Single Cell Proteomics Using a Trapped Ion Mobility Time-of-Flight Mass Spectrometer Provides Insight into the Post-translational Modification Landscape of Individual Human Cells

Benjamin C. Orsburn, Yuting Yuan & Namandjé N. Bumpus

Department of Pharmacology and Molecular Sciences, Johns Hopkins University; Baltimore, MD, USA, 21205

Corresponding:
Benjamin C. Orsburn
The Johns Hopkins University School of Medicine
Department of Pharmacology and Molecular Sciences
725 North Wolfe Street
Baltimore, MD 21205
Phone: 717-638-8726

Email: borsbur1@jhmi.edu
ORCID: 0000-0002-0774-3750

Namandjé N. Bumpus
The Johns Hopkins University School of Medicine
Department of Pharmacology and Molecular Sciences
Physiology 312
725 North Wolfe Street
Baltimore, MD 21205
Phone: (410)955-0562

Email: nbumpus1@jhmi.edu
ORCID: 0000-0002-6096-3824

Author Contributions:
Conceptualization: NNB, BCO
Methodology: BCO
Investigation: NNB, BCO, YY
Funding acquisition: NNB
Writing – original draft: BCO
Writing – review & editing: NNB, YY, BCO
Competing Interest Statement: We have no competing interests to disclose.

Keywords: Single cell proteomics, pasefRiQ, $KRAS^{G12C}$, post-translational modifications, phosphorylation, acetylation

Abstract
Single cell proteomics is a powerful tool with potential for markedly enhancing understanding of cellular processes. Previously reported single cell proteomics innovations employ Orbitrap mass spectrometers. In this study we describe the development, optimization, and application of multiplexed single cell proteomics to the analysis of human-derived cells using trapped ion mobility time-of-flight mass spectrometry. This method, denoted as pasefRiQ is an advance as it allows accurate peptide quantification as demonstrated by a two-proteome standard model, with little variation in accuracy in samples at picogram peptide concentrations. When employing a peptide carrier channel to boost protein sequence coverage, we obtain over 40,000 tandem mass spectra in 30 minutes, providing unprecedented sequence coverage of each identified protein. Using NCI-H-358 cells, which are a human bronchioalveolar carcinoma and $KRAS^{G12C}$ model cell line, we demonstrate that the level of coverage achieved using this method enables the quantification of up to 1,255 proteins per cell and the detection of multiple classes of post-translational modifications in single cells. Further, when the cells were treated with AMG510, a $KRAS^{G12C}$ covalent inhibitor, pasefRiQ revealed cell-to-cell variability in the impact of the drug on the NCI-H-358 proteome. In fact, when examining proteins that changed in protein abundance in response to AMG510 we can identify those that are disproportionately increased or decreased in a small number of cells relative to the total population of cells that were cultured and treated simultaneously. These results suggest single cell proteomics through pasefRiQ can provide powerful insight for cellular and drug mechanism studies.

Summary
This work describes the establishment of a single cell proteomics method using a time-of-flight mass spectrometer. Through this approach, we demonstrate the confident identification of post-translational modifications in single human-derived cells. Additionally, using a $KRAS^{G12C}$ covalent inhibitor as a model compound we show that this method can be used to understand pharmacological responses of single human-derived cultured cells.

Main Text

Introduction

Single cell RNA sequencing has markedly advanced understanding of biology at the level of individual cells.[1] While an unquestionably powerful tool, a major limitation of all RNA based technology is the lack of correlation between transcript abundance and protein abundance in human systems.[2] In addition, cellular function is often influenced by protein processing events such as proteolytic cleavage or post translational modifications (PTMs).[3] As such, the direct analysis of the protein themselves in single cells is a promising avenue. [4]

Single cell proteomics using liquid chromatography mass spectrometry (LCMS) is a newly emerging field led by parallel improvements in both instrumentation and methodology.[5] The majority of studies published to date have focused on the essential method development and proof of concept work necessary to set the stage for applications of the technology.[6,7] The most promising biological works described to date have utilized multiplexing reagents to obtain quantitative proteomic data on multiple cells in each single LCMS experiment. Multiplexing has several advantages, most notably allowing enough cells to be analyzed per study for sufficient statistical power to draw biological conclusions. In recent works, independent teams using these approaches have demonstrated the ability to study heterogeneity in macrophage differentiation [8] and to identify proteome level differences between individual cancer cells of disparate origins.[9]

Today single cell proteomics has demonstrated the ability to quantify hundreds of proteins per cell, largely driven by quantifying a relatively small number of peptides per protein.[10] While accurate

quantification of proteins can be derived from measurements of individual peptides, higher sequence coverage is required for the identification of many protein features. For example, protein post-translational modifications such as phosphorylation and acetylation are only detected in proteomics studies where high relative sequence coverage is obtained or offline chemical enrichment is performed.[11]

To date, all multiplex single cell proteomics studies have utilized various iterations of hybrid Orbitrap mass spectrometers (MS). Orbitraps are popular MS systems due to their relatively high mass accuracy and resolution, characteristics that are largely obtained at the consequence of relative scan acquisition rate compared to other mass analyzers. In contrast, Time of Flight (TOF) systems are characterized by higher relative scan acquisition rates, and correspondingly lower resolution and mass accuracy. New advances in ion accumulation prior to TOF have successfully circumvented these traditional limitations allowing new equilibriums between sensitivity and speed to be leveraged in LCMS workflows. In this study we explore the capabilities of a third generation trapped ion mobility time of flight mass spectrometer (TIMSTOF Flex) for sensitive and accurate multiplexed proteomics through a combination of parallel accumulation serial fragmentation for reporter ion quantification (pasefRiQ). As an application of pasefRiQ, we describe the analysis of single cells from the $KRAS^{G12C}$ model lung cancer cell line NCI-H-358 (H358). We find that the number of proteins quantified in individual cells are sufficient to cluster each cell by their relative cell cycle stage. Importantly, we describe the first observations of multiple classes of protein PTMs by LCMS in single human cells and find that the quantification values for PTMs of well-studied proteins correlate with other phenotypic observations.

To explore the application of single cell proteomics in the context of molecular pharmacology, we treated cells with the $KRAS^{G12C}$ covalent inhibitor, AMG510.[12] We find that when the proteomes of individual cells are handled as biological replicates during data analysis, AMG510 treatment largely mimics the effects observed in studies based on the proteomics of bulk cell homogenates. Single cell proteomics provides additional insight into these systems by allowing us to directly elucidate the cell-to-cell heterogeneity in response to inhibitor treatment. With this additional data we find evidence that the proteins displaying the largest differential response to AMG510 are disproportionately impacted in a relatively small number of individual cells. Taken together, these results demonstrate a powerful role for single cell proteomics in understanding cell-to-cell variability including in drug response.

Results and Discussion

Practical intrascan linear dynamic of pasefRiQ approaches 3-orders

To evaluate the practical dynamic range of pasefRiQ we prepared a 4-order dilution series of a commercial K562 cell line digest with TMTPro 9-plex reagent. As shown in **Fig. 1A** the mean reporter signal for each channel from all identified peptides maintained near linearity across the entire dilution series, with an R^2 of 0.982. While the number of missing reporter ions increases as the peptide concentration decreases, we do not observe a marked increase in the number of missing reporter values until we exceed an intrascan dilution of three orders. In the reporter ion channel that contained 200 picograms of K562 digest standard compared to a channel of 200 nanograms, the number of reporter ions detected decreased by 53.2%. Surprisingly, not all reporter ion signal is lost at even 10-fold below this level as 10.2% of peptides contained reporter ion signal at a level approximating 10 picograms on column **(Fig. 1B).** These data suggest that further investigation is warranted as pasefRiQ may not demonstrate the same limitations of intrascan dynamic range and subsequent ratio distortions recently described as the “carrier proteome effect” as Orbitrap systems.[13]

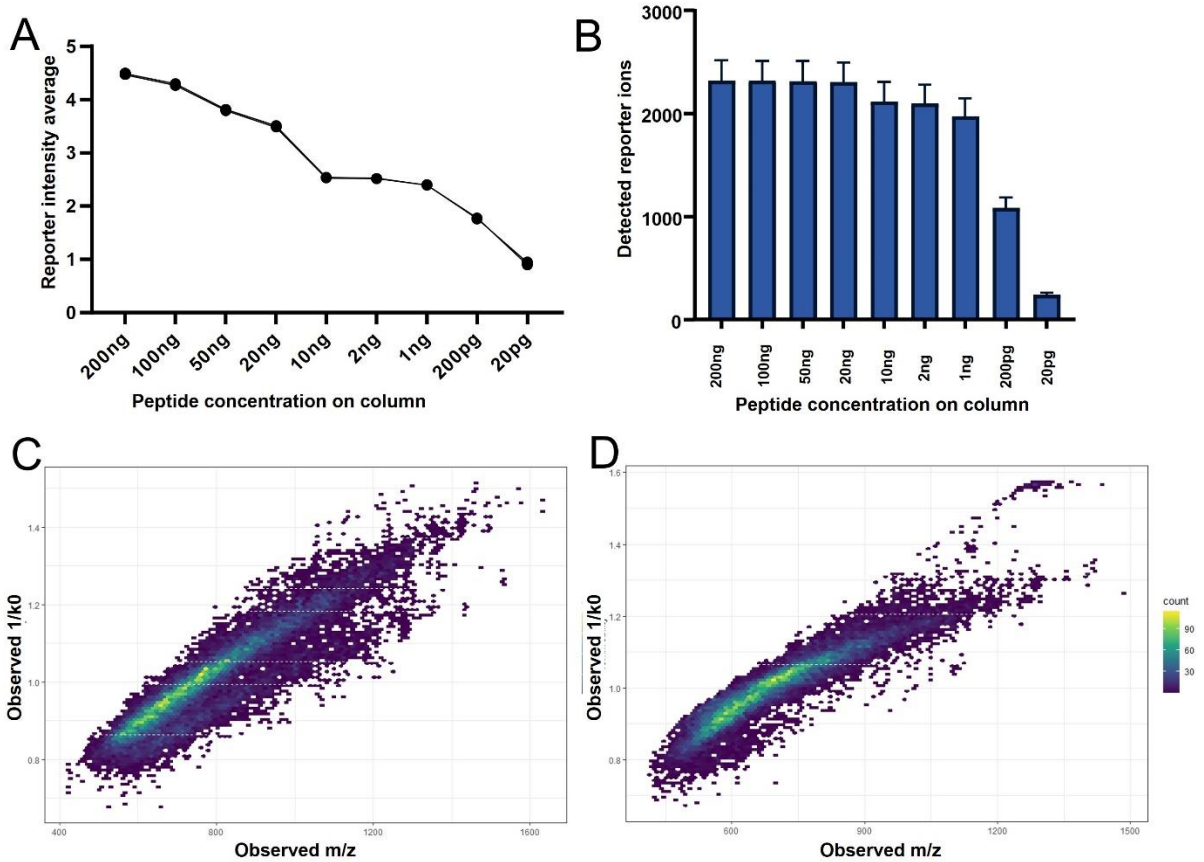


Fig. 1. A. The log10 converted average intensity of each reporter ion in a TMTPro 9-plex dilution series. **B.** A comparison of the number of detected reporter ions at each concentration to evaluate the number of relative missing values across the dilution series. **C.** A plot of the distribution of m/z and 1/k0 values for unlabeled peptides**. C.** A plot of the same concentration of sample of peptides labeled with TMTPro reagent.

Ion mobility optimization reduces co-isolation interference

To evaluate the level of background interference in pasefRiQ we utilized a well-characterized yeast triple knock out TMT standard (TMT-TKO). [14] In this standard, biological replicates of a parent yeast strain and three separate transgenic strains with single genes removed are labeled with three TMT 11-plex labels. The TMT-TKO standard was analyzed with pasefRiQ using a consistent 60-minute LC gradient with adjustments made to the quadrupole isolation and TIMS ramp settings in each iteration of the respective methods. Reductions in the quadrupole isolation widths had minimal effects on the number of proteins and peptides identified per run, with the exception of a 1Da symmetrical isolation width which resulted in an approximate 18% loss in identified peptides identifications. We therefore chose to use the minimum quadrupole isolation width that did not lead to a loss in peptide identifications, which was a 1.5 Da symmetrical isolation. (**S. Data. 1**) In order to determine the appropriate ion mobility settings, we plotted the observed 1/k0 values and m/z for all identified peptides from a K562 peptide mixture both unlabeled and labeled with TMTPro reagent. As shown in **Fig. 1C and 1D**, the majority of peptide signal is observed within a relatively narrow 1/k0 region. Peptides labeled with the TMTPro reagent exhibit a less symmetrical distribution than unlabeled peptides. By targeting the ion mobility range within the region of 0.8 -

1.3 1/k0 we obtain the highest level of improvement in co-isolation interference when using the TMT-TKO standard.

In order to multiplex more than 10 samples with commercially available reagents today, isobaric reporter reagents with alternating N^{15} and C^{13} isotopes must be used. The neutron mass discrepancy in these two isotopes leads to a separation of m/z of approximately 0.007 amu. To fully resolve these reporter ions, current generation Orbitrap instruments are equipped with an optimized resolution of 45,000 at an m/z of 200. Orbitrap systems operating at this resolution obtain substantially fewer MS/MS scans per experiment than typical label free experiments which obtain MS/MS scans at the much faster scans of approximately 15,000 resolution.[15] The ability to multiplex up to 18 separate samples simultaneously is an attractive return on this loss in data acquisition rate.[16]

The resolution of an Orbitrap mass analyzer is a function of the ion m/z with lower masses having higher resolutions within the same scan than ions with higher relative m/z.[17,18] In contrast, time of flight analyzers demonstrate comparable resolution throughout the mass range.[19] During the calibration and tuning process, we can obtain estimates on the TIMSTOF Flex mass resolution, which routinely achieves 40,000 at 1222 m/z. To determine the capacity of a TIMSTOF Flex to achieve higher multiplexing, we prepared commercially available human tryptic digests while using all 16 reagents following a manual calibration of the TOF resolution. As shown in **S. Fig. 2** we can obtain nearly complete baseline separation of the 127n and 127c reporter regions. Similarly, when evaluating the TMT-TKO standard following manual calibration of the flight tube we observe a decrease in protein signal in both the Δhis4 and Δura2 channels. However, closer evaluation of the reporter masses indicated the mass accuracy of the instrument in the low mass region was a clear limiting factor in the size of the integration windows utilized.

Evaluation of quantitative accuracy with a two-proteome labeled 9-plex standard

The use of two-proteome standards is a well-established method in LCMS based proteomics for the evaluation of quantitative accuracy.[20,21] To evaluate the accuracy of pasefRiQ we prepared 9 samples containing an identical concentration of a K562 commercial tryptic digest with an *E. coli* commercial tryptic digest at three separate concentrations. The samples were prepared in triplicate to evaluate the reproducibility within the experiment. The *E. coli* digest concentrations used are in a repeating dilution of 1:5:10. A minimum of 3 technical replicates were acquired using the optimized parameters obtained from the yeast digest standard experiments (**S. Data 1**). As shown in **S. Fig. 3A,** we observed ratios closely matched the expected ratios in pasefRiQ for proteins with high relative sequence coverage such as the TNAa protein. These results extend to the average intensities of all detected *E. coli* proteins in the study (**S. Fig. 3B**) and in both cases, the observed ratios more closely matched expected in pasefRiQ than a quadrupole Orbitrap system in our lab when analyzing the same sample and similar chromatography with an MS2 based approach.

Post-acquisition mass recalibration further improves isolation interference

To evaluate the effects of the reporter ion mass accuracy on quantification we developed an offline manual calibration tool, the *pasefRiQCalibrator*, that writes a new MGF file following a manual calibration adjustment across a user- specified low mass range. Following manual evaluation of multiple reporter ion masses, we can determine an appropriate adjustment factor for each file. Prior to manual calibration the most effective reporter ion isolation window for pasefRiQ files described in this study is approximately 50 ppm falling within previously reported mass accuracy estimations for the TIMSTOF.[22] Following manual adjustment, we can reprocess the same files using a 20 ppm mass tolerance window with no loss in reporter ions quantified, and a reduction in the mean signal intensity of the observed Δmet6 reporter ion of 56.4% (**S. Fig. 4)**. Remarkably, despite the lack of complete baseline resolution of the reporter ion regions in the Δhis4 and Δura2 channels, we observe clear downregulation in these channels following recalibration (**S. Fig. 5**)

While this is a promising development, the lack of baseline separation of reporter ions has been shown by others to result in the loss of accuracy in protein ratios with less extreme quantitative differences.[23] For this reason, we chose not to perform higher multiplexed quantification with pasefRiQ with this iteration of the commercial TIMSTOF hardware.

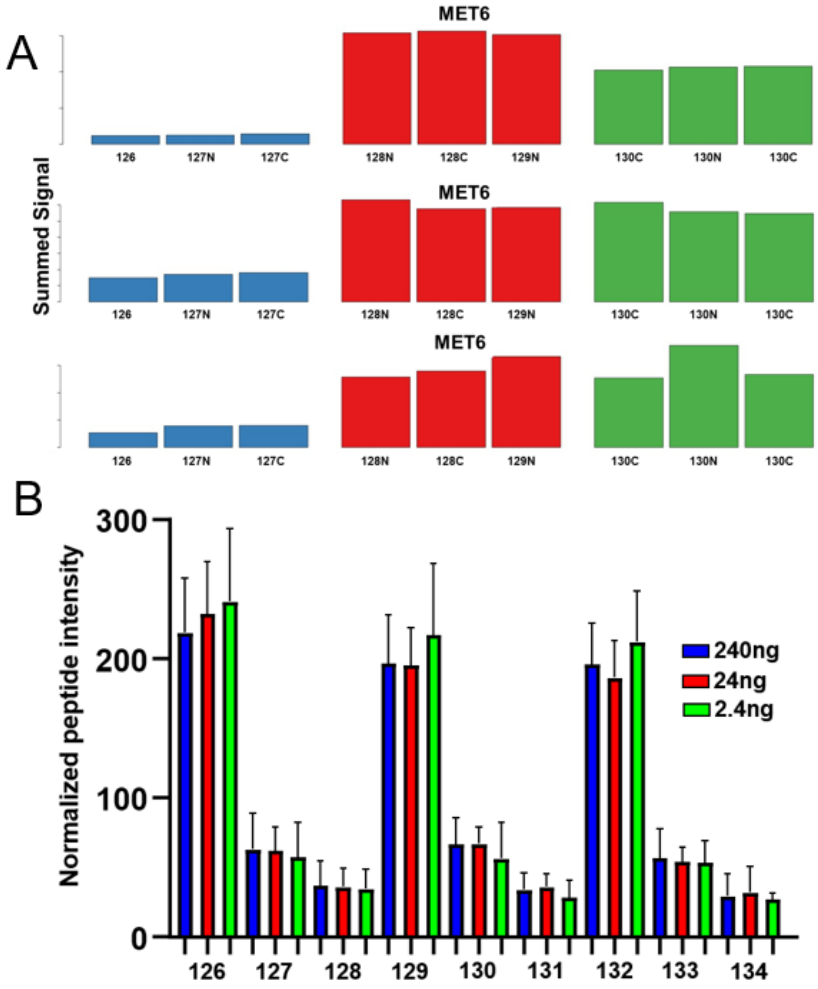


Fig. 2. Analytical figures of merit for pasefRiQ. A. A comparison of results from 3 instruments that following analysis of 200ng of the TMT-TKO yeast digest standard and visualized using the TVT web tool.[24] (Top) vendor example from an Orbitrap system using MS3 based quantification. (Middle) vendor example results from a quadrupole Orbitrap system using MS2 based quantification. (Bottom) Results from a pasefRiQ file. **B.** A graph demonstrating the normalized abundance of the *E. coli* protein TNAa in the two-proteome standard injected at different concentrations on column to illustrate the effects of sample dilution on ratio accuracy.

Quantitative accuracy of pasefRiQ is maintained at single cell relevant concentrations

To determine the relevant biological limits of detection of pasefRiQ we prepared serial dilutions of the two-proteome standards from 240 ng to 2.4 ng on column. While the number of peptides and proteins identified exclusively from MS/MS spectra at each subsequent dilution level decreased (**S.Data 1**), we observed no decrease in quantitative accuracy. The expected ratio of the diluted *E. coli* proteins in this standard is 1:5:10 with 3 intraexperiment technical replicates across the 9

channels. When averaging the ratios from three separate LCMS experiments, the most accurate ratios for the TNAa protein observed were from the 2.4ng injections which returned mean ratios of 1: 3.98: 7.43. The least accurate ratios were observed for 240ng injections on column with mean ratios of 1: 3.28: 6.09 (**Fig. 2B**). These results indicate that values observed from pasefRiQ at picogram levels of peptide load per channel can return reliable quantification values.

Over 1,000 proteins can be identified in single human cells using 30-minute LC gradients

To evaluate the capabilities of pasefRiQ for the analysis of single human cancer cells, individual NCI-H-358 (H358) cells were aliquoted by flow-based sorting into 96-well plates. The first well in each row was loaded with buffer with no cell by the core facility operator to provide a channel for intraexperiment estimation of background interference. The cells were lysed and labeled by hand using the SCOPE2 methodology. A bulk cell lysate of H358 cells labeled with channel 134 at an approximate concentration of 25 ng was used as the carrier channel to both resuspend and pool the single cell peptides as well as to increase the number of peptides triggered for fragmentation. Orbitrap based methods can multiplex 18 channels simultaneously while pasefRiQ is limited to a maximum of 10 channels cells per run. To maintain a comparable daily throughput of single cells with pasefRiQ we chose to use 30-minute gradients which are approximately one half the time of previous studies when including the 15 minute equilibration and reinjection time required by our HPLC system.

To characterize the relative performance of pasefRiQ, previously published files were obtained and processed with a well-established commercial software package using the same search engine, false discovery rate estimation tools and protein database. A summary of the results of representative samples from these studies are presented in **Table 1,** and as **S. Data. 2.** As shown, we can obtain comparable numbers of protein identifications to previously reported methods in approximately one-third of the total MS instrument acquisition time. In addition, we observe approximately 5x more fragmentation spectra per experiment and a corresponding increase in the number of peptides per experiment, leading to an overall increase in total sequence coverage for all identified proteins (**S. Data. 2**)

Study	MS2 scans	Peptides	Proteins per cell	Max # of cells	MS Time
This study	42555	8515	1044	6	30 min
SCOPE2	8418	2450	599	9	100 min
nPOP	8248	1822	775	14	105 min

Table 1. A comparison of pasefRiQ single cell results to Orbitrap MS2 based publicly available data

Analysis of 443 single cells in 50 hours of total instrument time

To further expand on this study, proteomic data was acquired on a total of 443 single H-358 cells using approximately 50 hours of LCMS instrument time, including all controls and inclusive of the 15 minutes of HPLC equilibration time between experiments. A full summary of the spectra, peptides and proteins identified in this study are provided as (**S. Data. 3**). Complete summary reports of all identifications are provided in the publicly available data that contain all peptides, proteins and quantification data from these cells and text summaries of one analysis are presented as **S. Data. 4.**

We observed considerable signal inflation in the 133n labeled single cell channel in all runs. The level of signal inflation appeared to be proportional to the level of labeling impurity of the 134-loading channel affecting 133n, as reported by the manufacturer. As such, we chose to exclude all 133n labeled single cells from downstream analysis. Following the removal of two LCMS runs which did not pass our automated quality control pipeline (data to be presented elsewhere) due to an

overall lack of signal, a total of 366 cells single H358 cells were considered for downstream analysis. In addition, the visual analysis of the raw log10 converted intensity values provided a valuable metric for interrogation of data quality (**S. Fig. 7B**).

The resulting files were processed with multiple well established search tools to obtain the most comprehensive interpretation of protein identifications and quantifications. The most conservative search tool, MaxQuant, identified approximately 1,600 protein groups at an average of 655.2 quantifiable proteins per cell with a maximum of 1,255 proteins per single cell. The number of proteins identified in this cell line is comparable to previous described iterations of the SCoPE-MS workflow reported by others when using cancer cell lines for analysis.[5,25]. All other search tools returned a higher number of peptide and protein identifications. A summary of the results from these tools are provided in **S. Data 3**. For all downstream interpretation, we will focus on the results obtained from iterative approach through Proteome Discoverer as supporting evidence for reporter ions, peptide sequence coverage and diagnostic fragment ions can be easily visualized for each peptide and protein identified.[27]

As described by others [28] unsupervised clustering of the proteomes of individual cells and comparisons of those clusters largely implicate cell cycle status as a major differentiating factor of the proteome. As expected, due to the flow cytometry gates employed to obtain living cells within particular size constraints, we observe a limited range of the overall cell cycle, with H358 cells clustering into two main groups (**Fig. 3B**) along cell cycle dependent lines largely driven by the abundance of ribosomal subunit proteins (**Fig. 3C**).

We have previously described a publicly available web application for the estimation and filtering of proteomics data against libraries of established cellular protein copy numbers.[29] Using this tool we find that proteins identified in H358 single cells closely track to the theoretical copy number for each individual protein. Over 95% of all proteins with quantifiable signal from individual cells in this study are from proteins estimated to possess more than 100,000 copies per cell. The median copy number for a protein identified in a single cell in this study is approximately 630,000. In contrast, the median calculated copy number of all proteins in this cell line is approximately 16,000 copies (**Fig. 3D**).

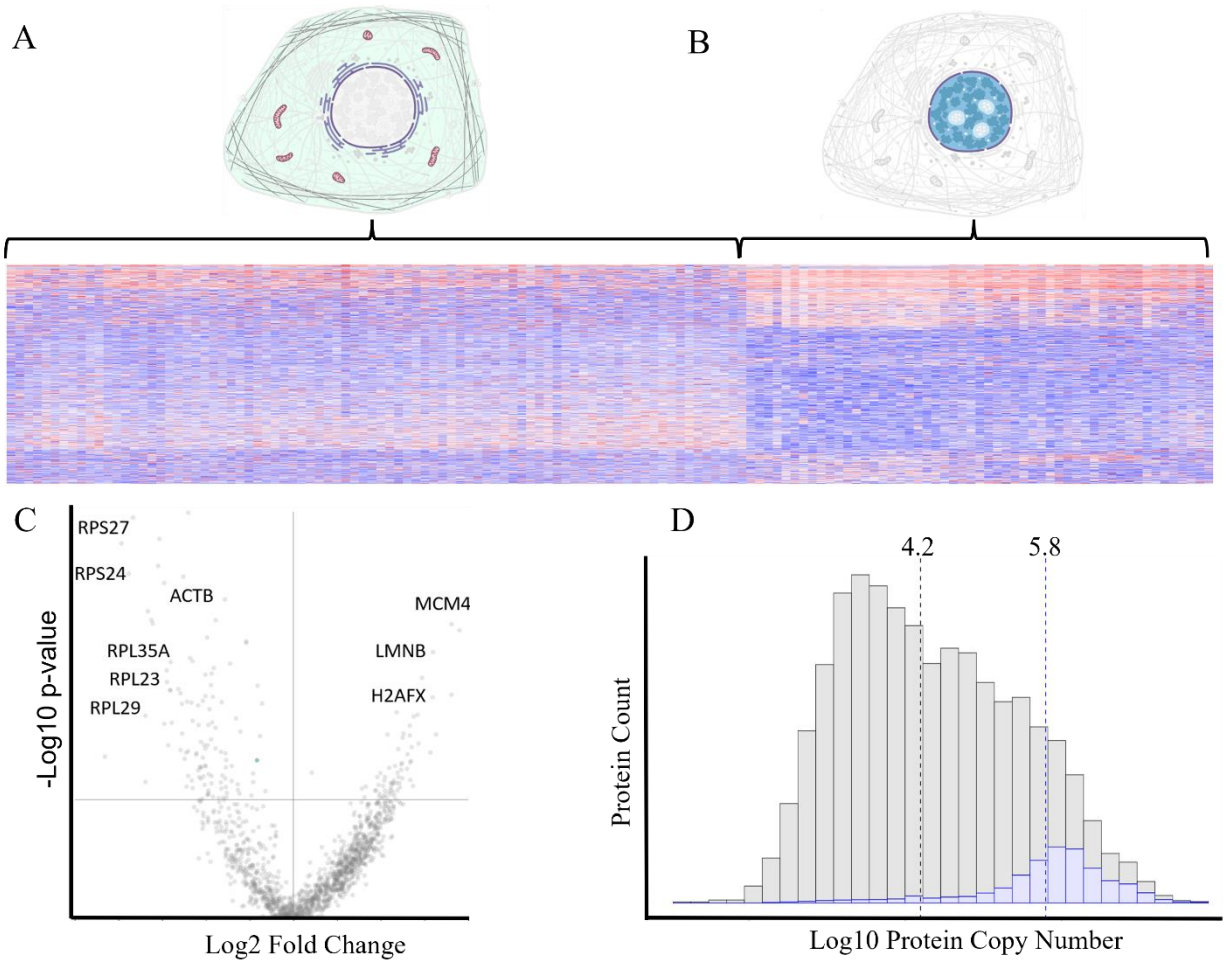


Fig. 3. Characteristics of proteins quantified in single H358 cells. A. Single cells cluster into two distinct groups based on cell cycle status with the largest group demonstrating relative upregulation of central metabolism and protein translation. **B**. A second cluster of cells is best characterized by upregulation of mitotic machinery and nuclear function. **C**. A heatmap demonstrating the strongest changes when single cells are clustered in two distinct groups with ribosomal subunits and nuclear proteins largely driving clustering. **D**. A histogram demonstrating the calculated log10 copy number for 14,178 proteins (grey) compared to the proteins identified in single cells in this study (blue). Dotted lines indicate the median for each group.

Protein post-translational modifications can be identified in single human cells

A large number of the unmatched MS/MS spectra in any proteomics workflow can be attributed to post-translational modifications (PTMs) that were not included in the data analysis.[30,31] To date, no single cell study using mass spectrometry has described the identification of PTMs. To evaluate if the increase in relative sequencing information in pasefRiQ could provide evidence of PTMs, we reanalyzed these files to search for phosphorylation on serine and threonine as well as acetylation on lysine residues. A summary of consensus proteins, phosphopeptides and acetylated peptides are provided as **S. Data. 4**.

The most common phosphorylated protein observed in single cells was prelaminin (LMNA), which was identified as phosphorylated in 153 of the cells analyzed in the total course of the study. The

most common sites observed on the protein are well-characterized serine sites Ser22, Ser390, Ser392, Ser404, and Ser406 which were observed in 114 individual cells in total with other sites on this highly modified protein distributed among observation in other cells. Sufficient fragmentation data was obtained for both confident identification and localization of the site of phosphorylation in peptides where multiple sites were possible. (**Fig. 4A and Extended Data 1**). In addition, for the most abundant phosphopeptides clear reporter ion signal can be observed for multiple single cells in each spectrum (**Fig. 4B**). Other phosphoproteins that were observed in a high number of cells were sites on the high cellular abundance proteins Titin, Albumin and Myosin-10 with phosphopeptides observed in 101, 97, and 78 cells, respectively. In addition, we observe phosphorylations on multiple proteins in the MAPK pathway that are well established as highly expressed in $KRAS^{G12C}$ mutant cell lines.[32]. When compared to the calculated copy number, phosphoproteins were found to have a median abundance of approximately 600,000 copies (**Fig. 4C**), a value only 25% higher than the median for identification of proteins in single cells. However, a clear difference was found when comparing sequence coverage observed for each protein from the two groups. Phosphopeptides were identified on proteins with approximately 10-fold more supporting peptide spectral matches than identified proteins as a whole demonstrating the necessity in protein sequence coverage in making these identifications (**Fig. 4D**).

Lysine acetylation is a post-translational modification with regulatory importance in a variety of cellular systems.[33] Notably, lysine acetylation plays a central role in the regulation of nuclear histone proteins.[34] We observed 23 high confidence acetylation sites in single cells, primarily localized on histone sequences and on actin polymerizing proteins, Thymosins 4 and 10. **Fig. 5** is an example of one such site from Histone 2.2. Peptides with an acetylated lysine commonly produce a diagnostic fragment ion during collision induced dissociation with a mass of 126.0913. Mass spectrometers with lower relative resolution or mass accuracy may not be able to accurately discern the 0.0364, or 29ppm mass difference between this diagnostic ion and the 126 TMT reporter ion. As shown in **Fig. 5C** we can confidently extract reporter ion signal from each H358 cell analyzed within that spectrum and observe no reporter ion signal from the 126 method blank control well. In addition, we can clearly discern a lysine acetylation diagnostic ion further supporting this peptide identification.

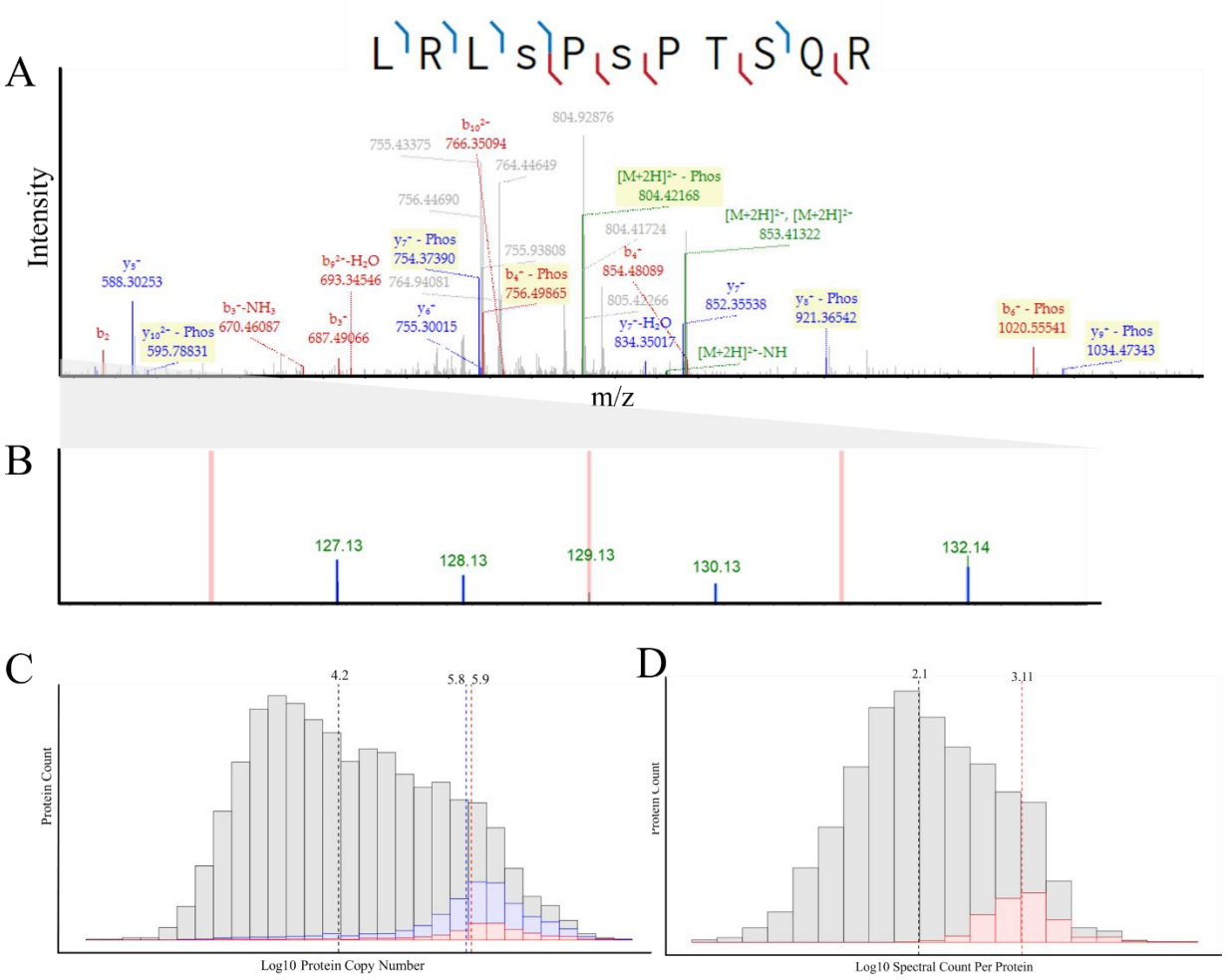


Fig. 4. Phosphopeptides and proteins from single cells. A. A fragmentation pattern for LMNA phosphopeptide with sufficient fragmentation data for confident site localization. **B**. Zoomed in reporter region for this phosphopeptide demonstrating a lack of reporter signal in the method blank region and quantifiable signal from multiple single cells. **C.** A histogram with all proteins identified in single cells (blue) and phosphoproteins (red) compared against the calculated copy number of all proteins with median values highlighted. **D.** A histogram of the displaying the number of peptide spectral matches for proteins identified in this study (grey) compared to the same of proteins where phosphopeptides were confidently identified. Dotted lines represent the median for each set of values.

Post-translational modifications are exclusively found on proteins with the highest intracellular copy numbers

The number of times peptides from a single protein are fragmented in an LCMS experiment, or spectral count, is a well-established function of the relative abundance of a protein in a given sample.[35] To determine spectral counts, the number of MS/MS spectra from all runs combined that can be attributed to a single protein or peptide spectral match (PSM) are typically normalized against the molecular weight or theoretically observable peptides from a protein.[36] We used the number of PSMs, percent sequence coverage and calculated theoretical protein copy numbers to help develop a profile of proteins where PTMs were discovered. The median percent coverage of

all protein in this study was 18.51%, resulting from 125.8 PSMs with quantitative reporter ions from single cells for each protein.

When directly comparing the distribution of spectral counts for all proteins identified in this study versus the spectral counts of all phosphoproteins, we find a 10-fold increase in the median of the latter (**Fig. 4D**). These results suggest that approximately 10 times more spectral data is required to confidently identify a single phosphorylation site in a protein than is required to confidently quantify an unmodified protein. The relationship between protein copy number, spectral counts and confident phosphorylation site location is highlighted by the multiple sites observed on PLEC and AHNAK. Over 1.9% of all MS/MS spectra acquired in this study (51,075/2,599,602) can be attributed to the 531 kDa PLEC cytoskeletal protein. The 628.7 kDa AHNAK protein comprised 1.3% of all total MS/MS spectra in this study. The 34,311 PSMs allow over 76% sequence coverage of this protein with quantifiable signal observed in nearly every cell in the study.

Similarly, the observed acetylation site on the relatively small protein Histone 2.2 is only one site of over 2,200 separate MS/MS spectra matched to this protein sequence. When considering the number of peptides identified compared to the length of the protein, Histone 2.2 likely has an even higher cellular abundance than PLEC that is 50 times larger, but only has 25-times more spectra identified. When compared to the calculated copy numbers of all proteins in a cell, acetylation sites were identified on proteins with a median log copy number of 6.38, or approximately 2.2×10^6 copies per cell, representing proteins in the top 2.6% of total abundance (**Fig. 5D**). The value of total protein sequence coverage is also apparent in the identification of acetylated proteins in single cells. The median total protein sequence coverage for a protein with a confidently identified acetylation site is 44.62%, compared to a median of 18.51% for all proteins identified in this study (**Fig 5E**).

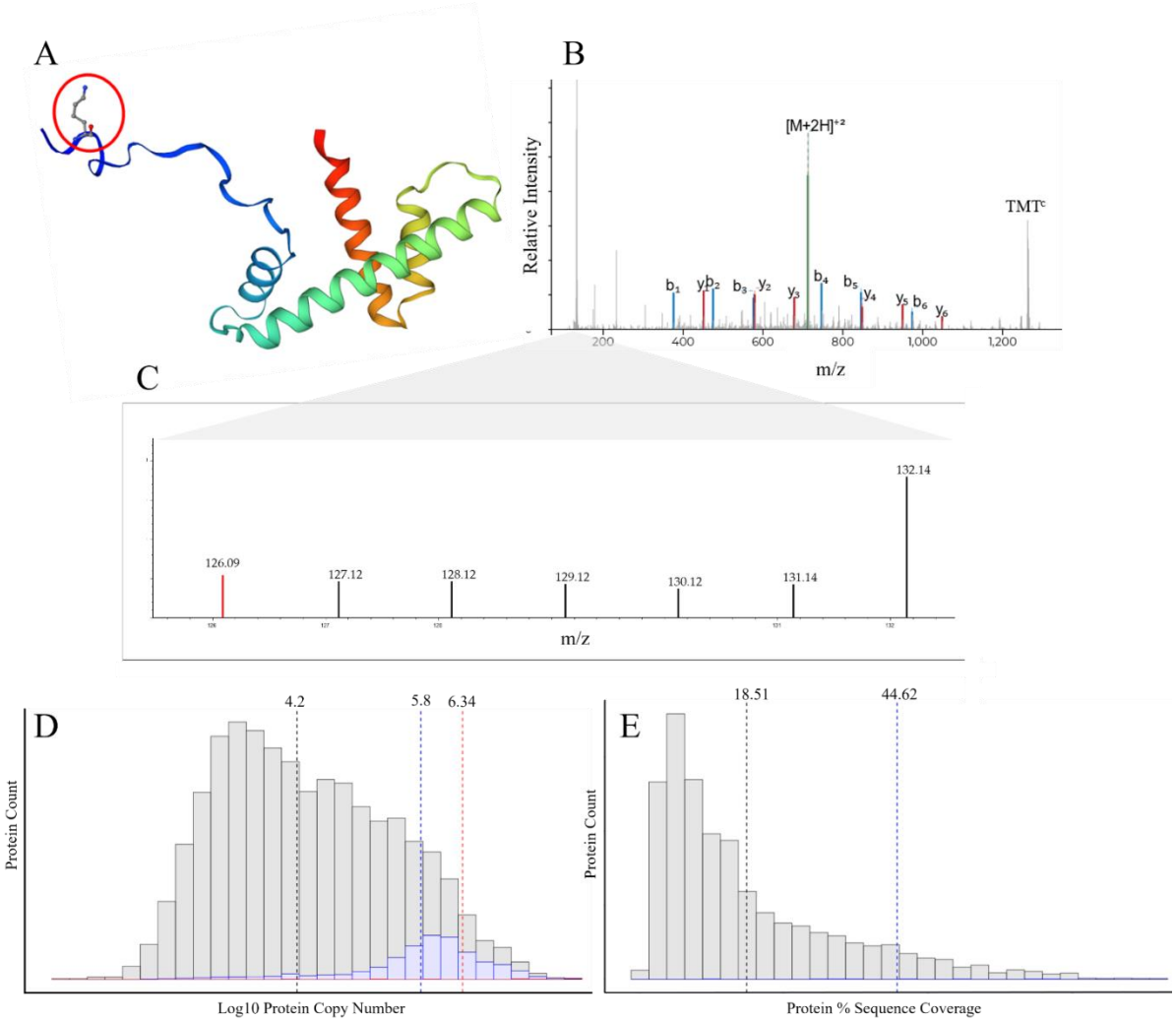


Fig. 5. Acetylated proteins detected in single H358 cells. A. The Histone 2.2 protein with annotated acetylation site highlighted. **B.** A fragmentation spectra for this acetylation site demonstrating complete sequence coverage for identification and localization of this site. **C.** A zoomed in reporter ion region demonstrating signal for this peptide is apparent in every individual cell in this spectrum. The 126.09 highlighted in red corresponds to a diagnostic ion for an acetylated lysine residue. **D.** A histogram illustrating the relative copy numbers of all proteins in this study (grey), all proteins found in single cells (blue) and acetylated proteins (red). Dashed lines indicate median values. **E.** A histogram displaying a comparison between the total protein sequence coverage of all proteins identified in single cells (grey) and the median sequence coverage of acetylated proteins (blue line).

Mitotic phosphopeptide abundance correlates with single cells expressing other mitosis markers

As demonstrated by others, unsupervised clustering of single H358 cells based on quantitative proteomic data was largely driven by proteins involved in cell cycle status.[37] Phosphorylations on nuclear laminin protein LMNA were identified in 42% of cells in this study. The distribution of those phosphorylation sites, however, were dispersed among at least 6 distinct sites including sites with well-characterized roles in nuclear division. Phosphorylation of Ser22 and Ser392 were detected

in 33 and 22 individual cells, respectively. These “mitotic sites” are essential for LMNA localization and actively promote depolymerization of the intact nuclear lamina to allow nuclear division. In contrast, phosphorylation of Ser390, Ser 404 and Ser406 largely considered later stage events, with Ser390 demonstrating high turn-over rates comparable to Ser22 and Ser392.[38] Proteins quantified in single cells in this study with annotated involvement in specific cell cycle stages were flagged and their relative abundance values were extracted. Pearson correlation coefficients were calculated using the relative abundance of each LMNA phosphopeptide in each cell against the abundance of each known cell cycle protein (**Fig. 6A**, **S. Data. 5**). Ser22 phosphorylation showed a strong correlation to multiple proteins involved in the G2/M transition. Likewise, Ser392 phosphorylation was found to possess a similarly strong correlation coefficient to many of these same proteins. In addition, Ser392 phosphorylation was also strongly linked to later stage mitotic events and demonstrated a strong negative correlation to ZNF259. Ser390 phosphorylation showed no correlation to mitotic protein abundance with the exception of a strong negative correlation to TUBB4B. Ser390 phosphorylation only demonstrated strong correlation to a single mitosis protein, the regulator of ploidy, LATS1 (r =0.9980) with a weaker correlation to and RAB8A (r = 0.78).

We extended this correlation analysis of these three phosphorylation sites to all proteins quantified in H358 single cells **(Fig. 6).** Phosphorylation on Ser22 and Ser392 showed strong correlations to ribosomal proteins and proteins in central metabolism in addition to mitotic proteins. In contrast, Ser390 phosphorylation demonstrated no meaningful level of correlation to any ribosomal protein. Phosphorylation of this site most strongly correlated with EIF2AK2 and PML, which are known to be involved in the shutdown process of protein translation. Additional proteins of high correlation were mitochondrial oxidative stress proteins, PPIF and UQCRQ and PEX14, which has a recently described role of protecting DNA upon nuclear breakdown at mitosis.[39]. Additionally, phosphorylation of Ser390 demonstrated a strong negative correlation to central metabolism proteins such as pyruvate dehydrogenase and S100P and mRNA transport proteins such as FXR1.

Ser22 phosphorylation demonstrated negative correlations with histone H2AX, MCM7 and PSMA1. Phosphorylations on LMNA mitotic sites Ser22 and Ser 392 were strongly correlative with other mitotic proteins as well as with central metabolism and translation, through both multiple components of ribosomal subunits and elongation factors (**Fig. 6B, 6C**). In later stages of nuclear disassembly, Ser22 phosphorylation is no longer necessary, suggesting that the role it plays is primarily in early-stage nuclear transport. Ser390 phosphorylation occurs only during nuclear division when all DNA is exposed, while the cell is minimizing metabolism and upregulating protection mechanism to prevent DNA damage. The prevalence of Ser392 phosphorylation in combination with other LMNA phosphorylations occur from the G2 through G1 transition indicating that it plays roles throughout the entire mitotic process.

Other phosphorylations of this protein were observed at high confidence but their overall mechanisms are less well established. These results lend support to both our identification and to the value that post-translational modifications can play in elucidating the intracellular environment in single cells. While considerable work has been performed using molecular biology approaches to the study of LMNA phosphorylation, we can find no published accounts detailing the interplay of these phosphorylation sites at the single cell level. LMNA dysregulation has been linked to a number of diseases, including a well-established role in cardiomyopathies and progeria [38] as well as recent evidence suggesting a role of Ser404 in Emery-Dreifuss muscular dystrophy.[40] Therefore, despite the clear requirements for extremely high intracellular protein copy number for PTM identification in single cells, we find there is much to be learned from these data.

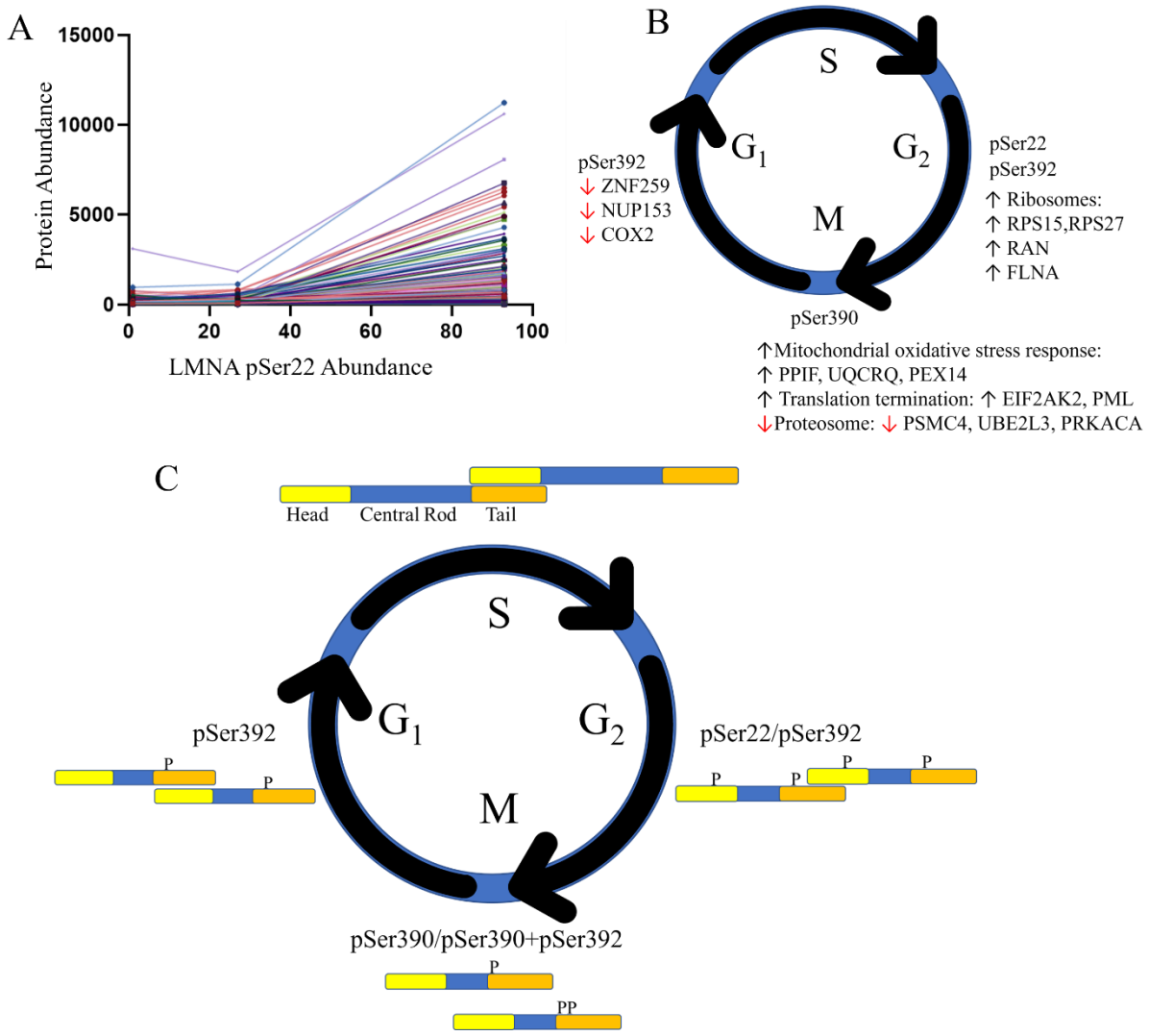


Fig. 6. Correlation analysis of LMNA phosphopeptides to cell cycle markers. A. Summary plot of correlation analysis for pSer22 to all other proteins identified in single cells. **B.** A representation of the cell cycle status of proteins most strongly correlating with each phosphorylation site on LMNA. **C.** A cartoon illustrating our interpretation of cell cycle status with LMNA phosphorylation and subsequent polymerization status

Single cells treated with a KRAS^{G12C} covalent inhibitor exhibit expected phenotypes while allowing additional insight into the heterogeneity of drug response

To explore the power of single cell proteomics toward drug mechanism studies, we treated H358 cells with the FDA approved KRAS^{G12C} covalent inhibitor, AMG510 using the same culture and dose concentrations described in a recent single cell RNASeq study of the same.[41]

When cells treated with AMG510 are analyzed as if they are technical replicates with peptide and protein abundances averaged, our results closely mimic the effects of this compound as established by others.[42] Following treatment with AMG510, we observe decreased abundance of other proteins well-established as direct interactors of KRAS following inhibition (**S. Fig. 8A**). Gene set enrichment analysis (GSEA) of proteomic alteration found the canonical VEGF pathway to be the single most altered mechanism upon AMG510 treatment (**S. Fig. 8A**), in line with previous observations. When phosphopeptide changes between treated and untreated were considered, the most enriched KEGG pathway observed was hsa0414 RAS signaling, which exhibited a 7.36-fold enrichment over the baseline observed proteome with p-value of $1e10^{-11}$.

A recent temporal proteomic and phosphoproteomic analysis of two cell lines treated with KRAS G12C covalent inhibitors identified a large list of proteins with differential expression following treatment. When treating single cells as replicates, we find our data to be in generally high concordance with these observations, despite the lower number of quantified proteins in single cells. When evaluating each cell in isolation, however, the heterogeneity in abundance raises questions about some of the most seemingly important protein alterations. One example is the chloride channel protein CLIC3. In both the bulk temporal proteomics and single cells when protein expression is summed, we observe marked differential abundance of CLIC3. Closer evaluation at the single cell level, demonstrates that this differential in abundance is driven by only 10 treated cells that demonstrate a high level of CLIC3 expression, while no quantifiable signal is observed for this protein in any other cell (**S. Fig. 8B**). In addition, the quantitative phosphoproteomics of bulk AMG510 treated H358 cells identified differential abundance of LMNA phosphorylation sites. Phosphorylation of mitotic sites Ser22 and S392 were demonstrated as downregulated following treatment in addition to sites on 392, 394 and 628. In contrast, phosphorylation sites on residues 277, 404, 406 and 638 were observed as upregulated in the bulk proteomic analysis. The previously noted cell cycle dependent phosphorylation effects on LMNA proteins observed across single cells suggests that these observations are more likely a reflection of cell cycle status than direct effects of AMG510 treatment. Due to the observed effects of $KRAS^{G12}$ inhibition on this cell line leading to the majority of cells into a quiescent state, these results appear to be supportive, however, further analysis will be required to better understand the interplay of these factors.

Conclusions

Single cell proteomics is a newly emerging field with tremendous promise as a complementary technology to both proteomics and other single cell technologies. The majority of studies described to date have necessarily focused on developing the sample preparation,[6,9,25] instrument methods [43,44] and informatics tools [29,45,46] necessary to lay the groundwork for later applications. Much of this valuable development has been performed using diluted bulk proteomics samples [13], frog embryos [47], or clonal populations of hundreds of cells harvested from single cell seeding [48] or laser capture microdissection.[28] Many of these studies have contributed key insights into the challenges LCMS faces at picogram concentrations of peptide. From both previous results and the protein identifications made in this study, it appears the limiting factor in single cell proteomics is primarily the total concentration of each protein in a single cell, despite the hardware configuration utilized.

In this study we have developed optimized methods for reporter-based quantification with a focus on reducing background coisolation interference and obtaining the highest quantitative accuracy. Through an extreme dilution series of labeled peptides, we observe that the functional intrascan linear dynamic range of this approach may exceed 3-orders of magnitude for many fragmentation scan on this instrument. At a 1,000-fold dilution we find that approximately half of the low intensity reporter ions are lost, which provides a useful metric for our functional operational range. Surprisingly, approximately 10% of identified peptides possess a measurable reporter ion at the lowest dilution level, a number 4-orders of magnitude lower than the highest peptide concentration

used. While a dilution series that causes 90% of quantification values to be missing values is considered a unit below the limits of detection and therefore unusable by LCMS proteomics operators[49–51], it is worth noting that this is a typical metric for single cell RNA-Seq. While single cell RNA-Seq is continually maturing with improved coverage and sensitivity, in most workflows in use today as many as 95% of transcript values from a single cell will be a zero, or a missing value. While the number of missing values may be inflated by technical issues [52], this is simply a function of the technique, and all informatics pipelines have ways to compensate or utilize zero values.[53–55]

Optimized methods for pasefRiQ can provide quantifiable data over an extremely wide dynamic range as demonstrated by the accurate quantification values obtained for the TNAa *E. coli* protein was observed in the 9 technical replicates of 2.4 ng on column. As each labeled peptide channel contributes to the total peptide load on column, approximately 267 picograms of peptide are present in each of the 9 TMT channels. The lowest dilution of the labeled *E. coli* peptides is approximately 27 picograms per channel in this highly complex standard. While human cells vary widely in both their total size and concentration, 267 picograms is a lower amount of protein than is contained in cells of interest in our lab, such as those of the liver which contain ~400 picograms of protein per cell.[56] These results strongly suggest that pasefRiQ can return quantifiable data from a variety of single human cell types

When applying pasefRiQ to single human cells we find clear indications of cell cycle mediated proteomic effects and estimate our current limit of detection appears to be proteins with approximately 100,000 protein copies per cell.[29] It is worth noting that all comparisons of single cell proteomics methods in this study exclusively used MS/MS based identifications and classic search tools. Single cell proteomics benefits from both advanced informatics workflows and from spectral matching that takes into account the unique properties of these spectra.[57] In addition, classical search tools for TIMSTOF data analysis have proven inferior to next generation engines such as MSFragger which can employ open search functionalities.[58] As nearly all tools described for single cell proteomics have been designed for Orbitrap systems, considerable adaptation will be necessary and is underway in our laboratory today. Even without the ability to leverage these tools the enhanced relative scan acquisition rate of pasefRiQ leads to higher protein sequence coverage than any previously described technique. This increase in relative sequence coverage can be leveraged to allow the confident identification of protein PTMs in single human cells. Peptides possessing biological PTMs such as phosphorylation and acetylation are well characterized as more challenging to identify in LCMS proteomics data, often due simply to the stoichiometry of modified peptides. [59] We do observe direct effects of this here as we exclusively identify PTMs on proteins of the highest intracellular abundance and sequence coverage.

Finally, we evaluate pasefRiQ for a drug mechanism study of single cells by treating a model $KRAS^{G12C}$ mutant cell line with the covalent inhibitor AMG510. AMG510 was approved by the FDA in early 2021 as the first in a line of similar covalent inhibitors currently in clinical trials and under development.[41] While possessing clear value as a first of its kind treatment, spontaneous resistance to this drug has been observed in both cell lines and patients.[60] Recent work using temporal proteomics identified multiple proteins upregulated following inhibitor treatment and demonstrated the value of combinatorial therapies that inhibited both $KRAS^{G12C}$ and these resistance mechanisms.[42] When comparing our results to the temporal bulk proteomics and other related studies of $KRAS^{G12C}$ inhibition, we find our results generally high concordance. We observe both alterations in the canonical VEGF and MAPK pathways that are hyperactivated by the perpetually GTP bound and active KRAS mutant protein.[61] When considering our cells individually, we find our results are less clear. While proteins in the MAPK pathway appear largely affected in similar manners, some protein level observations appear to be wholly driven by extreme protein differentials in small cell populations. Furthermore, we find that LMNA phosphorylations observed in the bulk cell proteomics are clearly confounded by cell cycle mediated effects where these

phosphorylations play a key role. Cell cycle status was recently implicated as a key driver in the development of AMG510 resistance, suggesting that further investigation into these mechanisms through single cell proteomics would be a valuable contribution to the understanding of inhibitor response and resistance. While a thorough analysis of KRAS inhibition is outside of the scope of the study presented here, we find these results highly promising as indicative of the potential single cell proteomics can play in molecular pharmacology studies.

Materials and Methods

Samples for optimization.

Human cancer cell line digest standards K562 (Promega) and diluted to 100 microgram/mL in 50mM TEAB and labeled with TMTPro reagents according to manufacturer instructions. For preparation of the TMTPro 16 and TMT9 standards, the unit resolution reagents 126, 134, 127n, 128n, 129n, 130n, 131n, 132n, and 133n channels were used in all experiments. The loading capacity of the LCMS system was determined empirically from dilution series of these standards. A TMT triple knock out (TMT TKO) yeast standard (Pierce) was prepared following manufacturer instructions and prepared in serial dilutions in 0.1% trifluoroacetic acid.[14]

Preparing bulk samples for spectral libraries and carrier channels

NCI-H-358 cells were obtained from ATCC (Catalog #5807) and were reconstituted and passaged according to the included instructions using 6 well culture dishes. For bulk cell experiments, cells were aspirated, washed with ice cold water which was rapidly aspirated prior to addition of S-Trap lysis buffer. All steps of the S-Trap mini protocol were performed according to manufacturer instructions (ProtiFi) with the exception that alkylation and reduction were not performed. Peptides for spectral library generation were labeled with the 128C reagent from the TMTPro reagents according to all manufacturer protocols. The peptides were fractionated by using high pH reversed phase spin columns (Pierce) and eluted peptides were centrifuged to near dryness prior by SpeedVac. Peptides for use as carrier channels were labeled with the 134N reagent from the TMTPro, lyophilized and the concentration was determined using a Qubit system (Thermo Fisher) following manufacturer instructions.[15]

Preparation of single cells

Single control cells were treated with DMSO or with 10 µm AMG510 (SelecChem S8830) prepared in DMSO as previously described.[41] AMG510 treated, and control cells were cultured alongside control cells for 40 hours prior to rapid washing of cells with ice cold magnesium and calcium free PBS (Fisher). Cells were sorted and aliquoted following a previously published protocol using flow cytometry. [62]. Briefly, cells trypsinized to circularize, washed and resuspended with trypsin inhibitor and rapidly taken to the Johns Hopkins University School of Public Health Cell Sorting and Sequencing Core on dry ice. Cells were stained with a viability marker and sorted directly onto microwell plates and were immediately transferred to dry ice prior to -80C storage. Following sorting 20,000 cells were sorted to provide more clear metrics of cell characteristics and visualization of gating parameters (Fig. S5).

Lysis, digestion, and labeling followed the SCoPE2 protocol previously described [8] with minor alterations. Briefly, single cells were removed from -80C storage and placed directly in a solid fitted heat block at 95C for 10 minutes. (Fisher) All manipulation of plates was performed with the author grounded by alligator clip to the bench surface to prevent static discharge removing the cells from the microwells. The protein from lysed cells were digested in 1 µL 10ng/µL of trypsin/LysC (Promega) in 100mM TEAB (ProtiFi). Digestion was performed at 37C for 3 hours in sealed plates on a revolving incubator. Plates were centrifuged at 4,000 x g at 4C every half hour to concentrate condensate. Following digestion, cells were labeled with previously aliquoted and stored TMTPro

reagent [63] resuspended in LCMS grade anhydrous acetonitrile (Fisher) to a total concentration of 44mM. 500nL of resuspended reagent was added to each cell as appropriate and labeling was performed at room temperature for 1 hour. Plates were centrifuged twice to concentrate condensate. TMT labeling was quenched by the addition of 500nL of 0.5% hydroxylamine and centrifugal shaking for 1 hour at room temperature. Well with single cells were resuspended with the serial addition of the TMT134N carrier channel at an approximate concentration of 100 ng.

LCMS settings for standards

All instrument settings are included within the Bruker. d files in the ProteomeXchange and have been uploaded to www.LCMSMethods.org as pasefRiQf_v1 and published as (dx.doi.org/10.17504/protocols.io.b4ifqubn). **Table 2** is a brief summary of these settings.

Instrument settings	
LCMS	EasyNLC1200
Column	IonOpticks Aurora 25cm
Trap	PepMap100 C-18 3cm 75µm
Flow rate	200nL/min
Buffers	0.1% formic acid, 80% acetonitrile 0.1% formic acid
Source	Bruker Daltronic CaptiveSpray
Column temperature	50C
ESI voltage	1650V
Number of Ramps	10 ramps, unless specifically noted
Cycle time	2.2 seconds for 10 ramp files, 1.1 s for 5 ramp files
Quadrupole Isolation Width	Default 2 Da for ions < 700, 3 Da for ions > 700
Optimized Isolation Width	1.5 Da for all masses
TIMS ramp, default	0.6 - 1.6
Optimized TIMS ramp TMT peptide IDs	0.8 - 1.3
Optimized TIMS ramp TMT protein IDs	0.7 - 1.5
TIMS Step 1 Fragmentation energy	75ev
TIMS Step 2 Fragmentation energy	25ev

Table 2. A summary of the instrument settings used in this study

Data conversion and processing

Vendor proprietary output (. d) files were converted to MGF using three separate solutions: ProteoWizard MSConvert Developer Build Version: 3.0.20310-0d96039e2 for both GUI and command line-based conversions. MSConvert parameters for default pasef MGF conversion were utilized through the GUI. MSConvert combined MS/MS spectra if all of the following criteria were met elution time ≤ 5 seconds, ion mobility tolerance ≤0.1 1/k0 and precursor m/z ≤0.05 Da. FragPipe 14.0 was used for the direct analysis with MSFragger 3.1, Philosopher 3.3.11 and Python 3.8.3.23 All calibrated MGF spectra were generated with MSFragger 3.1.1. which was released during the construction of this manuscript and enabled the recalibration functionality. All MSFragger settings were set at default for pasefDDA data closed search with the addition of the TMTPro reagents as modifications on the peptide N-terminus and on lysines. Conversion of data through Data Analysis 5.3 was performed via a Visual Basic script provided by the vendor.

Processing of pasefRiQ data in Proteome Discoverer

All MGF files were processed with Proteome Discoverer 2.4 using SequestHT and Percolator with the reporter ion quantification node. A scan filter was used to the unrecognized fragmentation of the TIMSTOF MS/MS spectra as HCD and the spectra as FTICR at a resolution of 35,000 in order to allow the visualization of all significant figures in downstream processing. To reduce the complexity of TIMSTOF MS/MS spectra a binning method was used that retained only the 12 most intense fragment ions from each 100 Da mass window within each spectrum. Spectra were searched with SequestHT using a 30ppm MS1 tolerance and a 0.05 Da MS/MS fragment tolerance. Static modifications of the corresponding TMT reagent and the alkylation of cysteines with iodoacetamide was employed in all searches, with the exception of PTM searches in which the TMTPro reagent and acetylation of lysine were added as dynamic modifications. The UniProt SwissProt reviewed FASTA was used for the corresponding organisms from downloads of the complete reviewed SwissProt library in January of 2021. The cRAP database (www.gpm.org) was used in all searches. For quantification of the TMTPro9 samples, a custom quantification scheme was built from the TMTPro defaults that disabled the C^{13} isotopes of each pair. The vendor default workflow for reporter-based quantification was used with the following alterations: only unique peptides were used for quantification, unique peptides were determined from the protein, not protein group identification, the default quantification scheme was intensity based, and the minimum average reporter intensity filter was set at 10. The protein marker node was used to flag and filter contaminants from the cRAP database and the result statistics node was added for post-processing. Data was viewed using the IMP-MS2Go (www.pd-nodes.org) release for Proteome Discoverer 2.5.[27]

Processing of pasefRiQ single cell data in MaxQuant and Proteome Discoverer

The sixty-three LCMS runs, containing a total of 441 single cells from this study, vendor proprietary. d files were converted to MGF using MSConvert using the default DDA pasef conversion method. The calibrated MGF files were processed in Proteome Discoverer 2.4/2.5 as described above. Percolator was used for FDR estimation, as well as for FDR estimation at the peptide and protein group levels. For normalization, the files were reprocessed with identical settings in two separate workflows. The first had the removal of the 134N and 133N reporter channels. This allowed the comparison between the signal and blank (126) channels but did not permit normalization due to the scaling of the signal in the blank channel. A second normalization and scaling using a total sum-based approach was used for final analysis. Entire LCMS runs or individual sample channels were removed from consideration as outliers in downstream data analysis when no reporter ion signal was obtained that exceeded that of the blank method blank channel. In addition, the 133N channel was excluded from all analysis due to significant ratio distortion which was identified as impurities in the 134N carrier channel from the TMT reagent kit. For MaxQuant analysis, the recommended settings described for pasefDDA [64] were used in version 1.6.17, with the addition of the 9 TMTPro tags which were manually added to the XML schema. For PTM analysis the spectra were searched with SequestHT in Proteome Discoverer 2.4 with lysine acetylation and the phosphorylation of serine and threonine considered as possible modifications. The IMP-PTMrs node was used for modification site localization scoring. For TIMSTOF data a precursor and fragment tolerance of 30ppm was used for all analyses.

Pathway and gene ontology analysis

Three commercial programs were utilized for pathway analysis, Protein Center (Thermo) and Ingenuity Pathways Analysis (Qiagen). For protein center analysis, differential proteins were selected from the normalized data in proteome discoverer and all proteins not meeting a cutoff of 2-fold at a p value <0.05 were excluded. The top pathways were determined by the number of remaining proteins that were identified within that group. For IPA analysis, the normalized ratios of all proteins with quantification of AMG510/control were exported as CSV and uploaded into IPA using core analysis. The following settings were used, Core Expression Analysis based on Expr Fold change utilizing z-scores for directional analysis. Files were compared against the Ingenuity

Knowledge Base at the gene level. Only experimentally observed relationships were used for pathway construction with filtering for human samples and cell lines. The SimpliFi cloud server (Protifi, Toronto, Canada) beta version was used for downsteam analysis and visualization of all data using the default interpretation settings for Proteomics data and a direct import of the MaxQuant output file.

Phosphopeptide quantification

Identified phosphopeptides were filtered using a log2 fold change of 1 using Boolean logical filters within Proteome Discoverer 2.4 and a minimum peptide confidence cutoff filter of approximately 0.05% and a localization confidence score from phosphoRS of 70%. This final list consisted of 1435 phosphopeptides, which were exported to CSV for normalization. The ratios of all proteins containing these phosphopeptides in addition to >1 unique unmodified peptide, as determined from the ungrouped protein level were exported to .csv. Phosphopeptide ratios were normalized against the total protein ratio of the 673 proteins they mapped to through use of an in house developed tool provided with this manuscript. Normalized phosphopeptides with ratios demonstrating a log2 fold change of greater than 2 were exported to CSV for pathway analysis.

Code and Availability

A summary of the Proteome Discoverer results for proteins and phosphopeptides can be downloaded at this direct link. Phosphopeptides ratios were normalized against protein abundances using LPN V1.0 which is available at https://github.com/orsburn/LazyPhosphoNormalizer and https://proteomicsnews.shinyapps.io/proj25_benjamin/ All copy number distribution plots were generated using the openly available Shiny tool, available at https://proteomicsnews.shinyapps.io/CopyNumber_plotterV2/ . All additional figures were generated in R Studio 1.3.1093 utilizing Tidyverse, ggplot and ggally using the most up to date versions available on 3/10/2021.

Data and Availability

Results from previous SCoPE-MS studies utilized in this study are available at www.ProteomeXchange.org via the following accession numbers: PXD008985, PXD011748 and PXD025387. Optimization files for background coisolation reduction and two proteome standards at 240ng are available at the following link: ftp://MSV000088757@massive.ucsd.edu with reviewer password: pasefRIQ. Following publication, the data will be permanently available at: ftp://massive.ucsd.edu/MSV00008875. Files used for optimization of pasefRiQ at dilution levels relevant to single cell loads are available at this link: ftp://MSV000088796@massive.ucsd.edu using the password: pasefRiQ. All vendors. d files, MGF peak lists used in the study, MaxQuant protein lists and Proteome Discoverer results for protein and phosphopeptide identifications in H358 single cells are available through ProteomeXchange as PXD028710. During review, these can be accessed via using reviewer credentials: MSV000088144 and Password: pasefSCOPE700N. Direct download is available through ftp://MSV000088144@massive.ucsd.edu. In addition, all filtered MGF files used for MSFragger open search, mutational assignment and processed data are available at MASSIVE through: ftp://MSV000088157@massive.ucsd.edu with reviewer credentials: MSV000088157 and password: pasefSCOPE700N.

Acknowledgments

We would like to thank Megan Rigby for help interpreting single cell RNASeq results and for advice on the NCI-H-358 cell line and Dr. Alexis Norris for assistance with custom R tools featured in this

work. We would also like to thank Conor Jenkins for access to unpublished Python scripts adapted to assess the quality of the LCMS files generated in this work.

Funding:

National Institutes of Health grant R01AG064908 (NNB)

National Institutes of Health grant R01GM103853 (NNB)

References

1. Hwang, B., Lee, J. H. & Bang, D. Single-cell RNA sequencing technologies and bioinformaticspipelines. *Experimental & Molecular Medicine* **50**, 1–14 (2018).

2. Ghazalpour, A. *et al.* Comparative analysis of proteome and transcriptome variation in mouse. *PLoS Genetics* (2011) doi:10.1371/journal.pgen.1001393.

3. Ponomarenko, E. A. *et al.* The Size of the Human Proteome: The Width and Depth. *International Journal of Analytical Chemistry* (2016) doi:10.1155/2016/7436849.

4. Marx, V. A dream of single-cell proteomics. *Nature Methods* (2019) doi:10.1038/s41592-019-0540-6.

5. Slavov, N. Unpicking the proteome in single cells. *Science* (2020) doi:10.1126/science.aaz6695.

6. Kelly, R. *et al.* Single Cell Proteome Mapping of Tissue Heterogeneity Using Microfluidic Nanodroplet Sample Processing and Ultrasensitive LC-MS. *Journal of Biomolecular Techniques : JBT* **30**, S61–S61 (2019).

7. Ctortecka, C. & Mechtler, K. The rise of single-cell proteomics. *Analytical Science Advances* (2021) doi:10.1002/ansa.202000152.

8. Specht, H. *et al.* Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. *Genome Biology* **22**, 50 (2021).

9. Hartlmayr, D. *et al.* An automated workflow for label-free and multiplexed single cell proteomics sample preparation at unprecedented sensitivity. *bioRxiv* 2021.04.14.439828 (2021) doi:10.1101/2021.04.14.439828.

10. Budnik, B., Levy, E., Harmange, G. & Slavov, N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biology* (2018) doi:10.1186/s13059-018-1547-5.

11. Bekker-Jensen, D. B. *et al.* An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell Systems* **4**, 587-+ (2017).

12. Gillson, J. *et al.* Small molecule KRAS inhibitors: The future for targeted pancreatic cancer therapy? *Cancers* (2020) doi:10.3390/cancers12051341.

13. Cheung, T. K. *et al.* Defining the carrier proteome limit for single-cell proteomics. *Nature Methods* (2021) doi:10.1038/s41592-020-01002-5.

14. Paulo, J. A., O'Connell, J. D. & Gygi, S. P. A Triple Knockout (TKO) Proteomics Standard for Diagnosing Ion Interference in Isobaric Labeling Experiments. *Journal of the American Society for Mass Spectrometry* (2016) doi:10.1007/s13361-016-1434-9.

15. Thompson, A. *et al.* TMTpro: Design, Synthesis, and Initial Evaluation of a Proline-Based Isobaric 16-Plex Tandem Mass Tag Reagent Set. *Analytical Chemistry* (2019) doi:10.1021/acs.analchem.9b04474.

16. Li, J. *et al.* TMTpro-18plex: The Expanded and Complete Set of TMTpro Reagents for Sample Multiplexing. *Journal of Proteome Research* **20**, 2964–2972 (2021).

17. Eliuk, S. & Makarov, A. Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annual Review of Analytical Chemistry* (2015) doi:10.1146/annurev-anchem-071114-040325.

18. Hecht, E. S., Scigelova, M., Eliuk, S. & Makarov, A. Fundamentals and Advances of Orbitrap Mass Spectrometry. in *Encyclopedia of Analytical Chemistry* (2019). doi:10.1002/9780470027318.a9309.pub2.

19. Mihaleva, V. V. *et al.* Accurate mass error correction in liquid chromatography time-of-flight mass spectrometry based metabolomics. *Metabolomics* (2008) doi:10.1007/s11306-008-0108-4.

20. Navarrete-Perea, J., Gygi, S. P. & Paulo, J. A. HYpro16: A Two-Proteome Mixture to Assess Interference in Isobaric Tag-Based Sample Multiplexing Experiments. *Journal of the American Society for Mass Spectrometry* **32**, 247–254 (2021).

21. O'Connell, J. D., Paulo, J. A., O'Brien, J. J. & Gygi, S. P. Proteome-Wide Evaluation of Two Common Protein Quantification Methods. *Journal of Proteome Research* **17**, 1934–1942 (2018).

22. Meier, F. *et al.* Online parallel accumulation–serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. *Molecular and Cellular Proteomics* (2018) doi:10.1074/mcp.TIR118.000900.

23. Grinfeld, D., Aizikov, K., Kreutzmann, A., Damoc, E. & Makarov, A. Phase-Constrained Spectrum Deconvolution for Fourier Transform Mass Spectrometry. *Analytical Chemistry* **89**, 1202–1211 (2017).

24. Gygi, J. P. *et al.* Web-Based Search Tool for Visualizing Instrument Performance Using the Triple Knockout (TKO) Proteome Standard. *Journal of Proteome Research* **18**, 687–693 (2019).

25. Leduc, A., Huffman, R. G. & Slavov, N. Droplet sample preparation for single-cell proteomics applied to the cell cycle. *bioRxiv* 2021.04.24.441211 (2021) doi:10.1101/2021.04.24.441211.

26. Zhang, Z. *et al.* Reverse and Random Decoy Methods for False Discovery Rate Estimation in High Mass Accuracy Peptide Spectral Library Searches. *Journal of Proteome Research* (2018) doi:10.1021/acs.jproteome.7b00614.

27. Orsburn, B. C. Proteome Discoverer—A Community Enhanced Data Processing Suite for Protein Informatics. *Proteomes* **9**, (2021).

28. Mund, A. *et al.* AI-driven Deep Visual Proteomics defines cell identity and heterogeneity. *bioRxiv* 2021.01.25.427969 (2021) doi:10.1101/2021.01.25.427969.

29. Orsburn, B. C. Evaluation of the Sensitivity of Proteomics Methods Using the Absolute Copy Number of Proteins in a Single Cell as a Metric. *Proteomes* vol. 9 (2021).

30. Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D. & Nesvizhskii, A. I. MSFragger: Ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nature Methods* (2017) doi:10.1038/nmeth.4256.

31. Solntsev, S. K., Shortreed, M. R., Frey, B. L. & Smith, L. M. Enhanced Global Post-translational Modification Discovery with MetaMorpheus. *Journal of Proteome Research* (2018) doi:10.1021/acs.jproteome.7b00873.

32. Ye, X. *et al.* Comparative proteomics of a model MCF10A-KRasG12V cell line reveals a distinct molecular signature of the KRasG12V cell surface. *Oncotarget* (2016) doi:10.18632/oncotarget.13566.

33. Choudhary, C. *et al.* Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* (2009) doi:10.1126/science.1175371.

34. Sarg, B. *et al.* Comparing and Combining Capillary Electrophoresis Electrospray Ionization Mass Spectrometry and Nano–Liquid Chromatography Electrospray Ionization Mass Spectrometry for the Characterization of Post-translationally Modified Histones. *Molecular & Cellular Proteomics* (2013) doi:10.1074/mcp.M112.024109.

35. Levin, Y. The role of statistical power analysis in quantitative proteomics. *Proteomics* (2011) doi:10.1002/pmic.201100033.

36. Zhang, Y., Fonslow, B. R., Shan, B., Baek, M. C. & Yates, J. R. Protein analysis by shotgun/bottom-up proteomics. *Chemical Reviews* (2013) doi:10.1021/cr3003533.

37. Brunner, A.-D. *et al.* Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *bioRxiv* 2020.12.22.423933 (2021) doi:10.1101/2020.12.22.423933.

38. Torvaldson, E., Kochin, V. & Eriksson, J. E. Phosphorylation of lamins determine their structural properties and signaling functions. *Nucleus (Austin, Tex.)* **6**, 166–171 (2015).

39. Yamashita, K. *et al.* Mitotic phosphorylation of Pex14p regulates peroxisomal import machinery. *The Journal of cell biology* **219**, e202001003 (2020).

40. Cenni, V. *et al.* Lamin A Ser404 Is a Nuclear Target of Akt Phosphorylation in C2C12 Cells. *Journal of Proteome Research* **7**, 4727–4735 (2008).

41. Xue, J. Y. *et al.* Rapid non-uniform adaptation to conformation-specific KRAS(G12C) inhibition. *Nature* (2020) doi:10.1038/s41586-019-1884-x.

42. Santana-Codina, N. *et al.* Defining and Targeting Adaptations to Oncogenic KRASG12C Inhibition Using Quantitative Temporal Proteomics. *Cell Reports* (2020) doi:10.1016/j.celrep.2020.03.021.

43. Schoof, E. M. *et al.* Quantitative single-cell proteomics as a tool to characterize cellular hierarchies. *Nature Communications* **12**, 3341 (2021).

44. Huffman, R. G., Chen, A., Specht, H. & Slavov, N. DO-MS: Data-Driven Optimization of Mass Spectrometry Methods. *Journal of Proteome Research* (2019) doi:10.1021/acs.jproteome.9b00039.

45. Vanderaa, C. & Gatto, L. Replication of single-cell proteomics data reveals important computational challenges. *bioRxiv* 2021.04.12.439408 (2021) doi:10.1101/2021.04.12.439408.

46. Chen, A. T., Franks, A. & Slavov, N. DART-ID increases single-cell proteome coverage. *bioRxiv* 399121 (2019) doi:10.1101/399121.

47. Lombard-Banek, C., Choi, S. B. & Nemes, P. Chapter Thirteen - Single-cell proteomics in complex tissues using microprobe capillary electrophoresis mass spectrometry. in *Enzyme Activity in Single Cells* (eds. Allbritton, N. L. & Kovarik, M. L. B. T.-M. in E.) vol. 628 263–292 (Academic Press, 2019).

48. Hartmann, F. J. *et al.* Single-cell metabolic profiling of human cytotoxic T cells. *Nature Biotechnology* **39**, 186–197 (2021).

49. Kalxdorf, M., Müller, T., Stegle, O. & Krijgsveld, J. IceR improves proteome coverage and data completeness in global and single-cell proteomics. *Nature Communications* **12**, 4787 (2021).

50. Yu, F., Haynes, S. E. & Nesvizhskii, A. I. Label-free quantification with FDR-controlled match-between-runs. *bioRxiv* (2020) doi:10.1101/2020.11.02.365437.

51. Vowinckel, J. *et al.* Cost-effective generation of precise label-free quantitative proteomes in high-throughput by microLC and data-independent acquisition. *Scientific Reports* (2018) doi:10.1038/s41598-018-22610-4.

52. Hicks, S. C., Townes, F. W., Teng, M. & Irizarry, R. A. Missing data and technical variability in single-cell RNA-sequencing experiments. *Biostatistics* **19**, 562–578 (2018).

53. Xu, L., Xu, Y., Xue, T., Zhang, X. & Li, J. AdImpute: An Imputation Method for Single-Cell RNA-Seq Data Based on Semi-Supervised Autoencoders . *Frontiers in Genetics* vol. 12 (2021).

54. Li, W. V. & Li, J. J. An accurate and robust imputation method scImpute for single-cell RNA-seq data. *Nature Communications* **9**, 997 (2018).

55. Linderman, G. C. *et al.* Zero-preserving imputation of single-cell RNA-seq data. *Nature Communications* **13**, 192 (2022).

56. Wiśniewski, J. R., Hein, M. Y., Cox, J. & Mann, M. A "proteomic ruler" for protein copy number and concentration estimation without spike-in standards. *Molecular and Cellular Proteomics* (2014) doi:10.1074/mcp.M113.037309.

57. Boekweg, H. *et al.* Features of Peptide Fragmentation Spectra in Single-Cell Proteomics. *Journal of Proteome Research* **21**, 182–188 (2022).

58. Yu, F. *et al.* Fast Quantitative Analysis of timsTOF PASEF Data with MSFragger and IonQuant. *Molecular and Cellular Proteomics* (2020) doi:10.1074/mcp.TIR120.002048.

59. Hogrebe, A. *et al.* Benchmarking common quantification strategies for large-scale phosphoproteomics. *Nature Communications* (2018) doi:10.1038/s41467-018-03309-6.

60. Hata, A. N. & Shaw, A. T. Resistance looms for KRASG12C inhibitors. *Nature Medicine* (2020) doi:10.1038/s41591-020-0765-z.

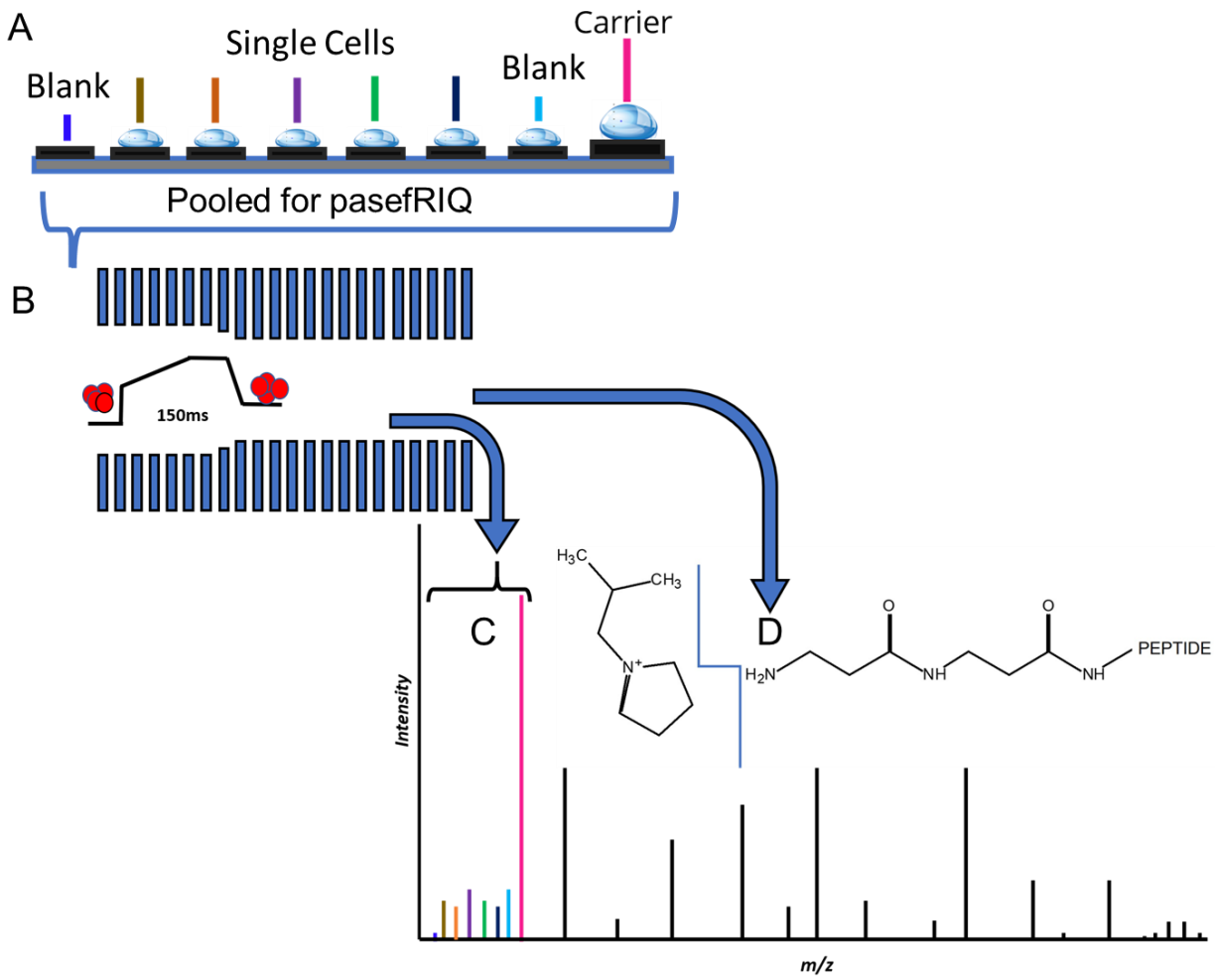
61. Drosten, M. & Barbacid, M. Targeting the MAPK Pathway in KRAS-Driven Tumors. *Cancer Cell* (2020) doi:10.1016/j.ccell.2020.03.013.

62. Kuboki, Y. *et al.* Single-cell sequencing defines genetic heterogeneity in pancreatic cancer precursor lesions. *The Journal of Pathology* **247**, 347–356 (2019).

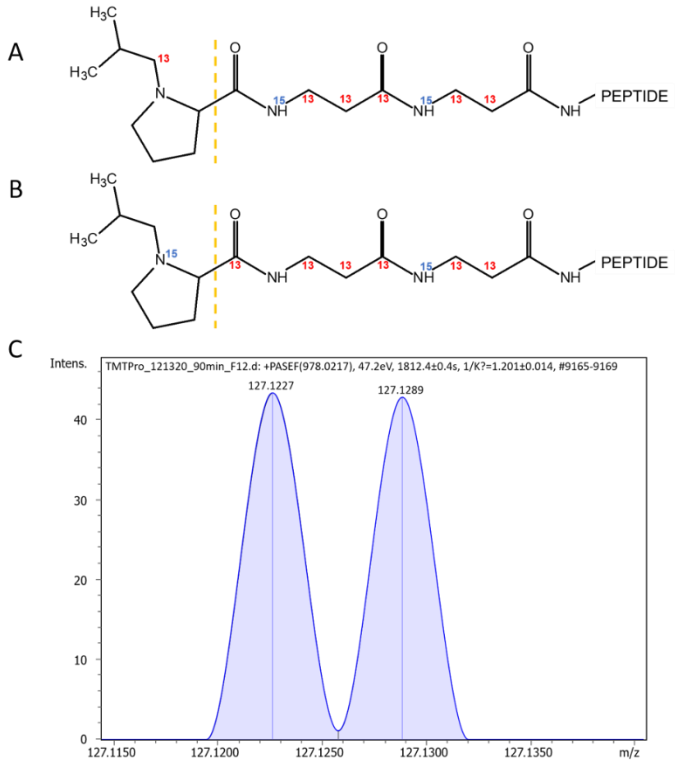
63. Yuan, Y. & Orsburn, B. C. Aliquoting of isobaric labeling reagents for low concentration and single cell proteomics samples. *bioRxiv* 2021.06.23.449560 (2021) doi:10.1101/2021.06.23.449560.

64. Meier, F. *et al.* Online parallel accumulation–serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. *Molecular and Cellular Proteomics* (2018) doi:10.1074/mcp.TIR118.000900.

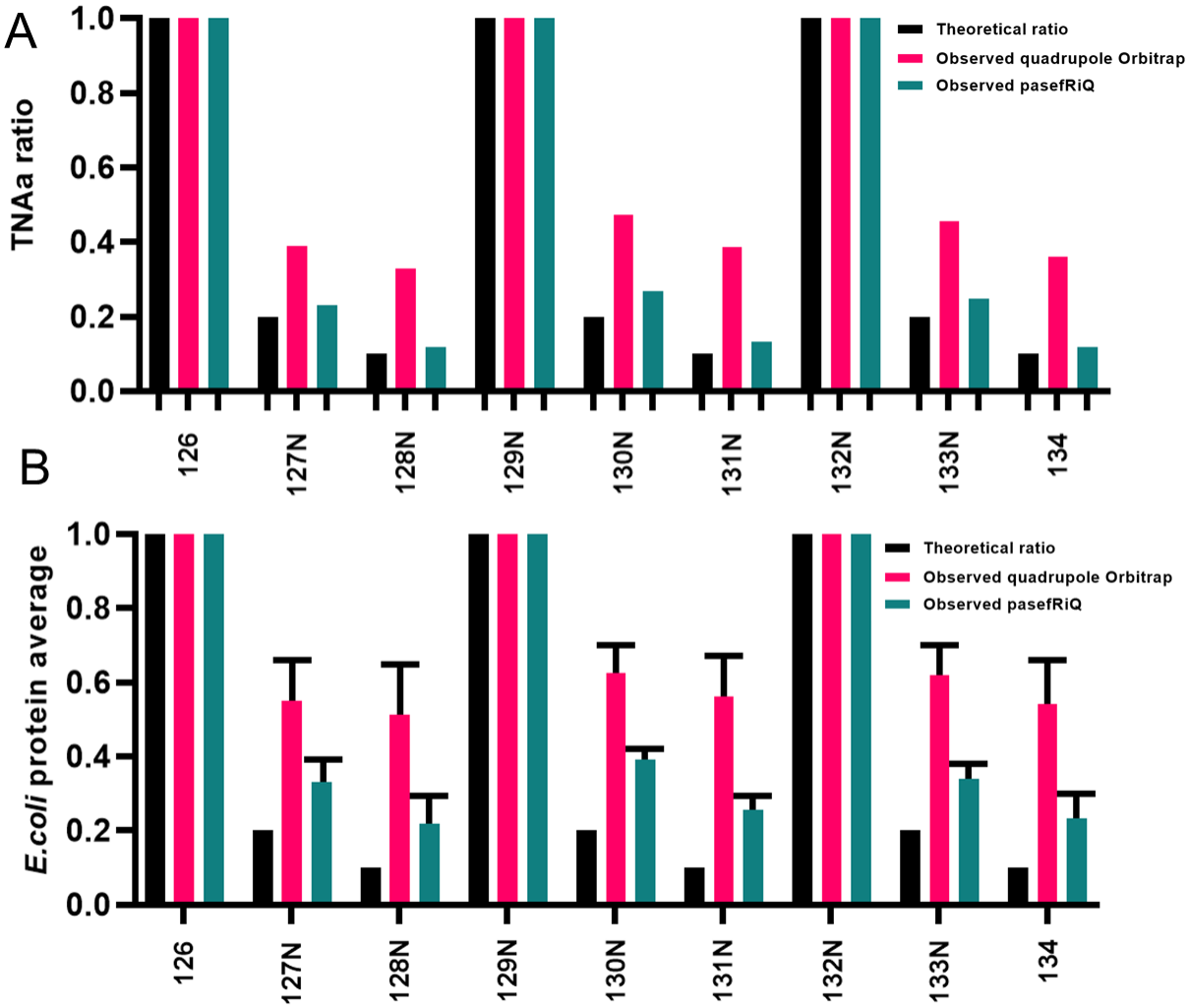
Supplemental Figures



S. Fig. 1. A cartoon illustrating the general strategy of analyzing single cells with pasefRiQ.

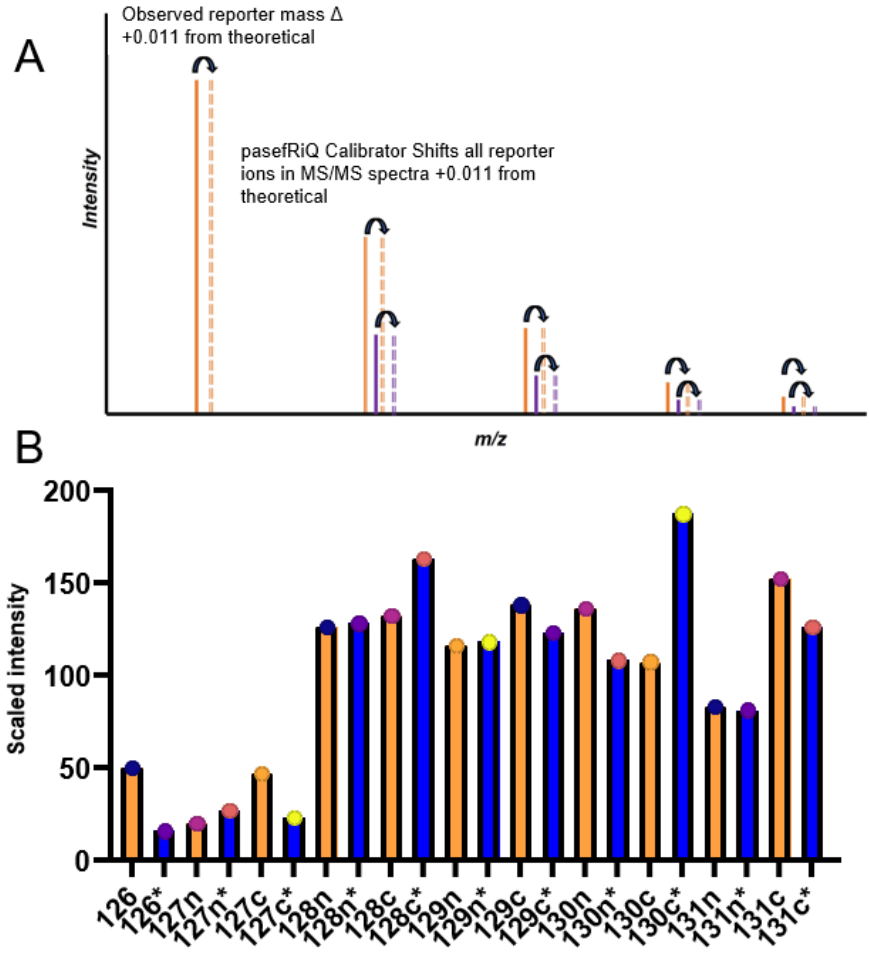


S. Fig. 2. **A.** The structure of the 127N reporter ion from the TMTPro reagent. **B.** The structure of the 127C reporter ion. **C.** A representative image from the liberated reporter ion regions of a labeled peptide digest.

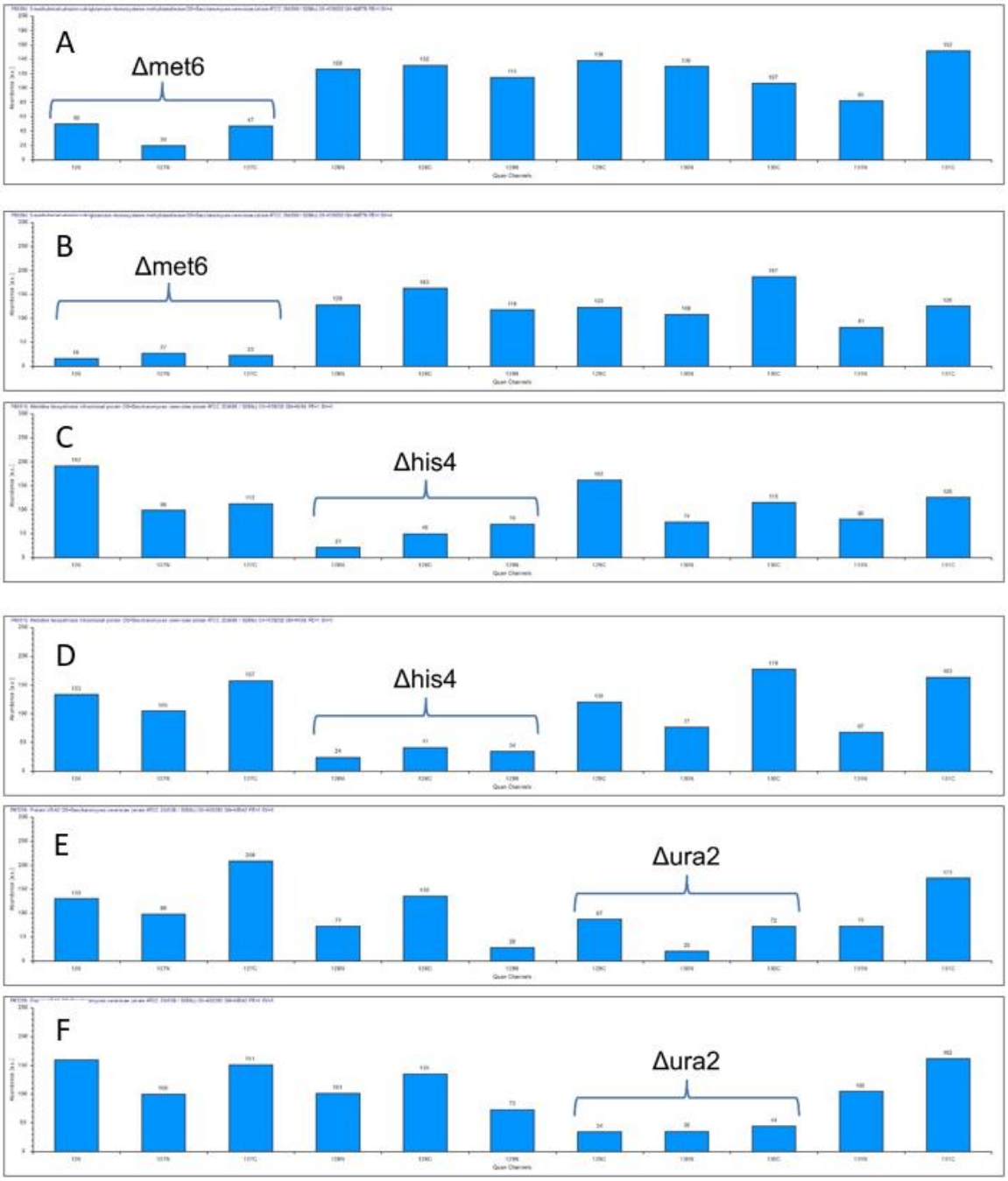


S.Fig. 3 A. A bar chart demonstrating the theoretical results and the observed results for the most abundant *E.coli* protein in the two proteome standard, with the highest *E.coli* concentration in each

three dilution series normalized to 1.0. **B.** The average of all identified *E.coli* proteins with two or more unique peptides per protein on each analyzer

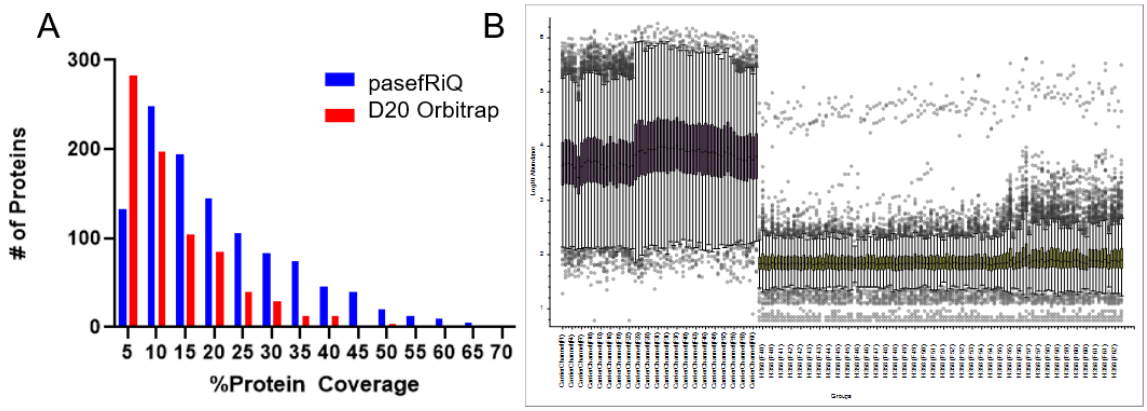


S.Fig. 4. **A.** An illustration describing the MS/MS recalibration of reporter ions with the pasefRiQ Calibrator. **B.** The scaled intensity of the met6 protein in the TMT-TKO standard where * denotes the values following one point recalibration

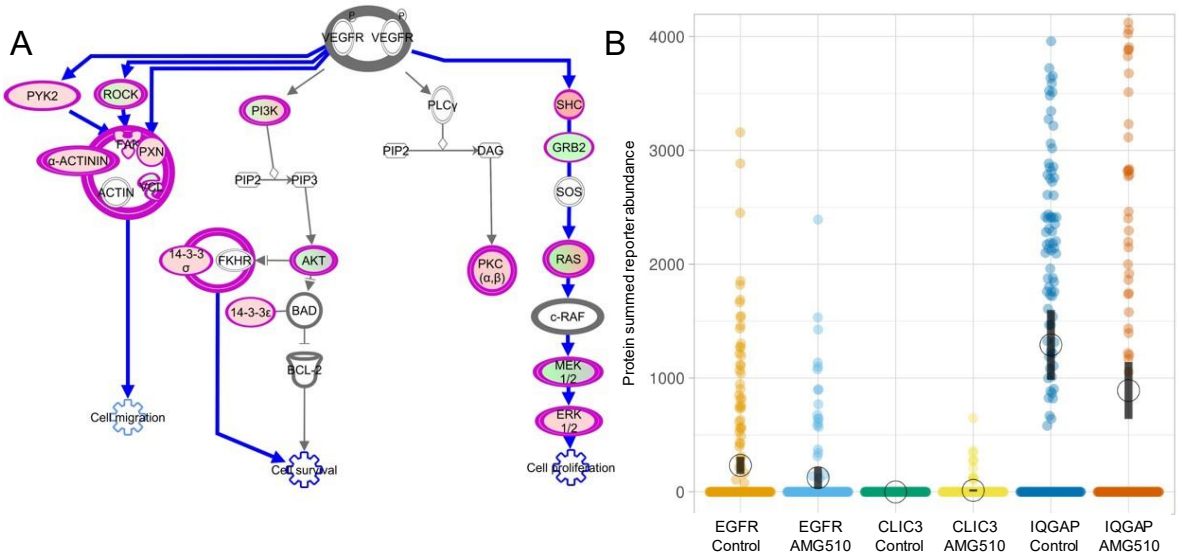


S. Fig. 5. .A.,C.,E. The scaled intensities of the Met6, His4, and Ura2 proteins from the TMT-TKO standard using an optimized pasefRiQ workflow integrated with a 20ppm mass tolerance window.

B.,D.,F. The scaled intensities of the same respective files following a 0.004 Da recalibration with the *pasefRiQCalibrator*.



S. Fig. 6. **A**. An overview of the relative protein sequence coverage obtained for each protein in representative pasefRiQ 30 minute experiment compared to an D20 Orbitrap based analysis described by Hartlmayer and Clortecka et al., **B.** An example loading plot demonstrating the raw log10 abundance values of carrier channels and single H358 cells demonstrating an approximate 100-fold signal difference.



S. Fig. 7. A. Pathway analysis indicated the VEGF pathway is the single most altered pathway in H358 cells following 40 hours of AMG510 treatment in single cells. **B**. The summed reporter abundance for three proteins demonstrated as having large ratio fold changes when cells are treated as technical replicates.