamic. MS1 summed isotope intensities were used for peptide quantification. Protein inference was set to Automatic and interference correction was enabled. The Spectronaut output file was exported as an Excel file and is included in the Supporting Information (Table S4). Protein intensities and normalization were calculated as described for the DDA analysis (see above). The normalized peptide and protein values were exported to an Excel file (Table S4).

**R-Script.** The analyses were performed using the R programming language ([www.r-project.org](http://www.r-project.org)), version 3.6. Our script accepts peptide- or protein-level intensity data, performs sum normalization on total peptide intensities, rolls up peptide-to-protein intensities via summation, runs statistical tests, and generates plots. *Note: Data imputation of missing values was not performed.* We have written a general user guide to download and run the script in RStudio or your choice of editor, which includes the necessary parameters for the input files (see below).

**Statistical Analysis.** Statistical analyses were performed on both protein- and peptide-level intensities using  $q$ -adjusted  $t$ -



tests, LIMMA, and ROTS. FDR was controlled via BH or Storey corrections.<sup>50,51</sup>

**Yeast Benchmarking Label-Free DDA and TMT Data Sets.** We analyzed a previously published data set from Gygi et al.<sup>2</sup> This benchmarking data set is very similar to our in-house generated sample except a yeast proteome spike-in was used instead of an *E. coli* spike-in. We analyzed both the DDA and TMT data sets using our analysis pipeline. The original Excel files were downloaded and minimally reformatted to conform to the data frame requirements of our R script. We only analyzed the “2x” versus the “1x” spike-in samples and did not include the “3x” samples. Normalization, peptide-to-protein rollup, and statistical analysis were performed as described above. The normalized peptide and protein values were exported to an Excel file (Table S4).

**Data and R Script Availability.** The normalized intensity values for all the experimental conditions are summarized in Table S4. The raw data, spectral libraries, and output files are publicly available on the ProteomeXchange data repository (PXD018408) and the MassIVE repository (MSV000085239). The R script and associated output files are available on GitHub (<https://github.com/DenuLab/LFQMagic>).

## ■ ASSOCIATED CONTENT

### ■ Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsomega.0c04030>.

FDR corrections, summary of normalized intensity values for peptide and protein analytes, MaxQuant output file, and Spectronaut output file (XLSX)

Protein and peptide standard deviation plots; yeast benchmark protein and peptide scatter plots; protein and peptide confusion tables; protein and peptide receiver operating characteristic curves; and yeast benchmark protein and peptide true and false positive stacked bar graphs (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

**James A. Dowell** – Wisconsin Institute for Discovery, University of Wisconsin–Madison, Madison, Wisconsin 53715, United States; [orcid.org/0000-0001-9358-7653](https://orcid.org/0000-0001-9358-7653); Email: [jadowell@wisc.edu](mailto:jadowell@wisc.edu)

**John M. Denu** – Wisconsin Institute for Discovery and Department of Biomolecular Chemistry, University of Wisconsin–Madison, Madison, Wisconsin 53715, United States; [orcid.org/0000-0001-9415-0365](https://orcid.org/0000-0001-9415-0365); Email: [john.denu@wisc.edu](mailto:john.denu@wisc.edu)

### Authors

**Logan J. Wright** – Wisconsin Institute for Discovery, University of Wisconsin–Madison, Madison, Wisconsin 53715, United States

**Eric A. Armstrong** – Wisconsin Institute for Discovery, University of Wisconsin–Madison, Madison, Wisconsin 53715, United States

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsomega.0c04030>

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We thank Anthony Gitter and Thevaa Chandereeng for useful discussions related to the statistical analysis of benchmarking data sets. This work was in part supported by the NIH grant R37GM059785 to J.M.D.

## ■ REFERENCES

- (1) Mann, M.; Kulak, N. A.; Nagaraj, N.; Cox, J. The Coming Age of Complete, Accurate, and Ubiquitous Proteomes. *Mol. Cell* **2013**, *49*, 583–590.
- (2) O’Connell, J. D.; Paulo, J. A.; O’Brien, J. J.; Gygi, S. P. Proteome-Wide Evaluation of Two Common Protein Quantification Methods. *J. Proteome Res.* **2018**, *17*, 1934–1942.
- (3) Li, Z.; Adams, R. M.; Chourey, K.; Hurst, G. B.; Hettich, R. L.; Pan, C. Systematic Comparison of Label-Free, Metabolic Labeling, and Isobaric Chemical Labeling for Quantitative Proteomics on LTQ Orbitrap Velos. *J. Proteome Res.* **2012**, *11*, 1582–1590.
- (4) Tebbe, A.; Klammer, M.; Sighart, S.; Schaab, C.; Daub, H. Systematic Evaluation of Label-Free and Super-SILAC Quantification for Proteome Expression Analysis. *Rapid Commun. Mass Spectrom.* **2015**, *29*, 795–801.
- (5) Ong, S.-E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol. Cell. Proteomics* **2002**, *1*, 376–386.
- (6) McAlister, G. C.; Nusinow, D. P.; Jedrychowski, M. P.; Wühr, M.; Huttlin, E. L.; Erickson, B. K.; Rad, R.; Haas, W.; Gygi, S. P. MultiNotch MS3 Enables Accurate, Sensitive, and Multiplexed Detection of Differential Expression across Cancer Cell Line Proteomes. *Anal. Chem.* **2014**, *86*, 7150–7158.
- (7) Ting, L.; Rad, R.; Gygi, S. P.; Haas, W. MS3 Eliminates Ratio Distortion in Isobaric Multiplexed Quantitative Proteomics. *Nat. Methods* **2011**, *8*, 937–940.
- (8) Thompson, A.; Schäfer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Hamon, C. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Anal. Chem.* **2003**, *75*, 1895–1904.
- (9) Pfammatter, S.; Bonnell, E.; Thibault, P. Improvement of Quantitative Measurements in Multiplex Proteomics Using High-Field Asymmetric Waveform Spectrometry. *J. Proteome Res.* **2016**, *15*, 4653–4665.
- (10) Schweppe, D. K.; Prasad, S.; Belford, M. W.; Navarrete-Perea, J.; Bailey, D. J.; Huguet, R.; Jedrychowski, M. P.; Rad, R.; McAlister, G.; Abbatiello, S. E.; Wouters, E. R.; Zabrouskov, V.; Dunyach, J.-J.; Paulo, J. A.; Gygi, S. P. Characterization and Optimization of Multiplexed Quantitative Analyses Using High-Field Asymmetric-Waveform Ion Mobility Mass Spectrometry. *Anal. Chem.* **2019**, *91*, 4010–4016.
- (11) Azimifar, S. B.; Nagaraj, N.; Cox, J.; Mann, M. Cell-Type-Resolved Quantitative Proteomics of Murine Liver. *Cell Metab.* **2014**, *20*, 1076–1087.
- (12) Wilhelm, M.; Schlegl, J.; Hahne, H.; Gholami, A. M.; Lieberenz, M.; Savitski, M. M.; Ziegler, E.; Butzmann, L.; Gessulat, S.; Marx, H.; Mathieson, T.; Lemeer, S.; Schnatbaum, K.; Reimer, U.; Wenschuh, H.; Mollenhauer, M.; Slotta-Huspenina, J.; Boese, J.-H.; Bantscheff, M.; Gerstmair, A.; Faerber, F.; Kuster, B. Mass-Spectrometry-Based Draft of the Human Proteome. *Nature* **2014**, *509*, 582–587.
- (13) Kim, M.-S.; Pinto, S. M.; Getnet, D.; Nirujogi, R. S.; Manda, S. S.; Chaerkady, R.; Madugundu, A. K.; Kelkar, D. S.; Isserlin, R.; Jain, S.; Thomas, J. K.; Muthusamy, B.; Leal-Rojas, P.; Kumar, P.; Sahasrabudhe, N. A.; Balakrishnan, L.; Advani, J.; George, B.; Renuse, S.; Selvan, L. D. N.; Patil, A. H.; Nanjappa, V.; Radhakrishnan, A.; Prasad, S.; Subbannayya, T.; Raju, R.; Kumar, M.; Sreenivasamurthy, S. K.; Marimuthu, A.; Sathe, G. J.; Chavan, S.; Datta, K. K.; Subbannayya, Y.; Sahu, A.; Yelamanchi, S. D.; Jayaram, S.; Rajagopalan, P.; Sharma, J.; Murthy, K. R.; Syed, N.; Goel, R.; Khan, A. A.; Ahmad, S.; Dey, G.; Mudgal, K.; Chatterjee, A.; Huang, T.-C.; Zhong, J.; Wu, X.; Shaw, P. G.; Freed, D.; Zahari, M. S.;

- Mukherjee, K. K.; Shankar, S.; Mahadevan, A.; Lam, H.; Mitchell, C. J.; Shankar, S. K.; Satishchandra, P.; Schroeder, J. T.; Sirdeshmukh, R.; Maitra, A.; Leach, S. D.; Drake, C. G.; Halushka, M. K.; Prasad, T. S. K.; Hruban, R. H.; Kerr, C. L.; Bader, G. D.; Iacobuzio-Donahue, C. A.; Gowda, H.; Pandey, A. A Draft Map of the Human Proteome. *Nature* **2014**, *509*, 575–581.
- (14) Hebert, A. S.; Richards, A. L.; Bailey, D. J.; Ulbrich, A.; Coughlin, E. E.; Westphall, M. S.; Coon, J. J. The One Hour Yeast Proteome. *Mol. Cell. Proteomics* **2014**, *13*, 339–347.
- (15) Bruderer, R.; Bernhardt, O. M.; Gandhi, T.; Xuan, Y.; Sondermann, J.; Schmidt, M.; Gomez-Varela, D.; Reiter, L. Optimization of Experimental Parameters in Data-Independent Mass Spectrometry Significantly Increases Depth and Reproducibility of Results. *Mol. Cell. Proteomics* **2017**, *16*, 2296–2309.
- (16) Aittokallio, T. Dealing with Missing Values in Large-Scale Studies: Microarray Data Imputation and Beyond. *Briefings Bioinf.* **2009**, *11*, 253–264.
- (17) Välikangas, T.; Suomi, T.; Elo, L. L. A Comprehensive Evaluation of Popular Proteomics Software Workflows for Label-Free Proteome Quantification and Imputation. *Briefings Bioinf.* **2017**, *19*, 1344–1355.
- (18) Cox, J.; Mann, M. MaxQuant Enables High Peptide Identification Rates, Individualized p.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification. *Nat. Biotechnol.* **2008**, *26*, 1367–1372.
- (19) Karpievitch, Y. V.; Dabney, A. R.; Smith, R. D. Normalization and Missing Value Imputation for Label-Free LC-MS Analysis. *BMC Bioinf.* **2012**, *13* Suppl 1 (Suppl 16). DOI: 10.1186/1471-2105-13-S16-S5.
- (20) Lazar, C.; Gatto, L.; Ferro, M.; Bruley, C.; Burger, T. Accounting for the Multiple Natures of Missing Values in Label-Free Quantitative Proteomics Data Sets to Compare Imputation Strategies. *J. Proteome Res.* **2016**, *15*, 1116–1125.
- (21) Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M. Y.; Geiger, T.; Mann, M.; Cox, J. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **2016**, *13*, 731–740.
- (22) Goeminne, L. J. E.; Argentini, A.; Martens, L.; Clement, L. Summarization vs Peptide-Based Models in Label-Free Quantitative Proteomics: Performance, Pitfalls, and Data Analysis Guidelines. *J. Proteome Res.* **2015**, *14*, 2457–2465.
- (23) Goeminne, L. J. E.; Gevaert, K.; Clement, L. Experimental Design and Data-Analysis in Label-Free Quantitative LC/MS Proteomics: A Tutorial with MSqRob. *J. Proteomics* **2018**, *171*, 23–36.
- (24) Ludwig, C.; Gillet, L.; Rosenberger, G.; Amon, S.; Collins, B. C.; Aebersold, R. Data-independent acquisition-based SWATH - MS for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **2018**, *14*, 1–23.
- (25) Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-Independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteomics* **2012**, *11*, O111.016717.
- (26) Selevsek, N.; Chang, C.-Y.; Gillet, L. C.; Navarro, P.; Bernhardt, O. M.; Reiter, L.; Cheng, L.-Y.; Vitek, O.; Aebersold, R. Reproducible and Consistent Quantification of the Saccharomyces cerevisiae Proteome by SWATH-mass spectrometry. *Mol. Cell. Proteomics* **2015**, *14*, 739–749.
- (27) Rosenberger, G.; Bludau, I.; Schmitt, U.; Heusel, M.; Hunter, C. L.; Liu, Y.; MacCoss, M. J.; MacLean, B. X.; Nesvizhskii, A. I.; Pedrioli, P. G. A.; Reiter, L.; Röst, H. L.; Tate, S.; Ting, Y. S.; Collins, B. C.; Aebersold, R. Statistical Control of Peptide and Protein Error Rates in Large-Scale Targeted Data-Independent Acquisition Analyses. *Nat. Methods* **2017**, 921–927.
- (28) Dayon, L.; Hainard, A.; Licker, V.; Turck, N.; Kuhn, K.; Hochstrasser, D. F.; Burkhard, P. R.; Sanchez, J.-C. Relative Quantification of Proteins in Human Cerebrospinal Fluids by MS/MS Using 6-Plex Isobaric Tags. *Anal. Chem.* **2008**, *80*, 2921–2931.
- (29) Pursiheimo, A.; Vehmas, A. P.; Afzal, S.; Suomi, T.; Chand, T.; Strauss, L.; Poutanen, M.; Rokka, A.; Corthals, G. L.; Elo, L. L. Optimization of Statistical Methods Impact on Quantitative Proteomics Data. *J. Proteome Res.* **2015**, *14*, 4118–4126.
- (30) Tyanova, S.; Cox, J. Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research. *Methods Mol. Biol.* **2018**, 1711, 133–148.
- (31) Clough, T.; Thaminy, S.; Ragg, S.; Aebersold, R.; Vitek, O. Statistical Protein Quantification and Significance Analysis in Label-Free LC-MS Experiments with Complex Designs. *BMC Bioinf.* **2012**, *13*, S6.
- (32) Oberg, A. L.; Vitek, O. Statistical Design of Quantitative Mass Spectrometry-Based Proteomic Experiments. *J. Proteome Res.* **2009**, *8*, 2144–2156.
- (33) Suomi, T.; Corthals, G. L.; Nevalainen, O. S.; Elo, L. L. Using Peptide-Level Proteomics Data for Detecting Differentially Expressed Proteins. *J. Proteome Res.* **2015**, *14*, 4564–4570.
- (34) Suomi, T.; Elo, L. L. Enhanced Differential Expression Statistics for Data-Independent Acquisition Proteomics. *Sci. Rep.* **2017**, *7*, 5869.
- (35) Choi, M.; Chang, C.-Y.; Clough, T.; Broudy, D.; Killeen, T.; MacLean, B.; Vitek, O. MSstats: An R Package for Statistical Analysis of Quantitative Mass Spectrometry-Based Proteomic Experiments. *Bioinformatics* **2014**, *30*, 2524–2526.
- (36) Goeminne, L. J. E.; Gevaert, K.; Clement, L. Peptide-Level Robust Ridge Regression Improves Estimation, Sensitivity, and Specificity in Data-Dependent Quantitative Label-Free Shotgun Proteomics. *Mol. Cell. Proteomics* **2016**, *15*, 657–668.
- (37) Baeza, J.; Dowell, J. A.; Smallegan, M. J.; Fan, J.; Amador-Noguez, D.; Khan, Z.; Denu, J. M. Stoichiometry of Site-Specific Lysine Acetylation in an Entire Proteome. *J. Biol. Chem.* **2014**, *289*, 21326–21338.
- (38) Riley, N. M.; Coon, J. J. Phosphoproteomics in the Age of Rapid and Deep Proteome Profiling. *Anal. Chem.* **2016**, *88*, 74–94.
- (39) Dittenhafer-Reed, K. E.; Richards, A. L.; Fan, J.; Smallegan, M. J.; Fotuhi Siahpirani, A.; Kemmerer, Z. A.; Prolla, T. A.; Roy, S.; Coon, J. J.; Denu, J. M. SIRT3 Mediates Multi-Tissue Coupling for Metabolic Fuel Switching. *Cell Metab.* **2015**, *21*, 637–646.
- (40) Humphrey, S. J.; Karayel, O.; James, D. E.; Mann, M. High-Throughput and High-Sensitivity Phosphoproteomics with the EasyPhos Platform. *Nat. Protoc.* **2018**, *13*, 1897–1916.
- (41) Bekker-jensen, D. B.; Bernhardt, O. M.; Högberg, A.; Verbeke, L.; Gandhi, T.; Reiter, L.; Program, P.; Sciences, M. *Rapid and Site-Specific Deep Phosphoproteome Profiling by Data-Independent Acquisition (DIA) without the Need for Spectral Libraries*; Nature Publishing Group, 2019.
- (42) Robinson, M. D.; Vitek, O. Benchmarking Comes of Age. *Genome Biol.* **2019**, *20*, 19–20.
- (43) Navarro, P.; Kuharev, J.; Gillet, L. C.; Bernhardt, O. M.; MacLean, B.; Röst, H. L.; Tate, S. A.; Tsou, C.-C.; Reiter, L.; Distler, U.; Rosenberger, G.; Perez-Riverol, Y.; Nesvizhskii, A. I.; Aebersold, R.; Tenzer, S. A Multicenter Study Benchmarks Software Tools for Label-Free Proteome Quantification. *Nat. Biotechnol.* **2016**, *34*, 1130–1136.
- (44) Ramus, C.; Hovasse, A.; Marcellin, M.; Hesse, A.-M.; Mouton-Barbosa, E.; Bouyssié, D.; Vaca, S.; Carapito, C.; Chaoui, K.; Bruley, C.; Garin, J.; Cianférani, S.; Ferro, M.; Van Dorssaeler, A.; Burlet-Schiltz, O.; Schaeffer, C.; Couté, Y.; Gonzalez de Peredo, A. Benchmarking Quantitative Label-Free LC-MS Data Processing Workflows Using a Complex Spiked Proteomic Standard Dataset. *J. Proteomics* **2016**, *132*, 51–62.
- (45) Beasley-Green, A.; Bunk, D.; Rudnick, P.; Kilpatrick, L.; Phinney, K. A Proteomics Performance Standard to Support Measurement Quality in Proteomics. *Proteomics* **2012**, *12*, 923–931.
- (46) Muntel, J.; Kirkpatrick, J.; Bruderer, R.; Huang, T.; Vitek, O.; Ori, A.; Reiter, L. Comparison of Protein Quantification in a Complex Background by DIA and TMT Workflows with Fixed Instrument Time. *J. Proteome Res.* **2019**, *18*, 1340–1351.

- (47) Shalit, T.; Elinger, D.; Savidor, A.; Gabashvili, A.; Levin, Y. MS1-Based Label-Free Proteomics Using a Quadrupole Orbitrap Mass Spectrometer. *J. Proteome Res.* **2015**, *14*, 1979–1986.
- (48) Peters, B.; Brenner, S. E.; Wang, E.; Slonim, D.; Kann, M. G. Putting Benchmarks in Their Rightful Place: The Heart of Computational Biology. *PLoS Comput. Biol.* **2018**, *14*, No. e1006494.
- (49) Wettenhall, J. M.; Smyth, G. K. LimmaGUI: A Graphical User Interface for Linear Modeling of Microarray Data. *Bioinformatics* **2004**, *20*, 3705–3706.
- (50) Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. Roy. Stat. Soc. B Stat. Methodol.* **1995**, *57*, 289–300.
- (51) Storey, J. D. Direct Approach to False Discovery Rates. *J. Roy. Stat. Soc. B Stat. Methodol.* **2002**, *64*, 479–498.
- (52) Rappsilber, J.; Mann, M.; Ishihama, Y. Protocol for Micro-Purification, Enrichment, Pre-Fractionation and Storage of Peptides for Proteomics Using StageTips. *Nat. Protoc.* **2007**, *2*, 1896–1906.
- (53) Cox, J.; Hein, M. Y.; Lubner, C. A.; Paron, I.; Nagaraj, N.; Mann, M. Accurate Proteome-Wide Label-Free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Mol. Cell. Proteomics* **2014**, *13*, 2513–2526.