Published in final edited form as:

Pancreas. 2017 ; 46(5): 690–698. doi:10.1097/MPA.0000000000000800.

Proteome-Wide Protein Expression Profiling Across Five Pancreatic Cell Lines

Joao A. Paulo, PhD^{1, #}, Joseph D. Mancias, M.D., PhD^{1, 2}, and Steven P. Gygi, PhD¹

¹Department of Cell Biology, Harvard Medical School, Boston, MA 02115, United States

²Division of Genomic Stability and DNA Repair, Department of Radiation Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, United States

Abstract

Objectives—Mass spectrometry-based proteomics enables near-comprehensive protein expression profiling. We aimed to compare quantitatively the relative expression levels of thousands of proteins across 5 pancreatic cell lines.

Methods—Using tandem mass tags (TMT10-plex), we profiled the global proteomes of 5 cell lines in duplicate in a single multiplexed experiment. We selected cell lines commonly used in pancreatic research: CAPAN-1, HPAC, HPNE, PANC1, and PaSC. In addition, we examined the effects of different proteases (Lys-C and Lys-C plus trypsin) on the dataset depth.

Results—We quantified over 8,000 proteins across the 5 cell lines. Analysis of variance testing of cell lines within each dataset resulted in over 1,400 statistically significant differences in protein expression levels. Comparing the datasets, 10% more proteins and 30% more peptides were identified in the Lys-C/trypsin dataset than in the Lys-C-only dataset. The correlation coefficient of quantified proteins common between the datasets was greater than 0.85.

Conclusions—We illustrate protein level differences across pancreatic cell lines. Additionally, we highlight the advantages of Lys-C/trypsin over Lys-C-only digests for discovery proteomics. These datasets provide a valuable resource of cell line-dependent peptide and protein differences for future targeted analyses, including those investigating on- or off-target drug effects across cell lines.

Keywords

TMT; mass spectrometry; pancreas; SPS-MS3; digest; Lys-C

Introduction

Pancreatic cancer with an average five-year survival rate of 8% is the fourth most common cause of cancer-related death in the United States ¹. Pancreatic ductal adenocarcinoma is the major form of pancreatic cancer. Tumor cells associated with this cancer can arise from

[#]Corresponding author: Joao A. Paulo, Department of Cell Biology, 240 Longwood Ave., Harvard Medical School, Boston, Massachusetts 02115, USA, joao_paulo@post.harvard.edu.

Conflicts of Interest The authors acknowledge no conflict of interest.

multiple differentiated pancreatic cell types^{2,3}. Delineating the molecular mechanisms regulating normal pancreatic cell function and alterations produced by oncogenic cellular stressors will add to our understanding of the disease and lead to potentially new therapeutic avenues⁴. As various cell types are available, data acquired from one cell line may not reproduce that from another cell line under the same conditions. An extensive number of proteomic studies with pancreatic cell lines have been undertaken recently⁵⁻⁹, however, global proteome analyses across different pancreatic cell lines have been less common. Just as patients with specific subtypes of pancreatic ductal adenocarcinoma have varying reactions to therapy¹⁰, different cell lines may have a wide-range of responses to drugs and oncogenic stressors, due in part to variations in their global proteomes.

We use a versatile mass spectrometry-based multiplexing strategy to investigate the global proteomic differences across five pancreatic cell lines. We aim to determine if certain classes of proteins differ in expression level across cell lines. In addition, we offer the community two datasets of proteins and associated peptides from five pancreatic cell lines - including a normal ductal epithelial cell line, three pancreatic cancer cells lines, and one pancreatic stellate cell line – which will provide an atlas of targets for future assays investigating the molecular pathogenesis of pancreatic adenocarcinoma.

We chose five commonly studied human pancreatic cell lines grown under identical conditions. CAPAN-1 is a human pancreatic ductal adenocarcinoma cell line capable of invading extracellular matrix^{11,12}. HPAC is a pancreatic adenocarcinoma epithelial cell line derived from a xenograft of a primary tumor removed from the head of the pancreas¹³. HPNE originated from human pancreatic ducts and was immortalized by transduction with a retroviral expression vector containing the hTERT gene¹⁴. PANC-1 is a cell line established from the pancreatic duct of a patient who suffered from epithelioid carcinoma and is commonly used as an *in vitro* model of non-endocrine pancreatic cancer for tumorigenicity studies¹⁵. Finally, in addition to these pancreatic duct cell lines, we included pancreatic stellate cells (PaSC). PaSCs are myofibroblast-like cells that reside in exocrine areas of the pancreas and are thought to intercalate duct cells¹⁶. The cell line used herein, RLT-PSC, was immortalized using an out-growth method¹⁷. Although originating from the same organ, we anticipate vast cell line-specific proteomic differences.

We used a mass spectrometry-based multiplexing strategy to quantitatively compare the global proteomic differences among five pancreatic cell lines. Multiplexing strategies in mass spectrometry-based quantitative analyses, such as tandem mass tags (TMT) and isobaric tags for relative and absolute quantitation (iTRAQ) have many advantages for whole proteome profiling^{18,19}. Such strategies allow for multiple samples to be analyzed simultaneously thereby reducing instrument time and costs, while producing fewer missing values between samples and permitting multiple comparisons in a single experiment.

Proteolytic digestion is a key aspect of any proteomic profiling experiment. Proteins must be cleaved into mass spectrometry-amenable peptides for accurate mass measurements from which protein identifications are later inferred²⁰. Trypsin, with its advantages of sensitivity, specificity, and relatively low cost, is the enzyme of choice for mass spectrometry-based proteomic analyses. Although trypsin is specific for cleaving after lysine and arginine

residues, proteolysis is typically inhibited if proline is the adjacent C-terminal residue²¹. Enterococcal Lys-C has similar benefits as trypsin, but cleaves only after lysine residues whether or not a proline is present. In addition, Lys-C is also advantageous as it is active in 4 M urea, and is often used to nick the protein prior to tryptic digestion²². As such, these two enzymes are typically used sequentially. We aimed to gain insight into the different characteristics of the TMT-labeled peptides produced by these two proteases. As such, we performed replicate TMT10-plex analyses on the five cell lines and varied the protease so as to compare a) Lys-C followed by trypsin and b) Lys-C-only digestion strategies.

We present a TMT10-plex workflow in which we determine the relative expression levels of thousands of proteins across the 5 pancreatic cell lines mentioned above: CAPAN-1, HPAC, HPNE, PANC1, and PaSC in biological duplicate. We show statistically significant differences in relative protein expression levels across cell lines and subject these proteins to gene ontology analysis to determine what specific pathways or protein types are enriched. Here we produce two TMT10-plex datasets to compare the five pancreatic cell lines under different digestion conditions. These datasets identified peptides and associated proteins that can be used for future development of quantitative, targeted assays.

Methods

2.1 Materials

Tandem mass tag (TMT) isobaric reagents were from Thermo-Fisher Scientific (Waltham, MA). Water and organic solvents were from J.T. Baker (Center Valley, PA). Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS) were from LifeTechnologies (Waltham, MA). Unless otherwise noted, all other chemicals, were from Sigma (St. Louis, MO).

2.2 Cell growth and harvesting

Methods of cell growth and propagation followed previously utilized techniques^{23,24}. In brief, cells were propagated in DMEM supplemented with 10% FBS. Upon achieving 80% confluency, the growth media was aspirated and the cells were washed 3 times with ice-cold phosphate-buffered saline (PBS). Cells were dislodged with a non-enzymatic reagent, harvested by trituration following the addition of 10 mL PBS, pelleted by centrifugation at $3,000 \times g$ for 5 min at 4°C, and the supernatant was removed. One milliliter of HBSp (50 mM HEPES, 150 mM NaCl, pH 8.0 supplemented with 1X Roche Complete protease inhibitors), and 2% SDS were added per each 10 cm cell culture dish.

2.3 Cell lysis and protein digestion

Cells were homogenized by 10 passes through a 21 gauge (1.25 inches long) needle and incubated at 4°C with gentle agitation for 30 min. The homogenate was sedimented by centrifugation at $21,000 \times g$ for 5 min and the supernatant was transferred to a new tube. Protein concentrations were determined using the bicinchoninic acid (BCA) assay (Thermo-Fisher Scientific, Waltham, MA). Proteins were subjected to disulfide bond reduction with 5 mM tris (2-carboxyethyl)phosphine (room temperature, 30 min) and alkylation with 10 mM iodoacetamide (room temperature, 30 min in the dark). Excess iodoacetamide was quenched

with 10 mM dithiothreitol (room temperature, 15 min in the dark). Methanol-chloroform precipitation was performed prior to protease digestion. In brief, 4 parts of neat methanol were added to each sample and vortexed, 1 part chloroform was added to the sample and vortexed, and 3 parts water was added to the sample and vortexed. The sample was centrifuged at 14,000 RPM for 2 min at room temperature and subsequently washed twice with 100% methanol. Samples were resuspended in 50 mM HEPES, pH 8.5 and digested at room temperature for 13 h with Lys-C protease at a 100:1 protein-to-protease ratio. To specified samples, trypsin was then added at a 100:1 ratio while to the Lys-C-only digests, a second aliquot of Lys-C was added at a 100:1 ratio. In both cases, the reaction was incubated 6 h at 37°C.

2.4 Tandem mass tag labeling

Approximately 50 µg of peptides from each sample were labeled with TMT reagent. TMT reagents (0.8 mg) were dissolved in anhydrous acetonitrile (40 µL) of which 10 µL was added to the peptides along with 20 µL of acetonitrile to achieve a final acetonitrile concentration of approximately 30% (v/v). Following incubation at room temperature for 1 h, the reaction was quenched with hydroxylamine to a final concentration of 0.3% (v/v) for 15 min. The TMT-labeled samples were pooled at a 1:1 ratio for all samples. The combined sample was vacuum centrifuged to near dryness and subjected to C18 solid-phase extraction (SPE) via Sep-Pak (Waters, Milford, MA).

2.5 Off-line basic pH reversed-phase (BPRP) fractionation

We fractionated the pooled TMT-labeled peptide sample using BPRP HPLC. We used an Agilent (Santa Clara, CA) 1100 pump equipped with a degasser and a photodiode array (PDA) detector (set at 220 and 280 nm wavelength) from Thermo Fisher Scientific (Waltham, MA). Peptides were subjected to a 50 min linear gradient from 5% to 35% acetonitrile in 10mM ammonium bicarbonate pH 8 at a flow rate of 0.8 mL/min over an Agilent (Santa Clara, CA) 300Extend C18 column (5 µm particles, 4.6 mm ID and 220 mm in length). The peptide mixture was fractionated into a total of 96 fractions, which were consolidated into 12. Samples were subsequently acidified with 1% formic acid and vacuum centrifuged to near dryness. Each consolidated fraction was desalted via StageTip, dried again via vacuum centrifugation, and reconstituted in 5% acetonitrile, 5% formic acid for LC-MS/MS processing.

2.6 Liquid chromatography and tandem mass spectrometry

Our mass spectrometry data were collected using an Orbitrap Fusion mass spectrometer (Thermo-Fisher Scientific, San Jose, CA) coupled to a Proxeon EASY-nLC II liquid chromatography (LC) pump (Thermo-Fisher Scientific, San Jose, CA). Peptides were separated on a 100 µm inner diameter microcapillary column packed with ~0.5 cm of Magic C4 resin (5 µm, 100 Å, Michrom Bioresources, Auburn, CA) followed by ~35 cm of GP-18 resin (1.8 µm, 200 Å, Sepax, Newark, DE). For each analysis, we loaded ~1 µg onto the column and separation was achieved using a 2.5 h gradient of 7 to 27% acetonitrile in 0.125% formic acid at a flow rate of ~550 nL/min. Each analysis used the MultiNotch MS3-based TMT method²⁵. The scan sequence began with an MS1 spectrum (Orbitrap; resolution 120,000; mass range 400–1400 m/z; automatic gain control (AGC) target 2×10^5 ;

maximum injection time 100 ms). Precursors for MS2/MS3 analysis were selected using a TopSpeed of 2 sec. MS2 analysis consisted of collision-induced dissociation (quadrupole ion trap; AGC 4×10^3 ; normalized collision energy (NCE) 35; maximum injection time 150 ms). Following acquisition of each MS2 spectrum, we collected an MS3 spectrum using our recently described method in which multiple MS2 fragment ions were captured in the MS3 precursor population using isolation waveforms with multiple frequency notches²⁵. MS3 precursors were fragmented by high energy collision-induced dissociation (HCD) and analyzed using the Orbitrap (NCE 55; AGC 5×10^4 ; maximum injection time 150 ms, resolution was 60,000 at 200 Th).

2.7 Data analysis

Mass spectra were processed using a Sequest-based in-house software pipeline²⁶. Spectra were converted to mzXML using a modified version of ReAdW.exe. Database searching included all entries from the human UniProt database. This database was concatenated with one composed of all protein sequences in the reversed order. Searches were performed using a 50 ppm precursor ion tolerance for total protein level profiling. The product ion tolerance was set to 0.9 Da. These wide mass tolerance windows were chosen to maximize sensitivity in conjunction with Sequest searches and linear discriminant analysis^{26,27}. TMT tags on lysine residues and peptide N termini (+229.163 Da) and carbamidomethylation of cysteine residues (+57.021 Da) were set as static modifications, while oxidation of methionine residues (+15.995 Da) was set as a variable modification.

Peptide-spectral matches (PSMs) were adjusted to a 1% false discovery rate (FDR)^{28,29}. PSM filtering was performed using a linear discriminant analysis, as described previously²⁶, while considering the following parameters: XCorr, Cn, missed cleavages, peptide length, charge state, and precursor mass accuracy. For TMT-based reporter ion quantitation, we extracted the summed signal-to-noise (S/N) ratio for each TMT channel and found the closest matching centroid to the expected mass of the TMT reporter ion. PSMs were identified, quantified, and collapsed to a 1% peptide false discovery rate (FDR) and then collapsed further to a final protein-level FDR of 1%. Moreover, protein assembly was guided by principles of parsimony to produce the smallest set of proteins necessary to account for all observed peptides.

Proteins were quantified by summing reporter ion counts across all matching PSMs using in-house software, as described previously²⁶. PSMs with poor quality, MS3 spectra with more than eight TMT reporter ion channels missing, MS3 spectra with TMT reporter summed signal-to-noise ratio less than 100, or no MS3 spectra were excluded from quantification³⁰. Protein quantitation values were exported for further analysis in Microsoft Excel or SAS JMP. Each reporter ion channel was summed across all quantified proteins and normalized assuming equal protein loading of all 10 samples.

2.8 Data access

RAW files will be made available upon request. Supplemental Tables 1 and 2 list the proteins for Lys-C/trypsin and Lys-C only, respectively, as well as normalized TMT reporter

ion intensities used for quantitative protein profiling. Supplemental Tables 3A, 3B, and 4 list the peptides quantified in the Lys-C/trypsin and Lys-C only datasets, respectively.

3. Results

3.1 Our mass spectrometry workflow quantified over 8,000 proteins across five cell lines in a single experiment

The experimental overview was diagramed in Figure 1. Five pancreatic cell lines - CAPAN-1, HPAC, HPNE, PANC1, and PaSC - were propagated in duplicate. The extracted proteins were digested with Lys-C followed by trypsin (dataset 1) or with only Lys-C (dataset 2). Following TMT labeling, the 10 samples were combined and fractionated by basic pH reversed-phase chromatography and recombined into twelve fractions which were subjected to MultiNotch-MS3 quantitative proteomic profiling ³⁰.

In total, we identified 8,832 non-redundant proteins and 89,877 unique peptides across both datasets (Table 1). The Lys-C/trypsin dataset consisted of 95,006 total peptides, of which 67,553 were unique and resulted in a total of 7,879 quantified proteins with ~1% protein FDR. In comparison, the Lys-C-only dataset consisted of 61,932 total peptides, of which 40,689 were unique and resulted in a total of 6,904 quantified proteins, again with ~1% protein FDR. Of the quantified proteins, 5,951 were in both datasets, while 1,928 were unique to Lys-C/trypsin and 953 were unique to the Lys-C-only dataset (Figure 2A). Less of an overlap was observed at the peptide level. Comparing all unique peptides, only 18,356 were quantified in both datasets, while 49,188 and 22,324 were unique to Lys-C/trypsin and Lys-C-only, respectively (Figure 2B). Such a low overlap may be expected as each digestion condition produced its own set of protease-specific peptides. That is, the Lys-C-only dataset was comprised of mainly of peptides ending in lysine, while the Lys-C/trypsin dataset included both lysine and arginine-terminating peptides.

We determined the relative protein expression levels across each TMT10-plex (TMT signal was represented as a proportion out of 100). We subsequently performed hierarchical clustering on both the Lys-C/trypsin (Figure 3A) and the Lys-C-only dataset (Figure 3B). In both datasets, this analysis revealed tight clustering between biological replicates. As expected, the dendrogram of the clustered cell lines were near identical in both the Lys-C-only and the Lys-C/trypsin datasets.

3.2 Several proteins demonstrated higher expression levels in select pancreatic cell lines

Using analysis of variance (ANOVA), we determined 825 proteins as having significantly different expression levels in the Lys-C/trypsin dataset, while 841 showed the same for the Lys-C-only dataset ($p < 0.01$). We performed K-means clustering analysis on subsets of proteins with significant differences in abundance in a particular cell line to determine protein class enrichment. For each dataset, five clusters were extracted. Proteins in each of these clusters showed higher expression levels in one of the cell lines investigated compared to all others.

We subjected the proteins in each of these clusters to Panther gene classification analysis to assign biological function ³¹. When comparing K-means clustering between datasets,

different numbers of proteins were observed per cluster, but as expected, similar Panther classification categories were enriched for each cell line irrespective of the enzyme used. CAPAN-1 showed 82 and 124 proteins as having significantly higher expression levels in the Lys-C/trypsin and Lys-C-only datasets, respectively. Significant proteins included those with roles in cell communication, structure and adhesion (Figure 4A). From examining the expression profile of HPAC, we note 79 and 38 proteins as having higher expression levels in the Lys-C/trypsin and Lys-C-only datasets, respectively. Significant proteins included those involved in lipid and fatty acid transport and metabolism (Figure 4B). Likewise, HPNE showed 152 and 136 proteins of higher abundance in the Lys-C/trypsin and Lys-C-only datasets, respectively. Significant proteins included those mediating cell adhesion (Figure 4C). Examining the protein expression profile of PANC1 revealed 79 and 64 proteins of higher abundance in the Lys-C/trypsin and Lys-C-only datasets, respectively. Significant proteins included those with functions in mitosis and cell cycle (Figure 4D). Finally, PaSC showed 142 and 252 proteins of higher abundance in the Lys-C/trypsin and Lys-C-only datasets, respectively. Significant proteins included those participating in cell communication, adhesion, and neurogenesis (Figure 4E). We observed that similar GO categories were enriched in both datasets. This result revealed that qualitatively, in a biological context, the data interpretation was not dependent on the enzyme used for mass spectrometric profiling.

In addition to specifying proteins with different relative expression levels among cell lines and associated GO annotations, we also noted proteins with unaltered expression levels across cell lines. In Table 2, we listed the GO annotations for the biological process, cellular component, and molecular function of proteins with coefficients of variation (CV) of <0.2 with respect to their relative expression level across all 5 cell lines. Proteins showing no significant differences in expression levels included those with biological processes of protein transport/localization, ribonucleoprotein complexes, and translation. Likewise, these proteins were enriched in the cellular components of cytosol, Golgi, plasma membrane and ribosome. In agreement, molecular function including RNA binding, ribosome structure and GTPase activity were enriched. In total, 484 proteins were determined to have $CV < 0.2$ and did not significantly change (Benjamini-Hochberg-corrected ANOVA $p > 0.01$) (Supplemental Table 5). These proteins represent a core subset of proteins that have unaltered expression levels across cell lines following nicotine treatment and can be beneficial for interrogating the effects of additional exogenous perturbations in different cell lines.

3.3 Similar figures of merit were observed between the two datasets

We investigated several features inherent to the peptides in each dataset, including peptide length, Sequest XCorr, mass accuracy (PPM, parts per million), precursor intensity, peptide mass, and charge state (Figure 5). Most of the peptide characteristics were similar between the two datasets, including average peptide length (Figure 5A), XCorr value (Figure 5B), mass accuracy (PPM) (Figure 5C), precursor intensity (Figure 5D) and peptide mass (Figure 5E). Of the peptide characteristics investigated, only charge state differed substantially between datasets (Figure 5F). While 60% of the peptide-spectral matches (PSM) were doubly charged in the Lys-C/trypsin dataset, only 35% were doubly charged in the Lys-C-

only dataset. Triply and quadruply-charged peptides were more common in the Lys-C-only dataset (50% and 10%, respectively) than the Lys-C/trypsin dataset (30% and <5%, respectively). We expected such a result as peptides were likely to have more internal charges in the Lys-C-only dataset as Lys-C, unlike trypsin, does not cleave after arginine residues.

3.4 The relative protein expression levels correlated well between the two datasets

We investigated the relative expression level of each protein across the 5 cell lines between datasets. For each cell line, we constructed a correlation plot and determined the Pearson correlation coefficients (r) of the average TMT relative abundance for the proteins in each cell line (Figure 6). All plots showed strong correlation with Pearson correlation coefficients >0.85 . The correlation coefficients were as follows for each cell lines: CAPAN, $r = 0.9041$ (Figure 6A); HPAC, $r = 0.9501$ (Figure 6B); HPNE, $r = 0.9137$ (Figure 6C); PANC1, $r = 0.8906$ (Figure 6D); and PaSC, $r = 0.8734$ (Figure 6E). Comparing the correlation of replicates within each dataset (Supplemental Figure 1), all replicates demonstrated correlation coefficients of approximately 0.9. As expected, these values were slightly greater than the inter-dataset comparisons (Figure 6) that ranged between 0.82 and 0.86.

Discussion

We compared the global proteomes of five pancreatic cell lines using a mass spectrometry-based TMT10-plex strategy. In addition, we investigated whether protein digestion prior to mass spectrometry with sequential Lys-C and trypsin treatment or only Lys-C resulted in significant differences at the protein or peptide level. Combining the datasets, we quantified over 8,800 proteins, for which nearly 6,000 were quantified in both. More specifically, the Lys-C-only digest resulted in 6,904 quantified proteins, while Lys-C/trypsin analysis resulted in 7,879 quantified proteins, with an overlap of 5,951 proteins. Quantitatively, we observed a strong correlation of the relative protein expression levels for replicates within and between datasets. As expected, similar quantitative data were acquired in both datasets.

Evidence for the biological differences among the five cell lines may be elucidated from the proteomic data analyzed herein. In Figure 4, K-means clustering analysis revealed subsets of proteins which are significantly up-regulated in specific cell lines. Many of these are cell communication and adhesion proteins that are specific to one cell line may provide insight into targeting molecules that, in turn, may have an enhanced effect on that particular cell line due to specific cell surface proteins. Another example of an important class of enriched proteins are several kinases and protein phosphatases that appear up-regulated in PANC1 cells. Modulation of phosphorylation signaling by certain kinase inhibitors may have specific effects on PANC1 compared to the other cell lines investigated and may warrant further study³². Likewise, the role of lipid/fatty acid transport, metabolism-related proteins and perturbations thereof, which are enriched in HPAC cells has not to our knowledge been defined and similarly merits additional investigation. As such, determining the biological causes for and implications of the proteomic differences discovered herein provides a resource for the development of hypothesis-driven studies.

On a global proteome level, we showed that sequential digest with Lys-C followed by trypsin resulted in a greater number of identified proteins and peptides. These data indicate that more mass spectrometry-amenable peptides were obtained with the Lys-C/trypsin digestion strategy. However, we deduce from the Venn diagram (Figure 2B) that unique populations of peptides are observed in both datasets. The Lys-C/trypsin digest produces more peptides overall, but using Lys-C alone may allow short tryptic peptides to be sequenced as a result of eliminating an arginine-specific cleavage site that may result in one or two small, unidentifiable peptides. The more mass spectrometry-amenable peptides would be advantageous in targeted assays^{33–35}. Additionally, Lys-C produces a less complex peptide population which, again, may be advantageous for such target-based assays in reducing potential interference due to co-eluting peptides. As such, the choice of enzyme, with respect to its advantages and disadvantages, should be considered on an application-specific basis.

When comparing Lys-C and Lys-C/trypsin-cleaved peptides, peptide length, mass, and PPM (parts per million) mass error distribution did not differ significantly. However, the Lys-C-only digest demonstrated an overall increase in charge state when compared to the Lys-C/trypsin digest. Specifically, while the median charge state is 2 for the Lys-C/trypsin combination, it is 3 for the Lys-C-only dataset. This result is expected as Lys-C cleaves only after lysines and thereby allows for peptides with internal basic arginine residues. Surprisingly, the overall length of peptides did not vary significantly between digestion conditions, with the Lys-C/trypsin dataset having a median of 10 residues and the Lys-C dataset with a median of 11. Altering instrumental settings to those more similar to middle-down proteomics^{36,37} - such as adjusting the collision energy, fragmenting peptides with charge states >2, and/or adjusting the S-RIG voltage - may be necessary to optimize data collection parameters to be more amenable for detecting longer, Lys-C-cleaved peptides.

The use of isobaric tag-based multiplexed proteomic techniques facilitates the high-throughput elucidation of multiple complex proteomes. Using the methodology outlined herein, numerous experiments can be designed. With this workflow, additional pancreatic cell lines can be compared, as well as cancer cell lines from other organs. In contrast to stable isotope labeling by amino acids in cell culture (SILAC)³⁸, TMT-based isobaric protein profiling can be applied to human biological specimens, allowing for the multiplexed proteomic profiling thereof. With additional enrichment steps, the strategy described here may also be applied to the exploration of signaling in basal and activated cell states, by examining post-translational modifications, such as phosphorylation, ubiquitylation, and acetylation. Assays may also be developed to target specific peptides – and by extension proteins³⁹. Likewise, the multiplexing strategy itself can be expanded by incorporating the previously-published 3×3+1 strategy to link multiple experiments^{40,41} or hyperplexing for higher-order multiplexing⁴². In summary, we have used a TMT10-plex strategy to compare quantitatively the proteomic profiles of five pancreatic cell lines. Moreover, we have assembled one of the largest catalogs of proteins and associated peptides – for both Lys-C/trypsin and Lys-C-based digests - from pancreatic cell lines to date in two quantitative mass spectrometry-based experiments.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would also like to thank members of the Gygi Lab and Harper Lab at Harvard Medical School. The PaSC cell line (RLT-PSC) was a gift from Dr. Ralf Jesnowski (German Cancer Research Center).

Source of Funding: This work was funded in part by an NIH/NIDDK grant K01 DK098285 (J.A.P.).

References

1. Hariharan D, Saied A, Kocher HM. Analysis of mortality rates for pancreatic cancer across the world. *HPB (Oxford)*. 2008; 10:58–62. [PubMed: 18695761]
2. Ying H, Dey P, Yao W, et al. Genetics and biology of pancreatic ductal adenocarcinoma. *Genes Dev*. 2016; 30:355–385. [PubMed: 26883357]
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin*. 2016; 66:7–30. [PubMed: 26742998]
4. Fuchs CS, Colditz GA, Stampfer MJ, et al. A prospective study of cigarette smoking and the risk of pancreatic cancer. *Arch Intern Med*. 1996; 156:2255–2260. [PubMed: 8885826]
5. Brandi J, Dando I, Palmieri M, et al. Comparative proteomic and phosphoproteomic profiling of pancreatic adenocarcinoma cells treated with CB1 or CB2 agonists. *Electrophoresis*. 2013; 34:1359–1368. [PubMed: 23463621]
6. Chen R, Dawson DW, Pan S, et al. Proteins associated with pancreatic cancer survival in patients with resectable pancreatic ductal adenocarcinoma. *Lab Invest*. 2015; 95:43–55. [PubMed: 25347153]
7. Kim MS, Zhong Y, Yachida S, et al. Heterogeneity of pancreatic cancer metastases in a single patient revealed by quantitative proteomics. *Mol Cell Proteomics*. 2014; 13:2803–2811. [PubMed: 24895378]
8. Liu X, Zhang M, Go VL, et al. Membrane proteomic analysis of pancreatic cancer cells. *J Biomed Sci*. 2010; 17:74. [PubMed: 20831833]
9. Paulo JA, Gaun A, Gygi SP. Global Analysis of Protein Expression and Phosphorylation Levels in Nicotine-Treated Pancreatic Stellate Cells. *J Proteome Res*. 2015; 14:4246–4256. [PubMed: 26265067]
10. Collisson EA, Sadanandam A, Olson P, et al. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nat Med*. 2011; 17:500–503. [PubMed: 21460848]
11. Deer EL, Gonzalez-Hernandez J, Coursen JD, et al. Phenotype and genotype of pancreatic cancer cell lines. *Pancreas*. 2010; 39:425–435. [PubMed: 20418756]
12. Fogh J, Fogh JM, Orfeo T. One hundred and twenty-seven cultured human tumor cell lines producing tumors in nude mice. *J Natl Cancer Inst*. 1977; 59:221–226. [PubMed: 327080]
13. Gower WR Jr, Risch RM, Godellas CV, et al. HPAC, a new human glucocorticoid-sensitive pancreatic ductal adenocarcinoma cell line. *In Vitro Cell Dev Biol Anim*. 1994; 30A:151–161. [PubMed: 25939163]
14. Lee KM, Yasuda H, Hollingsworth MA, et al. Notch 2-positive progenitors with the intrinsic ability to give rise to pancreatic ductal cells. *Lab Invest*. 2005; 85:1003–1012. [PubMed: 15924149]
15. Lieber M, Mazzetta J, Nelson-Rees W, et al. Establishment of a continuous tumor-cell line (panc-1) from a human carcinoma of the exocrine pancreas. *Int J Cancer*. 1975; 15:741–747. [PubMed: 1140870]
16. Wilson JS, Pirola RC, Apte MV. Stars and stripes in pancreatic cancer: role of stellate cells and stroma in cancer progression. *Front Physiol*. 2014; 5:52. [PubMed: 24592240]

17. Jesnowski R, Furst D, Ringel J, et al. Immortalization of pancreatic stellate cells as an in vitro model of pancreatic fibrosis: deactivation is induced by matrigel and N-acetylcysteine. *Lab Invest.* 2005; 85:1276–1291. [PubMed: 16127427]
18. Ross PL, Huang YN, Marchese JN, et al. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics.* 2004; 3:1154–1169. [PubMed: 15385600]
19. Thompson A, Schafer J, Kuhn K, et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem.* 2003; 75:1895–1904. [PubMed: 12713048]
20. Paulo JA, Kadiyala V, Banks PA, et al. Mass spectrometry-based proteomics for translational research: a technical overview. *Yale J Biol Med.* 2012; 85:59–73. [PubMed: 22461744]
21. Rodriguez J, Gupta N, Smith RD, et al. Does trypsin cut before proline? *J Proteome Res.* 2008; 7:300–305. [PubMed: 18067249]
22. Jekel PA, Weijer WJ, Beintema JJ. Use of endoproteinase Lys-C from *Lysobacter* enzymogenes in protein sequence analysis. *Anal Biochem.* 1983; 134:347–354. [PubMed: 6359954]
23. Paulo JA, Urrutia R, Banks PA, et al. Proteomic analysis of a rat pancreatic stellate cell line using liquid chromatography tandem mass spectrometry (LC-MS/MS). *J Proteomics.* 2011; 75:708–717. [PubMed: 21968429]
24. Paulo JA, Urrutia R, Banks PA, et al. Proteomic analysis of an immortalized mouse pancreatic stellate cell line identifies differentially-expressed proteins in activated vs nonproliferating cell states. *J Proteome Res.* 2011; 10:4835–4844. [PubMed: 21838295]
25. McAlister GC, Nusinow DP, Jedrychowski MP, et al. MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal Chem.* 2014; 86:7150–7158. [PubMed: 24927332]
26. Huttlin EL, Jedrychowski MP, Elias JE, et al. A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell.* 2010; 143:1174–1189. [PubMed: 21183079]
27. Beausoleil SA, Villen J, Gerber SA, et al. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol.* 2006; 24:1285–1292. [PubMed: 16964243]
28. Elias JE, Gygi SP. Target-decoy search strategy for mass spectrometry-based proteomics. *Methods Mol Biol.* 2010; 604:55–71. [PubMed: 20013364]
29. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods.* 2007; 4:207–214. [PubMed: 17327847]
30. McAlister GC, Huttlin EL, Haas W, et al. Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. *Anal Chem.* 2012; 84:7469–7478. [PubMed: 22880955]
31. Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res.* 2013; 41:D377–386. [PubMed: 23193289]
32. Paulo JA, McAllister FE, Everley RA, et al. Effects of MEK inhibitors GSK1120212 and PD0325901 in vivo using 10-plex quantitative proteomics and phosphoproteomics. *Proteomics.* 2015; 15:462–473. [PubMed: 25195567]
33. Gallien S, Kim SY, Domon B. Large-Scale Targeted Proteomics Using Internal Standard Triggered-Parallel Reaction Monitoring (IS-PRM). *Mol Cell Proteomics.* 2015; 14:1630–1644. [PubMed: 25755295]
34. Gallien S, Bourmaud A, Kim SY, et al. Technical considerations for large-scale parallel reaction monitoring analysis. *J Proteomics.* 2014; 100:147–159. [PubMed: 24200835]
35. Peterson AC, Russell JD, Bailey DJ, et al. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol Cell Proteomics.* 2012; 11:1475–1488. [PubMed: 22865924]
36. Cannon J, Lohnes K, Wynne C, et al. High-throughput middle-down analysis using an orbitrap. *J Proteome Res.* 2010; 9:3886–3890. [PubMed: 20557100]
37. Cannon JR, Edwards NJ, Fenselau C. Mass-biased partitioning to enhance middle down proteomics analysis. *J Mass Spectrom.* 2013; 48:340–343. [PubMed: 23494789]

38. Ong SE, Kratchmarova I, Mann M. Properties of ¹³C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *J Proteome Res.* 2003; 2:173–181. [PubMed: 12716131]
39. Jedrychowski MP, Wrann CD, Paulo JA, et al. Detection and Quantitation of Circulating Human Irisin by Tandem Mass Spectrometry. *Cell Metab.* 2015; 22:734–740. [PubMed: 26278051]
40. Paulo JA, Gygi SP. A comprehensive proteomic and phosphoproteomic analysis of yeast deletion mutants of 14-3-3 orthologs and associated effects of rapamycin. *Proteomics.* 2015; 15:474–486. [PubMed: 25315811]
41. Paulo JA, McAllister FE, Everley RA, et al. Effects of MEK inhibitors GSK1120212 and PD0325901 in vivo using 10-plex quantitative proteomics and phosphoproteomics. *Proteomics.* 2015; 15:462–473. [PubMed: 25195567]
42. Dephoure N, Gygi SP. Hyperplexing: a method for higher-order multiplexed quantitative proteomics provides a map of the dynamic response to rapamycin in yeast. *Sci Signal.* 2012; 5:rs2. [PubMed: 22457332]

Figure 1. Experimental overview of the SPS-MS3 analysis

Five pancreatic cell lines were propagated in duplicate. Proteins were extracted and digested either with Lys-C and trypsin sequentially or Lys-C only (producing two separate datasets). The resulting peptides were labeled with TMT, pooled, and fractionated via basic pH reversed-phase high performance liquid chromatography (BPRP-HPLC) prior to MS3 analysis.

Figure 2. Comparing proteins and peptides between the Lys-C/trypsin and the Lys-C-only datasets

Venn diagrams illustrating the A) protein and B) peptide overlap for the Lys-C/trypsin and the Lys-C-only datasets.

Figure 3. Hierarchical clustering of relative protein expression levels across all ten samples
 The heat maps and associated dendrograms for A) Lys-C/trypsin and B) Lys-C-only datasets. Across each row of the heat map, the relative protein expression levels are displayed, such that each row sums to 100%. The scale corresponds to the percentage of total signal across all channels.

Figure 4. K-means clustering and associated Gene Ontology categories
We performed K-means clustering on statistically significant proteins (Benjamini-Hochberg-corrected ANOVA $p < 0.05$) and extracted 5 clusters, which represent proteins of relatively high abundance in each of the 5 cell lines: A) CAPAN-1, B) HPAC, C) HPNE, D) PANC1, and e) PaSC.

Figure 5. Peptide characteristics comparison between Lys-C/trypsin and Lys-C datasets
 We compared the A) Peptide length, B) XCorr score, C) PPM (parts per million mass deviation) per peptide, D) precursor intensity, E) peptide mass, and F) charge state between the Lys-C/trypsin versus the Lys-C-only datasets.

Figure 6. Correlation plots across datasets

Correlation plots and corresponding correlation coefficients (r) were determined for the average relative abundance value of replicates for each protein in all five cell lines: A) CAPAN-1, B) HPAC, C) HPNE, D) PANC1, and E) PaSC.

Table 1

Summary of mass spectrometry data

	Total peptides	Unique peptides	Proteins *	Proteins with altered abundance [‡]
Lys-C	61,932	40,689	6,904	841
Lys-C/Trypsin	95,006	67,553	7,879	825
Non-redundant total:		89,877	8,832	1,326

* Proteins quantified across all 10 TMT channels.

[‡] Proteins with altered expression levels with ANOVA p-values <0.01 following multiple testing correction across the five cell lines.

Table 2

Gene ontology category enrichment of unchanging proteins

Category	Term description	Count *	% †	Fold enrichment ‡	p-value §
GO: Biological Process					
GO:0008104	protein localization	58	13.55	2.72	9.02E-09
GO:0015031	protein transport	53	12.38	2.88	1.18E-08
GO:0030529	ribonucleoprotein complex	45	10.51	3.43	4.91E-10
GO:0016192	vesicle-mediated transport	41	9.58	2.94	2.49E-06
GO:0006412	translation	28	6.54	3.50	5.21E-05
GO: Cellular component					
GO:0005829	cytosol	69	16.12	2.03	4.76E-06
GO:0005794	Golgi apparatus	54	12.62	2.43	8.77E-07
GO:0012505	endomembrane system	48	11.21	2.41	1.15E-05
GO:0031410	cytoplasmic vesicle	35	8.18	2.14	1.62E-02
GO:0009898	plasma membrane	25	5.84	3.10	7.18E-04
GO:0005840	ribosome	19	4.44	3.46	3.85E-03
GO: Molecular function					
GO:0003723	RNA binding	46	10.75	2.63	2.09E-06
GO:0003735	ribosome structure	19	4.44	4.65	7.07E-05
GO:0003924	GTPase activity	18	4.21	3.50	7.99E-03

* Number of proteins in this specified category.

† Percentage of unaltered proteins in specified category.

‡ Fold enrichment = ((identified proteins in category)/((total identified proteins))/((proteins in database in a particular category)/(total proteins in database))).

§ Bonferroni-multiple testing-corrected Fisher Exact test to determine if the proportion of identified genes in each category is not a result of random chance.