Article

Inflect: Optimizing Computational Workflows for Thermal Proteome Profiling Data Analysis

Neil A. McCracken, Sarah A. Peck Justice, Aruna B. Wijeratne, and Amber L. Mosley*

Read Online

| ACCESS | | Metrics & More | | Article Recommendations | | Supporting Information |

ABSTRACT: The CETSA and Thermal Proteome Profiling (TPP) analytical methods are invaluable for the study of protein−ligand interactions and protein stability in a cellular context. These tools have increasingly been leveraged in work ranging from understanding signaling paradigms to drug discovery. Consequently, there is an important need to optimize the data analysis pipeline that is used to calculate protein melt temperatures (T_m) and relative melt shifts from proteomics abundance data. Here, we report a user-friendly analysis of the melt shift calculation workflow where we describe the impact of each individual calculation step on the final output list of stabilized and destabilized proteins. This report also includes a description of how key steps in the analysis workflow quantitatively impact the list of stabilized/destabilized proteins from an experiment. We applied our findings to develop a more optimized analysis workflow that illustrates the dramatic sensitivity of chosen calculation steps on the final list of reported proteins of interest in a study and have made the R based program Inflect available for research community use through the CRAN repository [McCracken, N. Inflect: Melt Curve Fitting and Melt Shift Analysis. *R package version 1.0.3*, 2021]. The Inflect outputs include melt curves for each protein which passes filtering criteria in addition to a data matrix which is directly compatible with downstream packages such as UpsetR for replicate comparisons and identification of biologically relevant changes. Overall, this work provides an essential resource for scientists as they analyze data from TPP and CETSA experiments and implement their own analysis pipelines geared toward specific applications.
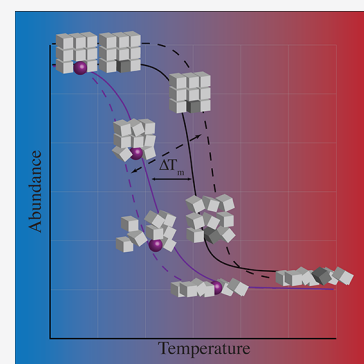
KEYWORDS: TPP, CETSA, protein stability, protein−protein interaction, LC-MS/MS, Inflect, proteomics

■ INTRODUCTION

Within the complex cellular milieu, there has long been an inability to screen for untargeted changes in protein stabilization or destabilization. The advent of Cellular Thermal Shift Analysis (CETSA)[2] and thermal proteome profiling (TPP)[3,4] has rapidly increased our ability to measure changes in protein stability within the context of the intact proteome. The CETSA/TPP workflow begins with cultures of cells exposed to different conditions such as those treated with a small molecule vs vehicle or that have different genetic backgrounds.[5,6] After culture, the cells are lysed in a nondenaturing extraction buffer, and the cellular debris is pelleted and discarded. The supernatant is decanted, aliquoted, and subsequently exposed to a thermal gradient (typically using a PCR machine) ranging from ambient temperature to 90 °C. Alternatively, intact cells can be exposed to a thermal gradient.[4] During this heat treatment, the bulk of the proteins in the solution unfold at a temperature range based on their inherent biophysical properties such as individual structure and their interactions with other partners (including proteins, small molecules, metabolites, etc.). As they unfold, proteins have a greater propensity for aggregation with nearby unfolded proteins and may also precipitate postaggregation. After the short thermal treatment, the heat-treated samples are centrifuged to pellet aggregated protein. The supernatant containing the soluble fraction is decanted once again, proteolytically digested, cleaned up, and labeled with isobaric chemical tags such as tandem mass tag (TMT) reagents for multiplexed analysis by LC-MS/MS. Relative peptide fragment abundance values from a LC-MS/MS experiment are analyzed using a proteomics search program, and the list of reported protein melt curves is further processed manually or by an analysis program to yield a list of proteins affected in the experiment and across replicates. Each step in this described computational process, while having an essential role in the execution of the assay, also has its own potential for adding variability to the final output and conclusions from the study. Challenges with accounting for variability were addressed in part by R- and Python-based pipelines that calculate the melt shift curves from a search algorithm data set.[7,8] The pipeline that accompanied the TPP method, hereafter described as

"TPP-TR," uses raw abundance values from a proteomics program and processes the data through several steps prior to calculating melt temperatures (T_m) and by comparison melt shift values. The operations used in these steps include data filtering, normalization, meltome quantification by curve fitting with correction, and individual protein melt fitting, along with melt temperature and shift calculations.

Despite the availability of resources like TPP-TR that do the heavy lifting in melt shift analysis, there has been no report to date that describes how the chosen analysis steps can impact the final study conclusions. Along with this void, additional challenges remain related to the computational analysis. In order to address these deficiencies related to the downstream computational data analysis, we investigated the existing TPP-TR workflow with aims to better describe and optimize the output of a TPP experiment. Herein, we describe the relative impact of each melt shift analysis step on the total number of proteins and melt shift standard deviations. Our analysis shows that the impact of the data analysis workflow on the results reported from a TPP experiment is significant and rivals the impact of the aforementioned technical issues on the output. We used our findings to develop an analysis workflow that acts as a complementary pipeline to the existing TPP-TR. Our R based analysis pipeline, named "Inflect," is publicly available to the research community so that it can be utilized to aid in the ease and accuracy of CETSA/TPP analysis and for comparison to results that can already be obtained with other analysis programs; furthermore, Inflect will be updated as we continue to update the data analysis workflow. Our findings summarized below will allow researchers to not only better leverage the results from the costly and time-consuming TPP experiments but also act as a resource for those who develop their own algorithms for analysis.

■ MATERIALS AND METHODS

Data Sets

The Peck Justice Data set—the first data set used in our analysis—is one where the investigators illustrated a novel approach for utilizing the TPP-TR workflow to understand the impact of genetic mutations on the melt of the proteome.[6] Their data set was generated from *S. cerevisiae* strains, with mutations in the ORFs encoding proteasome subunits Pup2 and Rpn5. The first and third data sets (p1 and p3) generated from a wild type (WT) strain and mutant pup2 and rpn5 cells were used in our pipeline analysis with the resulting data matrix allowing for comparison between replicates. Raw abundance values reported from a search in Proteome Discoverer were used in our analysis, and the raw data files are available from PRIDE Project ID PXD017222. The Perrin data set—the second data set used in the analysis—was reported by Perrin and co-workers using CETSA to identify targets of Panobinostat in organs and blood of rodents and humans, respectively.[9] The raw data files were analyzed in Proteome Discoverer Version 2.4. Files for the rat kidney and liver were obtained from PRIDE Project ID PXD015427 (sample IDs 02290_F1_R1_P0189540B, 02293_F1_R1_P0189550B, 02066_F1_R1_P0177049B, and 02065_F1_R2_P0177039B), while files for the human PBMC and whole blood data sets were obtained from PRIDE Project ID PXD015373 (files 02032_F1_R1_P0175529B and 02604_F1_R1_P0204098E). Proteome Discover searches for the rat data set were searched against *Rattus norvegicus* NCBI

062312, using the trypsin enzyme setting, a precursor mass tolerance of 20 ppm, and a fragment mass tolerance of 0.5 Da. Regression settings for the search used nonlinear regression with coarse parameter tuning. The same search settings were used for the human blood data sets, except the *Homo sapiens* (092919) database from Uniprot was used.

Search results from human whole blood raw data sets were used as the "WT" data sets in the pipelines, while the peripheral blood mononuclear cell (PBMC) data sets were designated as the "mutant" data sets. Melt shifts (DT_m) were calculated by subtracting the "mutant" melt temperature from the "WT" melt temperature, and destabilized proteins were those with positive shifts while stabilized proteins were those with negative shifts. After searching against the *Rattus norvegicus* proteome, the raw abundance values for each protein were also analyzed using both the TPP-TR and our pipeline in R. The kidney data sets were set as the "mutant" strains, while the liver data sets were designated as "WT" in order that each of these two organs could be compared against the liver data sets. The analysis workflows were the same as those that were previously described for the Peck-Justice data set.[6]

JMP Analysis

The statistical analysis software JMP Pro 14 and Pro 15 were used to randomly vary five of the factors used in the custom TPP analysis workflow along with their respective two-level ranges to generate a full factorial design of experiments (DoE). A total of 32 experiment conditions were created, and each of the 12 data sets discussed in this report were used to evaluate the performance of each combination of steps (288 total experiments). The outputs of the workflow analyses were both the total number of reported significant proteins along with the standard deviation of the observed melt temperatures. These outputs not only allowed for an understanding of the "signal" that came out of each workflow combination but also gave an appreciation of the level of uncertainty from the overall data. Desirable conditions were those where there were high levels of proteins reported with low levels of standard deviation. A definition of significance was used to find melt curves with a calculated R^2 greater than or equal to 0.95 and melt shifts with values greater than 2 standard deviations from the mean melt shift.

R Analysis

Code development and execution were done in RStudio version 1.3.1056. R programs were used first for the development of the TPP analysis pipeline that we describe. The optimized workflow "Inflect" was also coded in R and RStudio.[10] The current version of Inflect contains several functions including readxl,[11] writexl,[12] optimr,[13] data.table,[14] plotrix,[15] tidyr,[16] and ggplot2.[17] Inflect currently analyzes biological replicate data sets separately from each other but summarizes the results from all replicates in several files that describe not only the melt temperatures but which proteins had significant melt shifts across the replicates. The data matrix output can be used directly as an input for UpsetR.[18] R programs were also coded for the multivariate analyses used to determine the relative impact of analysis steps on the final TPP pipeline outputs. Various diagrams shown in the figures were also generated within RStudio. GraphPad Prism 8 was also used for the generation of plots.
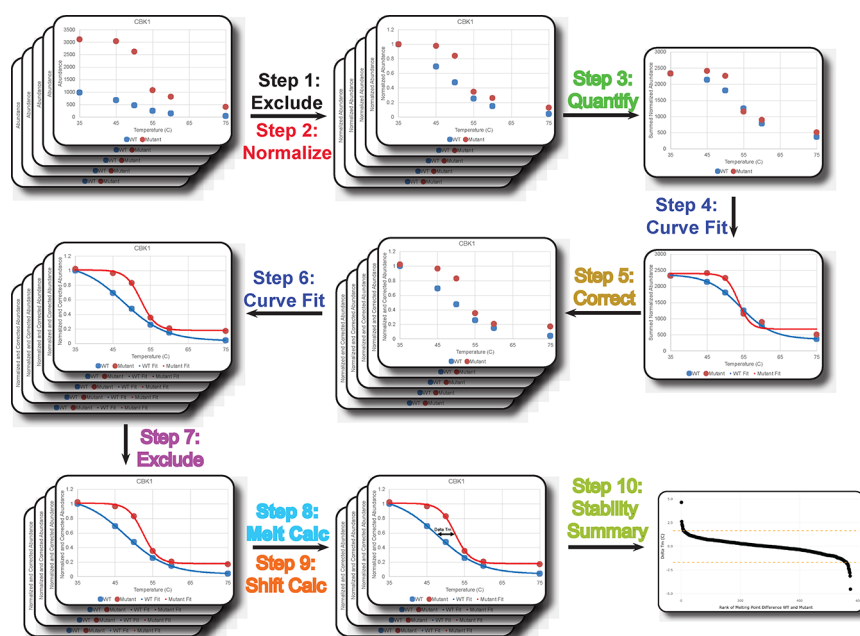
Figure 1. Pictorial representation of the general data analysis workflow from a TPP experiment. Step 1 in the pipeline excludes protein abundance data that do not meet certain criteria, while step 2 normalizes the abundance values for each protein. Step 3 quantifies the total protein meltome (i.e., statistical functions like median abundance in the case of the current TPP-TR package), and the curve fit routine in step 4 uses nonlinear equations to describe the meltome shape. Step 5 calculates correction values. Curve fitting occurs for individual proteins in step 6, while exclusion is used again in step 7 to remove more proteins that do not meet fit quality criteria. The calculation of the melt temperature for each protein occurs in step 8 after which the melt shift is calculated in step 9. The final step 10 involves the summary of the proteins with significant stabilization or destabilization.

Inflect Accessibility

The Inflect code is available through the CRAN repository.[1] The function processes data for each set of replicates that are specified by the user ("Control" and "Condition"). The outputs of this program are as follows: The Results.xlsx file lists the calculated melt shifts and related data for each protein regardless of the criteria (R^2 and standard deviations). The SignificantResults.xlsx file lists the calculated melt shifts and related data for each protein that was considered significant by the criteria above. The Curves folder contains the melt curves (in pdf format) for each protein regardless of the significance of the curve. The Significant Curves folder contains the melt curves (in pdf format) for significant proteins only. The Normalized Condition and Control result files contain the normalized abundance values for each protein and at each temperature. The Waterfall plot shows the calculated melt shifts across the proteome in the study. The melt shifts are plotted in order of value (from highest to lowest). A PDF version of this plot is created in the Curves folder. Outputs also include summary data matrix files that list melt shifts for each significant protein calculated across the replicates (if applicable); these files are amenable to further analysis as desired (i.e., Venn diagramming applications and UpsetR[18]). Example files for Inflect analysis have been included as Supporting Table 1 and Supporting Table 2 for Control 1.xlsx and Condition 1.xlsx from a biological replicate from the pup2-ts studies from prior work.

■ RESULTS AND DISCUSSION

Protein melt shift calculation of TPP experimental data can be delineated into 10 steps, which are summarized pictorially in Figure 1. Step 1 excludes data that do not meet predefined quality control criteria, followed by step 2 that normalizes the abundance values for each protein to the lowest temperature abundance. Step 3 uses statistics to quantify the total protein meltome, and the curve fit routine in step 4 uses nonlinear equations to describe the meltome shape. Step 5 calculates correction values based on the actual and predicted curve fit values, after which constants are then used to correct normalized abundance values for each protein. Curve fitting again occurs in step 6 but on each individual protein abundance that has been corrected. The computationally laborious step 6 fits nonlinear equations to each of the thousands of individual protein melt curves followed by another exclusion in step 7 to remove proteins that do not meet another set of quality control criterion. The calculation of the melt temperature for each protein occurs in step 8, after which the melt shift is calculated in step 9. The final step 10 involves the summary of the proteins with significant stabilization or destabilization based on the shift of all calculated proteins. We utilized two published CETSA/TPP studies[9,19] with a total of 12 separate replicate data sets to define and quantify the relative impact of each data analysis step on the output from an experiment. The Peck-Justice experiments investigated the impact of genetic mutations in proteasome subunits Pup2 and Rpn5 on protein interactions in *S. cerevisiae* on protein thermal stability and protein−protein interactions. The Perrin experiments focused on the use of CETSA to find Panobinostat targets in human blood and rat organs while also using a 10-temperature gradient (without drug) to probe for interactions in crude cell and tissues. These data sets were chosen because they represent more recent executions of the CETSA/TPP workflows, have publicly available raw data that have not been previously normalized, and also feature the use of TMT label sets. The analysis was
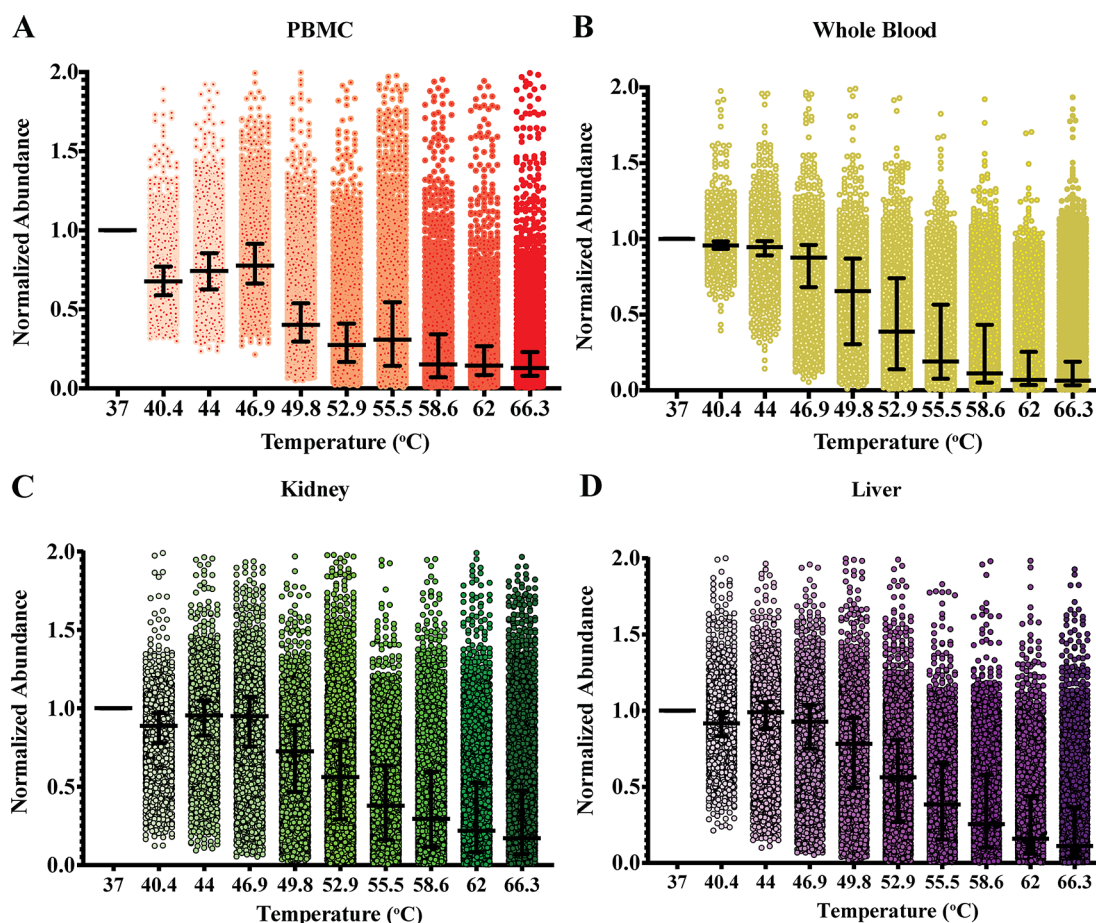
Figure 2. Dot plots of normalized abundance at each temperature relative to the abundance at the lowest temperature. The values are from the R2 data sets for PBMC (A) and human whole blood (B) Perrin data sets. The values are from the R1 and R2 data sets for rat kidney (C) and rat liver (D) Perrin data sets. Median and interquartile ranges are shown as box and whiskers with the maximum axis value at 2 for easier viewing.

demarcated into 10 steps with some key parameters discussed in more detail below.

Data Exclusion, Normalization, and Quantification

Raw abundance values reported by a proteomics search algorithm consist of the relative number of ions detected from a peptide homologous with an associated protein. Prior to beginning the analysis with the TPP-TR pipeline, proteins can be excluded or filtered based on predetermined quality control criteria. The purpose of the filtering is to address the technical variability that is present from the sample harvesting to LC-MS/MS analysis. One criterion used is whether a protein of interest is present in both data sets used to calculate the melt shift. In the event that a protein is observed in only one of the two conditions, the protein will be filtered from the analysis and will not be included in downstream analysis. It is also possible for proteins that are not present in all biological replicates to be excluded from further analysis so as to eliminate low abundance proteins. This step is not unlike data preprocessing done for other types of quantitative proteomics studies to deal with the challenges of missing values from multiple MS-acquired data sets.

The raw untreated data from the proteomic analyses for each of the 12 data sets were analyzed to determine the total number of proteins that were present in each condition so that the level of exclusion could be quantified. We calculated the number of proteins that were present in each condition (i.e., mutant, WT), along with the number of proteins that were not

present in the compared group. In the case of the Peck-Justice data sets, this analysis involved comparing the WT and *rpn5-ts* and *pup2-ts* mutant data sets from two biological replicates. The analysis done on the Perrin data sets compared the melt shift between the proteins in the PBMC and the whole blood data sets. The rodent organ data sets from Perrin were used by comparing the melt shift between the kidney and the liver data. While between tissue melt shifts were not specifically reported by Perrin et al., other studies have recently reported similar types of analyses.[20]

Between 7 and 32% of the total number of proteins would not be used in further analysis if an exclusion step were to be used in the pipeline (Supporting Figure 1). Upon further examination, it was confirmed that the reason for proteins being exclusive to only one of the two data sets (i.e., mutant or WT) was due to a generally low abundance for the protein of interest in the systems studied such that it was not detected across the MS runs that were being compared. These low abundance values have the impact of lowering the statistical values used to describe the total proteome or meltome (downstream analysis). The impact to the quantified meltome would be even more noticeable with the use of the mean function, which is more sensitive to data skew. It is important to note as well that the use of newly reported mass spectrometry techniques like the use of a isobaric carrier channel[21,22] would impact this step in the analysis pipeline and whether proteins have sufficient abundance to be included in
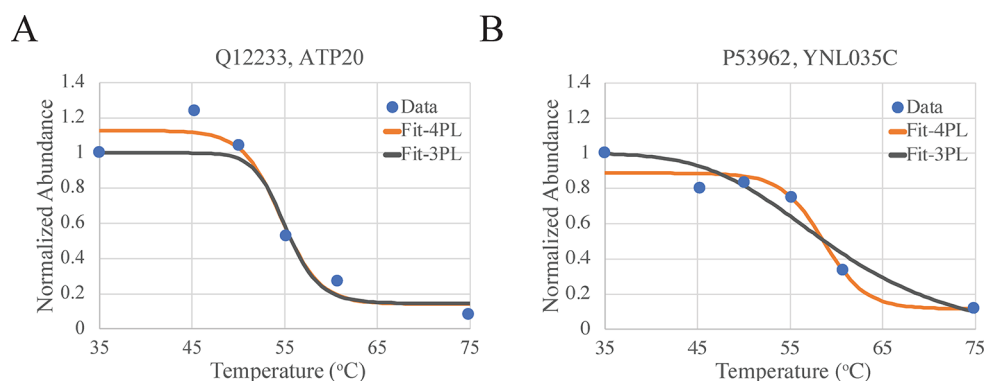
Figure 3. Comparison of 3PL fit and 4PL fit using normalized abundance at each temperature relative to the abundance at the lowest temperature. Data presented are from the Pup2-p1 data set. The 3PL and 4PL equations were used to fit the normalized abundance vs temperature for each protein reported. (A) Q12233 has an R^2 of 0.93 for the 3PL and 0.96 for the 4PL fit. (B) P53962 has an R^2 of 0.93 for the 3PL and 0.96 for the 4PL fit.

further analysis. Additionally, this clearly shows that efforts to increase the overall protein depth of coverage across samples is an important metric for this TPP/CETSA analysis, as it is for global proteomics studies.

After low abundance proteins are excluded, all protein abundance values at each temperature in the heat treatment are divided by the abundance observed at the lowest temperature in the heat treatment. This normalization step not only sets a reference of abundance to the lowest heat treatment temperature (or the theoretical max protein abundance) but it also converts protein abundance values to an equivalent scale so that they can be compared between different conditions. Results from abundance normalization for the Perrin and Peck-Justice data sets are shown in Figure 2 and Supporting Figure 2, respectively. Not only do the data in these sets of dot plots show the spread of abundance values that has been observed to occur at each temperature but the median bars in the plots demonstrate examples of the general departure from ideal sigmoidal shape that can occur. The sources of variability in a multiplexed workflow like this have previously been described to be due to a host of challenges ranging from technical differences to TMT label variation.[23−25]

Post normalization, abundance values across all proteins under each condition are then quantified statistically by use of mean or median functions. The calculated statistic and the corresponding proteome melt curves numerically describe the total abundance of proteins for a particular treatment or mutation. Curve fitting methodologies (described in the next section) are afterward utilized to describe and predict the total protein abundance as a function of heat treatment temperature. Differences between actual and predicted protein abundance are used to calculate correction constants for each heat treatment temperature. The correction process adjusts the abundance of each protein at each temperature for any departures of the global meltome from expected melt behavior. The statistic chosen to describe and create the global melt curve will have an impact on the correction constants calculated for each condition and will consequently have a downstream impact. For example, if the global protein abundance distribution is skewed to lower levels and the mean statistic is used to describe the global protein abundance, the correction constant may end up being a higher value than if the median is used. Curve fitting, used a second time (per the next section), is then used to describe the normalized and

corrected abundance of each individual protein as a function of heat treatment temperature.

Curve Fitting

A melt curve with its sigmoidal shape can be described mathematically by a logistic expression. Two nonlinear equations were used in our evaluation to determine which is optimal for TPP/CETSA studies: a three-parameter log fit (3PL) and a four-parameter log fit (4PL). The 3PL equation, which is solely used in TPP-TR, uses three calculated constants a, b, and Pl to describe the abundance as a function of temperature, T, whereas 4PL uses an extra constant to describe the variability. The 4PL constants a, b, c, and d are equal to the slope at the inflection point, the inflection point, lower plateau, and maximum plateau, respectively. The normalized abundance vs temperature for two proteins in the Peck-Justice data set are shown in Figure 3, and these two curves provide insight into the impact of fit equation on the melt curve. First, the curve fits for the two selected proteins are just below and above our commonly used cutoff criteria of R^2 (0.95) depending on which equation was used. In the case where the 3PL is used, the goodness of fit is below these criteria, whereas the 4PL fit results in a better fit that would be quantified as significant by the workflow (Figure 3A). Another point from this analysis is that the fitting equation can also contribute to the curvature of the melt plot. In the case of Figure 3B, the curvature of the melt is much steeper for the 4PL than for the 3PL fit. The steeper 4PL curve has a more clearly defined inflection (point on the line where curvature changes direction) than the 3PL fit and would have a more defined melt temperature if the inflection point definition were to be used. The 3PL fit, however, has a higher top plateau (crosses the y axis at 1 instead of 0.9) and has a shallower curve down the heat treatment. The impact of the more "stretched out" melt curve could affect the defined melt temperature depending on where the lower plateau of the 3PL curve levels off (greater than 75 °C in Figure 3B).

The overall ability for a mathematical expression to describe the observed variability in a series of data points is commonly done using the coefficient of determination, R^2. This coefficient used in linear systems equals the percentage of variability that is described by the independent variable. The sum of the residual squared error and regression squared error in a nonlinear or logistic system, on the other hand, does not necessarily equal the sum of squared total error, and therefore
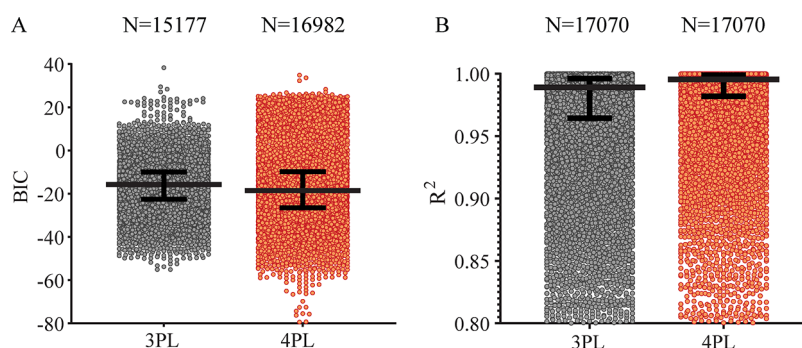
Figure 4. BIC and R^2 for the two fit methods using all data from Peck-Justice data sets. Medians along with interquartile ranges are shown in the box and whisker plots. A portion of the points are shown in each data set; some points were outside of the y-axis range. Statistical analyses utilized unpaired t test with Welch's correction. (A) The BIC for the two fitting methods where medians are −15.4 and −18.7 for 3PL and 4PL, respectively. Maximum values are 24.6 and 228.4 for 3PL and 4PL, respectively, while minimum values are −55.1 and −79.4, respectively. There is a statistically significant difference between the two data sets with $p < 0.0001$. The BIC could not be calculated for all conditions based on the curve shape, thus the difference in N for the two methods. (B) The R^2 for the two fitting methods where medians are 0.970 and 0.995 for 3PL and 4PL, respectively. Maximum values are 1.00 for both methods, and minimum values are −3.383 and −3.614, respectively. There is a statistically significant difference between the two data outputs with $p < 0.0001$.
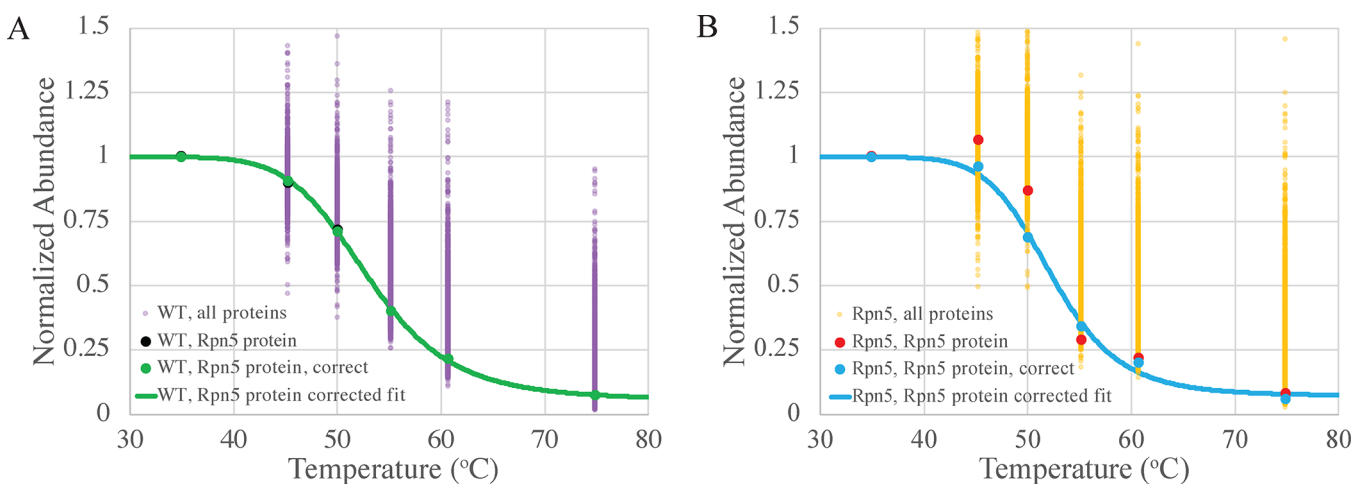

Figure 5. Global normalized abundance at each temperature relative to the abundance at the lowest temperature for both p1 data sets for (A) WT and (B) mutant Rpn5. Individual purple or yellow data points are for the individual proteins in each data set. In the case of panel A, the black and green dots are the Rpn5 protein normalized abundance values in the WT-p1 data set, before and after data correction, respectively. In the case of panel B, the red and blue dots are the Rpn5 protein normalized abundance values in the Rpn5-p1 data set, before and after data correction, respectively. The green line in panel A is the best fit line for Rpn5 protein in the WT-p1 data set, while the blue line in panel B shows the best fit line for the Rpn5 protein the Rpn5-p1 data set.

the R^2 can lie outside of the range of 0 to 1. The limitation makes the determination coefficient a poor measurement of fit for a nonlinear model.[26−28] One measurement of fit that was proposed[29] and evaluated[27] as a more suitable comparator for nonlinear fit than R^2 is the Bayesian information criterion (BIC). The BIC is a quantitative evaluation of fit where more negative values indicate a more optimal regression between conditions. The WT-p1 data set from the Peck-Justice experiments was fit using both of the previously described 3PL and 4PL equations, and the quality of fit was quantified using both the BIC and R^2 (Figure 4). While results shown in this figure indicate that the BIC has a smaller fit distribution for 3PL, the median BIC value was lower for the 4PL, suggesting improved fit with 4PL on average (Figure 4A). The results show that the 4PL fit provides models with a comparatively higher R^2 than the 3PL fit, again suggesting better modeling of the experimental curves with 4PL (Figure 4B). More detailed investigation needs to be done to understand the meaning of the BIC especially as it relates to

comparing results between different curve fitting methods and if specific protein melt properties are better captured by a specific combination of analysis procedures.

After the melt curves are described by their respective equations in step 4 in the analysis workflow, a single melt curve from one of the two conditions is used to calculate the correction factor for all conditions (Figure 1, step 5). The curve and corresponding condition with the best fit, as measured by the R^2, is the condition that is used to calculate the correction constant for both conditions. An example of a result from this TPP normalization is shown in Figure 5 for the Rpn5 protein (Uniprot accession: Q12250) using the Rpn5-p1 data set. These plots show how the melt curve changes as a result of the correction step. The purple (Figure 5A) and yellow (Figure 5B) data points for each of the two plots show the normalized abundance for each protein at each temperature in the WT and mutant data sets, respectively. The green (Figure 5A) and red (Figure 5B) data points are the normalized abundance values for the Rpn5 protein prior to
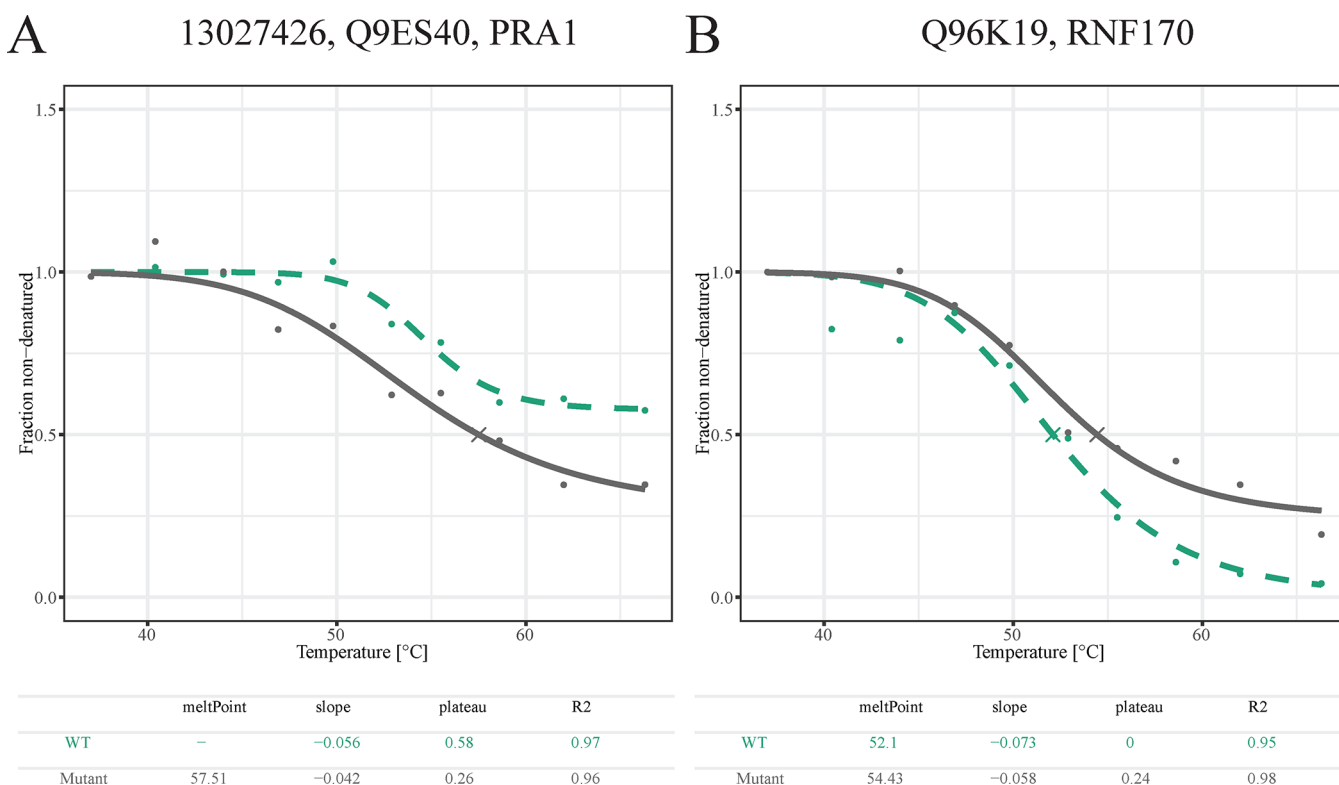
	meltPoint	slope	plateau	R2
WT	−	−0.056	0.58	0.97
Mutant	57.51	−0.042	0.26	0.96

	meltPoint	slope	plateau	R2
WT	52.1	−0.073	0	0.95
Mutant	54.43	−0.058	0.24	0.98

Figure 6. Examples of melt curves from TPP-TR method using Perrin data sets where calculated melt temperature is limited by definition of melt (i.e., temperature at which 50% soluble protein is lost). (A) PRA1 or Q9ES40 from the rat kidney vs liver set #1 and (B) RNF170 or Q96K19 from the blood data sets. The curves generated using this method either did not generate melt temperatures for all of the conditions due to curves not crossing 0.5 (A) or had a calculated melt shift significantly impacted by the definition of the melt temperature (B).

correction, while the black (Figure 5A) and orange (Figure 5B) points are the values post correction. The three parameter fits to the corrected points in A and B are shown in green and blue, respectively, and show no significant correction in the case of the WT data set. The curve shift for the three-parameter fit that resulted from the correction to the mutant data set, on the other hand, was larger than the one observed for the WT data set. These data are informative in a couple ways. First, it is important to note that correction can help to abrogate overall differences in curve shape that likely result from technical variation between samples, making it key for direct sample comparison and for reproducibility analysis across replicates. Second, the data sets included together in the analysis pipeline should be limited to data that are directly being compared, as direct comparison in the correction steps can impact the downstream calculation protein melt temperatures.

Melt Temperature Calculation and Exclusion

The melting point, T_m, of any protein is defined as the temperature at which a protein unfolds from its native state. Due to the fact that all proteins in solution do not unfold en bloc, the melt point is often defined as a transition point in one of two ways. While some sources have defined the melt as the point at which 50% of the protein remains folded,[30] others consider this transition to be the inflection point in a melt curve.[31,32] In order to study the impact of melt definition on the analysis output, the normalized abundance values for the proteins at the highest treatment temperature across all experiments were plotted (Supporting Figure 3). Each of the values at the highest temperature in these plots correspond to the bottom plateau of each melt curve and should ideally cross

at the value of 0 if all protein is denatured and is separated from the liquid. These plots show that despite the median abundance being near zero, there are a large number of proteins in each data set with normalized abundance values which remain above 0.5 at the highest temperature. In the case of the data sets evaluated in this report, up to 22% of proteins were in a replicate data set (control and condition) where the highest temperature abundance was greater than 0.5. The conclusion from these findings is that there are a large number of proteins in our evaluated data sets that departed from expected behavior and could result in a lack of T_m calculation simply because less than half of the starting protein abundance value is lost as a consequence of heat treatment. To ensure that proteins with a variety of biophysical properties are considered within CETSA/TPP workflows, these data suggest that a reconsideration of the chosen melt shift definition is appropriate. It is important to note that this conclusion is supported by others that have even considered use of variables other than melt point in TPP experiments to characterize changes in protein stability.[8] However, melt point calculation has value for the analysis of protein biophysical state across studies and cellular systems including but not limited to comparisons across species (or other systems) or when considering a biophysical state change as a consequence of genetic encoded protein sequence changes.[8,20]

Two of the melt curves from the TPP-TR pipeline analysis are plotted in Figure 6, and the example data in Figure 6A reinforce why some of the proteins do not have calculated melt temperatures as a consequence of the melting point definition alone. The melt curve for protein PRA1 (Uniprot accession: Q9ES40) in the "mutant" condition has a curve with a low
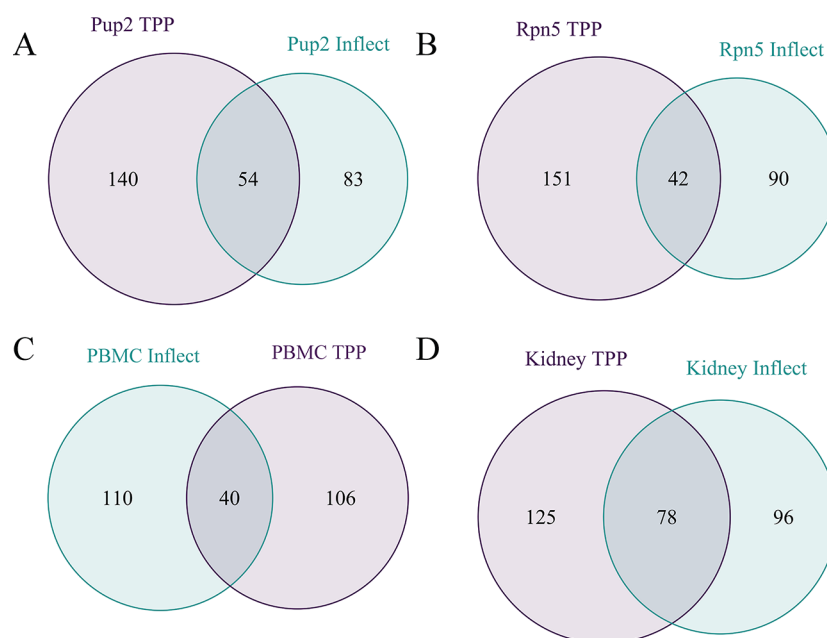
Figure 7. Melt shift evaluations from TPP-TR and our workflow. Comparison of significant proteins observed between the two pipelines. The Peck-Justice data sets for Pup2 and Rpn5 relative to WT are shown in A and B. The Perrin data sets are shown in C through D. Human PBMC relative to human whole blood is shown in C while rat kidney relative to rat liver is shown in D. Note that these Venn diagrams describe common and unique proteins without differentiating between stabilization and destabilization.

plateau higher than 0.5 and consequently does not have a defined melt temperature with a loss of 50% of protein abundance definition. A lack of melt temperature from one of the two conditions being compared ("WT" or "mutant") results in no melt shift calculation, which reduces the overall number of comparative values being reported. In the case of Figure 6B, the definition of the melt being equal to 0.5 can also have a more subtle impact on the calculated shift. The "mutant" curve for RNF170 (Uniprot accession: Q96K19) crosses the 50% loss of protein abundance point at 53 °C, while the inflection point of this same curve is closer to 51%. The melt shift based on a definition of where the curves equal 0.5 can result in a different shift than if the definition is based on the inflection point of the curves. These data clearly show that inflection point is favorable to a 0.5 loss in summed protein signal and may have significant impact on protein inclusion in the downstream data set as well as melt temperature values in final data sets. The definition of T_m could have impactful consequences in data sets in which small molecule/drug treatment or protein sequence changes lead to stabilization or improved solubility of specific proteins as shown in Figure 6A. As a consequence, we strongly recommend that inflection point be considered as advantageous for many CETSA/TPP studies.

Melt Shift Calculation and Stability Summary

Each of the data sets collected by Peck-Justice and Perrin were analyzed using both the Childs et al. TPP-TR workflow and our Inflect workflow in order to understand the relative quantity of significantly stabilized and destabilized proteins, and a comparison of our findings is shown in Figure 7. Our assessment used the same goodness of fit criteria (melt curve R^2 of 0.95) and melt shift significance cutoff (2 standard deviations from mean, a 95% confidence interval) for both pipelines. In the case of the Pup2 and Rpn5 data sets, each curve was compared with its corresponding WT data set in

order to calculate melt shifts. In the case of the Perrin human blood data set, PBMC values were used as the experimental condition while whole blood results were used as the control data set. This analysis was not carried out in the original published report but was used in our analysis for comparison to illustrate potential shifts in melt between a specific fraction of blood and the bulk of human blood matrix. Additionally, the rodent organ data analysis was done by comparing each of the kidney with the liver data sets with the goal of showing comparative protein stabilization/destabilization between each organ and the liver. Once each data set was evaluated using the two workflows, the number of proteins with significant melt shift results were compared. The first comparison (Figure 7A) illustrates the amount of overlap in significant proteins between the two outputs regardless of whether they were stabilized or destabilized. The replicates for the data sets (i.e., Pup2_p1 and Pup2_p3) were also combined in these respective diagrams. These diagrams show that while 54 protein T_m changes were shared between the two methods, there were an even greater number of proteins in each case that were not observed as significant by the other corresponding analysis pipeline. Similar observations were obtained for the other data sets under investigation (Figure 7B–D). These results reflect the strong sensitivity of the analysis output to the selection of data processing steps as described above.

To illustrate the overall positive relationship between calculated melt shift temperatures (between −20 °C and +20 °C) calculated using TPP-TR vs Inflect, we have plotted the correlation between the T_m values from both methods (Figure 8). We would expect most of the findings to be similar between TPP-TR and Inflect for analysis of the same data set, which is what is observed. While there are numerous melt shift temperatures that are clearly similar between the two methods, there are also a significant number of proteins with drastically different values calculated between TPP-TR and Inflect. The Pearson correlation coefficients for the Pup2 and Rpn5 data
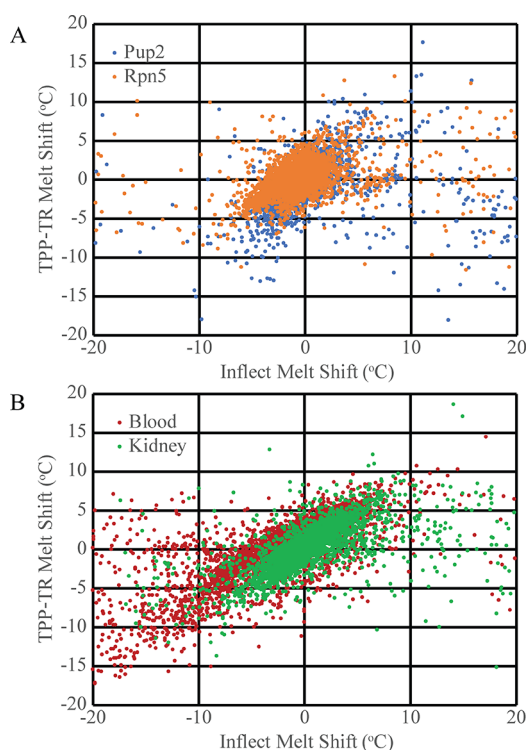
Figure 8. Melt shift values calculated from TPP-TR package vs Inflect package for both (A) Pup2 and Rpn5 data sets along with (B) human blood and rat kidney data sets. Melt shift values between −20 and 20 °C are shown, and 0 values are excluded.

between −20 °C and +20 °C are 0.31 and 0.37, respectively (Figure 8A), while those for the blood and kidney data between −20 °C and +20 °C are 0.76 and 0.60, respectively (Figure 8B).

Not only were there were a large number of proteins that were found to be uniquely significant in the Inflect workflow but we also found that many of these proteins were relevant to the question being asked in the original data set. In the case of the Pup2 data set, Pre1, a component of the 26S proteasome, was reported as destabilized in our data set but was not found to be significant using the same criteria for significance in the TPP-TR workflow. The TPP-TR workflow neglected to find this shift as significant due to the fact that the melt shift for the wild type condition was just below the fit quality criteria of 0.95. As we have shown in Figure 2, the fit quality as determined by R^2 is greatly improved using Inflect as a consequence of 4PL fitting. This protein is of interest due to the fact that the strain used in the reported experiment leveraged a mutation to the Pup2 gene, another component of the 26S proteasome and thereby a potential protein−protein interaction partner.[6] The negative shift in the melt temperature indicates that the *PUP2* mutation resulted in a destabilization of the Pre1 protein with other proteins, potentially those in the 26S proteasome. Other proteasome or ubiquitin-related proteins that were observed as significant in our workflow are shown in Table 1, and the associated melt curves for some of these proteins are displayed in Supporting Figure 4. A similar trend was observed using our workflow to analyze the Rpn5 data sets where a significant number of proteasome subunits were observed with significant melt shifts, which was not uncovered in the initial published study. The proteins of interest are shown in Table 2, with some of the melt shifts

Table 1. Summary of Proteasome Related Proteins (Based on Information from Uniprot) That Had Significant Temperature Shifts from the Pup2 Datasets Using Our Workflow but Were Not Observed to Be Significant in the TPP-TR Workflow

entry	protein names	gene names
P22141	proteasome subunit beta type-4	PRE1 YER012W
P53044	ubiquitin fusion degradation protein 1	UFD1 PIP3 YGR048W
Q12229	UBX domain-containing protein 3	UBX3 YDL091C
P28263	Ubiquitin-conjugating enzyme E2−24 kDa	UBC8 GID3 YEL012W

Table 2. Summary of Proteasome Related Proteins (Based on Information from Uniprot) That Had Significant Temperature Shifts from the Rpn5 Datasets Using Our Workflow but Were Not Observed to Be Significant in the TPP-TR Workflow

entry	protein names	gene names
P40087	DNA damage-inducible protein 1	DDI1 VSM1 YER143W
P53044	ubiquitin fusion degradation protein 1	UFD1 PIP3 YGR048W
P22141	proteasome subunit beta type-4	PRE1 YER012W
P38624	proteasome subunit beta type-1	PRE3 YJL001W J1407
P30657	proteasome subunit beta type-7	PRE4 YFR050C
P30656	proteasome subunit beta type-5	PRE2 DOA3 PRG1 YPR103W P8283.10
P21243	proteasome subunit alpha type-1	SCL1 PRC2 PRS2 YGL011C
P21242	probable proteasome subunit alpha type-7	PRE10 PRC1 PRS1 YOR362C O6650
P23638	proteasome subunit alpha type-3	PRE9 PRS5 YGR135W
P23724	proteasome subunit beta type-6	PRE7 PRS3 PTS1 YBL041W YBL0407
P32379	proteasome subunit alpha type-5	PUP2 DOA5 YGR253C G9155
P25043	proteasome subunit beta type-2	PUP1 YOR157C

from these Rpn5 proteins shown in Figure 9. We observed that in the case of more than one of the "WT" data sets, the melt curves had a higher than average inflection point, which suggests that these proteins have higher than average thermal stability in WT cells. It has already been shown by others that the proteasome and ubiquitin have higher melt temperatures than the average protein and therefore would have implied greater than average thermal stability.[33,34] It is possible that proteins with higher than average intrinsic thermal stability may represent data set outliers; however, to facilitate the development of TPP analysis methods that facilitate biological discovery, it is important that optimized analysis pipelines consider proteins with a wide array of biophysical properties.

PBMCs, or peripheral blood mononuclear cells, are by definition the fraction of blood that play a significant role in the immune response and are enriched in T-cells, B-cells, NK cells, and monocytes.[35] Proteins from the PBMC vs whole blood data that were observed as significant in our workflow but not significant in the TPP-TR workflow while also being relevant to a hematological or immunological function are highlighted in Table 3. The melt curves for three of these proteins with their corresponding comparisons between the two data analysis pipelines are shown in Figure 10. *Rattus norvegicus* TPP experiments that were executed by Perrin et al.
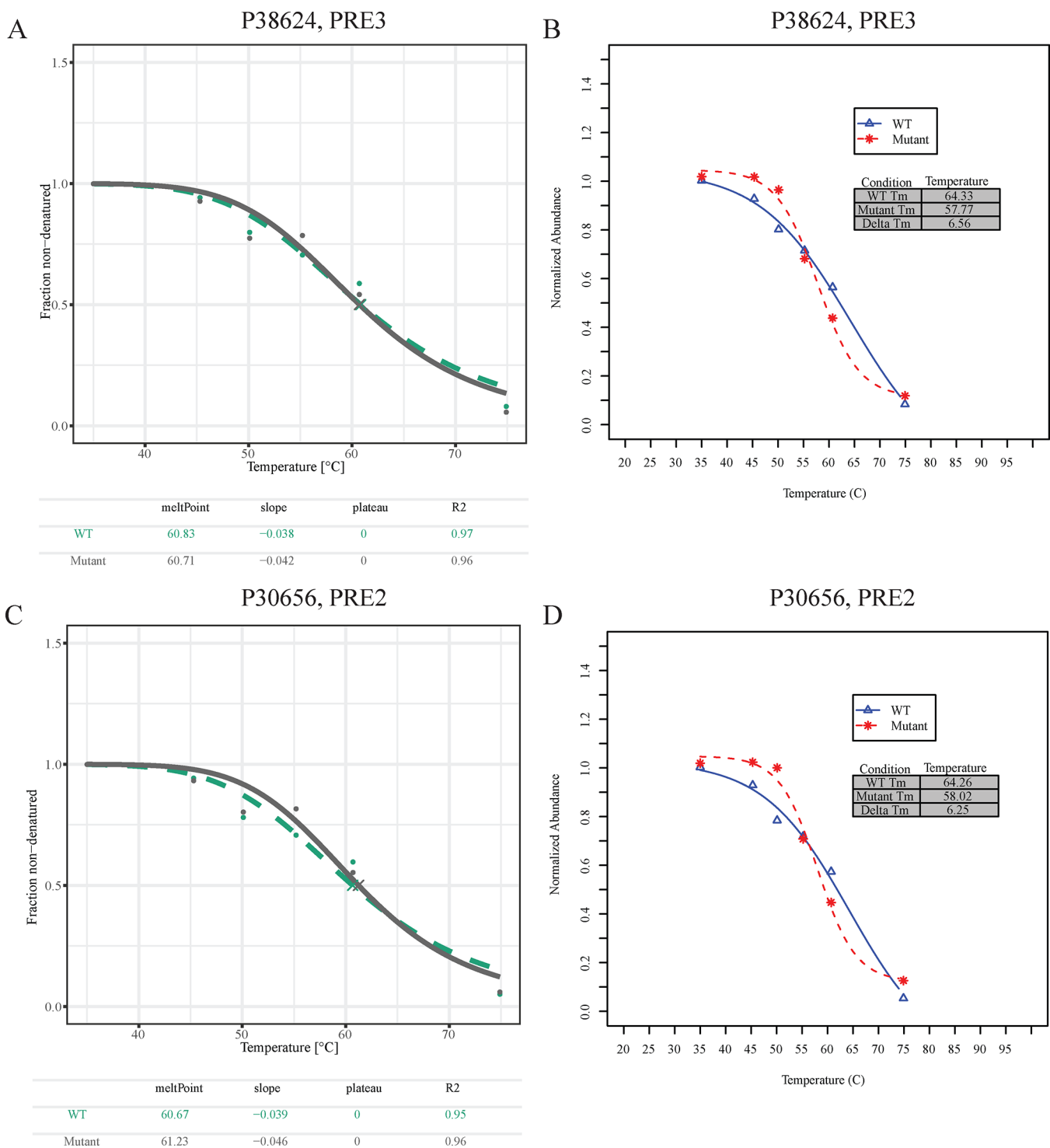
Figure 9. Rpn5_p1 data set melt shifts from Peck-Justice data sets that were reported significant in our workflow but not significant in the TPP-TR workflow. Melt curves generated from the proteasome subunit beta type-1, Pre3 using (A) TPP-TR and (B) our workflow. Melt curves generated from the proteasome subunit beta type-5, Pre2, using (C) TPP-TR and (D) our workflow.

were also examined for relevance to the organ being studied. In the case of the kidney data that were compared with the liver data set, there were several proteins that were unique to our workflow output. Many of these proteins (Table 4) have reported specificity for the kidney based on their functional annotations found within the Uniprot database[36] and thus suggest biological relevance for proteins found by our workflow. Examples of compared melt curves are shown in

Supplemental Figure 5, and each of these curves emphasize unique causes for results not being observed as significant in the TPP-TR workflow as we have discussed throughout this work.

Biological and/or Technical Replicate Analysis

TPP experiments should always include biological replicates to facilitate discovery of reproducible changes that occur across experiments. Our analysis in this article treats replicate

Table 3. Summary of Proteins Related to Blood or Leukocytes or Expressed in Blood Cells (Based on Information from Uniprot) That Had Significant Temperature Shifts from the Perrin PBMC vs Whole Blood Datasets Using Our Workflow but Were Not Observed to Be Significant in the TPP-TR Workflow

entry	protein names	gene names
Q96EK5	KIF-binding protein	KIFBP KBP KIAA1279 KIF1BP
O14672	disintegrin and metalloproteinase domain-containing protein 10	ADAM10 KUZ MADM
P05556	integrin beta-1	ITGB1 FNRB MDF2MSK12
P78325	disintegrin and metalloproteinase domain-containing protein 8	ADAM8MS2
Q96LC7	sialic acid-binding Ig-like lectin 10	SIGLEC10 SLG2 UNQ477/ PRO940
Q8WU39	marginal zone B- and B1-cell-specific protein	MZB1MEDA7 PACAP HSPC190
Q9NR28	diablo homologue, mitochondrial	DIABLO SMAC
Q9NP99	triggering receptor expressed on myeloid cells 1	TREM1
Q15631	translin	TSN
O76031	ATP-dependent Clp protease ATP-binding subunit clpX-like, mitochondrial	CLPX
O95232	Luc7-like protein 3	LUC7L3 CREAP1 CROP O48
Q68CP4	heparan-alpha-glucosaminide N-acetyltransferase	HGSNAT TMEM76

experiments individually and does not use mathematical operations to group replicates prior to the aforementioned TPP analysis workflow, which is also allowed in TPP-TR. One of the reasons for individual analysis of replicates is that proteins may not always be detected across biological replicates and thereby not pass filtering criteria as a consequence of a lack of detection in separate mass spectrometry experiments rather than a lack of change in thermal stability. We are working to address this challenge in other work using isobaric trigger channels to increase protein detection across biological replicates; however, that work is outside the scope of this current study.[21,22] In this implementation, the Inflect workflow processes each replicate data set separately (i.e., control and condition) through the analysis pipeline. Results from the function are saved to data matrix files that can be opened in programs such as Excel that summarize the melt shifts for each protein across each of the replicate experiments. The reporting of melt shifts by our function allows for the user to understand the reproducibility of the resulting melt shifts for each protein and allows user defined cutoffs for downstream significance reporting. Proteins with common stabilized or destabilized proteins between replicates provide further evidence that the change in stability is less transient in nature and potentially more significant. An example of the comparison of significant proteins between replicates is shown in the UpsetR generated plot[18] in Figure 11, which is used in our typical workflows to identify overlap of significant changes identified by TPP experiments. As shown, there were 15 destabilized proteins that were common between the two Pup2 replicates and only one protein commonly stabilized between the two replicate sets. The

data matrix output is formatted for direct use in the UpsetR function from the Inflect workflow.

Multivariate Analysis to Assess the Impact of Each Step on End Result

The melt shift analysis pipeline involves a series of steps that are used to prepare the raw abundance data for analysis, describe the prepared data using fitting routines, and calculate melt temperature shifts from each protein. In order to ascertain the relative influence of each step on the output of the analysis pipeline, a program was written in R which allowed for a multivariate analysis to be executed using various combinations of each step to be run in series with the goal of quantifying the respective output. All 12 data sets described in this article were used in the evaluation. The results from the analysis of the conditions were analyzed using the Fit Model routine in JMP. Fit Model (using a factorial to second degree) was used to describe the observed variability in the outputs as a function of the five factors (step 1, exclusion; step 3, total quantitation; step 4, curve fit; step 6, curve fit; step 8, melt definition) also using the experiment as a specific factor. The relative impact of each pipeline factor along with the experiment being evaluated on the two outputs was quantified by comparing the scaled estimates of each factor in the model. The fit of each of the four models was good (>90% of the variability was described in our model) and the order of the effects with respect to importance to the models is summarized in Table 5.

Results from our evaluation show that along with the standard deviation of the melt shift, the second curve fitting routine in step 6 is the most important of the variables studied in affecting the number of observed proteins and the standard deviation of the melt shift values. The initial curve fit equation had the next level of importance in the results from our experiment. Interestingly, the use of the 3PL fit for the initial curve fit of the meltome was more beneficial for increasing the number of proteins with significant melt shifts, while in the case of the second curve fit, the 4PL was more beneficial for detecting proteins. This finding of a step-specific benefit for different curve fitting routines indicates a need for a better understanding of how curve fitting equations affect the T_m and melt shift calculations. The definition of the melt temperature (50% reduction in protein abundance vs inflection point) and the statistical quantitation of the proteome (specifically, the use of mean or median to describe the total abundance) had less of an impact on these parameters but were still statistically significant. It should be noted that while the statistics indicate that there is a slight benefit to using the 0.5 definition over the inflection point in the number of proteins observed, we determined that the use of the inflection point over the 50% value allows for analysis of proteins that have nontraditional melt curves (where the lower plateau is not equal to 0). The exclusion step used in our *in silico* experiments did not have a statistically significant impact on the number of proteins or the standard deviation of the melt shifts.

■ CONCLUSIONS

Our group investigated the TPP melt shift analysis workflow and the evaluation found that it is beneficial to use the 4PL curve fit over the 3PL fit in order to define proteins with significant melt shifts. To facilitate comparison of our workflow with other data processing pipelines for TPP/CETSA, we have developed the R-based program Inflect. We also show that the number of equation parameters used in curve fitting can
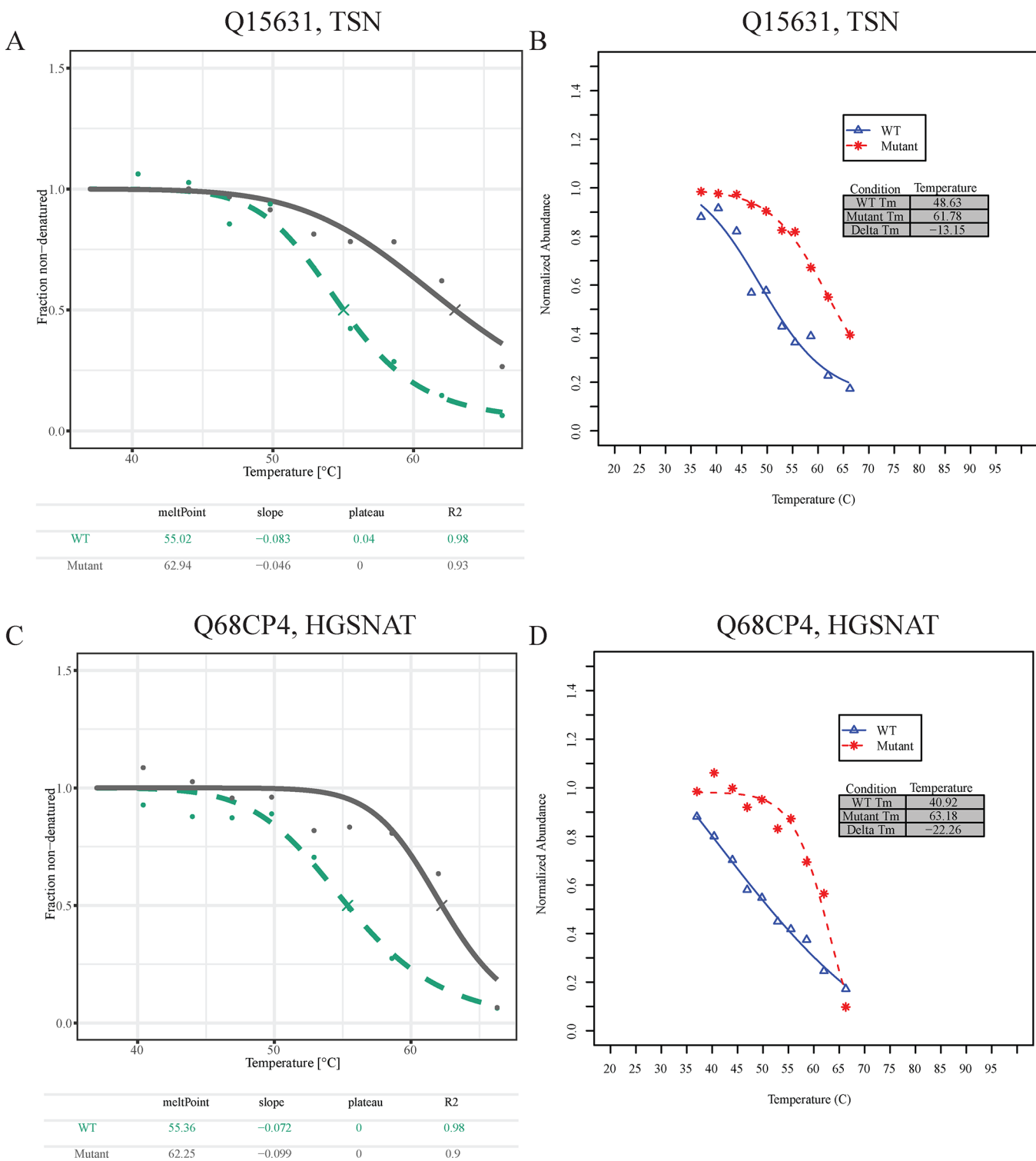
Figure 10. Example human PBMC vs whole blood melt shifts from Perrin data sets that were reported significant in our workflow but not significant in the TPP-TR workflow. Melt curves generated from the Translin, TSN, using (A) TPP-TR and (B) our workflow. Melt curves generated from the Heparan-alpha-glucosaminide N-acetyltransferase, HGSNAT, using (C) TPP-TR and (D) our workflow.

dramatically affect the number of proteins observed in the melt shift analysis.

Results from our optimized workflow show that the second curve fitting routine in step 6 is the most important of the variables studied in affecting the output of the analysis pipeline (number of observed proteins and the standard deviation of the melt shifts). The initial curve fit equation had the next level

of importance in the results from our experiment. These findings reflect on how critical the choice of fitting algorithm and melt curve equation are to the results of a TPP study. It is also important to note that it is possible that four parameter log fit equations are preferable to three parameter log fits in multiple cases including the first curve fitting step. Chosen optimization criteria (R^2 vs BIC) may also be important for

Table 4. Summary of Proteins Related to Kidney Function (Based on Information from Uniprot) That Had Significant Temperature Shifts from the Perrin Rat Kidney vs Liver Datasets Using Our Workflow but Were Not Observed to Be Significant in the TPP-TR Workflow

GI number	entry	protein names	gene names
74229032	Q9WTN5	two pore calcium channel protein 1	Tpcn1 Tpc1
19705453	P08011	microsomal glutathione S-transferase 1	Mgst1 Gst12
117647218	P38718	mitochondrial pyruvate carrier 2	Mpc2 Brp44
158534075	O88370	phosphatidylinositol 5-phosphate 4-kinase type-2 gamma	Pip4k2c Pip5k2c
31543514	Q9R0J8	legumain	Lgmn Prsc1
40786432	Q10758	keratin, type II cytoskeletal 8	Krt8 Krt2−8
40018562	Q6P756	adaptin ear-binding coat-associated protein 2	Necap2
164519066	P14270	cAMP-specific 3′,5′-cyclic phosphodiesterase 4D	Pde4d
32189350	Q09073	ADP/ATP translocase 2	Slc25a5 Ant2
29789307	O88658	kinesin-like protein KIF1B	Kif1b
21489985	Q8VHE9	all-trans-retinol 13,14-reductase	Retsat Ppsig Rmt7

Table 5. Summary of Factors in Order of Their Relative Importance in Describing the Variability in Number of Significant Proteins and Standard Deviation of Melt Temperatures

relative importance	impacting total number of significant proteins	impacting standard deviation of observed melt temperatures
1 (higher)	step 6, protein curve fit	step 6, protein curve fit
2	step 4, protein curve fit	step 4, protein curve fit
3	step 3, statistical quantitation of proteome	step 8, melt definition
4 (lower)	step 8, melt definition	step 3, statistical quantitation of proteome

impact on the number of proteins observed; however, these parameters have been shown to have a strong impact on specific proteins (Figure 6).

While our work provides extensive insight into the data analysis from TPP experiments, there is still ample opportunity for improvement. As more TPP experiments are executed, the experimental procedure will improve along with the methodology for isobaric labeling. As technical improvements occur, the data analysis pipeline should also be evaluated to determine whether the steps used are most appropriate and beneficial for maximizing the number of biologically relevant results.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jproteome.0c00872.

> Supplemental Figure 1, number of quantitated proteins in each of the 12 experiments analyzed in this report; Supplemental Figure 2, dot pots of normalized abundance at each temperature relative to the abundance at the lowest temperature from the Peck-Justice data sets; Supplemental Figure 3, comparison of normalized abundance at each temperature relative to the abundance at the lowest temperature for each of the highest temperatures in the Peck Justice data sets; Supplemental Figure 4, Pup2_p1 data set melt shifts from Peck-Justice data sets that were reported significant in our workflow but not significant in the TPP-TR workflow; and Supplemental Figure 5, example rat kidney vs liver (replicate 1) melt shifts from Perrin data sets that were reported significant in our workflow but not significant in the TPP-TR workflow (PDF)
> Supplemental Table 1: Control 1.xlsx file for example analysis using Inflect. This file contains data from the wild-type TPP experiment 1 data set from Peck Justice et al., *Journal of Biological Chemistry* **2020** (XLSX)
> Supplemental Table 2: Condition 1.xlsx file for example analysis using Inflect. This file contains data from the pup2-ts mutant TPP experiment 1 data set from Peck Justice et al., *Journal of Biological Chemistry* **2020** (XLSX)
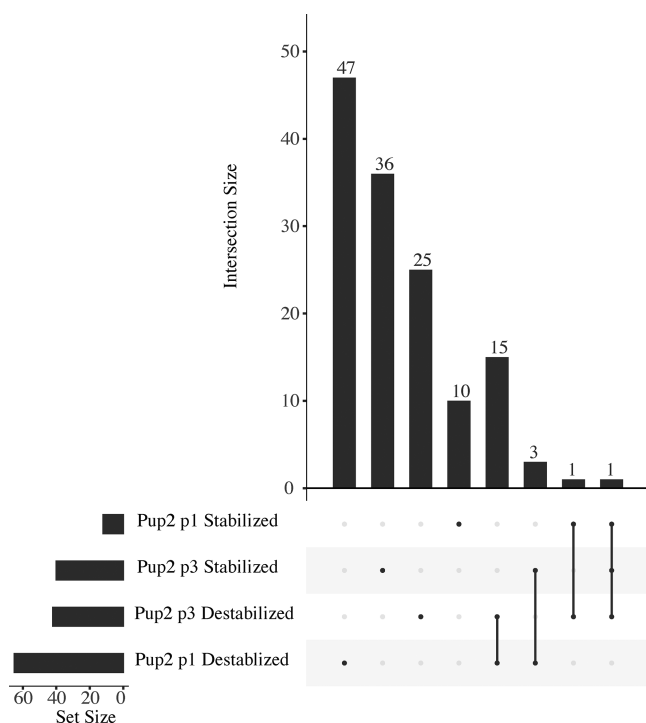

Figure 11. Upset plot generated using UpsetR function in R.[18] Significant proteins from Pup2 p1 and p3 (melt curve $R^2 > 0.95$ and melt shift greater than 2 standard deviations from mean melt shift). Number of proteins common to the replicates or stabilization/destabilization state are highlighted in the Upset plot using dots and lines between the rows.

future development of analysis algorithms as current significance criteria leverage linear system-based functions like the coefficient of determination. The definition of the melt temperature and total quantitation statistics had lessof an

■ AUTHOR INFORMATION

Corresponding Author

Amber L. Mosley − *Department of Biochemistry and Molecular Biology and Center for Computational Biology and Bioinformatics, Indiana University School of Medicine,*

Indianapolis, Indiana 46202, United States; ⓘ orcid.org/0000-0001-5822-2894; Email: almosley@iu.edu

Authors

Neil A. McCracken − *Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, Indiana 46202, United States*

Sarah A. Peck Justice − *Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, Indiana 46202, United States*

Aruna B. Wijeratne − *Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, Indiana 46202, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jproteome.0c00872

◼ ABBREVIATIONS

CETSA, Cellular Thermal Shift Assay; TPP, Thermal Proteome Profiling; TR, temperature range

◼ REFERENCES

(1) McCracken, N. Inflect: Melt Curve Fitting and Melt Shift Analysis. *R package version 1.0.3*, 2021.

(2) Jafari, R.; Almqvist, H.; Axelsson, H.; Ignatushchenko, M.; Lundback, T.; Nordlund, P.; Martinez Molina, D. The cellular thermal shift assay for evaluating drug target interactions in cells. *Nat. Protoc.* **2014**, *9* (9), 2100−2122.

(3) Franken, H.; Mathieson, T.; Childs, D.; Sweetman, G. M.; Werner, T.; Togel, I.; Doce, C.; Gade, S.; Bantscheff, M.; Drewes, G.; Reinhard, F. B.; Huber, W.; Savitski, M. M. Thermal proteome profiling for unbiased identification of direct and indirect drug targets using multiplexed quantitative mass spectrometry. *Nat. Protoc.* **2015**, *10* (10), 1567−93.

(4) Savitski, M. M.; Reinhard, F. B. M.; Franken, H.; Werner, T.; Savitski, M. F.; Eberhard, D.; Molina, D. M.; Jafari, R.; Dovega, R. B.; Klaeger, S.; Kuster, B.; Nordlund, P.; Bantscheff, M.; Drewes, G. Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science* **2014**, *346* (6205), 1255784.

(5) Mateus, A.; Kurzawa, N.; Becher, I.; Sridharan, S.; Helm, D.; Stein, F.; Typas, A.; Savitski, M. M. Thermal proteome profiling for interrogating protein interactions. *Molecular Systems Biology* **2020**, *16* (3), e9232.

(6) Peck Justice, S. A.; Barron, M. P.; Qi, G. D.; Wijeratne, H. R. S.; Victorino, J. F.; Simpson, E. R.; Vilseck, J. Z.; Wijeratne, A. B.; Mosley, A. L. Mutant thermal proteome profiling for characterization of missense protein variants and their associated phenotypes within the proteome. *J. Biol. Chem.* **2020**, *295* (48), 16219−16238.

(7) Childs, D.; Kurzawa, N.; Franken, H.; Doce, C.; Savitski, M.; Huber, W. TPP: Analyze thermal proteome profiling (TPP) experiments. *R package version 3.14.0*, 2019.

(8) Childs, D.; Bach, K.; Franken, H.; Anders, S.; Kurzawa, N.; Bantscheff, M.; Savitski, M. M.; Huber, W. Nonparametric Analysis of Thermal Proteome Profiles Reveals Novel Drug-binding Proteins. *Mol. Cell Proteomics* **2019**, *18* (12), 2506−2515.

(9) Perrin, J.; Werner, T.; Kurzawa, N.; Rutkowska, A.; Childs, D. D.; Kalxdorf, M.; Poeckel, D.; Stonehouse, E.; Strohmer, K.; Heller, B.; Thomson, D. W.; Krause, J.; Becher, I.; Eberl, H. C.; Vappiani, J.; Sevin, D. C.; Rau, C. E.; Franken, H.; Huber, W.; Faelth-Savitski, M.; Savitski, M. M.; Bantscheff, M.; Bergamini, G. Identifying drug targets in tissues and whole blood with thermal-shift profiling. *Nat. Biotechnol.* **2020**, *38* (3), 303−308.

(10) R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/.

(11) Bryan, H. W. a. J. readxl: Read Excel Files. *R package version 1.3.1.* https://CRAN.R-project.org/package=readxl.

(12) Ooms, J. writexl: Export Data Frames to Excel 'xlsx' Format. *R package version 1.3.1*, 2020.

(13) Nash, J. C. {optimr}: A Replacement and Extension of the 'optim.' https://CRAN.R-project.org/package=optimr.

(14) Srinivasan, M. D. a. A. data.table: Extension of 'data.frame.' *R package version 1.13.2.* https://CRAN.R-project.org/package=data.table.

(15) Lemon, J. Plotrix: a package in the red light district of R. *R-News* **2006**, *6* (4), 8−12.

(16) Wickham, H. tidyr: Tidy Messy Data. *R package version 1.1.2.* https://CRAN.R-project.org/package=tidyr.

(17) Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, 2016.

(18) Gehlenborg, N. *UpSetR: A More Scalable Alternative to Venn and Euler Diagrams for Visualizing Intersecting Sets*, 2019.

(19) Peck Justice, S. A.; Qi, G.; Wijeratne, H. R. S.; Victorino, J. F.; Simpson, E. R.; Wijeratne, A. B.; Mosley, A. L. Temperature sensitive Mutant Proteome Profiling: a novel tool for the characterization of the global impact of missense mutations on the proteome. *bioRxiv* **2019**, 2019.12.30.891267.

(20) Jarzab, A.; Kurzawa, N.; Hopf, T.; Moerch, M.; Zecha, J.; Leijten, N.; Bian, Y.; Musiol, E.; Maschberger, M.; Stoehr, G.; Becher, I.; Daly, C.; Samaras, P.; Mergner, J.; Spanier, B.; Angelov, A.; Werner, T.; Bantscheff, M.; Wilhelm, M.; Klingenspor, M.; Lemeer, S.; Liebl, W.; Hahne, H.; Savitski, M. M.; Kuster, B. Meltome atlas-thermal proteome stability across the tree of life. *Nat. Methods* **2020**, *17* (5), 495−503.

(21) Chua, X. Y.; Mensah, T.; Aballo, T.; Mackintosh, S. G.; Edmondson, R. D.; Salomon, A. R. Tandem Mass Tag Approach Utilizing Pervanadate BOOST Channels Delivers Deeper Quantitative Characterization of the Tyrosine Phosphoproteome. *Molecular Cellular Proteomics* **2020**, *19* (4), 730−743.

(22) Justice, S. A. P.; McCracken, N. A.; Victorino, J. F.; Wijeratne, A. B.; Mosley, A. L., Boosting detection of low abundance proteins in thermal proteome profiling experiments by addition of an isobaric trigger channel to TMT multiplexes. *bioRxiv* **2020**, 2020.12.30.424894.

(23) Maes, E.; Valkenborg, D.; Baggerman, G.; Willems, H.; Landuyt, B.; Schoofs, L.; Mertens, I. Determination of Variation Parameters as a Crucial Step in Designing TMT-Based Clinical Proteomics Experiments. *PLoS One* **2015**, *10* (3), e0120115.

(24) Christoforou, A. L.; Lilley, K. S. Isobaric tagging approaches in quantitative proteomics: the ups and downs. *Anal. Bioanal. Chem.* **2012**, *404* (4), 1029−37.

(25) Karp, N. A.; Huber, W.; Sadowski, P. G.; Charles, P. D.; Hester, S. V.; Lilley, K. S. Addressing accuracy and precision issues in iTRAQ quantitation. *Mol. Cell Proteomics* **2010**, *9* (9), 1885−97.

(26) Colin Cameron, A.; Windmeijer, F. A. G. An R-squared measure of goodness of fit for some common nonlinear regression models. *Journal of Econometrics* **1997**, *77* (2), 329−342.

(27) Spiess, A.-N.; Neumeyer, N. An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach. *BMC Pharmacol.* **2010**, *10*, 6.

(28) Wang, X.; Jiang, B.; Liu, J. S. Generalized R-squared for detecting dependence. *Biometrika* **2017**, *104* (1), 129−139.

(29) Schwarz, G. Estimating the Dimension of a Model. *Ann. Statist.* **1978**, *6* (2), 461−464.

(30) Ku, T.; Lu, P.; Chan, C.; Wang, T.; Lai, S.; Lyu, P.; Hsiao, N. Predicting melting temperature directly from protein sequences. *Comput. Biol. Chem.* **2009**, *33* (6), 445−450.

(31) Grøftehauge, M. K.; Hajizadeh, N. R.; Swann, M. J.; Pohl, E. Protein-ligand interactions investigated by thermal shift assays (TSA) and dual polarization interferometry (DPI). *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2015**, *71*, 36−44.

(32) Fluidigm. *Melting Curve Analysis, Instruction Guide for Biomark System.* Fluidigm; Vol. PN 68000118 E1.

(33) Lenkinski, R. E.; Chen, D. M.; Glickson, J. D.; Goldstein, G. Nuclear magnetic resonance studies of the denaturation of ubiquitin. *Biochim. Biophys. Acta, Protein Struct.* **1977**, *494* (1), 126−30.

(34) Volkening, J. D.; Stecker, K. E.; Sussman, M. R. Proteome-wide Analysis of Protein Thermal Stability in the Model Higher Plant Arabidopsis thaliana. *Mol. Cell Proteomics* **2019**, *18* (2), 308−319.

(35) Champion, L.; Pawar, S.; Luithle, N.; Ungricht, R.; Kutay, U. Dissociation of membrane-chromatin contacts is required for proper chromosome segregation in mitosis. *Mol. Biol. Cell* **2019**, *30* (4), 427−440.

(36) Consortium, T. U. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47* (D1), D506−D515.