Development of Data Independent Acquisition Techniques for the Analysis of Protein Mixtures by Tandem Mass Spectrometry

Jarrett D. Egertson

A dissertation submitted in partial fulfillment of the
requirements for the degree of:

Doctor of Philosophy

University of Washington

2013

Reading Committee:

Michael J. MacCoss, Chair

James Bruce

William Noble

Program Authorized to Offer Degree: Department of Genome Sciences

University of Washington

Abstract

Development of Data Independent Acquisition Techniques for the Analysis of Protein Mixtures by Tandem Mass Spectrometry

Jarrett D. Egertson

Chairperson of the Supervisory Committee:
Associate Professor Michael J. MacCoss
Department of Genome Sciences

Novel algorithms and data acquisition methods designed to improve the ability to identify and quantify proteins in complex biological mixtures using tandem mass spectrometry are presented. An algorithm for *de novo* correction of mass-to-charge measurements in MS/MS spectra acquired in a shotgun proteomics workflow is described. The correction of *m/z* measurements makes the interpretation of MS/MS spectra easier and can increase the number of peptides identified in a bottom-up shotgun proteomics experiment. The technique is described as *de novo* because it can detect systematic mass measurement error in a collection of tandem mass spectra without any input besides the spectra themselves. To improve quantitation of peptides, a multiplexed data independent acquisition (DIA) technique is presented. Data acquired using DIA can be queried for data on virtually any protein. These data can be used to compare the abundances of proteins in multiple samples with high sensitivity (often higher than using MS data). However, DIA techniques have typically had low precursor selectivity compared to popular data dependent acquisition (DDA) techniques resulting in mixed MS/MS spectra that are difficult to interpret and prone to chemical interference. A method for overcoming this limitation of DIA by multiplexing (and computational demultiplexing) is described. DIA and DDA techniques are used to study the response of the budding yeast *S. cerevisiae* to treatment

with rapamycin. The response to rapamycin is of interest because it extends lifespan in a wide range of organisms. Spectral counting for peptide quantitation using DDA data is compared to quantitation using MS/MS fragment ion chromatograms integrated over time with DIA data.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would never be in a position to write this thesis without the patience, encouragement, and support of my parents, sister, and the rest of my family. There were many teachers early on in my academic career that fed my curiosity and pushed me to meet my potential including Sandy Ruddick, Joe Skinner, Craig Williams, David Fleischman, and Mary Ann Rall. I of course need to thank my first true lab partner Andrew Redman. Our experiments performed in the chemistry laboratory, as well as in his garage, were some of my earliest and most creative. On attending UCLA, I was lucky enough to be surrounded by the best friends, mentors, and teachers imaginable. Dr. Matteo Pellegrini gave me my first project as an undergraduate researcher that was entirely mine and provided my first taste of being an independent researcher. I have to thank Dr. Parag Mallick for always having time to talk science with me, but most importantly for always caring about where I ended up and being my friend to this day. Without the encouragement of Dr. Pamela Hurley at the Department of Molecular, Cellular, and Developmental Biology to nurture my inquisitive nature, I would have never made it to graduate school. Thank you for taking the time to listen to a confused undergraduate considering his options and the courage to offer your sage, simple advice ("You belong in graduate school.")

It is almost impossible for me to thank everyone that had an impact on me during my pursuit of a degree in Genome Sciences at the University of Washington. Dr. Michael J. MacCoss was the perfect mentor for me. He came into the laboratory every day to have fun doing science. He was always patient and cared about his graduate students beyond just the work they did in the lab. I have to acknowledge the many members of the MacCoss lab I have worked with. They were always supportive and extremely patient in teaching me the way of the

lab. My thesis committee deserves special recognition for offering guidance beyond what I could have ever expected. From the first committee meeting, I realized that they were "on my side" and were only there to make sure I got the most out of my time as a graduate student. I enjoyed all of our meetings and feel grateful to have had the time to squeeze as much knowledge out of them over the course of my graduate studies as possible. The rest of the Department of Genome Sciences: the faculty, the IT team, and the administrative team are all exceptional. My fellow graduate students and friends within and outside the department deserve special recognition. Graduate school can be trying at times and my friends were there to help me deal with it time and time again.

Dedication

This work is dedicated to Mom, Dad, and Jessica.

Chapter 1 – Introduction

Section 1: The Proteome

The normal functioning of a cell is intimately tied to the normal functioning of the proteins within the cell. This was demonstrated in dramatic fashion with the description of the first "molecular disease" by Linus Pauling in 1949 in which a change in a single protein (haemoglobin) was identified as the cause of a debilitating human illness: sickle-cell anaemia[1]. This insight was made through direct measurement of the physical characteristics of haemoglobin by electrophoresis in normal and affected individuals. Later work revealed that the disease was caused by a single amino acid substitution which altered the kinetics of oxygen binding in haemoglobin[2, 3]. This striking example of the close link between protein function and phenotype is one of many motivating direct observation of the protein complement of an organism, commonly referred to as the "proteome"[4].

Proteins are the primary functional unit of the cell, performing diverse functions such as providing cellular structure, catalysing chemical reactions, mediating transport across the plasma membrane, intra, and extra-cellular signalling. The 20 primary amino acids that proteins are built from have a wide range of biochemical characteristics (i.e., hydrophilic, hydrophobic, basic, acidic, neutral) which give rise to a class of macromolecules with an even wider range of biochemical and structural characteristics required for such broad cellular functionality.

The vast range of protein biochemistry, structure, and abundance makes the comprehensive measurement of proteins in a complex sample a daunting task. In human plasma, the most abundant protein is 10^{10} times more abundant than the least abundant protein[5]. Contrast this

with messenger RNA, which has a dynamic range of ~100 and can be amplified. Additionally, protein solubility ranges from extremely insoluble membrane proteins, to very soluble cytosolic proteins. This range of solubility makes it difficult to manipulate proteins in solution. Finally, proteins show extreme variation in size, ranging from hundreds to millions of Daltons.

With the emergence of the capability to make comprehensive genomic, and nearly comprehensive transcriptome measurements, it may be tempting to use DNA or RNA measurements as a proxy for protein function. Unfortunately, these measurements are poor proxies for protein function. Protein and mRNA transcript abundances are correlated, but only modestly. Absolute measurements indicate that mRNA and protein abundance is correlated by ~40% in prokaryotes and eukaryotes[6]. Protein abundance can vary post-transcriptionally due to differential rates of translation and degradation. Post-translational modification (i.e. phosphorylation, acetylation, glycosylation) of proteins can quickly alter protein function and is often used in cell signalling (e.g. receptor tyrosine kinase cascades) and transcriptional regulation (e.g. histone post translational modification). Finally, protein localization can play a large role in the functioning of a protein. For example, the IFI16 protein detects pathogenic DNA and stimulates an immune response in mammals. When the protein is localized to the nucleus, it acts as a sensor for herpes virus; when localized to the cytoplasm, it senses cytoplasmic DNA viruses. The localization of the protein is modulated by acetylation of a nuclear localization sequence on the protein[7]. Understanding of this phenomenon was only made possible by direct observation of the protein.

Mass spectrometry has emerged as the most effective technology for measuring proteins in

complex biological mixtures. Most often, the biochemical complexity of the mixture is reduced by enzymatic digestion of proteins to peptides. These peptides are then ionized and their mass-to-charge (*m/z*) measured in the mass spectrometer. By measuring the *m/z* of peptides and, often, collision-induced fragments of those peptides, thousands of proteins can be identified in a sample over ~4 orders of magnitude of abundance[8, 9] with sensitivity in the attomolar range for some proteins. In order of increasing difficulty, differential abundance of proteins (increased or decreased abundance), relative abundance (amount of increase/decrease), and absolute abundance of proteins can be measured. Post translational modifications can be enriched and measured. Protein localization can also be detected when mass spectrometry is paired with cellular fractionation techniques. Mass spectrometry is easily multiplexed, allowing modern mass spectrometers to measure tens, hundreds, or thousands of proteins in a single one hour experiment depending on the application. The flexibility, sensitivity, dynamic range, and multiplexed-nature of mass spectrometry have made it the primary tool for proteomics research.

Section 2: Proteomic Analysis by Tandem Mass Spectrometry

This thesis focuses on "bottom-up" proteomics, in which proteins are digested to peptides using one or more proteolytic enzymes prior to mass analysis. Most of the digestion described in this thesis is done using trypsin, which cleaves specifically at the carboxyl side of lysine or arginine[10]. Tryptic digestion reduces the mass range of the analytes from hundreds to millions of Da for proteins to ~500-5000 Da for peptides, which simplifies mass analysis. Additionally, the digested peptides are easier to solubilize and separate chromatographically, which extends the range of proteins that can be analysed in a single

sample.

Prior to mass analysis, peptides must be converted to gas-phase ions and introduced into the mass spectrometer. Electrospray ionization (ESI) is a popular technique due to its compatibility with on-line liquid chromatography separation (discussed later), high sensitivity, high mass range for ionization, and its low incidence of peptide fragmentation during the ionization process. In nanoESI, the solvated peptides flow through a column (fused-silica is common) at nanoliter per minute flow rate and emit from a microcapillary with a needle-like tip (~5-10 μM in diameter) separated from the inlet of the mass spectrometer by a small distance (millimetres). A voltage applied at the emitter causes the emission of solvated peptide droplets which are drawn to the inlet of the mass spectrometer by an electrostatic field. The droplets are continuously de-solvated, eventually releasing gas-phase peptide ions as they travel toward the inlet of the mass spectrometer.

A mass spectrometer contains a mass analyser which separates the peptide ions by mass-to-charge (*m/z*) using magnetic and/or electrical fields as well as a means for detecting the signal from a population of ions at a particular *m/*z. The mass spectrometer measures a mass spectrum showing the ion signal across a range of *m/*z values. Only low complexity samples are practical for direct analysis without fractionation or separation. A protein digest from a biological sample is far too complex to be analysed in this manner; the dynamic range in peptide abundance would far exceed the dynamic range of ionization by ESI and detection by the mass spectrometer. A common technique to reduce the complexity of the sample being analysed by the mass spectrometer at any point in time is to couple reversed-phase high performance liquid chromatography (RP-HPLC) separation to the ESI

emitter. Using this technique, peptides are gradually emitted into the mass spectrometer over time in order of increasing hydrophobicity.

The peptides are first loaded onto a liquid chromatography column packed with a solid stationary phase of inert beads coated with alkyl chains (C12 and C18 are popular choices). The LC column is coupled directly to the ESI emitter. During mass analysis, a linear gradient of gradually increasing organic buffer content (e.g. acetonitrile) flows through the column causing peptides of increasing hydrophobicity to be eluted over time (up to 4 hours) as MS spectra are continually acquired. Nano-flow liquid chromatography (nanoLC) uses sub-µL/minute flow rates allowing analysis of small sample amounts (ng-µg of protein) and improved sensitivity when coupled with nanoESI.

The complexity of biological protein digest is problematic when trying to identify peptides based on MS scans alone. Modern, commercially-available mass analyzers such as the Fourier transform–ion cyclotron resonance analyzer (FT-ICR), orbitrap, and time of flight (TOF) analyzer can measure peptide *m/z* with accuracy of ~1-5 ppm and resolving power >35,000. With such high resolution and mass accuracy, the charge of the peptide can be determined and the monoisotopic mass deduced. Unfortunately, even with very accurate monoisotopic mass measurements, peptides cannot be uniquely identified based on their intact mass alone in high-complexity samples such as cell lysates. An extreme example is isobaric peptides (those with the same mass) which are impossible to differentiate by MS regardless of the accuracy and resolving power of the mass analyser. For this reason, tandem mass spectra (MS/MS) are acquired along with MS spectra to provide the increased selectivity required for peptide identification.

MS/MS spectra analyse the masses of fragments of a peptide, rather than the intact peptide. First, a population of intact peptide precursor ions is isolated. Next, the population of precursor ions is activated by collisions with an inert gas (e.g., helium or nitrogen) to generate fragment ions in a process known as collision-induced dissociation (CID, one of many fragmentation techniques). Finally, an MS/MS spectrum is generated by measuring the *m/z* and intensity of the resultant fragment ions.

When using CID, the majority of fragment ions are formed by breakage of a single peptide bond along the backbone of a peptide ion resulting in the formation of b and y-ions (Figure 1.1). The MS/MS spectrum of a peptide can be used to deduce the partial or entire amino acid sequence of the peptide by analysing the spacing of b or y-ions formed by fragmentation at successive locations along the peptide backbone. The additional selectivity of LC-MS/MS makes it possible to identify thousands of peptides in a "discovery" experiment, or to quantify 10-100 peptides with high-accuracy and sensitivity in a "targeted" experiment.

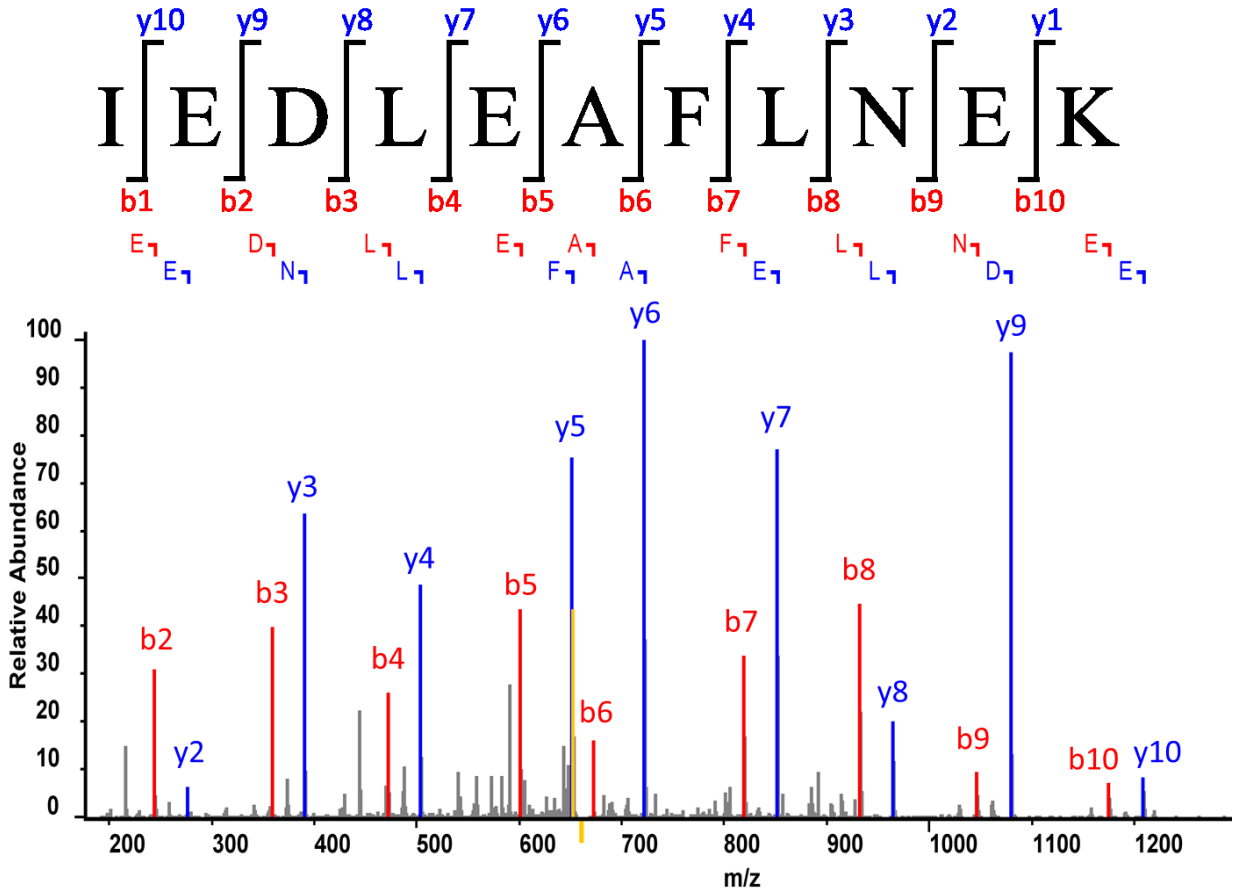


Figure 1.1: An Annotated MS/MS Spectrum.

An annotated MS/MS spectrum acquired on the peptide IEDLEAFLNEK++ is shown. The cleavage locations along the peptide backbone are shown with the resulting singly-charged fragment ions highlighted in the MS/MS spectrum.

Section 3: Discovery and Targeted LC-MS/MS Experiments

LC-MS/MS approaches can be categorized into "discovery" and "targeted" approaches. Discovery approaches attempt to identify as many proteins in a single sample as possible, as well as changes in abundance across samples. Thousands of proteins can be identified in a complex sample using these techniques. Targeted experiments trade comprehensiveness for sensitivity and accuracy of quantitation by constraining the LC-MS/MS analysis to 10-100 predetermined peptide targets. The workflows of discovery and targeted approaches differ

primarily in 1) MS/MS acquisition scheme, 2) instrumentation, and 3) post-acquisition data analysis.

Data dependent acquisition (DDA; Figure 1.2) coupled with database searching of MS/MS spectra has been the workhorse of discovery proteomics since it was pioneered in 1994[11]. DDA methods use information from MS scans to direct acquisition of MS/MS scans toward high-abundance intact peptide precursors. These MS/MS spectra can be analysed by a database search algorithm to confidently assign an amino acid sequence to a subset of these spectra. In HPLC-MS/MS, peptides usually elute off of the chromatography column over a period of 15-45 seconds. This is a long time relative to the scan speed of modern instruments which can acquire spectra at ~10 Hz. Because the instrument scans quickly relative to peptide elution time, multiple scans can be acquired before the composition of peptides eluting from the LC column changes significantly. DDA takes advantage of this by first acquiring an MS "survey scan" to determine the *m/z* of precursors eluting off of the column and then chooses 5-15 of these precursors to acquire MS/MS spectra on in subsequent scans. Despite the rapid scan speed of the instrument, the complexity of biological samples is high enough that many precursors are never sampled by MS/MS. DDA preferentially targets highly abundant precursors, which often generate high quality MS/MS spectra. However, low abundant precursors may never be sampled.

Instrumentation capable of acquiring high mass accuracy MS scans and rapid MS/MS scans are well-suited for DDA. Examples of this type of instrumentation are quadrupole-time-of-flight (Q-TOF) and hybrid instruments combining a fourier-transform mass analyser (i.e. FT-ICR, or Orbitrap) for MS acquisition with an ion trap analyser for MS/MS

acquisition. On modern instrumentation of these types, it is not uncommon to acquire more than 20,000 MS/MS spectra in a single sample analysis. Having so many spectra necessitates automated algorithms for determining peptide sequence from MS/MS spectra and determining the statistical confidence in these assignments.

Database search algorithms such as SEQUEST, Mascot, and X! Tandem use genomic information to generate theoretical MS/MS spectra that can be matched to acquired spectra. A library of open reading frame (ORF) sequences for the organism being analysed is digested *in silico*, and a theoretical fragmentation spectrum is generated for every potential peptide. The top scoring candidate spectrum matching each experimental spectrum (peptide-spectrum-match; PSM) is reported by the database searching algorithm. An empirical null distribution of PSM scores can be generated by performing a search using the same parameters against a decoy database of randomized or reversed ORF sequences[12]. The distributions of PSM scores searched against the target and decoy database can be used to assign measures of statistical significance (p-value or q-value[13]) to each PSM or each peptide identified. Having accurate precursor mass information improves the results of database searching by narrowing down the set of potential candidate spectra for each potential MS/MS spectrum.

Once peptides have been identified, their abundance can be compared to other samples by comparing the number of MS/MS spectra matched to a peptide (spectral counting), or the area under the elution curve of intact peptide precursor in the MS scan. These “label-free” relative quantification strategies can only provide estimates of changes in peptide abundance. The absolute abundance of peptides cannot be determined from these data because peptides

have variable ionization efficiencies which can be impacted by co-ionizing species (matrix effects). Additionally, the estimate of relative change in peptide abundance can be confounded by matrix effects and the potential of one or both of the peptide measurements being outside the linear range of quantitation, which varies by peptide. However, label-free data is still useful for detecting peptides that increase or decrease in abundance between samples.

Targeted approaches acquire MS/MS scans exclusively on a set of peptide targets determined prior to acquisition. These approaches are useful in situations where an investigator has interest in 10-100 peptides, for example, when studying a particular pathway or set of candidate biomarkers. Selected reaction monitoring (SRM; Figure 1.3) using triple-quadrupole instrumentation is the most common technique for targeted experiments[14]. Triple-quadrupole mass spectrometers are beam-type instruments which filter out all but a small window (0.2-0.7 *m/z*) of precursor mass from the ionized peptide beam in the first quadrupole, collisionally activate that precursor ion beam in the second quadrupole, and then filter a product-ion mass in the third quadrupole. The filtered fragment ion beam is allowed to strike the detector for a pre-determined "dwell" time. To quantify peptides by SRM, a handful of precursor-product ion pairs are selected for each peptide and monitored sequentially in a repeated cycle. A cycle time of 1-3 seconds is required so peptides can be sampled multiple times as they elute. These data are analysed by plotting the signal from each fragment ion of a peptide over time (extracted ion chromatogram; XIC). When the peptide elutes, the fragment ions of that peptide will co-elute in a roughly Gaussian-shaped peak. The total area under this peak is integrated to measure the signal for that peptide. The

high selectivity and sensitivity (down to <100 copies/cell in *S. cerevisiae* total lysate[15]) of SRM makes it a popular choice for quantifying small sets of peptides in complex matrices.

Both DDA and SRM share the same drawback: non-comprehensive sampling of peptides by MS/MS. In targeted SRM experiments, this is by design: MS/MS spectra are only acquired for the peptides of interest. In DDA, this happens because there is only time to stochastically sample a subset of the MS mass range. In both cases, the data is only useful for making inferences about the peptides that were sampled by MS/MS. This is particularly inconvenient because an investigator may develop a hypothesis about particular proteins, but find that they cannot obtain information from their data on these proteins because they were never sampled by MS/MS. With DDA the lack of signal for a peptide can be difficult to interpret. It can be because the peptide is not present in the sample, the peptide is below the limit of detection of the instrument, the peptide was not sampled by MS/MS, or the peptide was sampled but not identified with high enough confidence in the database search. Additionally, the stochastic sampling of DDA means that different MS/MS spectra will be acquired in replicate runs of the same sample, further confounding analysis. In replicate runs, up to 30% of peptides identified can differ[16].

Data independent acquisition (DIA) techniques (Figure 1.4) which comprehensively sample a wide *m/z* range by MS/MS, combines characteristics of targeted and discovery approaches. With these data, fragment ion chromatograms for peptides can be extracted and used for quantitation similar to a targeted experiment. Unlike targeted experiments, these fragment ion chromatograms can be extracted for any peptide within the wide m/z range sampled. The next section focuses on the potential advantages of DIA techniques for

discovery and targeted proteomics.

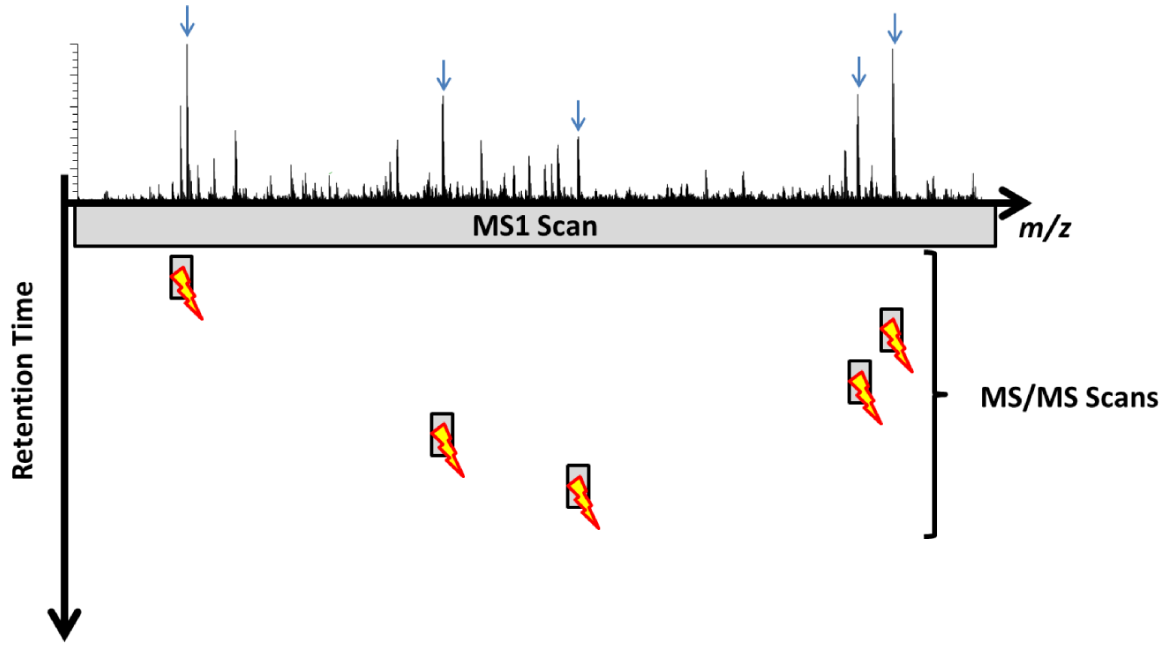


Figure 1.2: Data Dependent Acquisition.

In data dependent acquistion, an MS "survey" scan is acquired followed by 5-15 MS/MS scans targeting high abundance precursors (blue arrows) detected in the survey scan. The MS/MS scans normally have an isolation width between 1 and 3 *m/z*. This cycle is repeated throughout the course of sample analysis.

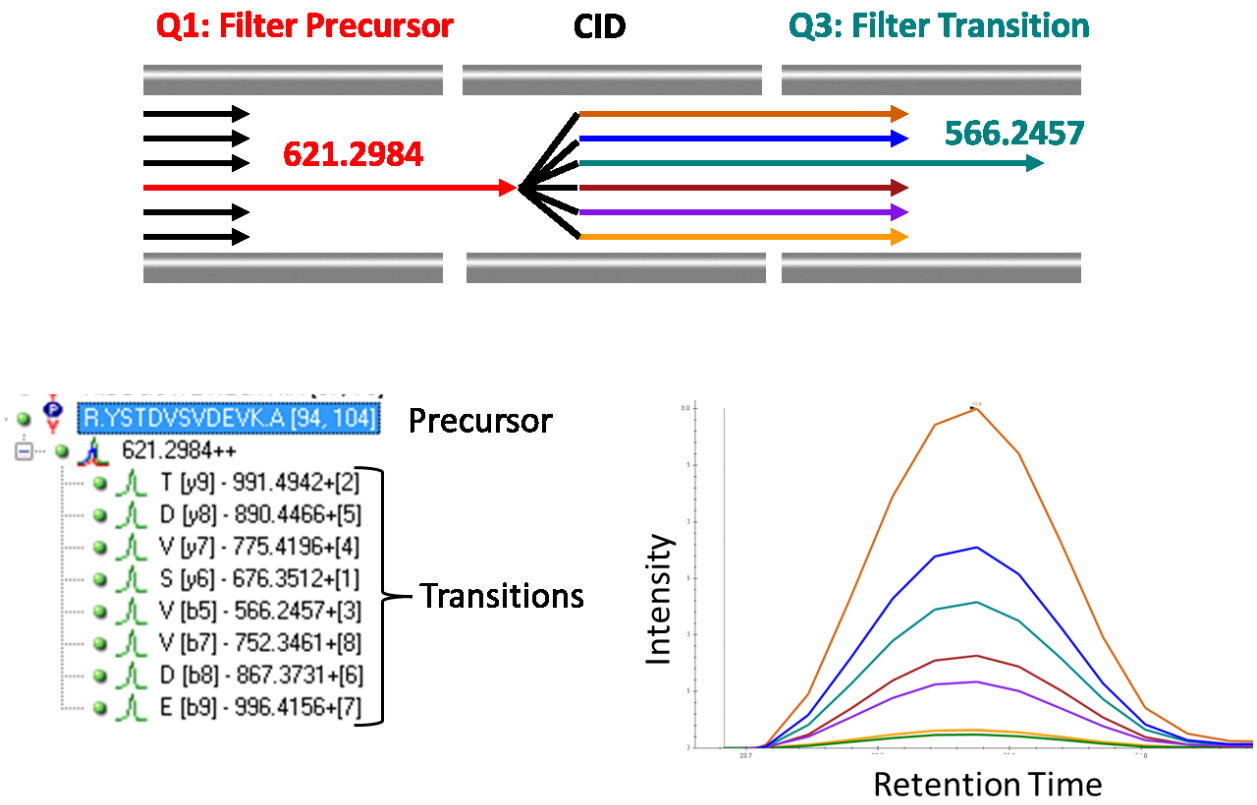


Figure 1.3: Selected Reaction Monitoring.

This figure illustrates selected reaction monitoring (SRM) on a triple-quadrupole mass spectrometer. The first quadrupole mass filter (Q1) allows only a small *m/z* range (0.2-0.7 *m/z*) centered around a predetermined precursor *m/z* to pass through. The precursor is fragmented by collision induced dissociation (CID) in the next quadrupole. Finally, another quadrupole mass filter (Q3) filters a small window around a pre-selected fragment ion *m/z*. The filtered ion beam strikes a detector for a pre-determined "dwell time" (10-100 ms) to detect the signal (ions/second). Multiple precursor->transition pairs (10s-100s) are measured are repeatedly measured in a cycle throughout and instrument run. Extraction of the signals for transitions of a particular precursor allows the elution of the peptide to be detected as co-eluting transition signals. The integrated area under these co-eluting peaks can be used for quantitation.

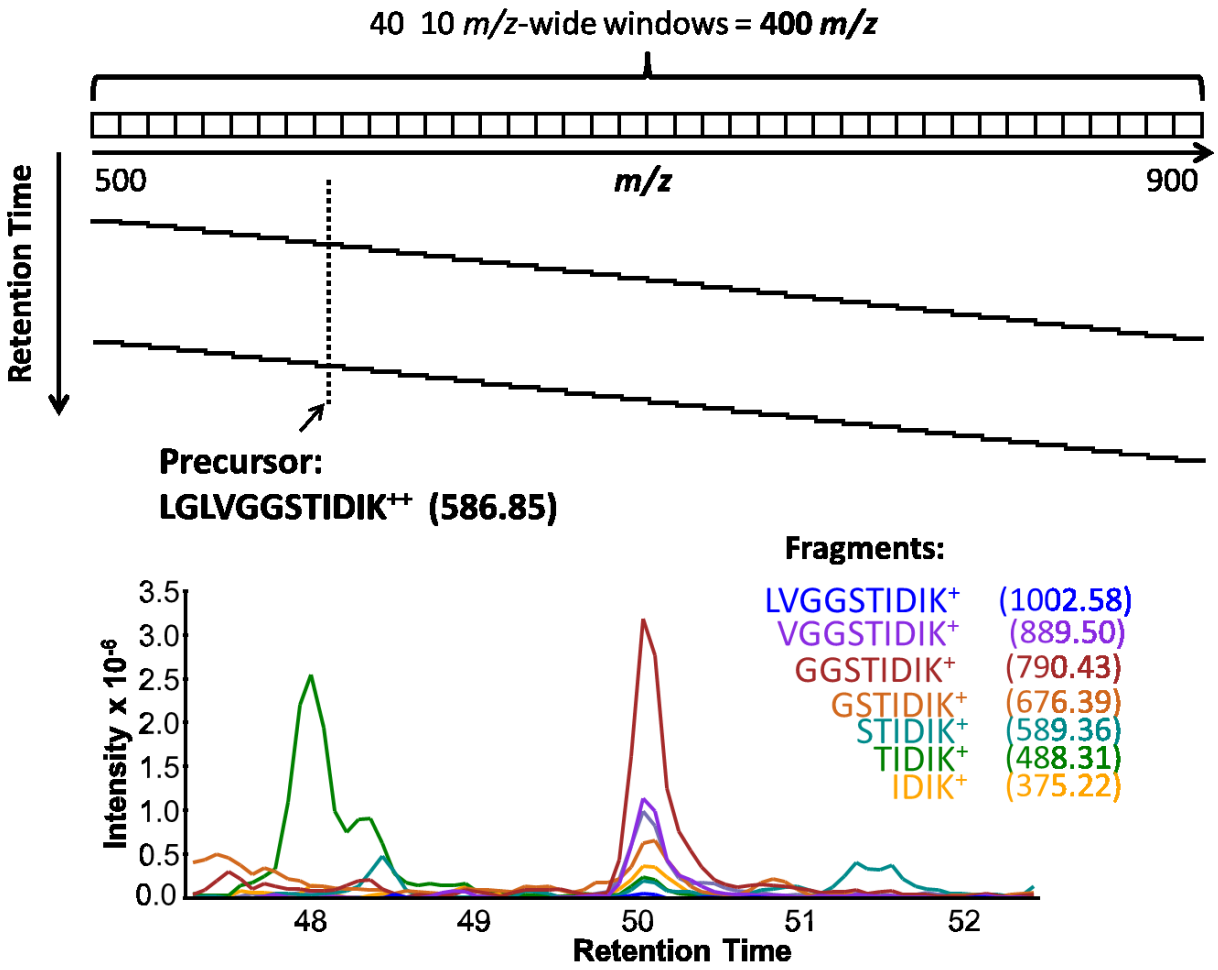


Figure 1.4: Data Independent Acquisition.

In data independent acquisition (DIA), a series of pre-determined targeted MS/MS scans are acquired in a repeated cycle. In this case, 40 consecutive targeted MS/MS scans with 10 *m/z* wide isolation windows are acquired in a cycle to cover the *m/z* range from 500-900 *m/z* comprehensively. From these scans, fragment ion chromatograms can be extracted and used for quantitation similar to SRM data (Figure 1.3). The precursor LGLVGGSTIDIK is fragmented multiple times as it elutes by the MS/MS scan isolating *m/z* 580-590. Fragment ion chromatograms for this peptide are extracted from this scan event and plotted.

Section 4: Data Independent Acquisition

The "data independent" in data independent acquisition refers to the fact that MS/MS spectra are acquired independently of any information from an MS scan. In fact, MS scans are not even required for DIA LC-MS/MS runs, although they are useful. An early appearance of DIA in the literature is the "Shotgun CID" technique pioneered by the Dave

Goodlett lab at the University of Washington[17]. In this technique, no precursor selection is used for fragmentation scans, so the entire precursor mass range is fragmented together using in-source fragmentation or fragmentation in a collision cell. This method makes a large trade-off in precursor selectivity for comprehensive MS/MS sampling.

In 2004, Venable *et. al.* demonstrated a DIA method on a linear ion trap mass spectrometer (LTQ, Thermo Fisher Scientific, San Jose) where 20 10 *m/z* wide windows covering 900-1100 *m/z* were acquired repeatedly in a 4 second cycle[18]. These isolation windows are much wider than used for typical SRM (0.2 – 0.7 *m/*z) and DDA (1-3 *m/*z) experiments. With a 4 second cycle time, any peptide between 900 and 1100 *m/z* would be sampled by MS/MS frequently enough to resolve an elution profile for that peptide using the fragment ion chromatograms of the peptide, similar to SRM data. Venable *et. al.* demonstrated improved signal-to-noise for quantitation using the fragment ion signal for quantitation than with the area under the precursor peak. When analysing complex samples, there is often substantial chemical interference in the MS signal for a peptide. In these cases, quantitation using the MS/MS signal can be beneficial. MS/MS spectra will have a reduced signal for the peptide of interest because of ion losses in the fragmentation process and the separation of the signal for the peptide from one measurement (intact peptide *m/z*) into multiple fragment ion measurements. However, the improved selectivity of the MS/MS scan will often reduce chemical noise much more than the reduction in signal resulting in higher signal-to-noise[19]. Low resolution ion trap precursor measurements in this thesis are particularly susceptible to chemical noise. It is difficult to identify peptides using the DIA spectra acquired this way because a 10 *m/z* range of candidate peptide precursor mass has to be considered in the

database search. Additionally, most database search algorithms assume that each MS/MS spectrum contains fragment ions from only one spectrum, which is virtually never the case for scans with such a wide isolation window (Figure 1.5, Figure 1.6). Database search and post-processing algorithms which do not make this assumption have since been developed and used to increase the number of peptides that can be identified in DIA[20-27].

The MS^E technique addressed some of the challenges in identifying peptides from DIA data[28]. In MS^E, a Q-TOF instrument acquires scans on the *m/z* range from 300-2000 *m/z* alternating between low and high collision energy, similar to the shotgun CID approach. The low collision energy scans provide *m/z* measurements of intact precursors, while the high collision energy scans provide measurements of the fragment ions from all of these precursors. To identify peptides, spectra are deconvolved using information from the elution profile of peptides. Clusters of fragment ions that co-elute with a precursor are likely from the same peptide and are grouped together in a single, deconvolved spectrum which can be searched using a standard database searching algorithm. Deconvolution by elution profile information is also applicable to DIA data acquired in the manner of the Venable study[27].

The FT-ARM technique combines a wide isolation window (up to 100 *m/z*), high resolving power (up to 30,000) MS/MS scans with a novel technique for peptide identification[29]. Traditional database searching seeks to find the best candidate peptide match for each spectrum. FT-ARM does the opposite by searching for the best spectrum match to each candidate peptide using a dot-product based score. This "peptide-centric" approach works well for DIA data where virtually every MS/MS spectrum contains information from multiple peptides, most of which is ignored in a traditional database search which only keeps

the best peptide match for each spectrum. This "peptide-centric" approach was later applied in the "SWATH" technique where a cycle of 32 25 *m/z* wide MS/MS windows are acquired over the *m/z* range from 400-1200 *m/z*[30].

The ability to extract MS/MS information from any peptide within a wide sampled mass range is a huge advantage of DIA techniques over DDA techniques. The hypotheses an investigator can ask of the data is no longer biased by the acquisition of the data itself. With DDA data, the investigator is limited to only querying the high-abundance subset of peptides that were sampled by MS/MS because the MS data alone is not selective enough for peptide identification. With SRM data, the investigator can only query the 10-100 peptides that were targeted in the approach. As hypotheses change, the investigator can continue to query the same DIA data. The extracted MS/MS data for a peptide will be selective enough to confidently identify the queried peptide in many cases as well as quantify it by integrating the MS/MS signal over time as the peptide elutes, similar to SRM studies. Quantitation using the MS/MS signal rather than the MS signal is advantageous in complex mixtures where the improved selectivity of MS/MS reduces chemical noise and improves sensitivity[19]. Statistical measures of confidence for the presence or absence of each query peptide can be determined. Because the data is acquired in the same manner for every run, the reproducibility is also improved. If there is no signature for the peptide in the data, it is because the peptide was below the limit of detection of the instrument, not because it was potentially not sampled by MS/MS.

To achieve comprehensive MS/MS sampling, all of the described DIA methods use wider MS/MS isolation windows (10 – 1700 *m/z*) than used in DDA (1-3 *m/z*) and SRM (0.2-0.7

m/z) approaches, amounting to a trade-off in MS/MS precursor selectivity for wider sampling. Reducing precursor selectivity increases the chemical noise (noise from other components in the sample that interfere with the signal from the desired analyte). As the isolation window becomes wider, more precursors are isolated and co-fragmented. With the co-fragmentation of precursors, there is a greater chance of co-fragmented peptides having fragment ions with signals overlapping that of the desired analyte. This impacts the ability to both identify and quantify peptides.

If database searching is used to identify the peptides, the wide precursor window means that a larger number of candidate spectra need to be considered as potential matches for the spectrum which decreases sensitivity due to multiple hypothesis testing. Additionally, the MS/MS spectrum may not match the theoretical spectrum as well due to fragment ion interference.

If a "peptide-centric" search such as with FT-ARM or SWATH is used to identify the peptide, a wider isolation window makes it more likely that there will be multiple high-scoring spectra for a particular query peptide. This happens when a peptide with a similar fragmentation pattern to the query peptide falls within the isolation window of the query peptide (Figure 3.1). One common cause of such interference is a peptide with a post-translational modification. For example, a triply-charged peptide containing a methionine will differ by 5.3 *m/z* from the same peptide with an oxidized methionine, and the fragmentation pattern will show 50% similarity. With the 25 *m/z* wide windows of SWATH, any modification within 25 Da, 50 Da, or 75 Da for +1, +2, and +3 charged peptides will be co-isolated. The situation is even bleaker when virtually no precursor isolation is used in

methods such as MS^E and shotgun-CID. Improved MS/MS resolving power will not help in this situation because the overlapping fragment ions have the exact same *m/z*. If MS data is available, it can help to resolve between the modified and unmodified forms.

The increased chemical noise due to wide isolation windows will negatively impact peptide quantitation by MS/MS by causing fragment ion interference. Additionally, wider windows can impact the dynamic range of quantitation due to ion counting statistics. Consider MS/MS acquisition on a trapping analyzer such as an orbitrap. The orbitrap can be filled to a maximum of ~10^6 ions before space charging starts to impact mass accuracy. With wide isolation windows, this 10^6 fill target is almost always reached. This means that the same number of ions will be accumulated in the orbitrap regardless of whether these ions are from a 20 *m/z* or a 10 *m/*z isolation window. The same number of ions is spread amongst more analytes in the 20 *m/z* case resulting in a reduced number of ions being collected for any particular analyte which reduces sensitivity and accuracy of quantitation. Instruments with TOF detectors will be similarly impacted as the intra-scan dynamic range of the analytes is more likely to exceed the intra-scan dynamic range of detection.

The issue of reduced precursor selectivity on the MS/MS scans can be overcome in part by increasing the resolution of mass analysis, utilizing retention time or MS information, or adding additional separations such as ion mobility[31].

This thesis focuses on improving the outcome of discovery-based proteomics experiments, with a focus on DIA. The first section demonstrates an algorithm for *de novo* calibration of MS/MS spectra. The following section presents a novel multiplexed MS/MS

acquisition technique for DIA which improves precursor selectivity without reducing the *m/z* range sampled by MS/MS or increasing duty cycle time.

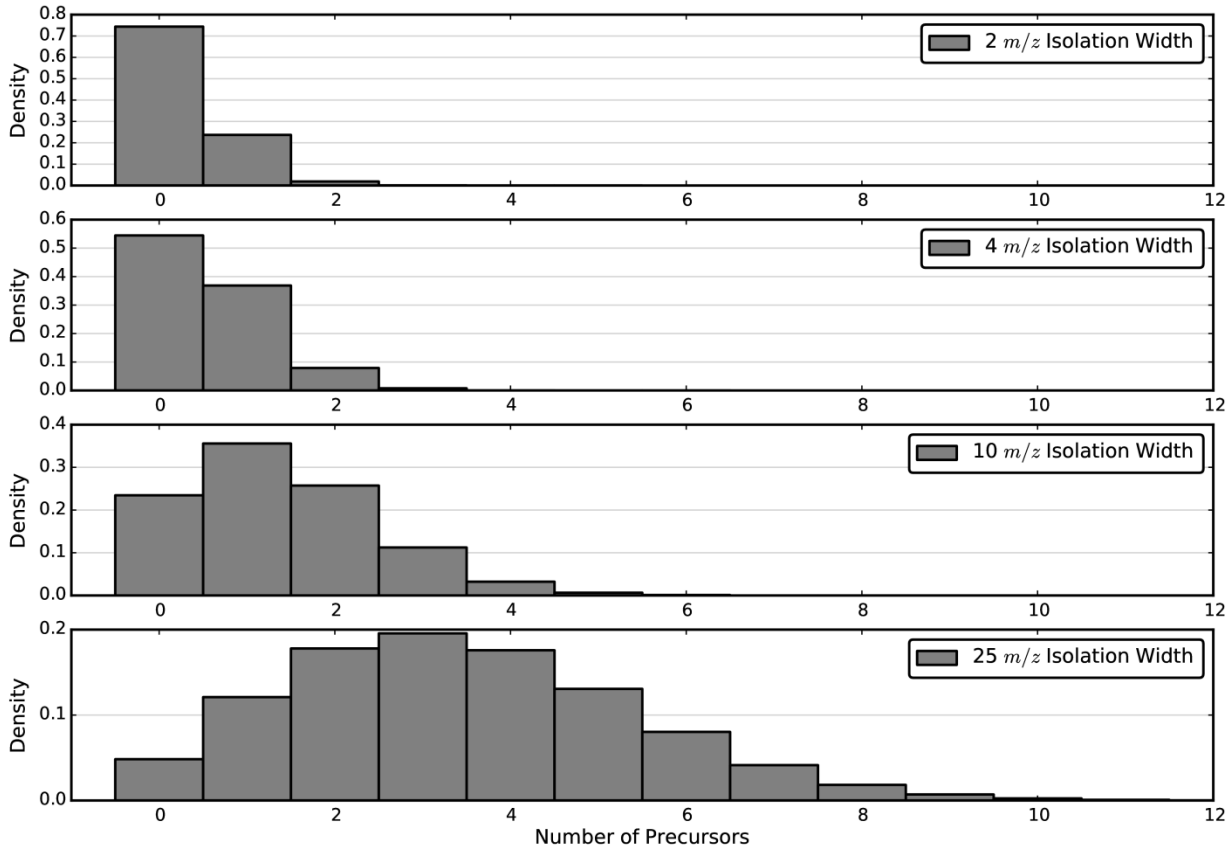


Figure 1.5: Simulation of the Impact of Isolation Width on Co-fragmentation.

MS1 scans from 6 replicate analyses of *S. cerevisiae* lysate acquired on a Q-Exactive (Thermo Scientific, Bremen, Germany) were analyzed using Hardklör (Hoopmann M.R., *et. al.* Analytical Chemistry, 2007) and Krönik to detect persistent peptide isotope distributions (PPID). Next, isolation was simulated at every point in time of the instrument run where peptides elute (between 25 to 90 minutes) for isolation windows centered at integer values between 525 and 875 *m/z*. In each panel, the simulation is done with a different isolation width, and a histogram of the number of peptides co-fragmented is plotted. When 2 *m/z* wide isolation windows are used, more than 70% of the isolation windows do not contain a peptide precursor detected in the MS scan. When 25 *m/z* wide windows are used, ~85% of the windows isolate and fragment two or peptide precursors and on average, 3.4 peptide precursors are co-fragmented in each window.

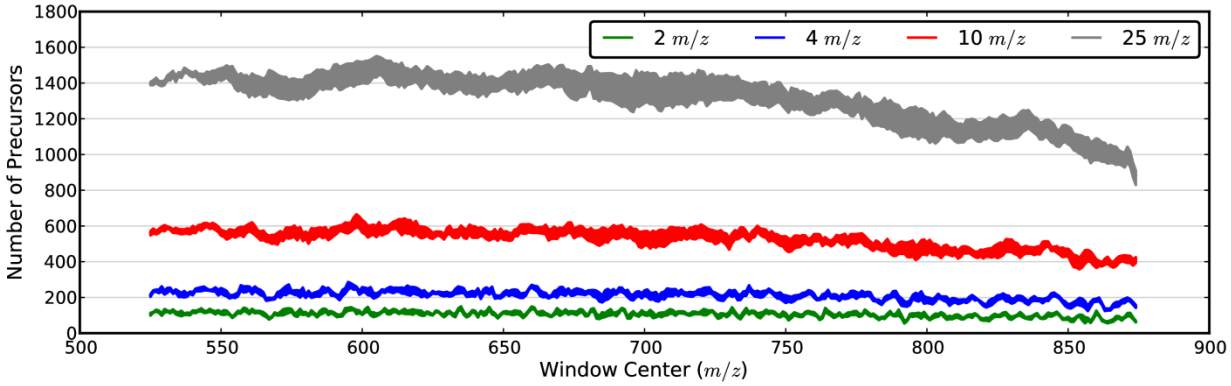


Figure 1.6: Number of Persistent Peptide Isotope Distributions Detected in 2, 4, 10, and 25 *m/z*-wide Isolation Windows.

MS1 scans from 6 replicate analyses of *S. cerevisiae* lysate acquired on a Q-Exactive (Thermo Scientific, Bremen, Germany) were analyzed to detect persistent peptide isotope distributions (PPID) in the same manner as in Figure 1.5. The number of PPID's detected in simulated isolation events with retention time between 25 and 90 minutes using windows of widths 2, 4, 10, and 25 *m/z* centered at every integer *m/z* between 525 and 875 is plotted. The upper and lower bounds of the lines for each isolation width is the maximum and minimum number of PPID's observed out of the six MSX runs analyzed.

Chapter 2 – *De novo* Correction of Mass Measurement Error in Low Resolution Tandem MS Spectra for Shotgun Proteomics

Section 1: Introduction

The field of proteomics has experienced significant growth in the past decade due to advances in mass spectrometer instrumentation and computational tools for data interpretation. Instrument scan speed, dynamic range, sensitivity, resolution, and mass measurement accuracy (MMA) continue to improve, allowing for more comprehensive analysis of complex protein digests.

Systematic mass measurement error (SMME) is typically corrected by routine instrument calibration. External calibration is performed by analyzing a standard with molecules of known elemental composition and calibrating the instrument to match the measurements of the calibrant ions to their known mass-to-charge (*m/z*). Ion species covering a wide range of *m/*z values are often used for calibration because systematic mass measurement error can vary with *m/z*. Over time, the mass calibration drifts, requiring periodic recalibration. Even the most sophisticated instruments will have some systematic and/or random mass measurement error if they are not properly calibrated.

Mass measurement accuracy can be further improved by analyzing calibrants spiked into each sample (i.e., internal calibration). Internal calibration is usually coupled with high-resolution mass analyzers (e.g., TOF[32, 33] , FTICR[34, 35], Orbitrap[36]) where instrument factors such as space charge effects, electric fields, peak intensity, and mass analyzer temperature vary during the course of an analysis. These factors ultimately cause mass measurements to deviate. On trapping instruments, the implementation of automatic gain control helps alleviate the mass

deviations due to space charge by controlling for total ion population[37]. As a result, internal calibration is not necessarily needed in these instruments to achieve high MMA and can be detrimental to performance due to the loss in sensitivity and dynamic range associated with introducing a calibrant ion.

Computational calibration techniques aim to match the robustness of internal calibration without requiring the addition of specific calibrants. To calibrate precursor data, ion species known to be present in a sample can be used as internal calibrants in lieu of spiked in calibrants. Peptides expected to be present in a data set *a priori*[33, 38, 39], peptides confidently identified by database searching[40-44], or commonly observed contaminants can be used in this manner[45]. These techniques improve mass measurement accuracy but are likely less effective with low resolution data due to the difficulty of unambiguously mapping theoretical ion species to low resolution precursor features. Charge state pairs can be used to detect frequency shifts in precursor spectra without *a priori* knowledge of sample content[46-48]. Unfortunately, the reliance of this technique on resolving charge-state precludes its application to low-resolution data where resolving power is too low to resolve isotopic peaks for charge state determination. A newer technique uses fixed mass differences between fragment ions to calibrate MS/MS data and does not have the same resolution requirements[49]. Monoisotopic, singly-charged peak masses in peptide mass fingerprinting data collected on MALDI-TOF instruments can be calibrated without *a priori* knowledge of the sample content[50-52].

Fewer computational techniques exist for the calibration of MS/MS spectra. Fragmentation spectra can be calibrated *de novo* by detection of type-1 peak edges (peaks in an MS/MS spectrum that differ by the mass of one amino acid)[53] or by analysis of trypsin auto-lysis

products[54]. Confident peptide-spectrum matches from a database search can be used to calibrate MS/MS data as well[55], although such methods can be computationally expensive and require protein annotations from a genome sequence.

Herein, we demonstrate that low resolution ion trap instruments can be subject to systematic mass measurement error even after external calibration. We implement a novel method for *de novo* calibration of peptide MS/MS data collected on low resolution instruments capable of improving mass measurement accuracy and mass measurement precision quickly (< 1 minute) without the need for protein annotations or even knowledge of the organism(s) being analyzed. The algorithm is implemented in a freely available, open-source software package named "FineTune". We demonstrate improved mass measurement accuracy after calibration with FineTune for MS/MS spectra acquired on both an LTQ and LTQ-Velos mass spectrometer by analyzing mass measurement error in confident peptide-spectrum matches pre and post-calibration. The robustness of FineTune is demonstrated by successfully calibrating data sets with only a small percentage of the total spectra used as input for the calibration. The impact of calibrating MS/MS spectra on database search results is tested using the Mascot, SEQUEST, and X!Tandem search algorithms.

Section 2: Materials & Methods

Collection of data for testing the *de novo* calibration algorithm

A *S. cerevisiae* sample digest was analyzed by LC-MS/MS on an LTQ-Orbitrap-Velos (Thermo Fisher Scientific) hybrid mass spectrometer. Peptides were separated by reversed-phase high performance liquid chromatography (RP-HPLC) across a 100 minute linear acetonitrile gradient on a 40 cm column with 75 μM inner-diameter. Mass spectra were

collected using a top-13 data dependent acquisition scheme with precursor scans acquired in the orbitrap (60,000 Resolving Power @ 400 *m/*z) in profile mode. Peptide fragmentation by resonance collision induced dissociation (CID) and subsequent mass analysis was executed in the dual pressure linear-ion trap in centroid mode. Dynamic exclusion was enabled with a 50 entry exclusion list and 180 second exclusion time.

A *S. cerevisiae* sample digest was analyzed in a similar manner on an LTQ-FTICR (Thermo Fisher Scientific) hybrid mass spectrometer. A 125 minute linear acetonitrile gradient was used. Mass spectra were collected using a top-5 data dependent acquisition scheme with precursor scans acquired on the FT-ICR (50,000 resolution @ 400 *m/z*) in profile mode. Fragmentation spectra were acquired by resonance CID in the linear ion trap in centroid mode. Dynamic exclusion was enabled with a 50 entry list and 30 second exclusion time.

Generation of a theoretical fragment ion map for MS/MS spectra

Previous studies demonstrate that polypeptide masses occupy "allowable regions" of the mass range of width ~0.2 *m/z* spaced apart by about 1.00045475 *m/z* (β)[56]. We made a similar observation by analyzing fragment ion masses in spectra from our *C. elegans* Bibliospec library[57].

Every MS/MS spectrum and matching peptide sequence was extracted from the *C. elegans* Bibliospec library (Version 5.1) of confident peptide spectrum matches. For each of these spectra, a theoretical MS/MS spectrum was generated using monoisotopic amino acid masses for singly-charged b and y ions. Intensities of theoretical fragment ions were matched to the intensity of the nearest peak in the experimental spectrum within +/- 0.5 *m/z*. The spectra were then binned with a bin width of 1/β *m/z* and summed to generate a theoretical fragment ion

map (Figure 2.1). This bin width was selected to be amenable to applying the discrete fast Fourier transform (FFT) to model the location of the peaks in the theoretical ion map using a sinusoid function (see *De novo calibration of MS/MS data using the theoretical fragment ion map*). Using a bin width of 1/β *m/z* ensures that one of the bins of the discrete FFT will correspond to the component of the data with a period of β *m/z* which is the average distance between peaks in the theoretical fragment ion map[56, 58-62].

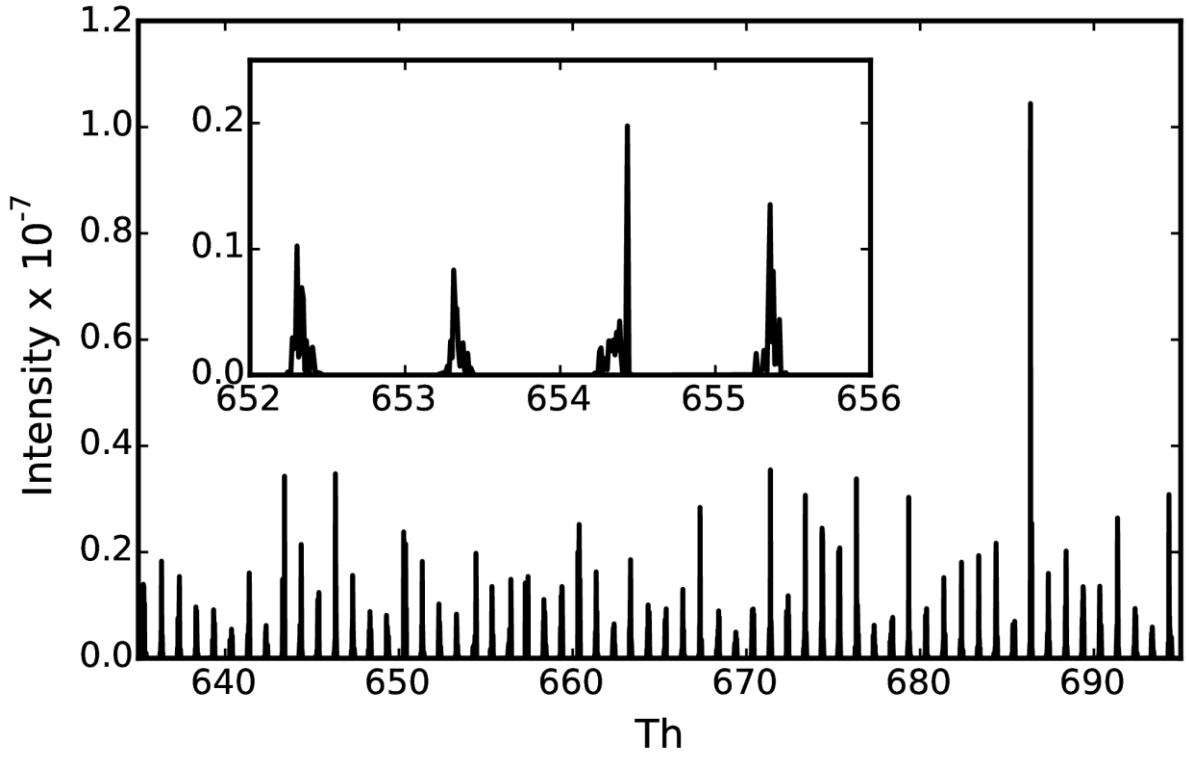


Figure 2.1: Theoretical fragment ion map.

The theoretical fragment ion map is plotted for m/z 635-695. The map was generated by summing over 100,000 fragmentation spectra. The peaks repeating every ~1.0005 *m/z* represent regions where a singly charged b or y-ion is theoretically likely to occur in a peptide fragmentation spectrum. The inset is a zoomed view of a subset of the fragment ion map. Notably, there are repeating regions in *m/z*-space where these peaks do not occur because they are not possible with normal peptide elemental compositions.

De novo calibration of MS/MS data using the theoretical fragment ion map

The technique used to calibrate MS/MS spectra is summarized in Figure 2.2. MS/MS spectra are binned with a bin width of 1/β *m/z* and summed to generate the observed fragment ion map. Misalignment between the observed and theoretical fragment ion maps is due to systematic mass measurement error (SMME). Therefore, the systematic mass measurement error can be determined by calculating the *m/*z shift required to align the observed map to the theoretical one. Because SMME can vary with respect to *m/z*, the SMME is calculated at an interval of every 20 *m/z*.

To calculate the SMME at *m/z* α, a subset of both fragment ion maps are analyzed between α-ε and α+ε. The value ε is the minimum value for which the total ion current in the experimental fragment ion map between α-ε and α+ε exceeds 2.5 x 10^9 ions/sec and ε ≥ 20 *m/z*. If the values α-ε or α+ε are outside of the range of observed *m/z* values, the center of the window is shifted. Adjusting the window width in this manner compensates for variability in signal intensity with respect to *m/z*.

To reduce the impact of noise and multiply-charged peaks on the algorithm, a 2^{18} point discrete fast Fourier transform is used to determine the phase of the signal component with period β *m/z* for both the theoretical (Θt) and experimental (Θe) ion map subsets. This is essentially a computationally efficient method of fitting a sinusoid to the observed signal. The SMME is the difference in the phase of these two frequency components (Θe - Θt). After SMME has been calculated for every 20 *m/z* interval, the points are interpolated linearly to allow for the reporting of SMME at any *m/z*.

FineTune was coded in C++ and compiled with GNU gcc-4.3.3 on a 64-bit system running Linux kernel v. 2.6.29.6. FineTune uses the Boost Build system to allow for cross-

platform compilation (tested on Linux and Windows). FineTune can read and write mzML, mzXML, MGF, MS2, CMS2, and BMS2 formats (http://proteowizard.sourceforge.net/formats.shtml). Additionally, the Windows version can read Agilent, Bruker, Thermo, Waters, and AB-Sciex vendor formats. Source code and binaries are available at the MacCoss lab website (http://proteome.gs.washington.edu/software/finetune).

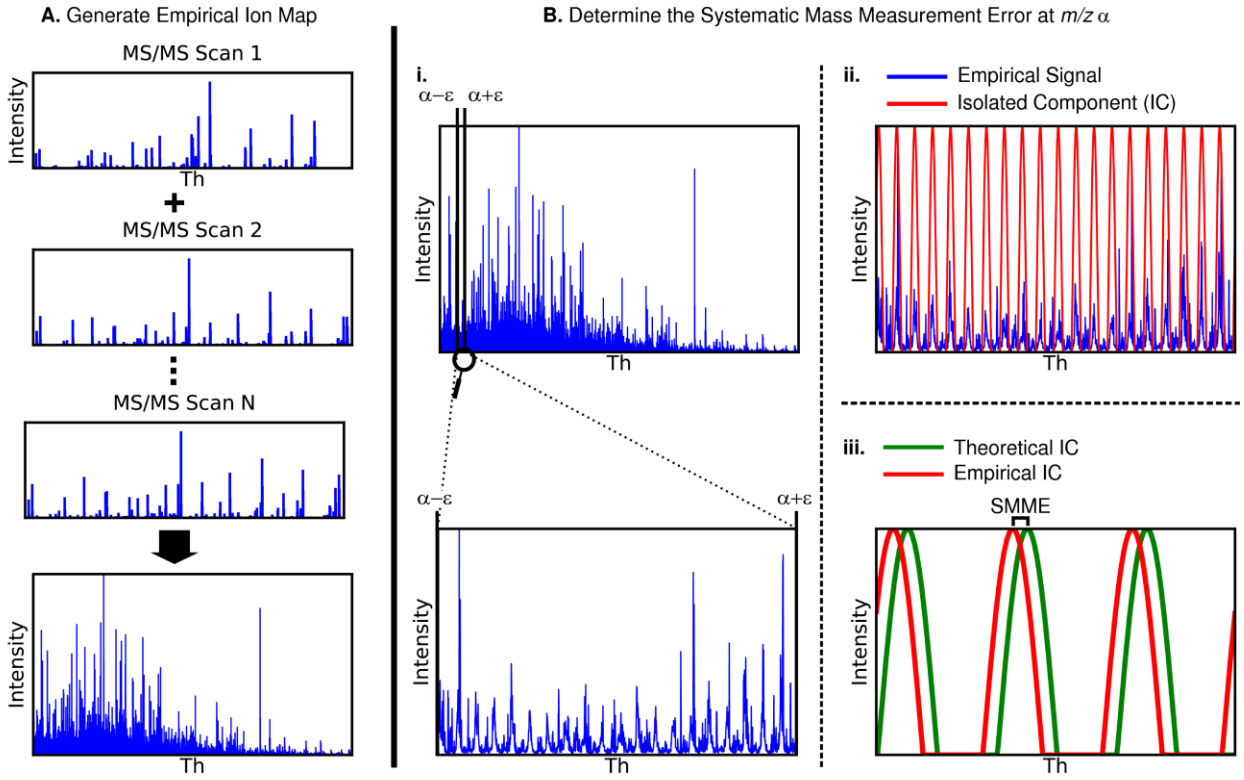


Figure 2.2: Steps for de novo calibration of MS/MS spectra.

A) All MS/MS spectra in the input file are binned and added to each other to generate an empirical fragment ion map for the file. This step only happens once for each file being calibrated. B) Systematic mass measurement error (SMME) is detected in 20 *m/z* intervals along the m/z range of the empirical fragment ion map.

To detect the SMME at *m/z* α:

- i) Analyze a slice of the empirical ion map with boundaries α +/- ε such that there is adequate signal for detection of SMME. The full empirical ion map (top) and a zoomed in slice of the empirical ion map (bottom) are shown.
- ii) Apply a discrete fast Fourier transform (DFFT) to this slice to isolate the frequency component of the signal with period 1.00045475 *m/z* (red). This isolated sinusoid component fits the observed peak clusters (blue).

- iii) Use the same technique as in steps i and ii to isolate the same frequency component from the theoretical ion map. The difference in phases of the empirical and theoretical frequency components is the systematic mass measurement error at *m/z* α.

Once SMME has been detected at each 20 *m/z* interval, use linear interpolation to determine the SMME at any *m/z* between these intervals.

Calculation of mass measurement error by database searching

The systematic mass measurement error detected by FineTune is compared to that detected by analyzing confident peptide spectrum matches (PSMs) identified by SEQUEST[11] using a target-decoy strategy, and post-processed by Percolator (v 1.14)[25]. If precursor scans were collected on a high-resolution instrument, Bullseye[63] is used to determine more accurate precursor masses prior to database searching. A threshold for peptide-level FDR of ≤0.01 is applied to the set of PSMs. For each PSM, the mass measurement error for each theoretical singly charged b- and y-ion with a matching peak within +/- 0.5 *m/z* is determined. Mass measurement accuracy is reported as the mean of the mass measurement error and its 95% confidence interval. When comparing mass measurement error before and after *de novo* calibration, the same spectra analyzed prior to calibration are analyzed post-calibration.

Section 3: Results

A theoretical fragment ion map for peptide fragmentation spectra

Mass excess is the nominal mass (i.e., mass number) of an atom subtracted from the exact mass. The twenty amino acids coded by the standard genetic code have a very similar mass excess which causes the masses of polypeptides to group together into regions of the mass range spaced roughly 1.0005 *m/z* apart[60] (Figure 2.1). The regions of the mass range between these mass clusters have been termed "forbidden zones"[61] because it is theoretically impossible for the mass of a polypeptide built from these twenty amino acids to fall in these zones.

To visualize the distribution of "allowable" fragment ion masses, MS/MS spectra in the *C. elegans* BiblioSpec spectrum library were summed (Figure 2.1). Only annotated MS/MS peaks were included in the summation, and each peak's *m/z* was corrected to match its

theoretical *m/z*. Figure 2.3A shows a "zoomed out" view of the summed spectrum. Fragment ions between 200 and 800 *m/z* contribute roughly the same intensity to the summed spectrum throughout this range despite the greater number of fragment ions at the lower end of this range (Figure 2.3B). Above 800 *m/z*, the summed intensity of fragment ions gradually decreases with increasing *m/z*.

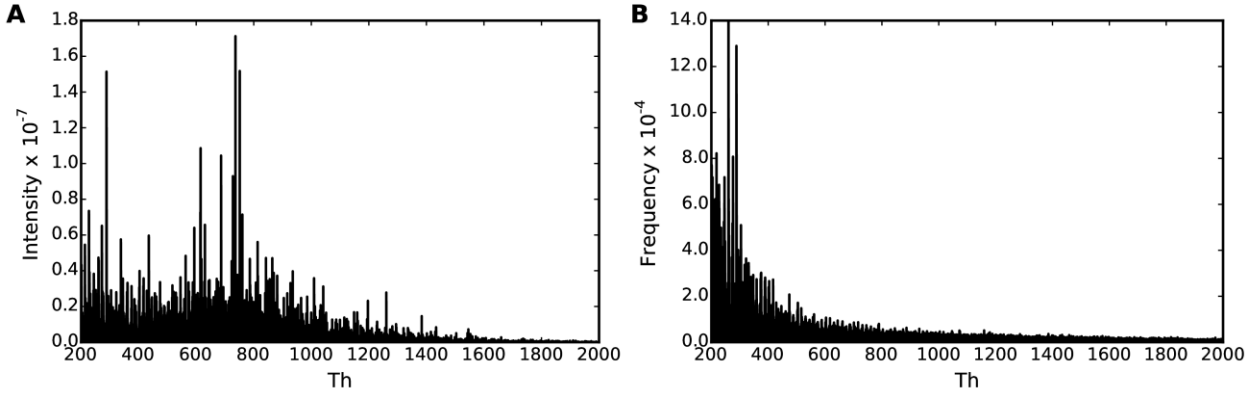


Figure 2.3: The theoretical ion map.

Two theoretical ion maps were generated using monoisotopic, singly charged b and y fragment ion masses from peptide-spectrum matches in the *C. elegans* BiblioSpec library. In both maps, peaks were binned into 0.01 *m/z* wide bins. In A) the m/z of each bin (x-axis) is plotted against the summed intensity of the peaks falling into each bin (y-axis). In B) the mass-to-charge (*m/z*) of each bin (x-axis) is plotted against the number of peaks falling into each bin (y-axis).

De novo calibration corrects systematic mass measurement error

FineTune was tested on 44,944 low resolution MS/MS spectra acquired by a shotgun LC-MS/MS analysis of *S. cerevisiae* lysate on an LTQ-Orbitrap-Velos (Thermo Fisher Scientific, San Jose, CA) mass spectrometer. The recalibration determined *de novo* matches very closely to that determined by analyzing 6,628 confident ($q \leq 0.01$) peptide spectrum matches from a database search (Figure 2.4A-B). The mean mass measurement error prior to *de novo* calibration is -0.1776 ± 0.0010 *m/z* (95% confidence interval) due to calibration drift. After calibration with FineTune, the mean mass measurement error is 0.0078 ± 0.0006 *m/z*, thus the precision is

improved by about half in addition to the improvement in mass accuracy (Figure 2.4A-B). We selecterd a data set with uncharacteristically poor calibration to illustrate the capabilities of FineTune. This poor mass calibration is caused by detector and thus automatic gain control miscalibration resulting in significant space charge effects (Figure 2.4A). In a more common scenario, the mass error does not vary so drastically with *m/z* (Figure 2.5A) and thus the improvement in the mass error spread will be more subtle. Additionally, the mass error between the caffeine and MRFA ions as well as between the MRFA and the first ultramark ion on the LTQ-Velos (Figure 2.4A, Figure 2.5A) is a result of the absence of *m/z* calibrants in that region and a non-linear response between the RF ejection frequency and *m/z*. The mass calibration has since been improved by the instrument manufacturer using a new calibration procedure and will be available in future releases of the Velos and Velos Pro Tune software (personal communication Jae Schwartz).

FineTune was also tested on 21,433 MS/MS spectra acquired on the linear ion trap of an LTQ-FTICR (Thermo Fisher) mass spectrometer, improving mean mass measurement error from 0.0444 ± 0.0011 *m/z* to 0.0055 ± 0.0011 *m/z* (Figure 2.4C-D). In contrast to the LTQ-Orbitrap-Velos data shown above, these data have very little systematic mass measurement error (SMME); the mass error distribution is centered close to zero and there is almost no variation in SMME with *m/z* (Figure 2.4C). FineTune corrects for the slight *m/z*-independent shift in SMME and importantly does not detect any false trends in SMME from noise in the data, indicating that FineTune is applicable to data with extreme or subtle SMME.

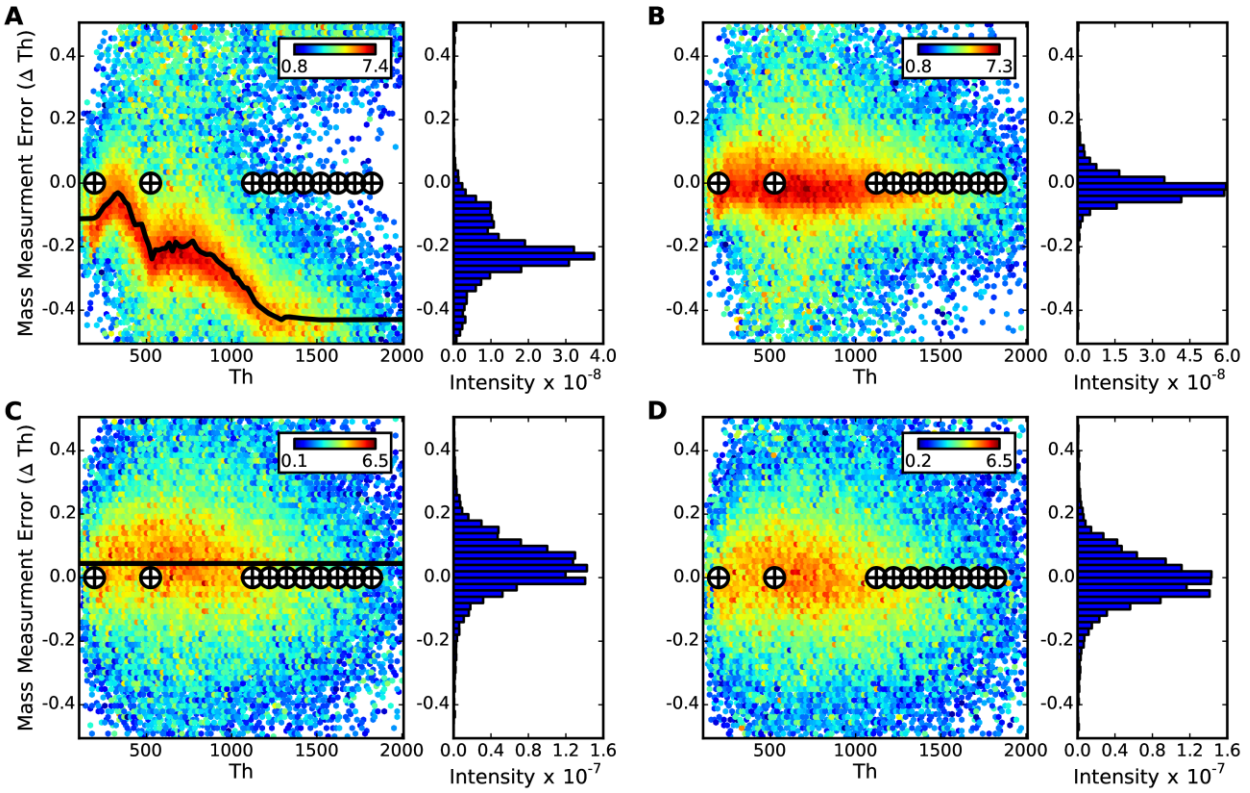


Figure 2.4: Robust de novo detection of systematic mass measurement error.

In A) and C) the mass measurement error detected by FineTune is compared to that detected by analyzing peptide-spectrum matches from a database search on the LTQ-Orbitrap-Velos (6,628 PSMs) and LTQ-FTICR (4,961 PSMs) data respectively. The heatmap in the background is a two-dimensional histogram of the mass error determined by comparing experimental fragment ion masses to theoretical masses from confident peptide-spectrum matches. The log10(total intensity) for peaks falling in any bin is mapped to color as indicated in the colorbar. The black curve on top of the heatmap is the systematic mass measurement error detected by FineTune. The cross hairs indicate the theoretical m/z of the calibrant ions in the manufacturer-supplied calibration mix. Next to each two-dimensional histogram is a one-dimensional histogram of mass measurement error detected from the confident peptide spectrum matches. B) and D) show the mass measurement error after using FineTune to calibrate the LTQ-Orbitrap-Velos and LTQ-FTICR data respectively. The data for these plots was generated using the same set of peptide-spectrum-matches as for A) and C) but with calibrated fragment ion masses.

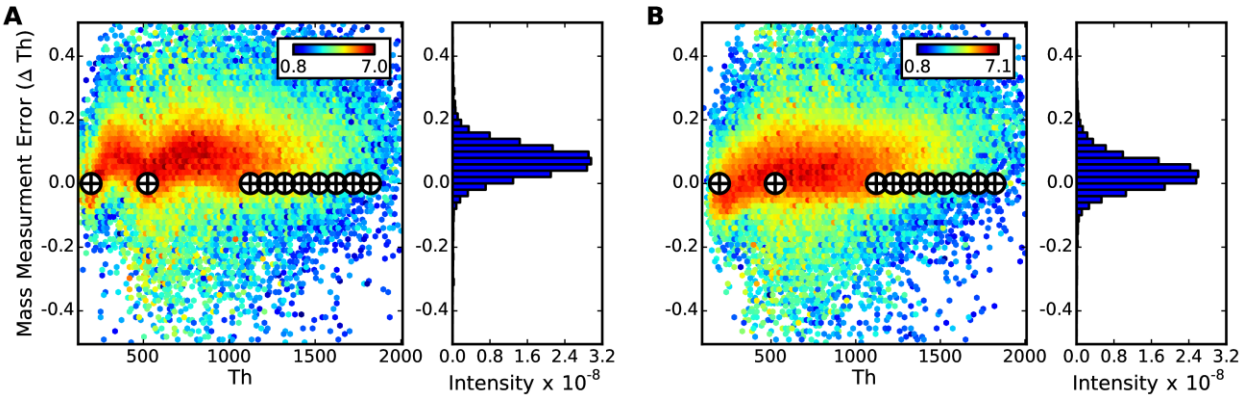


Figure 2.5: Improved calibration on the LTQ-Velos Pro.

MS/MS spectra were acquired analyzing C. elegans lysate on an A) LTQ-Velos (Thermo Fisher Scientific, San Jose, CA, USA) and B) LTQ-Velos Pro (Thermo Fisher Scientific, San Jose, CA, USA) followed by mass measurement error analysis by database searching. The new LTQ-Velos Pro calibration software (build 1083) corrected most of the non-linear variation in mass measurement error seen in the LTQ-Velos data. This new calibration routine is implemented on all future software releases for the LTQ-Velos, LTQ-Velos Pro and their respective Orbitrap hybrid mass spectrometers (personal communication Jae Schwartz, ThermoFisher Scientific).

De novo calibration is robust to a reduction in signal

To test the response of FineTune to a reduction in signal (i.e. few peptide MS/MS spectra), FineTune was applied to the LTQ-Orbitrap-Velos data with varying numbers of MS/MS spectra removed. For each MS/MS spectrum in the data, a random number between 1 and 100 was generated, and if the number was greater than or equal to a threshold, the spectrum was removed from the data set. Therefore, if a threshold of 60 is applied, the resulting data should contain about 60% of the original spectra.

Five "thinned" data sets were generated containing ~75%, 50%, 25%, 10%, and 1% of the 44,944 original MS/MS spectra from the LTQ-Orbitrap-Velos data (Figure 2.6). Even after removing 90% of the spectra, FineTune is extremely robust in high-signal regions of the data. In low-signal regions (e.g. $m/z > 1200$) SMME is still improved, albeit not completely corrected (Figure 2.6C-D). Once 99% of the spectra are removed, FineTune detects the average mass

measurement error, but not the *m/z*-dependent variation in systematic mass measurement error (Figure 2.6E-F). The mass accuracy is still improved, but the mass precision remains the same. The mean mass measurement error is -0.1776 ± 0.0010 for the uncalibrated data. The mean mass measurement error is improved in all of the aforementioned cases to 0.0162 +/- 0.0009, -0.0097 +/- 0.0006, 0.0078 +/- 0.0006, 0.0045 +/- 0.0006, 0.0064 +/- 0.0006, and 0.0078 +/- 0.0006 *m/z* when removing 99%, 90%, 75%, 50%, 25% and 0% of the spectra respectively (Figure 2.6G). Importantly, we have yet to find a case where FineTune negatively impacts the data.

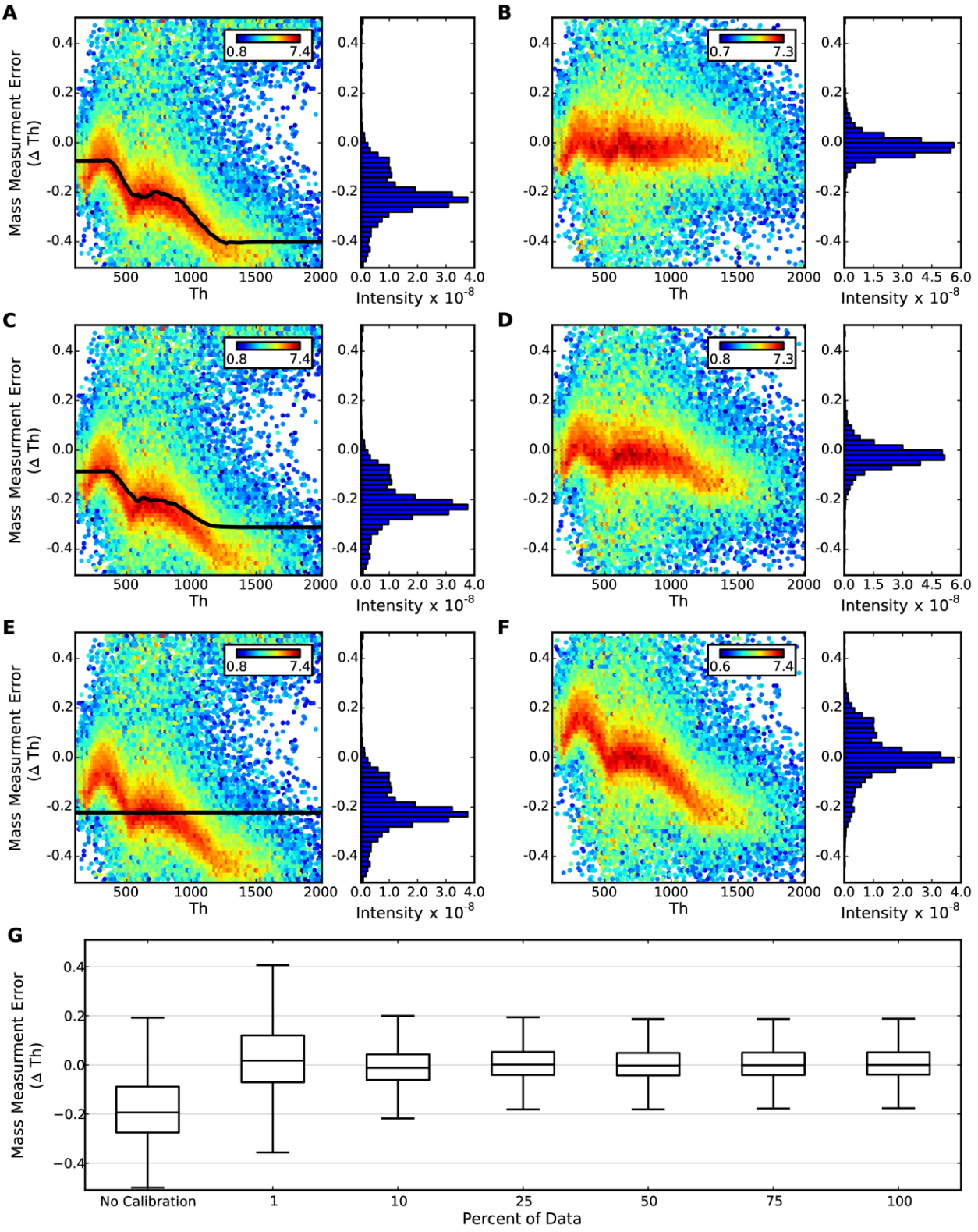

A
B
C
D
E
F
G
Mass Measurment Error
(Δ Th)
Th
Intensity x 10^-8
0.8
7.4
0.7
7.3
0.6
7.4
No Calibration
1
10
25
50
75
100
Percent of Data

Figure 2.6: *De novo* calibration is robust to a reduction in signal.

Mass measurement error heatmaps were generated using all 6,628 significant peptide spectrum matches in the LTQ-Orbitrap-Velos data as described in Figure 2.4 and Materials & Methods. A), C), and E) depict the systematic mass measurement error detected by FineTune using 25%, 10%, and 1% of the (44,944) MS/MS spectra from the original file, respectively; B), D), and F) depict the mass measurement error after correcting for the systematic mass measurement error detected in figures A), C), and E) respectively. G) shows the median mass measurement error from the 6,628 PSMs with no calibration and after calibration by FineTune with varying percentages of the original spectra as input. The whiskers are 1.5 times the inner quartile.

De novo calibration and database search results

The calibrated and uncalibrated (control) LTQ-Orbitrap-Velos data was searched using X!Tandem (Cyclone 2010.12.01.1). All data was first processed by Bullseye and then searched against *S. cerevisiae* sequence (target) and reversed sequence (decoy) databases with a precursor mass tolerance of 10 ppm. The fragment ion tolerance for the searches was varied between ±0.01Da and ±0.5 Da. The expectation values from target and decoy search results were used to determine the number of unique peptides at a false discovery rate (FDR) threshold of 0.01.

At fragment ion tolerances less than ±0.2 Da, the calibrated data returns many more peptide identifications than the data that has not been calibrated (Figure 2.7). For example, at a fragment ion tolerance of ±0.1 Da, the calibrated data returns 2.5 times more peptide identifications. However, at fragment ion tolerances greater than ±0.2 Da, calibrating the data does not increase the number of peptide identifications. Calibrated data had a smaller impact on Mascot search results. (Figure 2.8). Using calibrated data and the optimal fragment ion tolerance/binning parameters SEQUEST, Mascot and X!Tandem identify 3,755; 2,622; and 1,710 unique peptides respectively at a q-value ≤ 0.01.

SEQUEST was modified to allow for the modification of binning parameters (width and offset). During the preprocessing step for each spectrum, SEQUEST bins peaks by *m/z* using consecutive bins of uniform width. The binning offset modifies the phasing of the binning (e.g. a bin offset of 0.1 means that values between 0.1 and 1.1005 *m/z* will be assigned to bin 0). With

such a binning scheme, it is possible for mass measurement error to cause a peak to fall into a different bin than its theoretical bin and thus lower the similarity score between a spectrum and its true theoretical spectrum match. This case can be avoided by setting the bin width and offset such that peaks are more likely to fall in the center of a bin. This can be done by using a bin width of ~1.0005 *m/z* such that the edges of bins fall in the "forbidden zones" of the theoretical fragment ion map (Figure 2.1, Figure 2.3) and is the reason that a default bin width of 1.0005 *m/z* was reported in the original SEQUEST paper.

In Figure 2.9A, the bin width was fixed at 1.00045475 *m/z* and the offset was varied. For each binning offset, the LTQ-Orbitrap-Velos (Thermo Fisher Scientific, San Jose, CA, USA) data was searched using a target-decoy strategy and the number of unique peptides at a peptide level FDR threshold of 0.01 was determined. Peptide level q-values were calculated using the XCorr scores reported by SEQUEST.

The binning offset used by SEQUEST can have a large impact on the number of peptides identified. For the control data, the worst binning offset identifies only 1,845 peptides ($q \leq 0.01$) while the best offset identifies 3,007 peptides; a 63% improvement. After calibration with FineTune, the database search is robust across a wider range of binning offsets. The optimal binning offset for the calibrated data is 0.35.

A similar analysis was performed for bin widths 0.50028 (Figure 2.9B), 0.33348 (Figure 2.9C), 0.250113, and 0.200091 *m/z*. Bin widths 0.500028 *m/z* and 0.33348 *m/z* return more peptide identifications (3,755, and 3,745 for the calibrated data respectively) at their optimal offset than any other bin width. At this point in time, it is unclear why, at certain binning offsets, the uncalibrated data yields better results than the calibrated data. This could in part be due to

the half-height flanking peaks added adjacent to theoretical b and y ion peaks by SEQUEST during the preprocessing step.

It is clear that at certain binning offsets, a SEQUEST database search yields far fewer peptide identifications (~40%) than at the optimal binning offset (0.35 for bin width ~1.0005 *m/z*). For calibrated data, optimal binning offset/width combinations can be determined *a priori* by placing bin edges in the theoretical forbidden zones[61]. If data has not been calibrated, the optimal binning offset cannot be determined because systematic mass measurement error can change the optimal offset. Therefore, application of FineTune to a set of MS/MS spectra followed by a SEQUEST database search can avoid unnecessary losses in peptide identifications. While it is currently not possible to define the binning width and offset in non-modified versions of SEQUEST, the Crux[64] search algorithm produces very similar results to SEQUEST and enables these parameters to be defined.

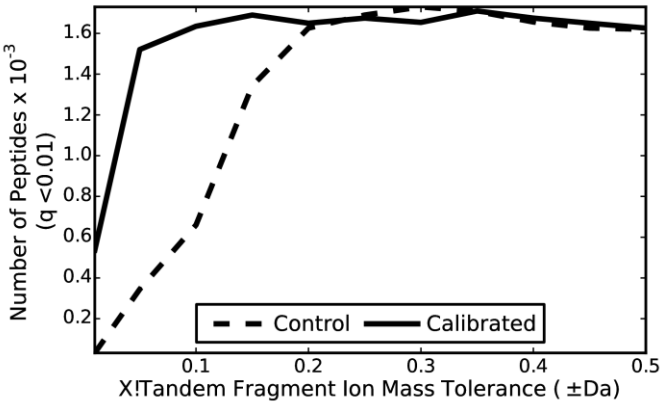


Figure 2.7: *De novo* improves X!Tandem search results.

The LTQ-Orbitrap-Velos data was searched with X!Tandem using a target-decoy strategy to determine q values for each unique peptide from reported expect scores. The number of unique peptides identified by X!Tandem for the control and calibrated data with $q \leq 0.01$ was compared using various fragment ion tolerances in X!Tandem.

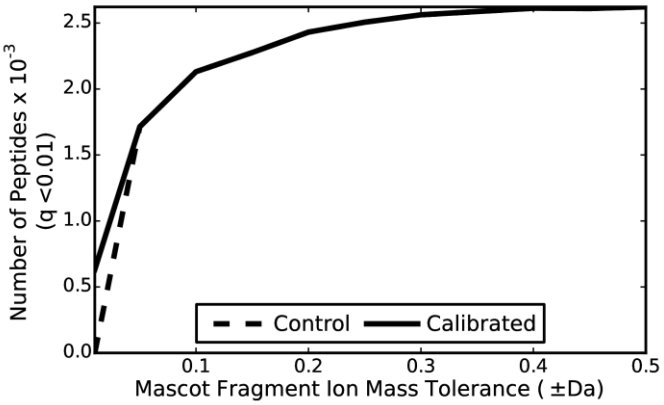


Figure 2.8: *De novo* calibration does not impact Mascot search results.

The LTQ-Orbitrap-Velos (Thermo Fisher Scientific, San Jose, CA, USA) data was searched with Mascot using a target-decoy strategy to determine q values for each unique peptide from reported expect scores. The number of unique peptides identified by Mascot for control and calibrated data with a FDR threshold of 1% was compared using various fragment ion tolerances in Mascot.

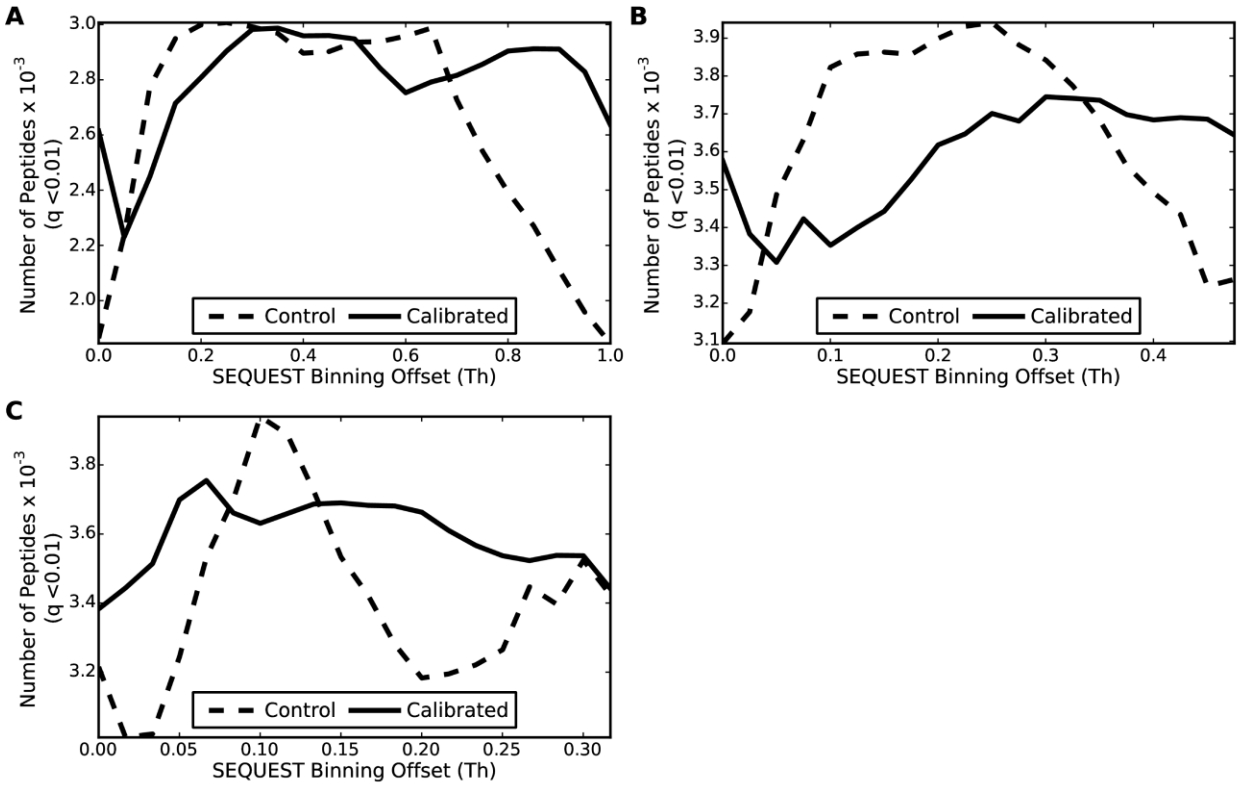


Figure 2.9: Bin offset and bin width impact Sequest search results.

The number of unique peptides identified by SEQUEST (q≤0.01) at different binning offsets was determined for bin widths A) 1.00045475 *m/z* , B) 0.50022738 *m/z* and C) 0.33348492 *m/z* using the LTQ-Orbitrap-Velos data before (Control) and after (Calibrated) *de novo* calibration.

Section 4: Discussion

FineTune enables the *de novo*, unsupervised calibration of MS/MS spectra acquired on low resolution instrumentation. The algorithm corrects non-linear systematic mass measurement error and does not negatively-impact a data set even when it is trained on only 1% of the spectra in the data set. This robustness, combined with the speed of the algorithm, makes it suitable for integration into an unsupervised data processing pipeline run on every data set acquired. Calibration can then be tracked over time which aids in the scheduling of instrument calibrations, especially for newer instruments which tend to drift out of calibration quickly. Additionally, the

algorithm can be applied retroactively to old data that a researcher may suspect was collected on an improperly calibrated instrument.

FineTune is compatible with MS/MS spectra enriched for post translational modifications (PTMs). Previous studies found that the inclusion or exclusion of PTMs does not alter the distribution of peptide mass significantly[58]. For example, theoretical modeling of spectra in which every serine, threonine, and tyrosine are phosphorylated does not have an appreciable impact on fragment ion mass clustering (data not shown). Additionally, we have used FineTune to successfully calibrate data enriched for phosphopeptides (data not shown).

Calibration of fragment ion masses had little impact on peptide identifications from three popular database searching algorithms. These algorithms are classified as database search algorithms due to their reliance on a database of candidate protein sequences to test each spectrum against. *De novo* search algorithms are designed to interpret spectra without a database of candidate sequences. Because FineTune does not rely on sequence information, it is a natural complement to *de novo* search algorithms often used when protein sequence information is unreliable, incomplete, or non-existent. *De novo* search algorithms can benefit from improved fragment ion mass accuracy[65, 66]. Thus, it would be worthwhile to test the impact of *de novo* calibration on *de novo* spectrum interpretation. MS/MS data acquired by data independent acquisition are often analyzed by extracting fragment ion chromatograms for a peptide from the data (see Chapter 1Section 4). Calibration of MS/MS spectra may improve the ability to identify and quantify peptides in these analyses.

FineTune corrects systematic mass measurement error reliably by only making corrections when enough signal is present in the data to justify these corrections. FineTune improved the

mass measurement accuracy of every data set tested to <0.01 *m/z*. This reliability makes FineTune suitable for use as an unsupervised preprocessor applied to any collection of MS/MS spectra prior to database searching. While improved mass measurement accuracy sometimes improves database search results (up to 40% increase in peptide identifications), the data presented herein indicates that database search algorithms could be optimized to take better advantage of more accurate fragment ion mass measurements. In future experiments, we look to expand the algorithm to be able to correct mass measurement error trends that vary with retention time as well as test the impact of calibration on *de novo* spectrum interpretation algorithms.

Chapter 3 : Multiplexed MS/MS for Improved Selectivity of Data Independent Acquisition

Section 1: Introduction

In proteomics experiments, tandem mass spectrometry (MS/MS) data are usually collected on peptides by automated data-dependent acquisition (DDA). Using DDA, mass information acquired on intact peptides in a previous MS spectrum is used to decide which subset of peptides will be targeted for acquisition of fragmentation (MS/MS) spectra necessary for sequence identification[67]. Although DDA is a powerful and versatile strategy, it suffers from several fundamental limitations.

When using DDA, the number of peptides sampled is limited by the MS/MS sampling speed regardless of the dynamic range and peak capacity of the mass analyzer. A single MS spectrum can contain over a hundred different molecular species, of which only a handful are analyzed by MS/MS prior to the next full scan. This sampling of peptides for MS/MS analysis follows a random sampling model[16] biased towards high-abundance peptides. Thus, in a complex protein digest, as much as 84% of peptides remain unsampled[68] and as many as 30% of the sampled peptides can vary between replicate analyses of the same sample[16].

DDA may compromise the sensitivity of MS/MS, because the full-scan mass spectrum used for the selection of the precursor ions typically contains greater chemical background interference than an MS/MS spectrum. A peptide could have abundance above the MS/MS detection limit but go unselected for fragmentation because the precursor is masked by background interference. Additional challenges are that 1) the MS/MS spectra acquired by DDA are rarely sampled at the optimal portion of the peptide elution profile[69], and 2) as many as 15 –

20% of the sampled MS/MS spectra from a complex mixture are chimeric (i.e. contain two or more coeluting molecular species within the isolation window)[70].

An alternative to DDA is data-independent acquisition (DIA). In DIA, MS/MS scans are collected systematically and independent of precursor information. This approach has seen many variations such as collecting fragmentation data without precursor ion selection[17], using ion mobility spectrometry-CID-time of flight mass spectrometry[31], using wide isolation windows[18], and using narrow isolation windows combined with many injections[71]. However, until recently, all of these methods have used a database searching strategy to assign peptide sequence to the fragmentation spectra. Because of the increased complexity of these MS/MS spectra, DIA methods have struggled to compete with the collection of MS/MS spectra using a traditional DDA strategy combined with database searching.

Recently, Gillet *et al.* reported an alternative strategy in which target peptides are queried against DIA data as opposed to trying to assign peptide sequences to every chimeric MS/MS spectrum acquired[30]. This strategy is similar to a targeted analysis using selected reaction monitoring (SRM) and fundamentally different from a discovery experiment that uses a database search engine to qualitatively profile the contents of peptides in a mixture. The appeal of the approach is that any peptide precursor and product ion data within the limit of detection of the instrument can be extracted from the collected datafiles. The relative fragment ion intensities, peptide precursor isotope peaks, and retention time of the extracted ion chromatograms are used to confirm the identity of the target molecular species just as with a targeted SRM experiment. Unlike SRM, different hypotheses can be tested on the data without having to perform another mass spectrometry experiment. However, with current mass spectrometers approaching

acquisition speeds of 10 Hz, a 20 *m/z* wide precursor isolation window is required to sample a 400 *m/z* range every 2 seconds. This wide isolation width is undesirable due to significantly increased fragment ion interference. For example, peptides and their modified forms (e.g. oxidized methionine) are difficult to distinguish if they are isolated in the same window due to similar fragmentation patterns (Figure 3.1).

We present a multiplexing strategy (MSX) where five separate 4 *m/z* isolation windows are analyzed per spectrum. These spectra are de-multiplexed into the five separate 4 *m/z* isolation windows using a novel strategy with similarities to Hadamard multiplexing[72] resulting in data with the sampling frequency of a DIA approach using 20x20 *m/z* wide windows but the selectivity of an approach using 100x4 *m/z* wide windows. Demultiplexing improves precursor selectivity by 1) narrowing down the range of potential precursors for an MS/MS spectrum from a 20 to 4 *m/z* window, and 2) generating the unmixed fragment ion spectrum with signal from only the 4 *m/z* window. We implement this de-multiplexing approach within the popular open source Skyline software tool[73] which provides a useful interface for the visualization and analysis of these data.

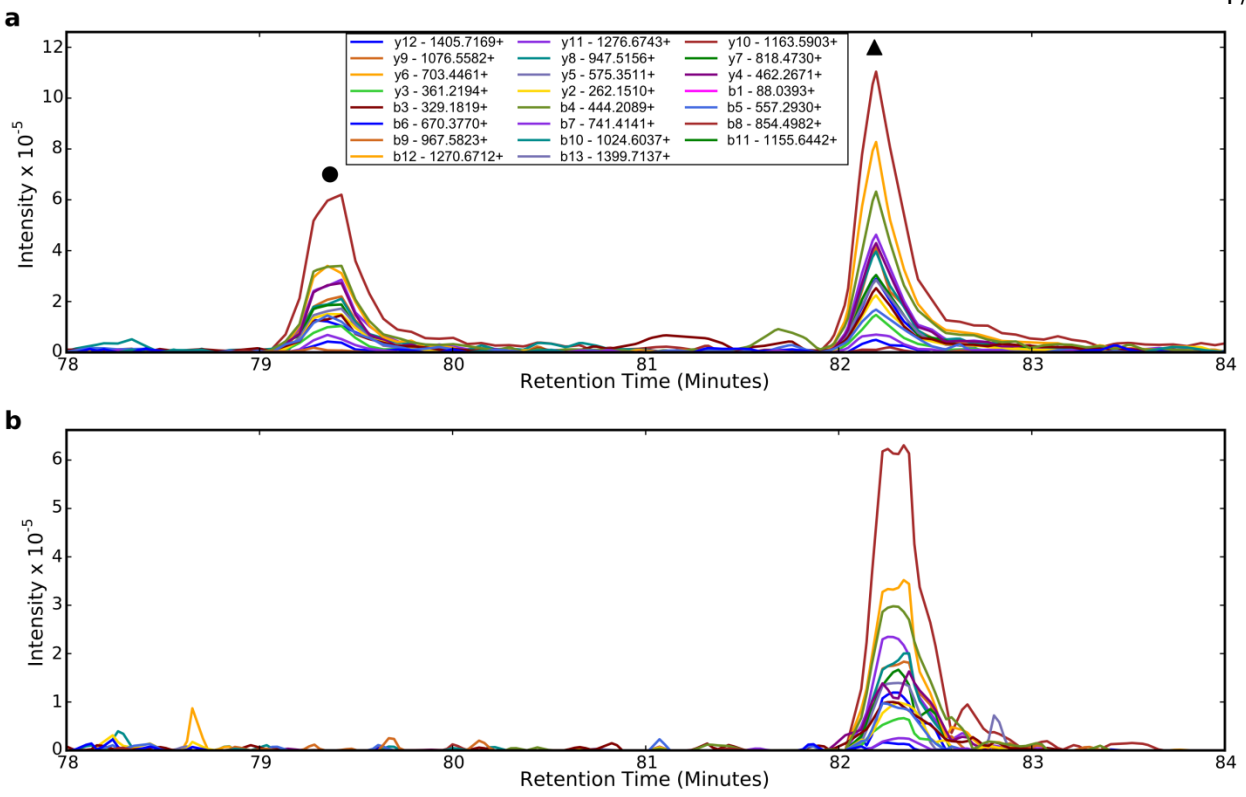


Figure 3.1: Improved Selectivity of MSX over 10 *m/z* DIA.

Both a) and b) show extracted fragment ion chromatograms for the peptide SLQDIIAILGMDELSEEDKLTVSR+++ from an analysis of a C. elegans soluble lysate sample on a Q-Exactive. Panel a shows the extracted fragment ion traces from a 10 *m/z* wide window DIA approach in which the mass range from 500 *m/z* to 900 *m/z* is analyzed with consecutive 10 *m/z* wide isolation window targeted MS/MS scans. The peak labeled with a triangle is the unmodified peptide, the peak labeled with the circle is the peptide with an oxidized methionine (validated using data from a DDA analysis of the same sample on the same column). Panel b shows the same fragment ion traces from an MSX experiment. The oxidized methionine peak has been removed due to the improved precursor selectivity of the MSX method.

Section 2: Materials & Methods

Implementation of Multiplexed MS/MS on a Q-Exactive Mass Spectrometer

In a multiplexed MS/MS (MSX) scan, fragments from multiple precursor isolation windows are stored together prior to mass analysis in the Orbitrap. All of the MSX data in this manuscript were collected using five 4 *m/z*-wide isolation windows per MS/MS scan, although different combinations of isolation windows per scan and isolation width can be used. The 5 windows isolated for each MS/MS scan are chosen randomly from a list of all 4 *m/z* isolation

windows to be analyzed in the experiment. In these experiments, the list is all 100 non-overlapping 4 *m/z* – wide windows between 500 and 900 *m/z*. The window width and position are shifted slightly from their integer values to reduce the likelihood of placing window edges where peptides are most likely to occur (Figure 3.2). For example, the first window is centered at 502.4783 *m/z* rather than 502 *m/z* and has a width of 4.002 *m/z*. This optimization is due to the fact that peptide masses are distributed across the mass range every ~1.0005 *m/z* in "allowable regions" interspersed with "forbidden zones"[61] where peptide masses do not occur. Because the allowable regions are spaced apart by 1.00045475 *m/z* on average[58], the isolation window width is set to be a multiple of this number. Additionally, windows are shifted so the edges fall in forbidden zones. All window edges are calculated using the equation 0.25 + 1.00045475*N where N is an integer. The offset value of 0.25 was determined to be optimal based on analysis of the distribution of peptide precursor masses (+1, +2, +3 charged) in the Bibliospec[57] spectral library. For window edges to be optimally placed, they must be a member of the series 0.25 + 1.00045475*N. For these data, the edges of the windows are N*4*1.00045475 + 0.25 where N is every integer in the range (inclusive) of 125-225.

Each scan is built by removing five random isolation windows from the isolation list. This process is repeated for each scan until the isolation window list is empty (20 scans in this case) at which point the list is repopulated. By sampling from the list without replacement, each isolation window is sampled every 20 scans (~3.5 seconds) on average and is guaranteed to be sampled at least every 40 scans. Since these data have been collected, an additional optimization avoids the selection of the same pairs of isolation windows in nearby scans. If selecting five isolation windows randomly from the isolation list results in pairing two windows together that

have been paired in the same scan recently, the random selection is repeated until this is no longer the case or too much time has passed. With the 5 x 4 *m/z* multiplexing scheme, it is typical to not see the same two isolation window pairs observed within 200 scans (~35 seconds) of each other.

Scans are defined prior to an experiment using this method in Skyline (v1.3) and exported as an inclusion list containing a sequence of 5,000 isolation windows (1,000 scans). During analysis, the Q-Exactive loops through this list sequentially choosing five isolation windows for each MSX scan. This capability is implemented in a firmware modification which keeps the mass list from being sorted by the instrument computer prior to data acquisition.

When an MSX scan is acquired, each isolation window is isolated, fragmented, and stored serially-in-time prior to mass analysis in the Orbitrap. The fill time for all isolation windows in a given scan are the same. If the fill times for each window were different, it would be impossible to calculate the ion current (charges/second) for a fragment ion population because the fill time for the precursor(s) of that fragment ion population would be unknown.

To implement automated gain control (AGC), an MS1 scan is acquired every 10^{th} scan. For each MSX scan, the most recent MS1 scan is analyzed to determine the summed ion current for all 5 precursor isolation windows in the scan. The summed ion current is determined by summing up the total MS1 extracted ion current for each of the five isolation windows. The target number of ions is divided by this summed ion current to determine the fill time (the same for each window). For each MS/MS scan:

1) The most recent MS scan is analyzed to determine the ion current (ions / second) for the 5 windows to be isolated. This analysis is done by summing up the total MS1 signal in the 5 discontiguous 4 *m/z* windows that will be isolated in the MS/MS scan.

2) The fill time is calculated by dividing the AGC target (# ions) by the ion current calculated in step 1 (ions / second). This is the amount of time that each of the 5 windows should be filled for. Each window is filled for this amount of time, with a maximum of 20 ms.

Using this approach, the total number of ions in the trap is fixed, but the number of ions from any of the five windows will vary. The maximum fill time for an isolation window is set to 20 milliseconds, meaning the maximum total fill time for a scan is 100 ms (5 windows x 20 ms). MS1 and MSX scans are acquired with resolving power 35,000 and 17,500 respectively. The AGC target for MS1 scans is set to 10^6 ions and 10^5 ions per isolation window for MS1 and MSX scans respectively.

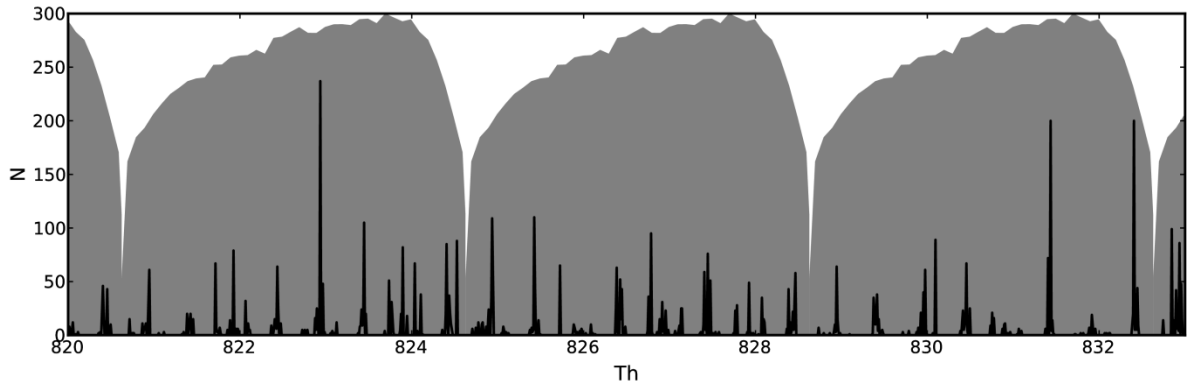


Figure 3.2: Optimized Placement of Isolation Window Edges.

As described in the Methods section, the edges of the isolation windows for DIA experiments were optimized to place the edges in regions where tryptic peptides are unlikely to occur to minimize splitting of precursor signal and inefficient isolation. The black line is a histogram of the number of precursor peptides (+1, +2, and +3 charged) in the Bibliospec spectral library. The grey overlay shows the 4 *m/z* windows targeting a subset of the 500 – 900 *m/z* range covered in our MSX experiment. The quadrupole mass filter in the Q-Exactive does not isolate with equal efficiency across an isolated *m/z* window (i.e. the efficiency near the edges is less than at the center). The quadrupole isolation efficiency on our Q-Exactive is superimposed in grey.

De-Multiplexing by Non-Negative Least Squares Optimization

Due to the random selection of isolation windows for each scan, any two spectra with an isolation window in common will not share any other isolation windows. This characteristic helps when de-multiplexing these spectra because overlapping fragment ion information is more likely to come from the single isolation window the spectra have in common than any of the others. In other words, spectra can be de-multiplexed by leveraging information (the isolation windows used and fragment ion intensities) from neighboring spectra. To do this, a system of equations is built describing each observed multiplexed spectrum as a linear combination of multiple unobserved single-precursor component spectra (Figure 3.3).

The system of equations is represented by matrix multiplication:

B = A * X

Matrix B contains the fragment ion intensities for the multiplexed spectra: each column is a fragment ion *m/z*, each row is a spectrum. Matrix A contains the isolation window pattern for each spectrum, each column is an isolation window, and each row is a spectrum. Isolation windows have the value 1 if they are present in a spectrum and 0 if they are absent. The matrix X is an unknown matrix containing the de-multiplexed component spectra. Each row is a de-multiplexed spectrum for an isolation window, each column is a de-multiplexed fragment ion intensity. The system of equations is solved by non-negative least squares[74]. For each spectrum to be de-multiplexed, this system of equations is built containing 140 consecutive MSX spectra with the spectrum to be de-multiplexed at the center of this window. 140 spectra are used for the 5 isolation window experiment, but for other methods, the number of spectra in the window is (# isolation windows total / isolation windows per scan) * (2 + isolation windows per scan). Setting

the number of spectra this way guarantees that the system of equations is not underdetermined. Using neighboring spectra to de-multiplex a spectrum has the negative consequence of reducing signal intensity due to the averaging of the de-multiplexed spectra with nearby spectra similar to a boxcar smooth. To minimize this, each row of matrix B and A are multiplied by the Savitzky-Golay smoothing coefficient (5 wide, second order) for that row. Finally, each de-multiplexed peak is normalized such that the sum of the intensities from each de-multiplexed component is equivalent to the intensity of the original observed peak. This de-multiplexing happens automatically on import into Skyline.

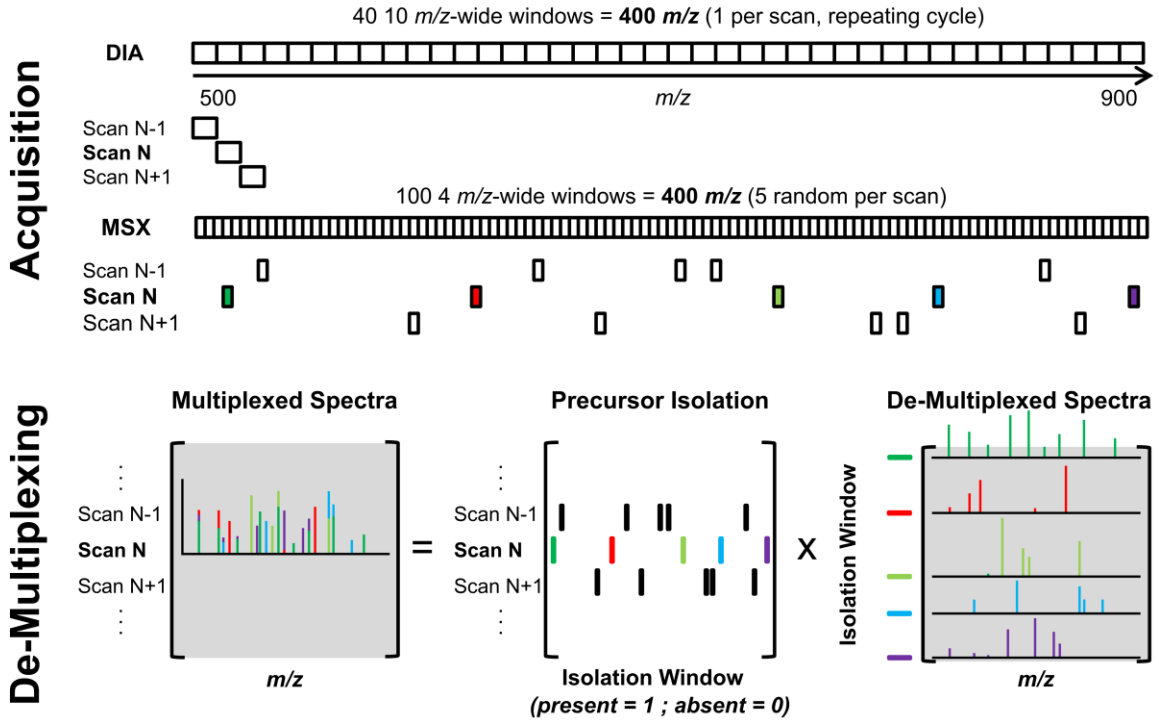


Figure 3.3: Multiplexed data independent acquisition (MSX).

A common implementation of data independent acquisition (DIA) is to use a repeated cycle of wide isolation window MS/MS scans to cover a mass range. In this example, the 500-900 *m/z* range is covered with 40 scans each sampling a single 10 *m/z* wide window. In multiplexed DIA (MSX), each scan isolates five 4 *m/z* wide windows prior to fragment ion mass analysis. The 5 windows isolated in each scan are chosen randomly from the set of 100 possible non-overlapping windows covering the 500-900 *m/z* range. Each mixed MSX spectrum is de-multiplexed into the 5 component spectra corresponding to each isolated window.

Evaluation of De-Multiplexing

A soluble *S. cerevisiae* lysate was reduced, alkylated, digested for an hour with trypsin, and cleaned by dual mode solid phase extraction (Oasis MCX cartridges, Waters Corporation). LC-MSX data were acquired on the Q-Exactive as described above on 1.2 µg of sample per injection. A 40 cm 75 µM fused silica column packed with reversed-phase C12 Jupiter resin (Phenomenex) was used to separate the sample across a 90-minute linear acetonitrile gradient from 0 to 25% Buffer B. Chromatography was performed using a Thermo EASY-nLC II system set to a flow rate of 250 nL/min. Buffer A was 2% ACN, 0.1% formic acid, and 97.9% water. Buffer B was 99.9% ACN and 0.1% formic acid. The data were analyzed using a modified version of Skyline which processed the data with and without de-multiplexing enabled.

Spike-in Experiment to Evaluate Quantitative Performance

An equimolar six protein digest (Bruker-Michrom) was spiked into a complex matrix (soluble *S. cerevisiae* lysate digest) over four orders of magnitude to test the quantitative performance of MSX. The yeast sample was reduced, alkylated, digested for an hour with trypsin, and cleaned by dual mode solid phase extraction (Oasis MCX cartridges, Waters Corporation) after digestion. Prior to the spike-in experiments, the complex matrix was run four times to condition the liquid chromatography column. Spike-in experiments with 0.6 µg of the complex matrix loaded on column with 0.05, 0.12, 0.31, 0.77, 1.92, 4.8, 12, 30, and 75 femtomoles of the bovine protein digest spiked in were run followed by blanks of the complex matrix and buffer A, followed by 0.07, 0.16, 0.41, 1.02, 2.6, 6.4, 16, 40, and 100 fmol spike-ins. Samples were run using the same chromatography conditions and column type as in the previous section “Evaluation of De-

Multiplexing," except using C18 Aqua resin (Phenomenex) instead of the C12 Jupiter reversed-phase resin.

36 peptides from 5 of the spiked-in proteins and not present in the background matrix were quantified (Appendix A). Peptides were quantified using the area under the curve of the M, M+1, and M+2 peaks for MS1, and a manually curated subset of the b- and y- ion series for MSX (Appendix A). The signal from each peptide was normalized by the signal from two highly-abundant peptides (DNSQVFGVAR++ and ESTLHLVLR++) from the background matrix. The lower limit of detection was determined for each peptide by manual inspection. For MSX data, the lower limit of detection was the lowest abundance where at least three transitions co-elute and have similar intensity ratios to that seen at higher abundance. For MS1 data, the lower limit of detection was the lowest abundance where at least the M and M+1 isotope peaks co-elute, and have the same rank-order as at higher abundances. The sensitivity of the two techniques was compared using a paired t-test on the log-transformed lower limits of detection for the 36 peptides (37 precursors) quantified. Regression lines were fit to the MS1 and MSX data for each peptide using all points at or above the lower limit of detection.

Section 3: Results

MSX LC-MS/MS data on a *S. cerevisiae* lysate were collected using five 4-*m/z* wide isolation windows per scan on a Q-Exactive (Thermo Fisher Scientific, Bremen, Germany) mass spectrometer. Due to the multiple fills per mass analysis, this multiplexing technique is best suited for instrumentation where isolation and collisional activation of peptides is fast relative to mass analysis. Five of the 100 possible 4 *m/z* isolation windows in the range of 500-900 *m/z* were selected randomly to be analyzed in each multiplexed scan (Figure 3.3, Materials &

Methods). The average time between repeated isolation of a 4 *m/*z window is 3.64 +/- 1.49 seconds (standard deviation). To analyze these spectra, Skyline was modified to detect MSX spectra and de-multiplex them automatically on import (available in v. 1.3). A Skyline document was generated containing peptides with spectra in the NIST *S. cerevisiae* Q-TOF and Ion Trap spectral libraries (5/24/2011 builds)[75]. The spectra were analyzed with and without de-multiplexing.

Figure 3.4 shows the full b- and y-ion series for the peptide GPLVLEYETYR with and without de-multiplexing. Without de-multiplexing, there are many intense fragment ion peaks present throughout the gradient which are fragments from other peptides (Figure 3.4a). After de-multiplexing (Figure 3.4b), the majority of these interfering peaks are removed because they do not originate from precursors in the same isolation window as GPLVLEYETYR. There is also interference in many of the fragment ion chromatograms that overlap in elution time with GPLVLEYETYR (Figure 3.4c). This interference has a characteristic "spike" shape because it originates from precursors in different isolation windows than the target peptide. Due to the random sampling of isolation windows for each scan (Materials & Methods, Figure 3.3), the window containing the interfering precursor is not isolated in consecutive scans containing the target isolation window. De-multiplexing removes the interfering signal while retaining the signal originating from the target isolation window (Figure 3.4d) resulting in a higher dot-product similarity (0.96 vs. 0.94) to a DDA spectrum for GPLVLEYETYR acquired on the same sample, under the same conditions, using the Q-Exactive with a 2 *m/z* wide isolation window. Figure 3.5 and Figure 3.6 show a comparison of data acquired by MSX (and demultiplexed) compared to data acquired using consecutive 40 10 *m/z*-wide isolation windows covering 500-

900 *m/z*. In both figures, the number of interfering peaks (peaks besides the target peptide) is greatly reduced in the MSX data compared to the 10 *m/z* DIA data.

To test the performance of this method for peptide quantification, a commercial 6 bovine protein digest (Bruker-Michrom) was spiked into a complex matrix (*S. cerevisiae* lysate, soluble fraction) in amounts ranging from 50 attomoles to 100 femtomoles on-column. MS1 and MSX data were collected simultaneously on each spike-in sample by acquiring MSX scans (R.P. 17,500) with MS1 scans (R.P. 35,000) interlaced every 10 scans. Five of the proteins were quantified using a total of 36 peptides (Figure 3.7, Appendix A).

The lower limit of detection for the 36 peptides quantified averaged 8.66 and 4.98 femtomoles for MSX and MS1 respectively. All peptides show a linear response above the limit of detection with R^2 values of the regression lines averaging 0.95 and 0.98 for MSX and MS1 respectively. Although the MSX method is less sensitive than the MS1 method on average ($p = 0.016$, paired t-test), the results are impressive given that the MSX data provide structural selectivity information in addition to quantification. The slight reduction in sensitivity of MSX relative to MS1 is not unexpected in the absence of chemical noise. MS1 is expected to be more sensitive in a simple mixture because in MS/MS the ion beam is split into multiple fragmented products, each a fraction of the original intensity[76]. Consider an ion beam for a peptide precursor ion of 10^6 ions/sec. If MS/MS is performed on this ion beam with 100% fragmentation of the precursor, zero losses, and the signal split between 20 equally abundant fragments, then in this absolute best case scenario each fragment would have an ion beam of $5x10^4$ ions/sec. If there is no chemical noise in the MS measurement, it will perform better because all of the ions are in a single beam. However, in more complex regions, with more chemical noise, chemical

interferences should decrease in intensity faster than the signal from the target analyte and MS/MS quantification should improve sensitivity.

7 of the 36 peptides suffered from interference in the MS1 signal, resulting in an average 3.4-fold improvement in sensitivity ranging from 1.3-8.3 fold by MSX quantification. Acquiring MS1 and MSX data simultaneously combines the high sensitivity of MS1 with the structural selectivity of MSX while theoretically improving quantitative precision by combining measurements from the precursor (MS1) and fragment ion peaks (MSX) for quantitation. In Figure 3.7, the data for a peptide (LVNELTEFAK++) where the improved selectivity of MSX results in greater sensitivity (LoD of 0.41 femtomoles) compared to MS1 (LoD of 1.02 femtomoles) is plotted. The slopes of the regression lines are 0.030 ± 0.004 (95% confidence interval) and 0.071 ± 0.002 for MSX and MS1 respectively. Figure 3.8 shows the signal used for quantitation from the MS1 (sum of M, M+1, and M+2 peaks) and MSX data (sum of fragment ion areas) at 1.02 femotomoles. The signal-to-noise is much improved in the MSX data. The CV of 18 replicate measurements of 6 background yeast peptides average 0.1 and 0.15 by MS1 and MSX respectively (Figure 3.9). The standard deviation in these measurements was 1.56 x 10^8 and 2.29 x 10^7 for MS1 and MSX respectively (Figure 3.10).

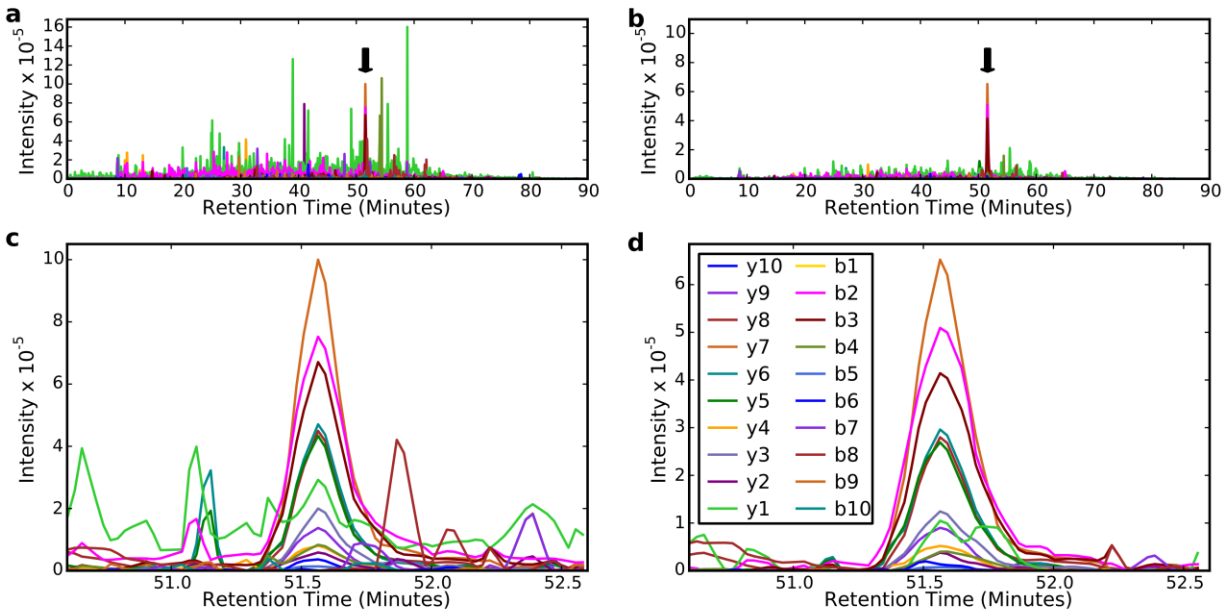


Figure 3.4: De-Multiplexing Reduces Chemical Noise and Improves Selectivity.

The full b- and y- ion series for the peptide GPLVLEVETYR are plotted from an MSX experiment on the soluble fraction of *S. cerevisiae* lysate prior to (**a, c**) and after (**b, d**) de-multiplexing. Prior to demultiplexing, there are many other peaks of similar or greater intensity than the peak for GPLVLEVETYR (indicated with an arrow in (**a**)). After de-multiplexing, peaks from other precursors are effectively removed and the true peak is by far the most intense (**b**). Additionally, the de-multiplexed peak (**d**) contains far less interference than the unprocessed peak (**c**). All peaks were extracted with a fragment ion tolerance of twice the full width at half maximum at a resolving power of 17,500 @ 200 *m/z*.

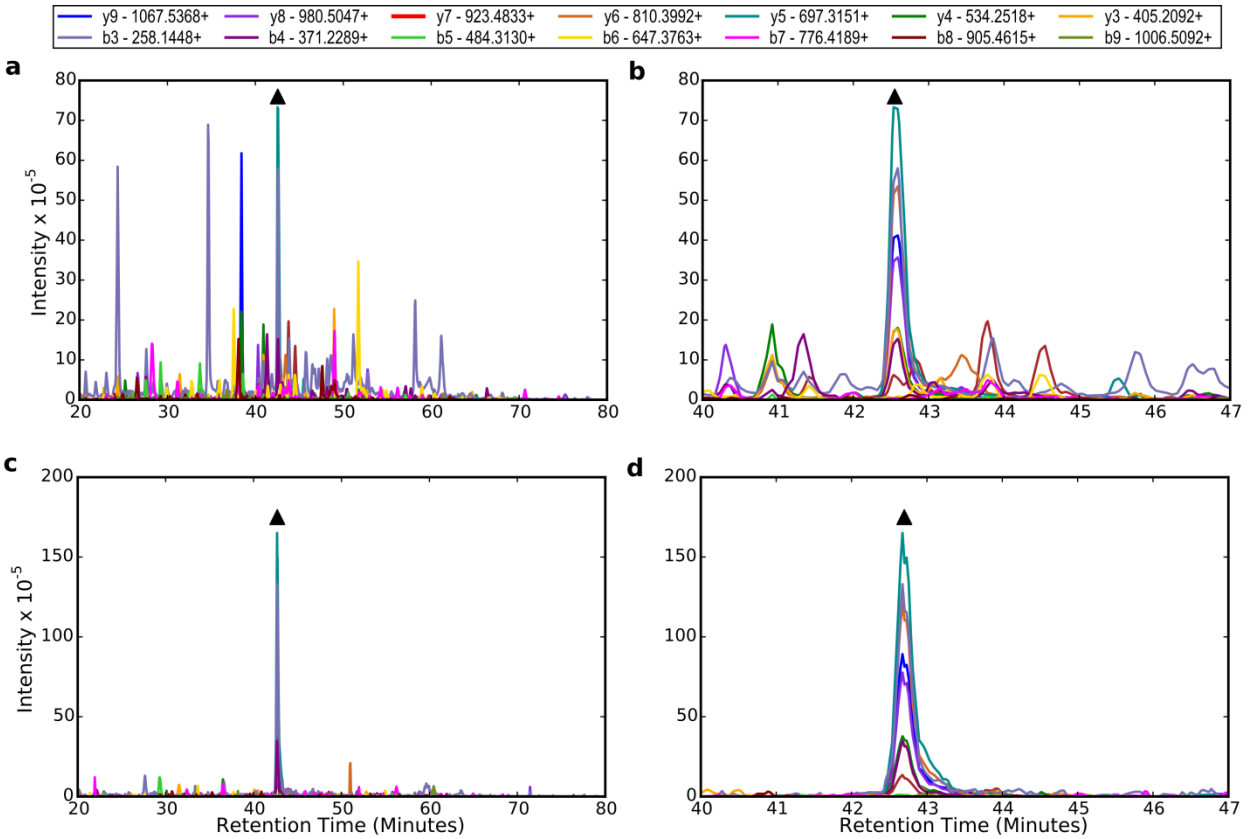


Figure 3.5: 10 m/z DIA and MSX Analysis of ISGLIYEETR++ peptide.

A *C. elegans* soluble lysate was analyzed on the Q-Exactive using a 10 *m/z* wide isolation window DIA approach in which the mass range from 500 *m/z* to 900 *m/z* is analyzed with consecutive 10 *m/z* wide isolation window targeted MS/MS scans and a 5 x 4 *m/z* MSX approach covering the same mass range. Fragment ion chromatograms for the peptide ISGLIYEETR++ were extracted using Skyline from both the wide-window DIA (**a** and **b**) and MSX data (**c** and **d**). The peak with the triangle hovering over it is the peak corresponding to this peptide as validated in a DDA analysis. In both the zoomed-out views (panels **a** and **c**) and zoomed-in views (panels **b** and **d**), the improved precursor selectivity of MSX results in much less non-target peaks.

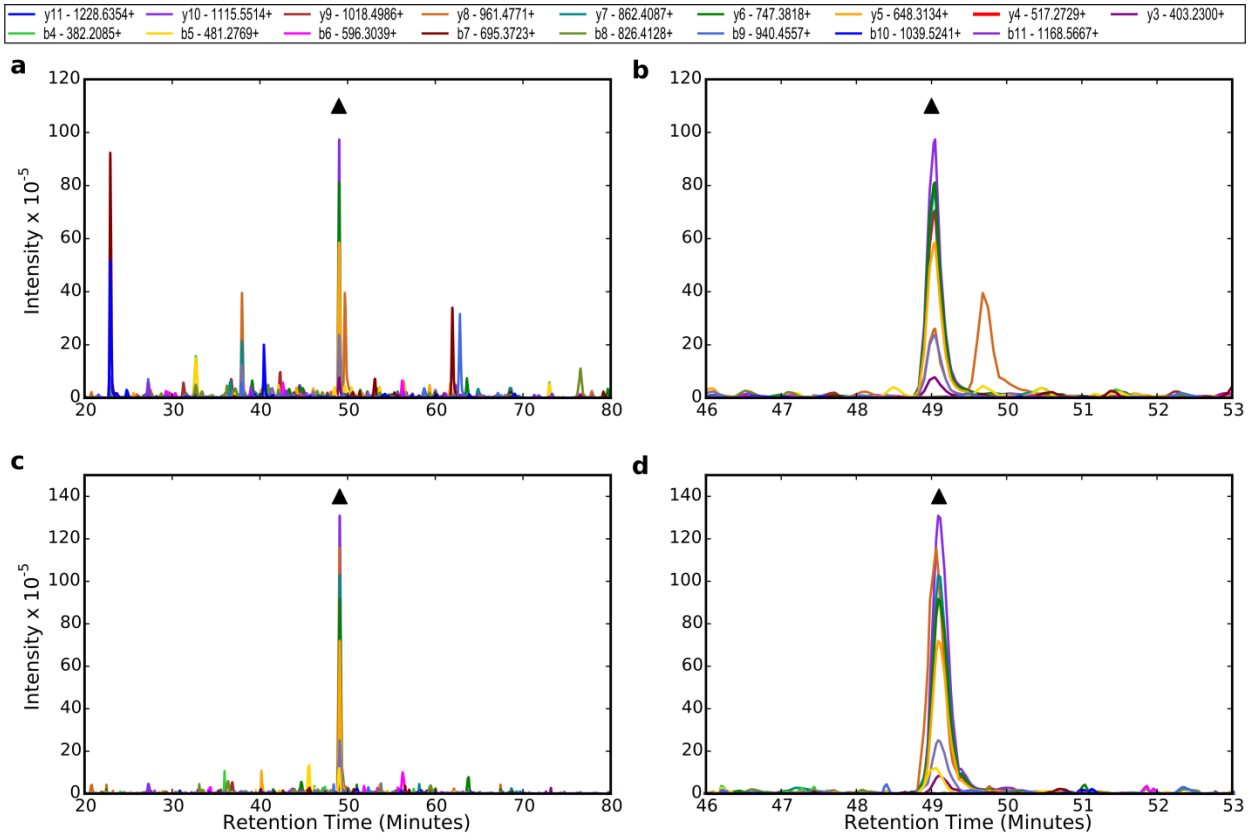


Figure 3.6: 10 m/z DIA and MSX Analysis of NIPGVDVMNVER++ peptide.

A *C. elegans* soluble lysate was analyzed on the Q-Exactive using a 10 *m/z* wide isolation window DIA approach in which the mass range from 500 *m/z* to 900 *m/z* is analyzed with consecutive 10 *m/z* wide isolation window targeted MS/MS scans and a 5 x 4 *m/z* MSX approach covering the same mass range. Fragment ion chromatograms for the peptide NIPGVDVMNVER++ were extracted using Skyline from both the wide-window DIA (**a** and **b**) and MSX data (**c** and **d**). The peak with the triangle hovering over it is the peak corresponding to this peptide as validated in a DDA analysis. In both the zoomed-out views (panels **a** and **c**) and zoomed-in views (panels **b** and **d**), the improved precursor selectivity of MSX results in much less non-target peaks.

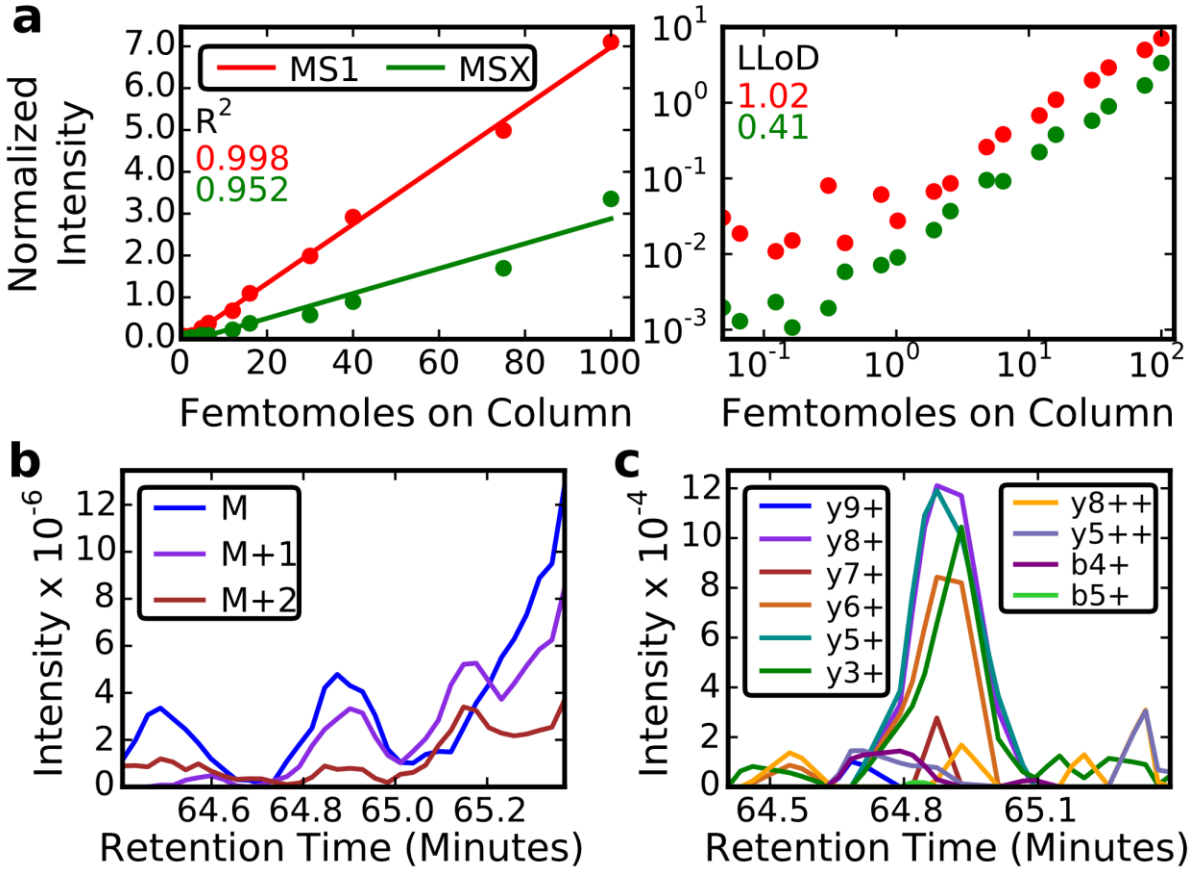


Figure 3.7: Quantitation of the LVNELTEFAK Peptide by MSX and MS1.

A commercial six protein digest was spiked into a complex matrix (soluble *S. cerevisiae* lysate) at amounts ranging from 50 attomoles – 100 femtomoles on column. MSX data were acquired with an MS1 scan interleaved every 10 scans. The signal intensity for each spike in point (normalized to two background peptides) for the peptide LVNELTEFAK++ is plotted in (**a**) with (right pane) and without (left pane) log scaling of the x and y axes. The lower limit of detection is 0.41 femtomoles and 1.02 femtomoles for MSX and MS1-based quantitation respectively. (**b**) and (**c**) show the MS1 and MSX signal at 1.02 femtomoles. Although on average MS1 quantitation is more sensitive than MSX, chemical noise hinders quantitation of this peptide by MS1 and MSX is more sensitive due to greater selectivity. For both MS1 and MSX measurements used, the *m/z* tolerance for extraction of signal was twice the full-width at half maximum (FWHM) for the peak being extracted. The ion tolerance for MS1 data is less because it was acquired at 35,000 resolving power @ 200 *m/z*, while the MS/MS data used for MSX quantitation was acquired at 17,500 resolving power @ 200 *m/z*.

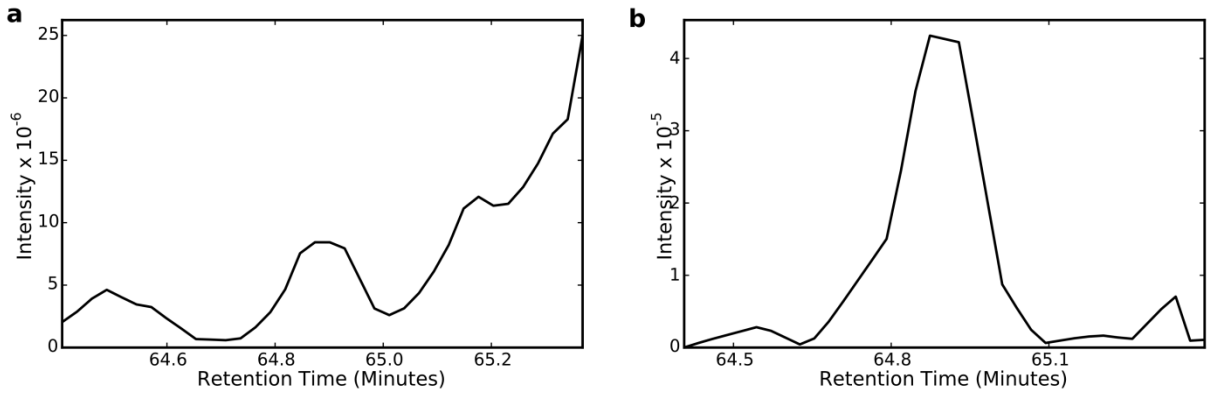


Figure 3.8: Summed Ion Current of MS1 and MSX Signals for LVNELTEFAK++ Peptide.

Serum Albumin (part of a commercial six protein digest) was spiked into a complex matrix (soluble *S. cerevisiae* lysate) at amounts ranging from 50 attomoles – 100 femtomoles on column. MSX data were acquired with an MS1 scan interleaved every 10 scans. The summed ion current for MS1 quantitation and MSX quantitation are plotted at the lower limit of detection for each technique. The summed ion current from the M, M+1, and M+2 masses used for MS1 quantitation is plotted in (**a**) at 1.02 femtomoles (the lower limit of detection by MS1). The summed ion current from the transitions used for quantifying the same peptide at the same spike in amount using MSX is plotted in (**b**).

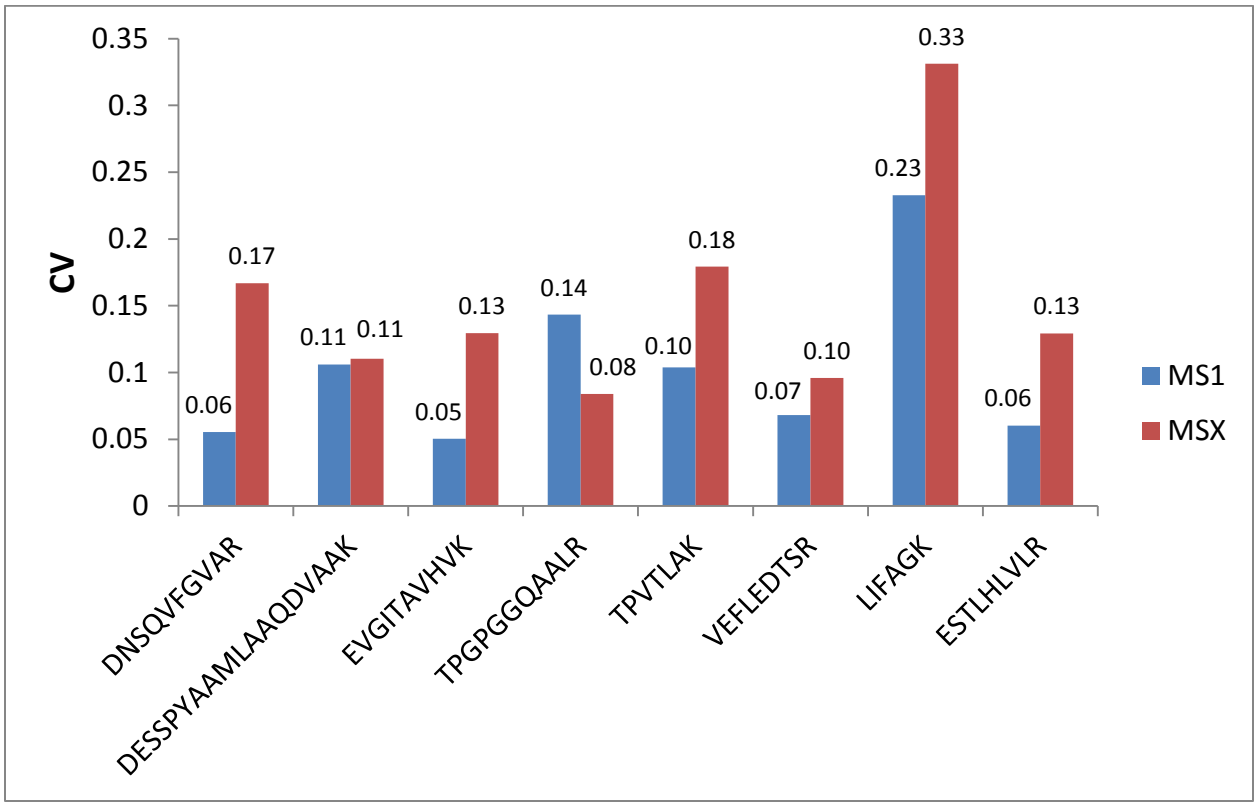


Figure 3.9: Reproducibility of Peptide Quantitation (CV).

The reproducibility of quantitation by MS1 is compared to MSX by calculating the CV in peak area for peptides in the *S. cerevisiae* background matrix over 18 injections. MS1 and MSX data were acquired in the same instrument runs. The average CV by MS1 and MSX is 0.1, and 0.15 respectively. The signal for MS/MS-based quantitation is reduced, causing an increase in the relative error (CV).

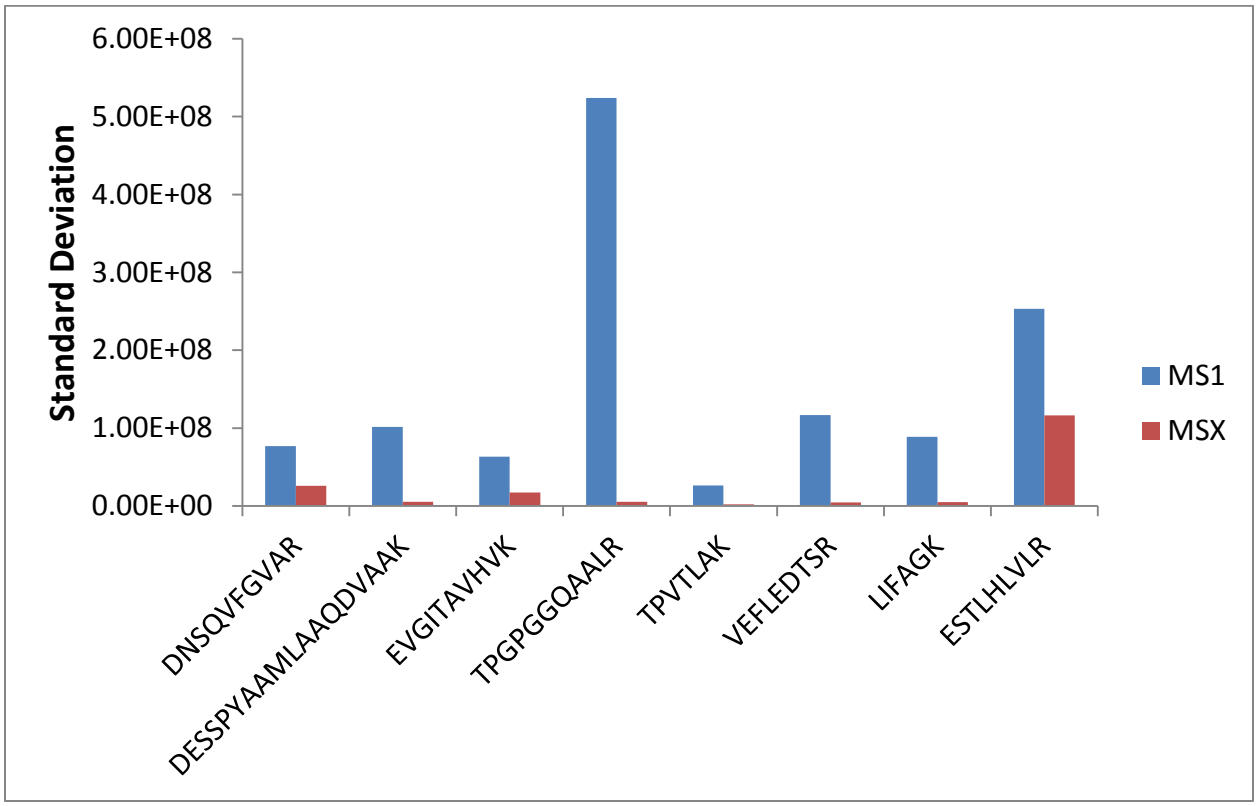


Figure 3.10: Reproducibility of Peptide Quantitation (StDev).

The reproducibility of quantitation by MS1 is compared to MSX by calculating the standard deviation in the peak area for peptides in the *S. cerevisiae* background matrix over 18 injections. MS1 and MSX data were acquired in the same instrument runs. The standard deviation in peak areas by MSX is lower as expected due to reduced variance in peaks of lower intensity. The average standard deviation is 1.56×10^8 and 2.29×10^7 for MS1 and MSX respectively.

Section 4: Discussion

Performing an MSX experiment only makes sense on hardware where the isolation and collisional activation of the peptides is fast relative to the mass analysis. In this case, we can isolate and activate 5 different precursor windows in the same time it takes to collect a spectrum at 17,500 resolving power in the Orbitrap. In contrast, this experiment would not make any sense to perform on a Q-TOF type hybrid instrument where the precursor isolation and activation is slow relative to the time-of-flight mass analyzer.

There is potential for optimization of certain parameters to improve the performance of this method. Currently, the orbitrap is underfilled in the majority of scans (Figure 3.11). It would be preferable to fill the orbitrap to improve sensitivity and dynamic range. If the width of isolation windows were increased, there would be a greater ion current, and the orbitrap would be filled more frequently. However, precursor selectivity would be reduced.

It may be possible to increase the resolution of MS/MS scans, but this is unlikely to markedly improve results. Product ion resolution is useful to some point but it is limited and in many ways less useful than precursor ion resolution. Unfortunately, a lot of peptides have similar sequence but different precursor mass. In these cases, no amount of product ion resolution can help because the product ion fragments have the exact same elemental composition. Additionally, on the Q-Exactive, increased resolving power comes at the cost of instrument scan speed. The maximum resolving power that could be used with 20 *m/z* windows is 35,000 on this instrument due to the necessity of having a duty cycle of 3.5 seconds or less. At 35,000 resolving power, the orbitrap takes 128 ms to acquire a transient. With the simplifying assumption that each scan takes 128 ms (no scan overhead time, AGC scan time unaccounted for), the amount of time it would take to perform the 20 MS/MS scans required to cover 500-900 *m/z* is 2.56 seconds. The next highest resolving power (70,000) would bring the duty cycle time up to 5.12 seconds at the minimum, which is much too slow on an HPLC time scale. It may be possible to acquire data at 35,000 R.P. rather than 17,500 as in this manuscript, but certainly not any higher.

It is important to note that an increase of resolving power to 35,000 is applicable to both the MSX and continuous window approaches. Both the MSX method, and the continuous 20 *m/z* wide isolation window approach isolate 20 *m/z* of the mass range per scan and therefore require

the same number of scans to cover the 500-900 *m/z* range. Additionally, for both approaches, the scan time is limited by the amount of time dedicated to mass analysis in the orbitrap since ion fills and mass analysis occur simultaneously. Therefore, the cycle time for both approaches should be the same, as well as the selectivity on the MS/MS level. However, the MSX technique would have improved precursor selectivity over the continuously acquired wide windows.

We covered the mass range from 500-900 *m/z* because previous work indicate that most of the best responding peptides from a tryptic digest fall in this range (Canterbury JD *et. al.*; ASMS 2010). Once the mass range is expanded outside of this range, there are diminishing returns in the number of peptides analyzed due to the non-uniform distribution of strong responding peptides across the *m/z* range. We reduced the mass range in comparison to Gillet *et. al.* as a compromise between mass range coverage, and duty cycle time. It would be possible to cover a wider mass range with these techniques by increasing precursor isolation window size (i.e. 5 x 6 *m/z* wide windows), increasing the number of precursors isolated per scan (i.e. 6 x 4 *m/z* wide windows), or simply accepting a slower duty cycle. However, selecting the optimal mass range to cover, isolation width, and number of isolation windows per scan is a non-trivial exercise which will depend on the complexity of the sample being analyzed.

As far as the required hardware configuration to perform a MSX DIA experiment, the Q-Exactive is currently the only commercially available instrument that can perform this experiment. We require an instrument that can isolate a precursor mass range, activate it, and store the fragments in an ion trap. The ion trap must be able to receive multiple fills with minimal ion losses of the ions already stored in the trap. This ion trap is then used to send the mixed ion population to a high resolution mass analyzer with a large dynamic range.

We have demonstrated the efficacy of multiplexed DIA in addressing the biggest remaining problem preventing wider adoption of these methods, low precursor selectivity. With wide isolation windows, there is an increased likelihood of relying on fragment ion data alone to discern between peptides with very similar fragmentation patterns (e.g. modified forms, or similar sequence). Even if a peptide can be unambiguously identified in these data, quantitation is negatively impacted by increased fragment ion interference. The improved precursor selectivity of the MSX technique thus improves identification and quantitation of peptides. Additionally, acquisition of MS spectra along with MS/MS spectra aides in identification of peptides; modified and unmodified peptides will have similar fragmentation patterns, but will have different precursor masses. The MS data also can be more sensitive than MS/MS for some peptides in the absence of chemical noise. With these obstacles addressed, other proven advantages of DIA – including increased dynamic range, and more complete and reproducible MS/MS sampling across technical replicates – may be employed. This multiplexing technique, coupled with recent improvements in the interpretation of DIA data[30], make this a practical technique for global, highly-specific, and reproducible relative quantitation of peptides in a shotgun experiment.

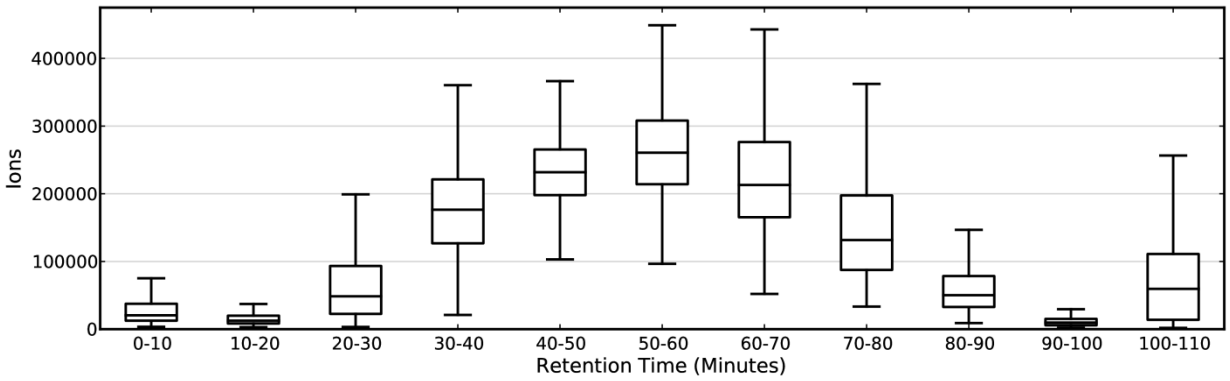


Figure 3.11: MSX Ion Counts.

The number of ions in the trap for MS/MS scans from an MSX analysis of the soluble *S. cerevisiae* lysate used as background in the spike-in experiment is plotted. Less ions enter the trap at lower and higher retention times due to less total ion current at the extremes of the acetonitrile gradient used to separate the peptides. The AGC target for these data is 500,000 ions. The whiskers are the most extreme data point within 1.5 times the inner quartile range.

Chapter 4 : Elucidating the Mechanism of Lifespan Extension by Rapamycin Treatment in Yeast.

Section 1: Introduction

In his 1991 book *Evolutionary Biology of Aging*, Michael Rose defined aging as "a persistent decline in the age-specific fitness components of an organism due to internal physiological deterioration"[77]. The underlying evolutionary cause is thought to be declining selective pressure post-reproduction combined with the benefits of early reproduction in the face of constant external threats such as predation or natural disaster. It is natural to assume that our aging is similar to that of a machine: an inevitable decline in function due to irreparable "wear and tear". However, the past 25 years of research into the mechanisms of aging indicate that aging is a regulated cellular process that can be modulated behaviorally (e.g. dietary restriction), genetically, and potentially pharmaceutically[78]. The ability to slow aging could have a major impact on human health. As we age, our likelihood of incurring a serious, potentially terminal, disease such as cancer or heart disease increases rapidly. If aging truly is the underlying cause, an intervention to delay aging would simultaneously reduce the impact of all of these ailments and improve human health dramatically. To slow aging in humans, more work must be done to understand the underlying mechanisms and causes of aging and determine if techniques that slow the aging process in model organisms will translate to humans.

Dietary restriction is the most broadly applicable and consistently effective technique to slow aging. Reducing caloric intake relative to *ad libitum* intake short of starvation increases the lifespan of a diverse group of organisms including yeast, nematodes, flies, fish, mice, rats, spiders, and potentially primates[79]. The increase in lifespan ranges from 30-50% in mice up to 3-

fold in yeast[80]. In addition to extending life span, dietary restriction improves the health of the animal. For example, caloric restricted mice show reduced neurodegeneration and are protected against cancer, diabetes, and cardiomyopathy[81]. This highly-conserved effect may have evolved as a response to periods of nutrient scarcity. When nutrients are scarce, it may be advantageous for the organism to divert energy into pathways to promote longevity until nutrients are more plentiful and the environment more fit for reproduction. The dauer life-stage of the nematode *C. elegans* is an extreme example of this behavior[82]. After the L1 larval stage, if environmental conditions are non-optimal (low population density, low nutrient availability, high temperature), the animal undergoes dauer arrest. The dauer form is a reduced metabolic state that is resistant to environmental insult and can persist for months prior to exit from the state and continuation of development into an adult.

The dietary restriction response is mediated by multiple nutrient-sensing pathways with the interacting Insulin/IGF-1 signaling (glucose response) and target of rapamycin (TOR; amino-acid sensing) pathways being particularly notable[83]. Reducing the activity of either of these pathways extends lifespan in a wide range of organisms from yeast to mice. Here, we focus on studying the highly-conserved TOR pathway which is unique because its activity can be reduced by treatment with rapamycin, a macrolide used in humans as an immunosuppressant and treatment for certain cancers[84].

Attenuation of TOR pathway signaling extends lifespan in *S. cerevisiae* (chronological and replicative), *C. elegans*, *D. melanogaster*, and *M. musculus*. In a mouse study, rapamycin treatment was found to delay aging even when administered to the mice at the age of 600 days[85]. Another mouse study found that rapamycin slowed age-related changes in heart, liver, adrenal glands, endometrium, and tendon[86]. The pathway plays a large role in the response to dietary

restriction. The replicative lifespan of *tor1* deletion *S. cerevisiae* is not further increased by dietary restriction. In *C. elegans*, the lifespan of *eat-2* mutant nematodes is not further extended by dietary restriction. In *D. melanogaster* TOR inhibition does not further increase lifespan in conditions of dietary restriction[84]. The pathway responds to amino acid content, growth factors, energy status (mediated by AMP-dependent kinase), and stresses such as hypoxia and high temperatures. Some of the downstream effects of TOR include reduced protein synthesis, a stress resistance response, increased autophagy, and changes in mitochondrial metabolism.

In this chapter, work on using mass spectrometry to study the response of *S. cerevisiae* to growth in rapamycin is reported. In S. *cerevisiae*, there are two TOR paralogs (Tor1 and Tor2) and two TOR complexes: TORC1 (sensitive to rapamycin) and TORC2. Tor1 has only been found in TORC1, Tor2 has been found in both. Using mass spectrometry to detect changes in yeast proteins in response to rapamycin treatment should provide additional insight into how reduced TOR signaling extends lifespan. A proteome study could capture post-translational changes missed in microarray studies[87].

There are already numerous examples of post-translational regulation downstream of the TOR pathway. The reduction of mRNA translation is one of the effects of reduced TOR signaling. When TOR signaling is reduced, ribosomal protein levels are reduced which leads to a reduction in translation[88]. However, although global mRNA translation is reduced, the abundance of the GCN4 transcription factor increases. The lifespan extension due to reduced TOR signaling appears to be partly dependent on induction of GCN4. Deletion of GCN4 reduces the change in replicative lifespan in *tor1* mutant yeast from ~55% to ~30%[89]. The differential translation of GCN4 is regulated by 5'UTR signaling. Additional examples are phosphorylation to regulate the activity of Sch9 and 4E-BP family of proteins[90, 91].

Section 2: Materials and Methods

S. cerevisiae strains and growth

The strains used were haploid wild type and *tor1* mutant *S. cerevisiae* strains from an ORF deletion collection[92] with the parental background BY4742 -- MATα hist3Δ1 leu2Δ0 lys2Δ0 ura3Δ0. Both strains were grown in YPD media in three biological replicates. The wild type strain was grown in the presence and absence of 10 nM rapamycin. Strains were grown to OD_{600} 0.6 prior to lysis and digestion.

Cells were lysed by bead beating in a lysis buffer of 50 mM ammonium bicarbonate at pH 7.8 with phosphatase inhibitors (Halt phosphatase inhibitor cocktail – Pierce). The lysis was performed by bead beating for one minute, followed by one minute on ice, repeated three times. With Rapigest(Waters) as a surfactant, the sample was reduced and alkylated by treatment with 5mM dithiothrietol and 15mM iodoacetamide. Finally, the sample was digested with modified sequencing grade trypsin (Promega) for one hour.

Mass spectrometry

1.2 μg of sample was loaded onto a 20 cm fused silica (75 μm inner diameter) liquid chromatography column. The column was pulled in-house using a CO_2 laser puller (P-2000; Sutter) and packed with C12 Jupiter reversed-phase resin (Phenomenex). A NanoAcquity HPLC system (Waters Corporation) was used to run a 60 minute linear mobile phase gradient of 2% to 32% buffer B (99.9% acetonitrile, 0.1% formic acid) in buffer A (0.1% formic acid in water) at a constant flow rate of 250 nL/min. The sample was eluted off of the column which was coupled on-line to a Q-Exactive (Thermo Scientific) mass spectrometer by an electrospray ionization (ESI) interface. Data were acquired using both a data dependent acquisition (DDA) method and a data independent acquisition (DIA) method with 20 *m/*z-wide windows covering 500-900 *m/z*.

The DDA method was a top-12 method with MS scans acquired with a resolving power of 35,000 @ 200 *m/z* and MS/MS scans acquired with a resolving power of 17,500 @ 200 *m/z*. The AGC target and maximum inject time for MS1 scans was 10^6 ions and 10 milliseconds respectively. The MS/MS scans had an AGC target of 5 x 10^5 ions and maximum inject time of 55 milliseconds. The MS/MS scans had a 2 *m/z* wide isolation window and normalized collision energy of 25. Dynamic exclusion was on with an underfill ratio of 2%, intensity threshold of 1.8 x 10^5, MIPS on, and a 20 second exclusion time. Precursors with charges that were unassigned, 1, or 5-8 were excluded from MS/MS acquisition.

In the DIA method, a cycle of 40 MS/MS scans total covering 500-900 *m/z* were acquired in a repeated cycle. The first 20 scans of the cycle consisted of 20 non-overlapping targeted MS/MS scans with 20 *m/z* windows together covering 500-900 *m/z*. The second 20 scans of the cycle were the same as the first, except with the target for each of the scans increased by 10 *m/z*. The scans from the second half of the cycle thus covered 510-910 *m/z*. The MS/MS scans were interlaced with one MS scan per 10 MS/MS scans. The MS scan was acquired with a resolving power of 35,000 @ 200 *m/z*, AGC target of 10^6 ions, and maximum fill time of 55 milliseconds. The MS/MS scans were acquired with a resolving power of 17,500 @ 200 *m/z*, AGC target of 10^6 ions, and maximum fill time of 55 milliseconds.

Analysis of Data Dependent Acquisition Data

DDA spectra were searched against a yeast open reading frame database (03/02/2011) downloaded from the Saccharomyces Genome Database (www.yeastgenome.org) using SEQUEST. Target and decoy (reversed) databases were searched with semi-tryptic enzyme specificity allowing for two missed cleavages and using a precursor *m/z* tolerance of 10 ppm and fragment ion tolerance of 0.36 *m/z*. The small precursor tolerance was used because the data

were pre-processed using Bullseye[63] prior to database searching. The data were searched with a fixed carbamidomethyl modification of cysteine. The target and decoy searches were analyzed using Percolator v 2.04 to assign a q-value to each peptide identification. Confident peptide identifications ($q<0.01$) were used to infer proteins using an algorithm based on the IDPicker algorithm[93] with each protein requiring one peptide for identification.

The proteins inferred by IDPicker were compared in the wild-type, wild-type + rapamycin, and *tor1* mutant yeast using spectral counting analysis. Spectral counts were normalized for each mass spectrometry run by the total number of spectra in that run. The spectral counts for each protein were standardized so the mean is zero and the standard deviation is one prior to clustering the data using hierarchical clustering.

Analysis of Data Independent Acquisition Data

DIA data were analyzed using Skyline[73]. All peptide precursors identified by DDA for each protein of interest with charge state +1, +2, or +3 and *m/z* between 500 and 900 were queried. The fragment ions extracted from the data were +1 and +2 charged y and b-ions from the 3rd ion in the series to the (last ion)-2 in the series. A chromatogram for each fragment ion was extracted from the data with a mass tolerance of twice the full width half max of the fragment ion peak assuming resolving power 17,500 @ 200 *m/z*. Additionally, chromatograms for precursor ions (M, M+1, and M+2) were extracted for the peptides with a mass tolerance of twice the full width half max of the precursor ion peak assuming resolving power 35,000 @ 200 *m/z*. Peptides were identified in the data by finding co-eluting fragment ion peaks that overlap or are very near to the retention time of identifications made in the DDA data. The integration boundaries for each peak were manually refined, and transitions with large amounts of interference were not integrated for quantitation. The integrated area for each peptide was

normalized by the total ion current for the entire mass spectrometry run. When peak areas were compared across runs, an independent Welch's t-test was used to determine a p-value for the significance of the abundance change between each yeast treatment (WT, WT + rapamycin, *tor1*) based on the variance across the three biological replicates for each.

Section 3: Results

Spectral Counting Analysis

DDA data were acquired on three biological replicates of wild type (WT; BY 4742) *S. cerevisiae*, wild type grown in the presence of 10 nM rapamycin (WT + rapamycin), and a *tor1* deletion mutant. The *tor1* mutant strain as well as WT + rapamycin are known to have an increased lifespan[84]. All three yeast cultures were grown to an OD_{600} of 0.6 prior to preparation for mass spectrometry analysis. The yeast grew ~33% slower in the presence of rapamycin, taking 8 hours to grow from OD 0.15 to 0.6 compared to 6 hours for the wild type and *tor1* yeast.

Peptides were identified using SEQUEST[11] combined with Percolator[25] for post-processing and assignment of peptide-level q-values. 5,288, 5,429, and 5,163 unique peptides were identified in the WT, *tor1*, and WT + rapamycin samples respectively ($q<0.01$). 711, 712, and 693 unique proteins were identified in these samples. Clustering of the samples by normalized spectral counts (Figure 4.1) indicates more similarity between the wild type and *tor1*samples than wild type and wild type + rapamycin samples. There appear to be few similarities between the long-lived WT + rapamycin and *tor1* samples.

Data Independent Acquisition Analysis of Mitochondrial Metabolism and Amino Acid Biosynthesis

Data were also acquired by DIA using 20 20 *m/z* wide windows covering 500-900 *m/z* on all of the samples (see Materials and Methods). From these data, fragment ion chromatograms were

extracted for 42 protein targets involved in mitochondrial metabolism (Table 4.1). Only peptides identified by DDA were used for quantitation in this analysis. 23 of the proteins had peptides associated with them identified by DDA (111 peptides total). Quantitation for each peptide was performed using the area under the curve of MS/MS fragment ion chromatograms when the peptide elutes (see Materials and Methods). Data on every peptide quantified can be found in Appendix B. Proteins involved in the tricarboxylic acid (TCA) cycle (CIT1, IDH2, KGD1, MDH1), the electron transport chain (COR1, COX4), and gluconeogenesis (PYC1, PYC2) were all found to have increased abundance in the rapamycin treated yeast relative to the wild type. However, these changes were not observed in the *tor1* yeast.

Additionally, targets of the GLN3 transcription factor, members of the TORC1 and TORC2 complexes, and proteins involved in stress response were analyzed (Table 4.2, Table 4.3, Table 4.4, and Table 4.5). Of the 31 proteins queried, 9 proteins had peptides identified in the DDA run and were compared (41 peptides total). Proteins involved in amino acid biosynthesis (ARG1, ARG4, ASN1, GLN1, HIS4) had higher abundance in the WT + rapamycin yeast compared to wild type, but not in the *tor1* strain. Additionally, IMD2, involved in GTP biosynthesis, had lower abundance in the WT + rapamycin sample, and potentially increased abundance in *tor1* yeast.

Spectral Counting Compared to Data Independent Acquisition

Spectral counting data was analyzed to find proteins with increased or decreased abundance in both WT + rapamycin and *tor1* samples relative to wild type. 19 proteins were found to have potentially increased abundance in both *tor1* and WT + rapamycin and 7 with potentially decreased abundance. Of these 26 proteins, 16 were detected in the DIA data (Table 4.6, Appendix B3), but only ATP2 was found to have decreased abundance in both *tor1* and WT +

rapamycin samples relative to wild type. All of the other proteins either did not have increased/decreased abundance in both the WT + rapamycin and *tor1* samples or had inconclusive results in the DIA data.

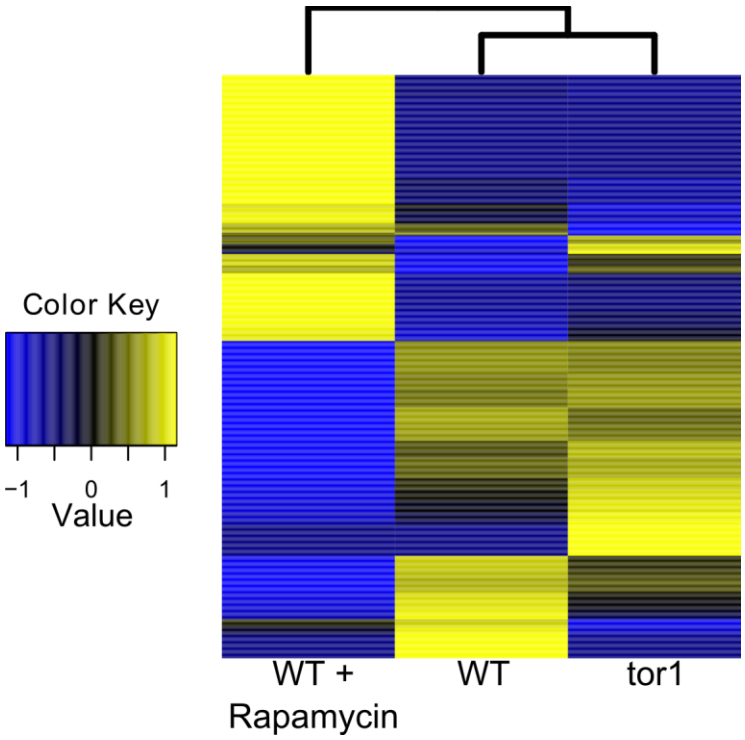


Figure 4.1: Spectral counting comparison of wild-type, rapamycin-treated and *tor1* yeast.

Normalized spectral counts for proteins detected in wild-type, wild-type yeast treated with rapamycin, and *tor1* mutant yeast are plotted. Each row is a protein, and each column is a sample. For each protein, the spectral counts are standardized to have a mean of zero and standard deviation of one. Hierarchical clustering was performed on the proteins, followed by clustering of the samples.

Protein Name	# Peptides	Protein Name	# Peptides
ACH1	1	GPD1	1
ACO1	9	GSY2	2
ACS1	0	GUT2	0
ADH2	2	ICL1	0
ADH6	2	IDH2	1
ALD6	15	KGD1	5
ATP1	5	LPD1	5
ATP2	8	LSC2	1
ATP3	0	MDH1	7
ATP4	0	MLS1	3
ATP5	0	NDE1	0
ATP7	0	NDI1	0
CIT1	6	PET9	1
COR1	1	PYC1	9
COX12	0	PYC2	10
COX4	2	QCR2	1
COX5A	0	QCR8	0
CYT1	0	RIP1	0
DLD1	0	SIT4	0
GCY1	0	STF2	0
GND1	14	YHI9	0

Table 4.1: Number of Peptides Detected for Proteins Targeted Involved in Mitochondrial Metabolism

Proteins targeted for quantitation using the area under the curve (AUC) of fragment ions from data independent acquisition data are listed with the number of peptides used for quantification for each. Only peptides identified by data dependent acquisition (DDA) were used for quantification of each protein. Peptides were only used for quantification if a signal for them could be detected in the DIA data.

Protein Name	# Peptides
ADH5	0
ARG1	2
ARG4	3
ASN1	15
CPA2	3
GGC1	0
GLN3	0
HIS4	7
IMD1	3
IMD2	4
SRY1	0
TMT1	1
UGA3	0

Table 4.2: Number of Peptides Detected for Proteins that are Targets of Gln3

Proteins targeted for quantitation using the area under the curve (AUC) of fragment ions from data independent acquisition data are listed with the number of peptides used for quantification for each. Only peptides identified by data dependent acquisition (DDA) were used for quantification of each protein. Peptides were only used for quantification if a signal for them could be detected in the DIA data.

Protein Name	# Peptides
KOG1	0
LTS8	0
TCO89	0
TOR1	0
TOR2	0

Table 4.3: Number of Peptides Detected for Proteins in TORC1

Proteins targeted for quantitation using the area under the curve (AUC) of fragment ions from data independent acquisition data are listed with the number of peptides used for quantification for each. Only peptides identified by data dependent acquisition (DDA) were used for quantification of each protein. Peptides were only used for quantification if a signal for them could be detected in the DIA data.

Protein Name	# Peptides
AVO1	0
AVO2	0
BIT61	0
LST8	0
SLM1	0
SLM2	0
TOR2	0
TSC11	0

Table 4.4: Number of Peptides Detected for Proteins in TORC2

Proteins targeted for quantitation using the area under the curve (AUC) of fragment ions from data independent acquisition data are listed with the number of peptides used for quantification for each. Only peptides identified by data dependent acquisition (DDA) were used for quantification of each protein. Peptides were only used for quantification if a signal for them could be detected in the DIA data.

Protein Name	# Peptides
GAP1	0
GLN1	3
HMS2	0
TAP42	0
URE2	0

Table 4.5: Number of Peptides Detected for Proteins Involved in Stress Response

Proteins targeted for quantitation using the area under the curve (AUC) of fragment ions from data independent acquisition data are listed with the number of peptides used for quantification for each. Only peptides identified by data dependent acquisition (DDA) were used for quantification of each protein. Peptides were only used for quantification if a signal for them could be detected in the DIA data.

Protein Name	# Peptides
APA1	1
ARD1	1
ATP1	3
ATP2	7
CCS1	0
GLO2	0
GUS1	9
HHF1	2
ILV5	11
NMD3	0
NPL3	2
PCS60	0
PMT2	0
PRE8	0
RHO1	0
RIB3	0
RIM8	1
RPL14A	3
RPL15A	5
RPL31A	4
RPN5	0
RPS20	4
RPS23A	2
SHM2	5
SOL3	1
UBP2	0

Table 4.6: Number of Peptides Detected for Proteins of Interest from Spectral Counting Analysis

Proteins that had a shared increase or decrease in WT + rapamycin and *tor1* yeast relative to WT according to spectral counting were quantified using DIA. Proteins targeted for quantitation using the area under the curve (AUC) of fragment ions from data independent acquisition data are listed with the number of peptides used for quantification for each. Only peptides identified by data dependent acquisition (DDA) were used for quantification of each protein. Peptides were only used for quantification if a signal for them could be detected in the DIA data.

Section 4: Discussion

Both the *tor1* and rapamycin treated yeast are long-lived compared to wild-type; however, clustering of the spectral counting data showed much greater similarity between the wild-type and *tor1* yeast than between the *tor1* and rapamycin-treated yeast. One potential explanation is that the rapamycin treatment induces a large amount of changes in the proteome in addition to the changes that cause an increase in lifespan. For example, the growth rate of rapamycin treated yeast is slowed, but not the *tor1* yeast. It could be that many of the changes detected are related to rapamycin-specific (such as the slowed growth rate) effects of which only a small portion are responsible for increased longevity.

Another potential explanation is that the *tor1* deletion and rapamycin treatment cause very similar changes, but with the rapamycin treatment causing the changes to a greater extent than the *tor1* mutation (i.e., the *tor1* mutation induces a more subtle version of the same changes caused by rapamycin treatment). Rapamycin targets the TOR1C complex in *S. cerevisiae* to reduce signaling through the TOR pathway[94]. Usually, TOR1 is found in this complex; however, TOR2 is sometimes found taking the place of TOR1. It is possible that in the *tor1* mutant, recruitment of TOR2 into the TOR1C complex partially compensates for the loss of TOR1 induced by the *tor1* deletion. If this were the case, it's possible that rapamycin-treatment is more effective at reducing TOR signaling because it targets the entire TOR1C complex, while a *tor1* deletion only targets a component of the complex which may be partially compensated for by recruitment of TOR2.

An increase in the abundance of proteins involved in the TCA cycle and the electron transport chain seem to indicate a shift toward cellular respiration in the rapamycin treated yeast but not the *tor1*and wild type yeast. A shift toward respiration in long-lived yeast seems

counterintuitive due to the generation of superoxide radicals in the mitochondria which may damage the cell. However, a shift toward respiration has been observed before in long-lived *S. cerevisiae* and this result is encouraging as a validation of the ability of DIA techniques to detect meaningful biological changes[95]. Similarly, the increased amino acid biosynthesis observed in the DIA data have also been observed in long-lived yeast and could be a by-product of increased GCN4 function[89].

Spectral counting data was used to find proteins that appeared to have either an increase or decrease in abundance in both the *tor1* and rapamycin yeast relative to wild type. Although the majority of proteins did not have this pattern, there were ~25 that did and could potentially be responsible for the increased longevity observed in both *tor1* and rapamycin treated yeast. 16 of these proteins were detected in the DIA data, and of these 16, only one (ATP2) had evidence of a shared change (decrease) in abundance relative to the wild type yeast. Such a large discrepancy between the DIA and spectral counting data was unexpected, but could be due to the low signal for the peptides being compared. Proteins such as ARD1, RIM8, and SOL3 had very low numbers of spectral counts (<10) which can be inaccurate. However, there are many cases with >50 spectral counts in which the trend observed in the DIA data does not match that from the spectral counting data. For example, spectral counting indicates a small increase in the abundance of the protein RPL31A in the *tor1* and rapamycin-treated yeast. The DIA data show a large decrease in RPL31A abundance relative to the wild-type sample. These differences could be due to bias in the sampling of MS/MS spectra by data dependent acquisition. DDA targets high abundance peptides preferentially. Because many peptides co-elute, the abundance of co-eluting peptides could bias the number of spectral counts acquired for a peptide of interest. For example, if the peptide being quantified does not change in abundance between two samples, but

a co-eluting peptide does, the sample with the more abundant co-eluting peptide could have a lower number of spectral counts for the quantified peptide due to the instrument preferentially acquiring MS/MS scans on the high abundance co-eluting peptide.

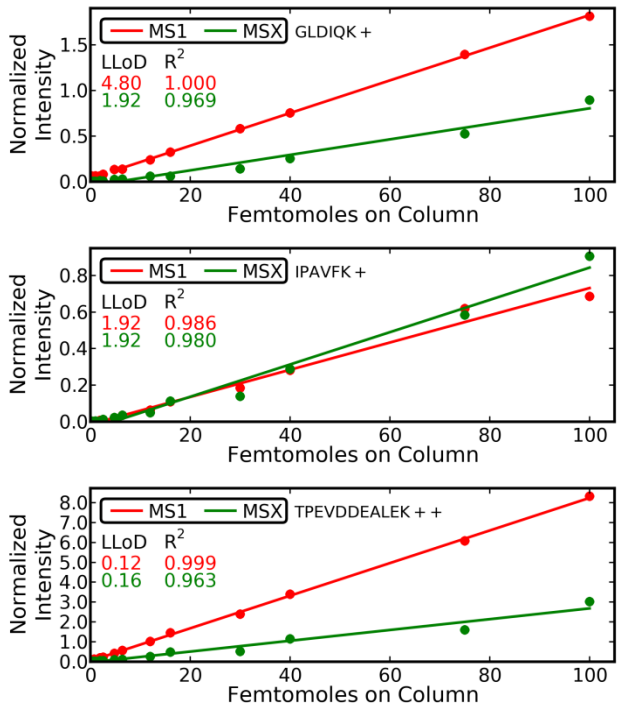
Data dependent acquisition combined with spectral counting is a useful technique due to its ability to quickly compare hundreds of proteins across many samples. However, discrepancies between MS/MS quantitation from DIA data and spectral counting indicate that spectral counting may not always be accurate. An alternative is to map MS/MS spectra for identified peptides to the precursor signal that triggered the MS/MS event, and integrate the area under the curve of this MS feature for quantitation. However, the MS measurements can often contain interference due to chemical noise. In a complex sample, the increased selectivity of MS/MS quantitation by DIA results in reduced chemical noise and more sensitive and accurate quantitation[19]. In this study, DDA and DIA data were both acquired on each sample to combine the ability of DDA to identify thousands of peptides with the robustness of quantitation by integrating an MS/MS signal over time in DIA data. However, future advances in the performance of mass spectrometers and algorithms for identifying peptides directly from DIA data may eventually make it unnecessary to acquire DDA data at all.

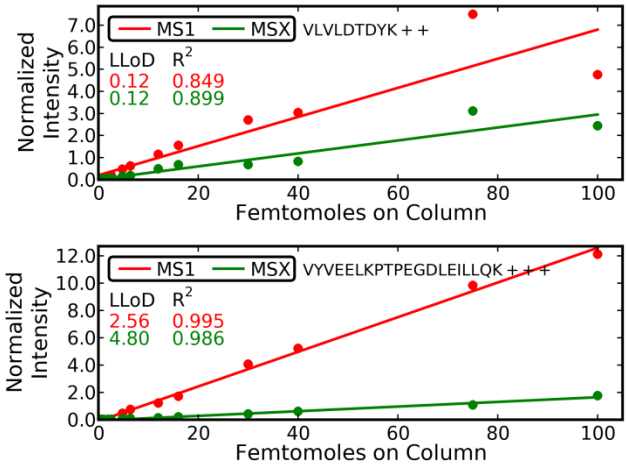
Appendix A: MSX Spike-In Peptides

A1 : Spike in data for 36 peptides (5 proteins)

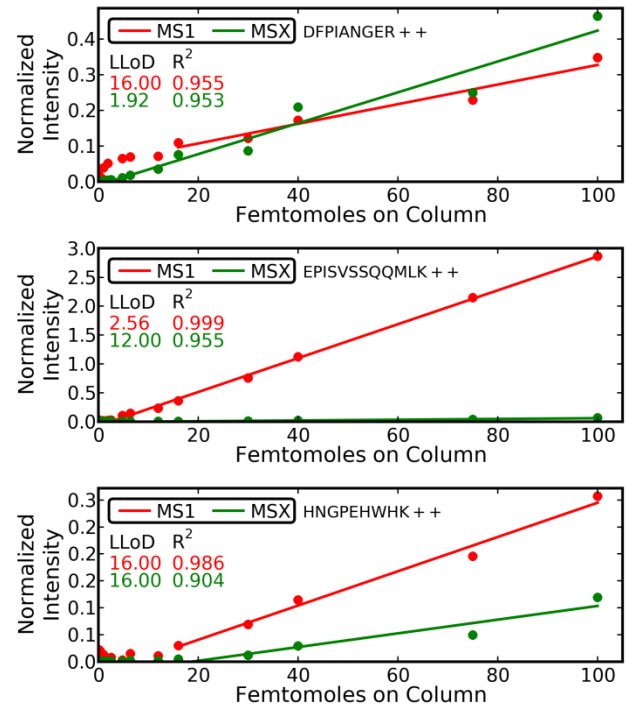
A commercial six protein digest was spiked into a complex matrix (soluble *S. cerevisiae* lysate) at amounts ranging from 50 attomoles – 100 femtomoles on column. MSX data were acquired with an MS1 scan interleaved every 10 scans. The normalized signal intensity for each spike in point is plotted for each peptide as well as the regression line fit to all points greater than or equal to the lower limit of detection.

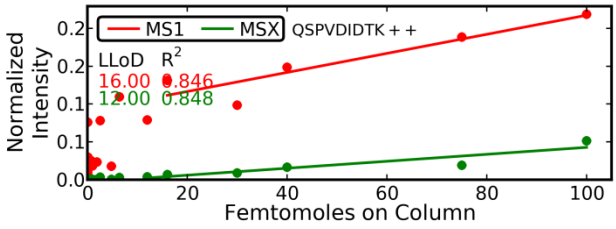
β-Lactoglobulin

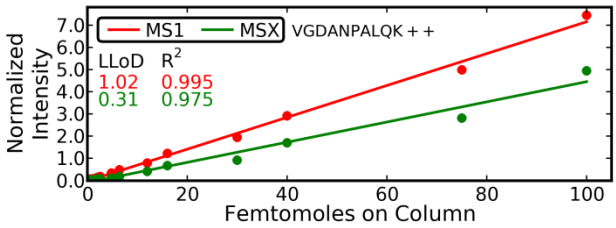




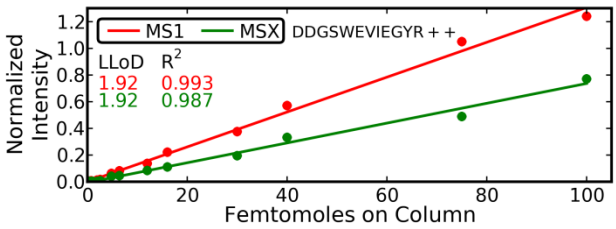
Carbonic Anhydrase

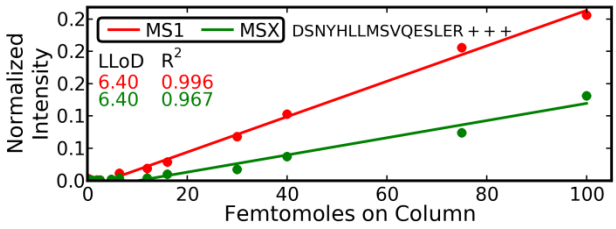


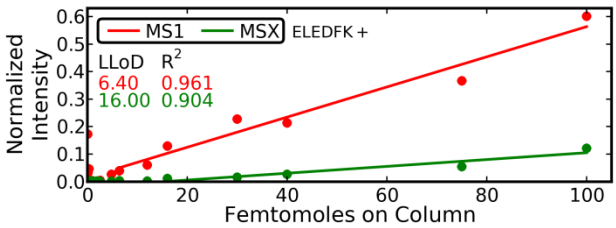


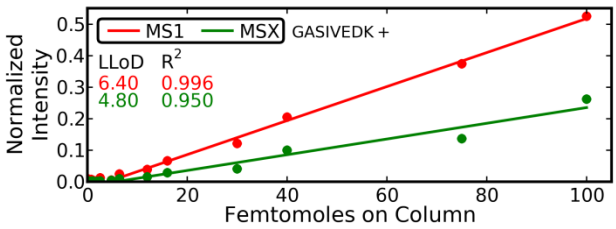


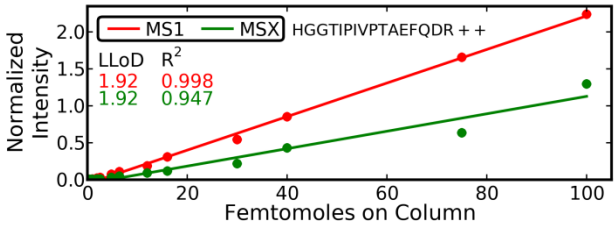
Glutamate Dehydrogenase

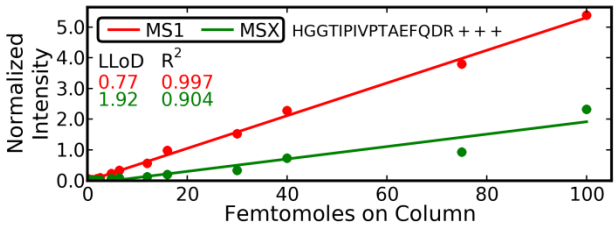


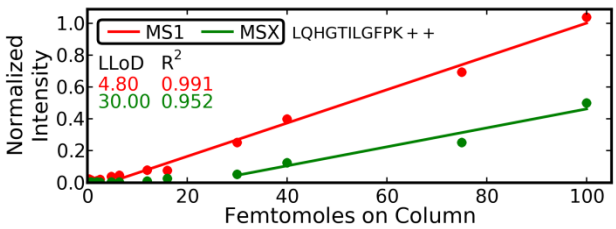


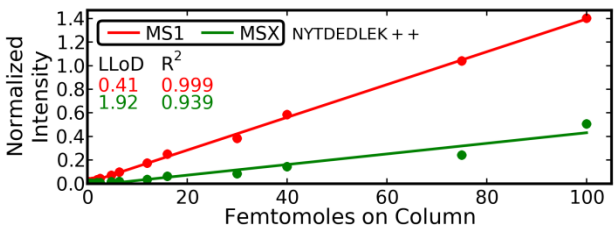


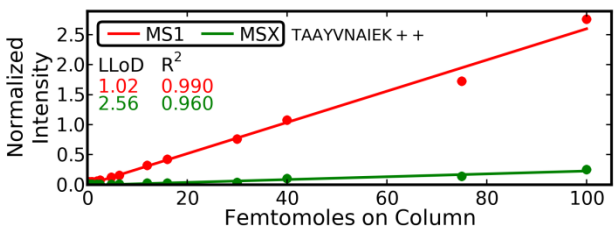


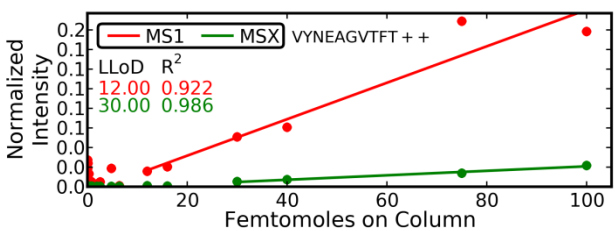

MS1
MSX
HGGTIPIVPTAEFQDR + +
LLoD R2
1.92 0.998
1.92 0.947
Normalized Intensity
Femtomoles on Column

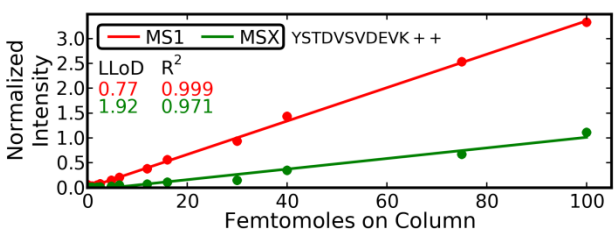

MS1
MSX
HGGTIPIVPTAEFQDR + + +
LLoD R2
0.77 0.997
1.92 0.904
Normalized Intensity
Femtomoles on Column


MS1
MSX
LQHGTILGFPK + +
LLoD R2
4.80 0.991
30.00 0.952
Normalized Intensity
Femtomoles on Column

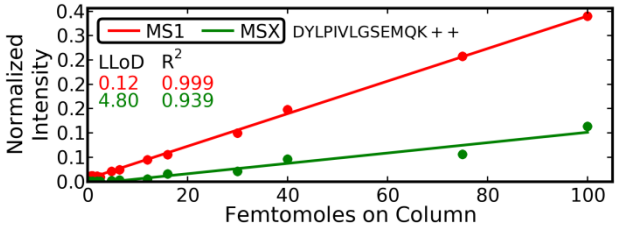

MS1
MSX
NYTDEDLEK + +
LLoD R2
0.41 0.999
1.92 0.939
Normalized Intensity
Femtomoles on Column

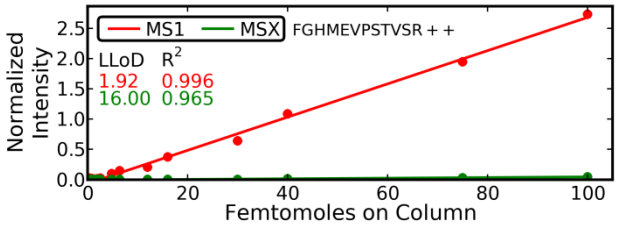

MS1
MSX
TAAYVNAIEK + +
LLoD R2
1.02 0.990
2.56 0.960
Normalized Intensity
Femtomoles on Column

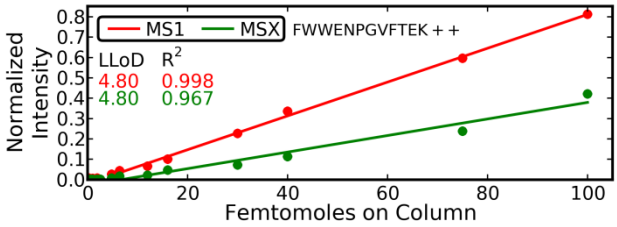

MS1
MSX
VYNEAGVTFT + +
LLoD R2
12.00 0.922
30.00 0.986
Normalized Intensity
Femtomoles on Column

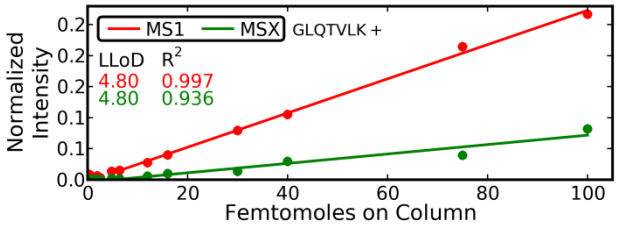

MS1
MSX
YSTDVSVDEVK + +
LLoD R2
0.77 0.999
1.92 0.971
Normalized Intensity
Femtomoles on Column

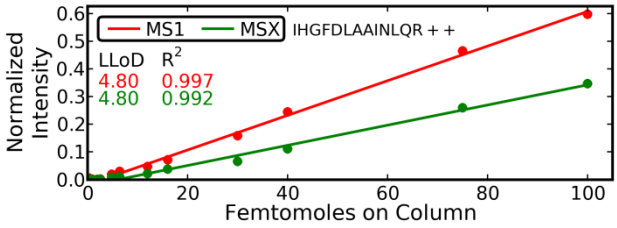
Lactoperoxidase

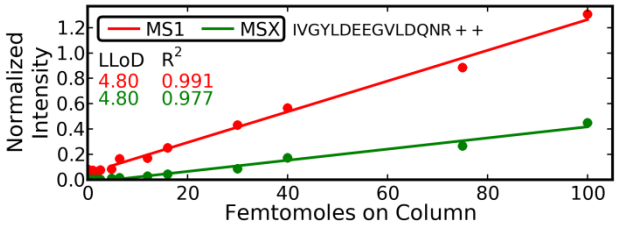




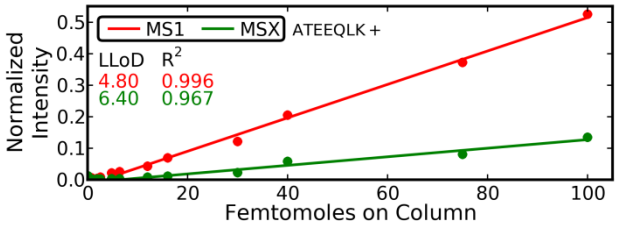


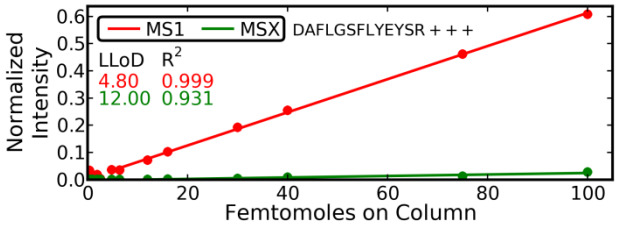


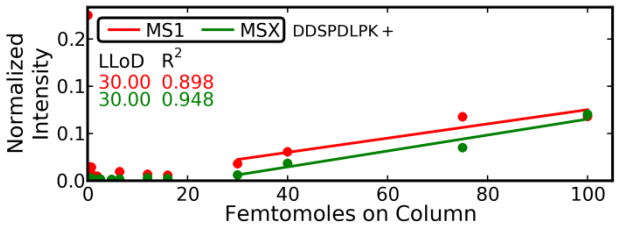


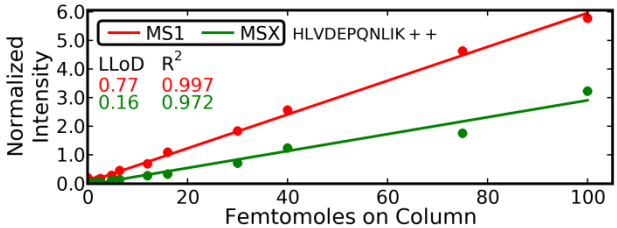


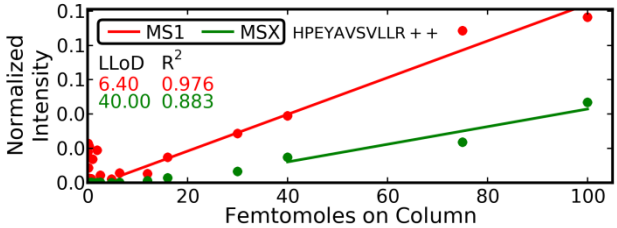
Serum Albumin

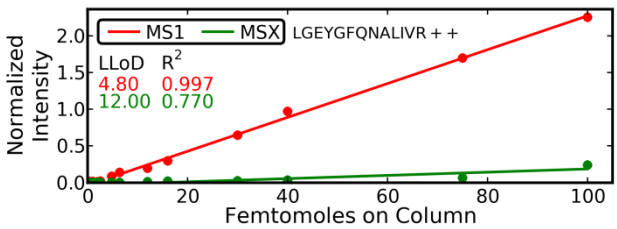


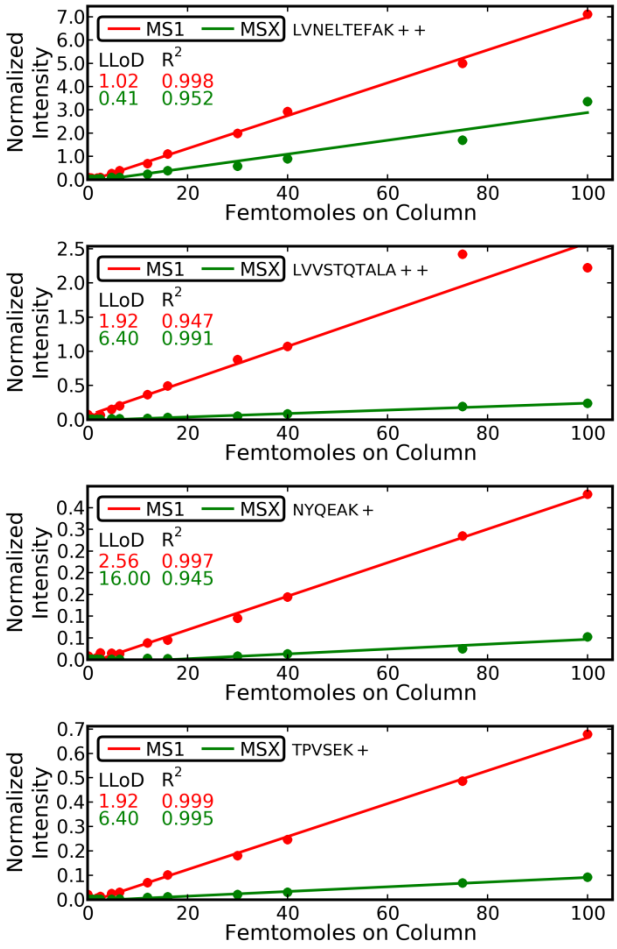












A2: Transitions Used For Bovine Spike-In Protein Quantification

β-Lactoglobulin

Peptide	Transitions
GLDIQK+	b+(4) y+(2-5)
IPAVFK+	b+(1,3-5) y+(2-5)
TPEVDDEALEK++	b+(6-7) y+(4,7-9) y++(10)
VLVLDTDYK++	b+(3,6) y+(4-8)
VYVEELKPTPEGDLEILLQK+++	b+(2) y+(3-7,9-10) y++(11-14)

Carbonic Anhydrase

Peptide	Transitions
DFPIANGER++	y+(3-7) y++(7)
EPISVSSQQMLK++	b+(5,8) y+(3) b++(4,10) y++(9)
HNGPEHWHK++	b+(1-2,7) y+(2-4,7-8) b++(6)
QSPVDIDTK++	y+(3,5-6) y++(7)
VGDANPALQK++	b+(5) y+(2-9) y++(5,9)

Glutamate Dehydrogenase

Peptide	Transitions
DDGSWEVIEGYR++	b+(3) y+(3-10) b++(9,11) y++(8)
DSNYHLLMSVQESLER+++	b+(4-8) y+(2,5-6,8)
ELEDFK+	b+(3-5) y+(2-5)
GASIVEDK+	b+(3,5-6) y+(2-5)
HGGTIPIVPTAEFQDR++	b+(1-5,8) y+(2-4,6,9-13) y++(8,11)
HGGTIPIVPTAEFQDR+++	b+(1-2,5) y+(2-3,6-9) b++(6) y++(6,8)
LQHGTILGFPK++	b+(2-6,8-9) y+(3-9) y++(9)
NYTDEDLEK++	b+(2-3) y+(3,6-8)
TAAYVNAIEK++	y+(2-6,8-9)
VYNEAGVTFT++	b+(2-3) y+(2) b++(9)
YSTDVSVDEVK++	b+(1,4) y+(2-5,7,9-10) b++(4)

Lactoperoxidase

Peptide	Transitions
DYLPIVLGSEMQK++	y+(3-10) y++(10)
FGHMEVPSTVSR++	y+(6-7) b++(8) y++(5,7,11)
FWWENPGVFTEK++	b+(2-4) y+(3-10) y++(3,7)
GLQTVLK+	b+(4) y+(3-6)
IHGFDLAAINLQR++	b+(3,6,8,10) y+(4-12) y++(12)
IVGYLDEEGVLDQNR++	b+(3) y+(3,8) b++(6,10,12-13)

Serum Albumin

Peptide	Transitions
ATEEQLK+	b+(4-6) y+(2,5-6)
DAFLGSFLYEYSR+++	b+(6-7,10) b++(6,10)
DDSPDLPK+	b+(1,5-6) y+(2-3,5-6)
HLVDEPQNLIK++	b+(1-7,10) y+(3-10) b++(6-7)
HPEYAVSVLLR++	y+(6,8-10)
LGEYGFQNALIVR++	y+(2-6,9-12)
LVNELTEFAK++	b+(4-5) y+(3,5-9) y++(5,8)
LVVSTQTALA++	b+(8) y+(9) b++(4)
NYQEAK+	b+(2-5) y+(3-5)
TPVSEK+	b+(3-5) y+(2-5)

Normalization Peptides (Background matrix)

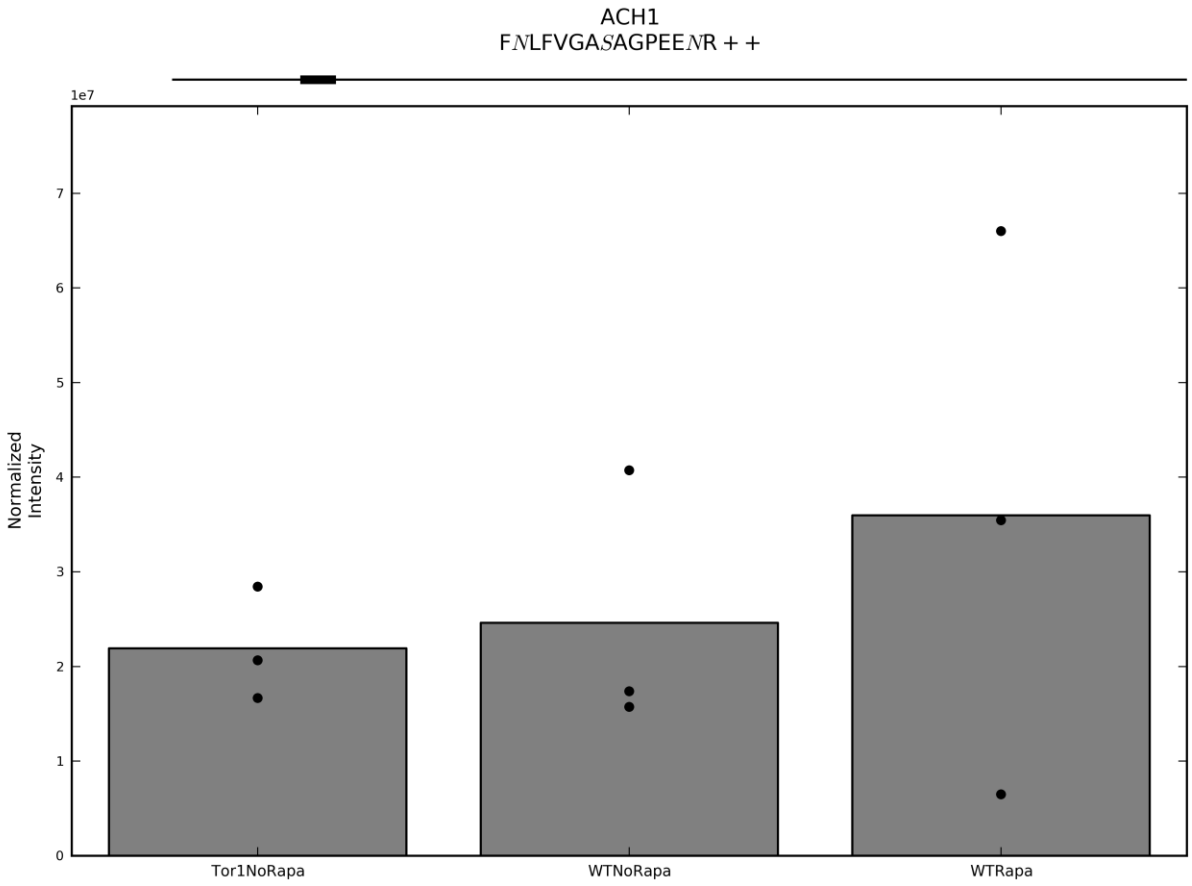
Peptide	Transitions
DNSQVFGVAR++	b+(2-3,5) y+(2-6)
ESTLHLVLR++	b+(2-3,5-7) y+(2-8) y++(5-8)

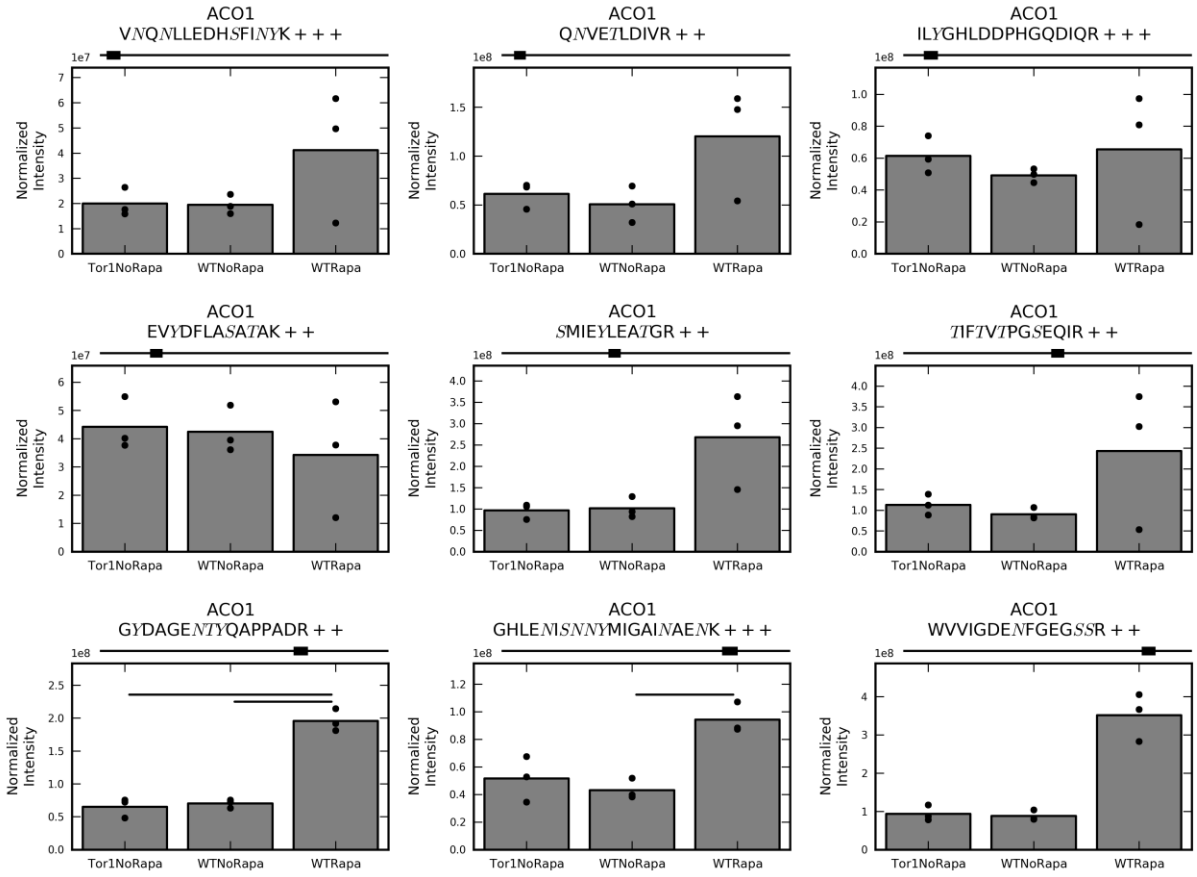
Appendix B: DIA Quantitation of Peptides in Yeast Rapamycin Experiment

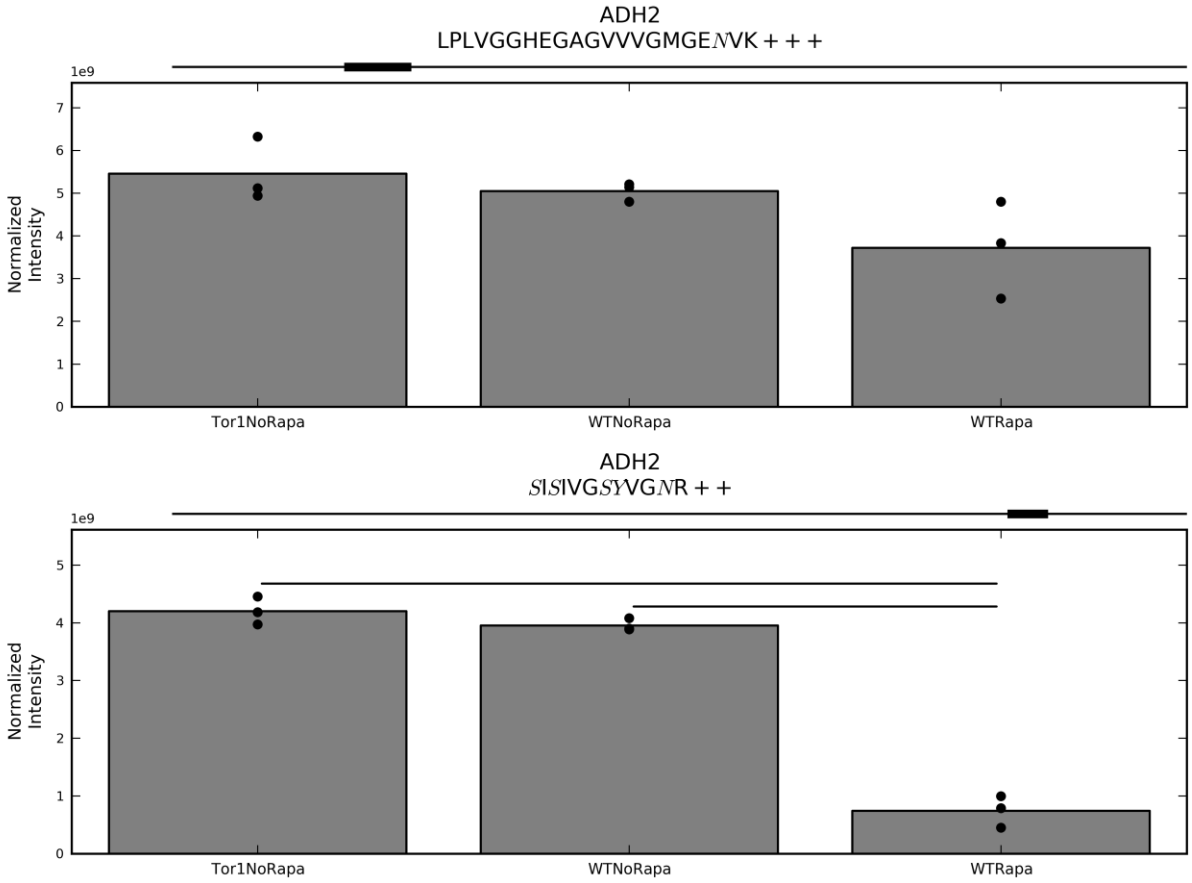
This section contains plots of peptide quantitation results from DIA data on proteins involved in mitochondrial metabolism (B1), targets of Gln3/stress response (B2), and proteins that appeared to have a shared increase or decrease relative to wild type in rapamycin treated and *tor1* yeast according to spectral counting data (B3). Each bar plot contains the name of the protein at the top, with the peptide being quantified below. Amino acids commonly modified are italicized. Below the peptide sequence, is a horizontal line representing the protein and the box overlaid on the line representing the location of the peptide on the protein. Note that this assumes that the protein being quantified is the same as the one in the SGD open reading frame database referenced, which may not always be the case. For example, the actual protein being quantified may be an isoform of that in the SGD database. There is a bar plot for every peptide quantified for a given protein. There is a bar for WT, WT + rapamycin, and *tor1* yeast representing the mean normalized intensity for that peptide across three biological replicates. The three dots overlaid on each bar are the normalized intensity measurements from each of the three biological replicates. A horizontal line is drawn above pairs of bars that show a significant change in abundance relative to each other (Welch's unpaired t-test; $p<0.01$).

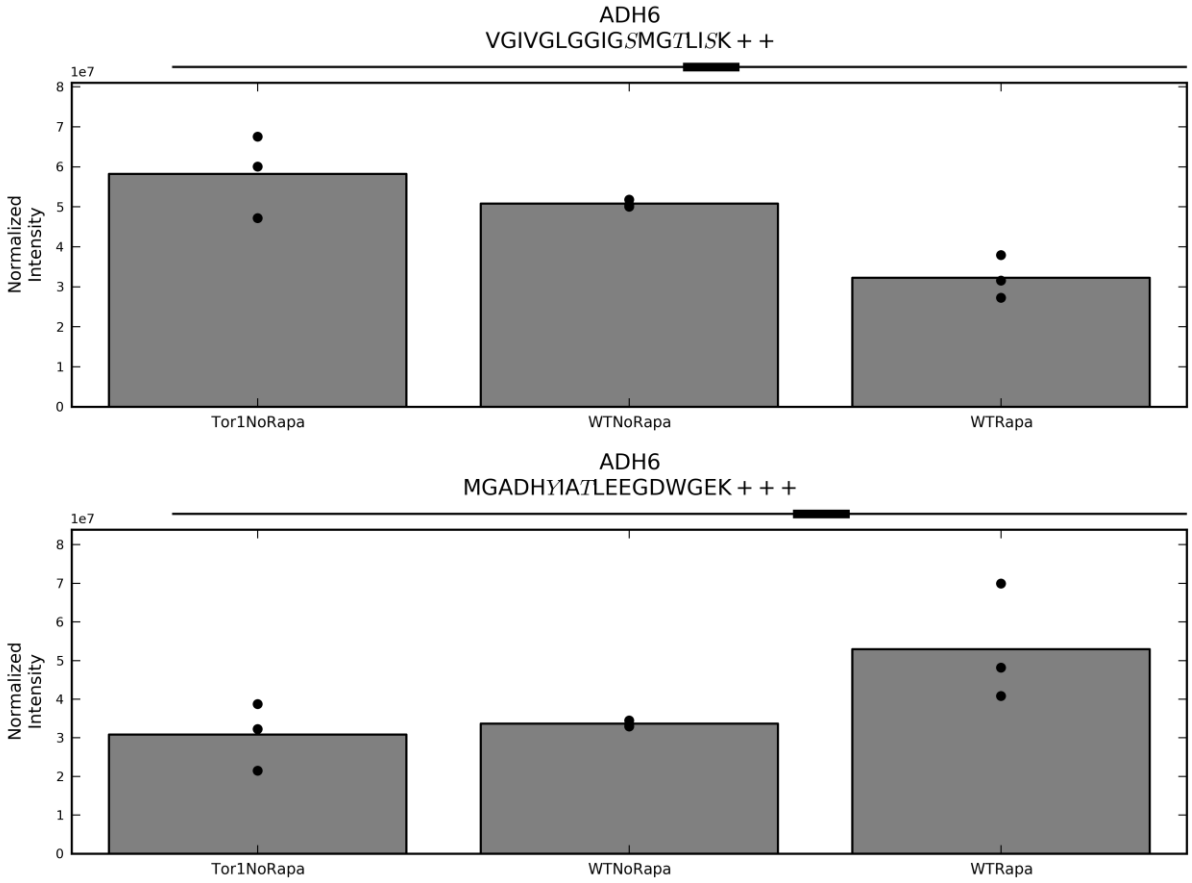
B1: Proteins Involved in Mitochondrial Metabolism

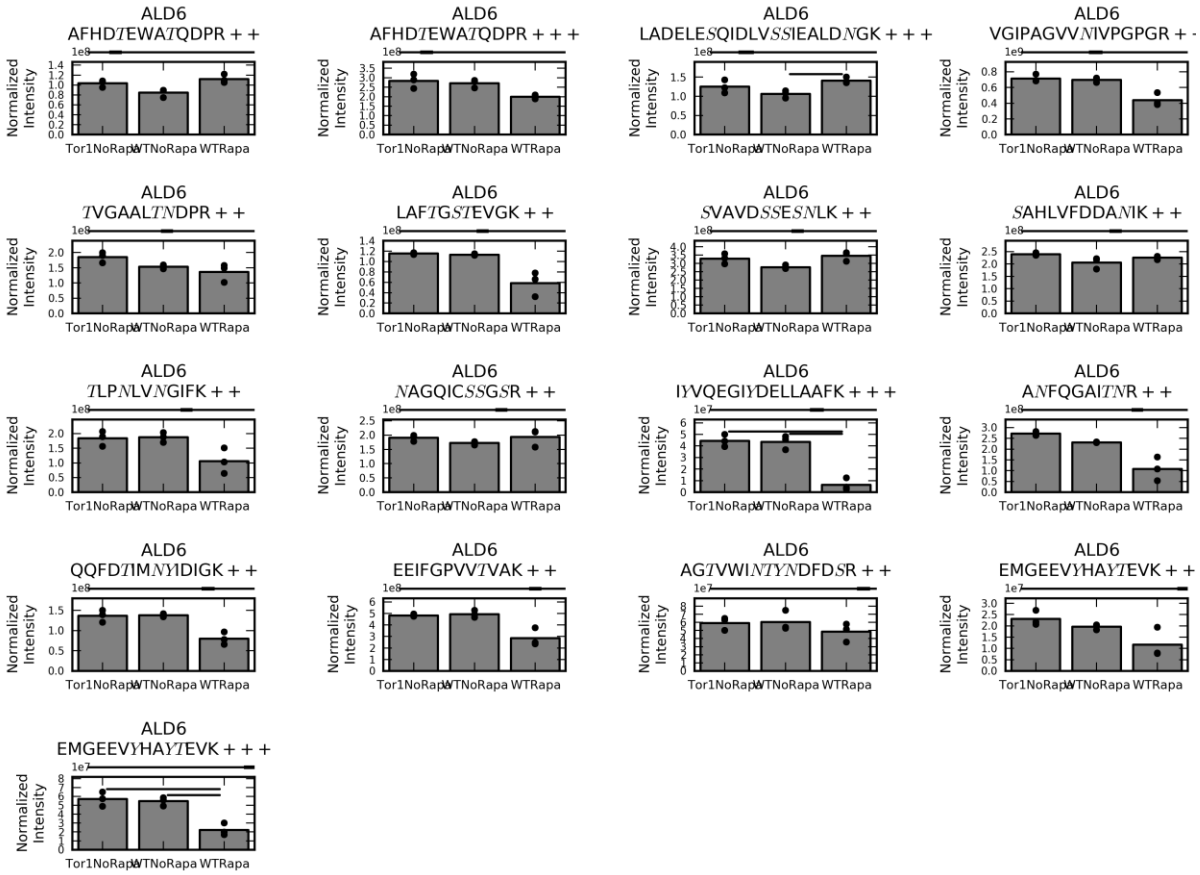
A description of these plots can be found at the beginning of Appendix B.

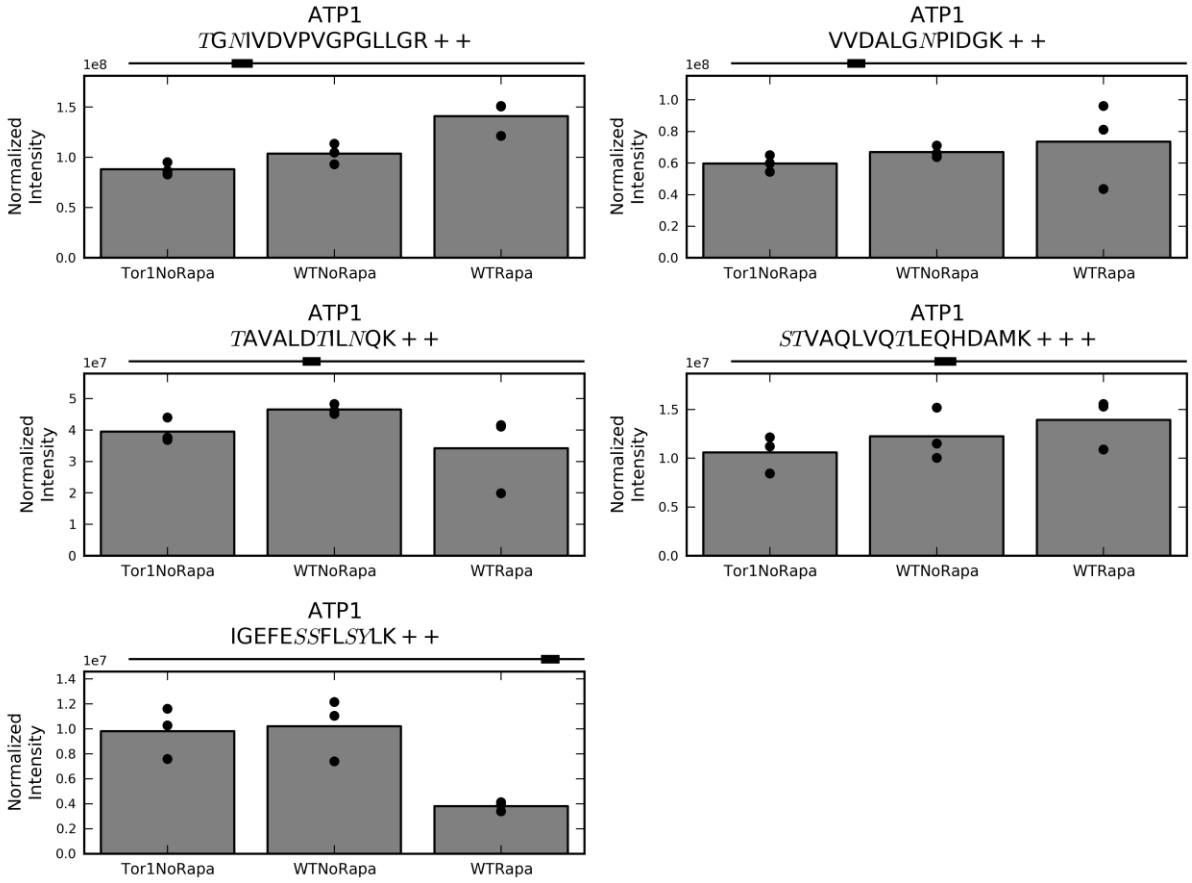


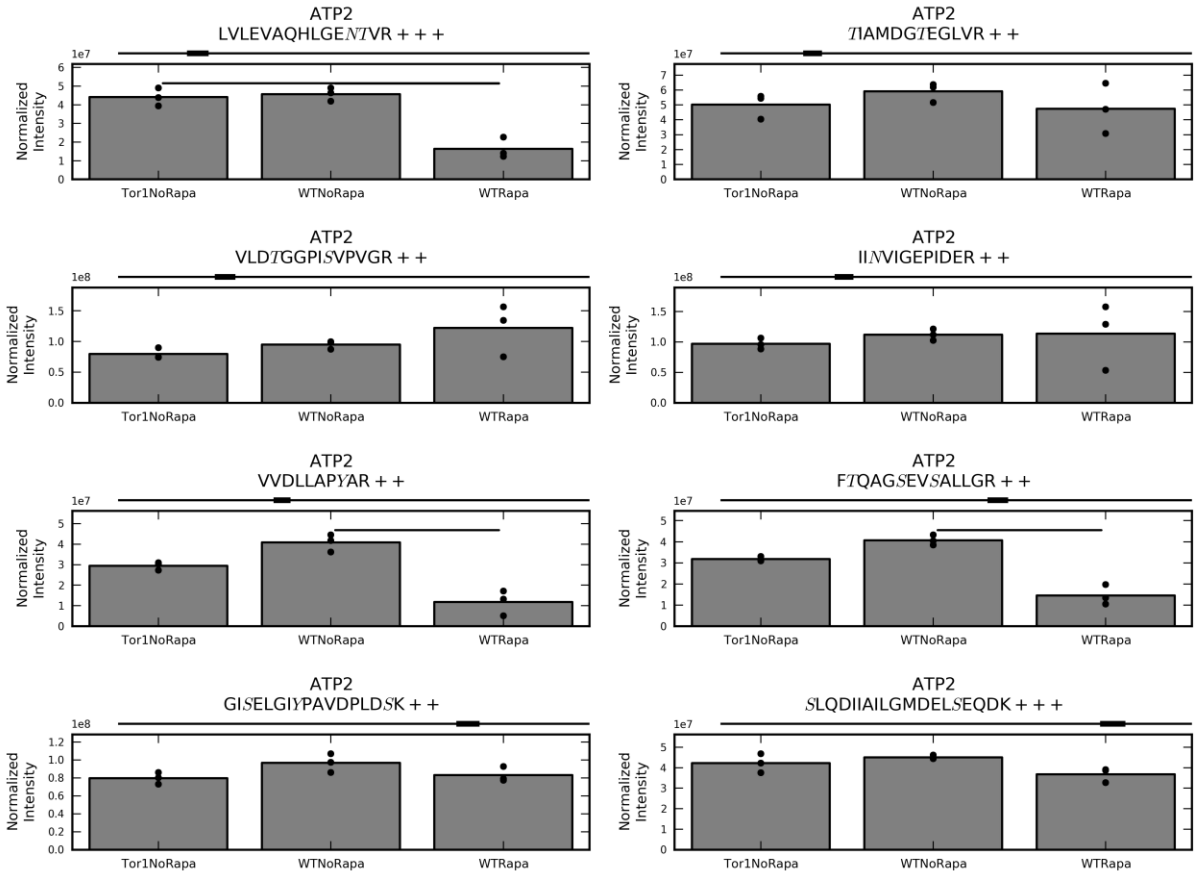

ACO1
VNQNLLEDHSFINYK + + +
ACO1
QNVETLDIVR + +
ACO1
ILYGHLDDPHGQDIQR + + +
ACO1
EVYDFLASATAK + +
ACO1
SMIEYLEATGR + +
ACO1
TIFTVTPGSEQIR + +
ACO1
GYDAGENTYQAPPADR + +
ACO1
GHLENISNNYMIGAINAENK + + +
ACO1
WVVIGDENFGEGSSR + +
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

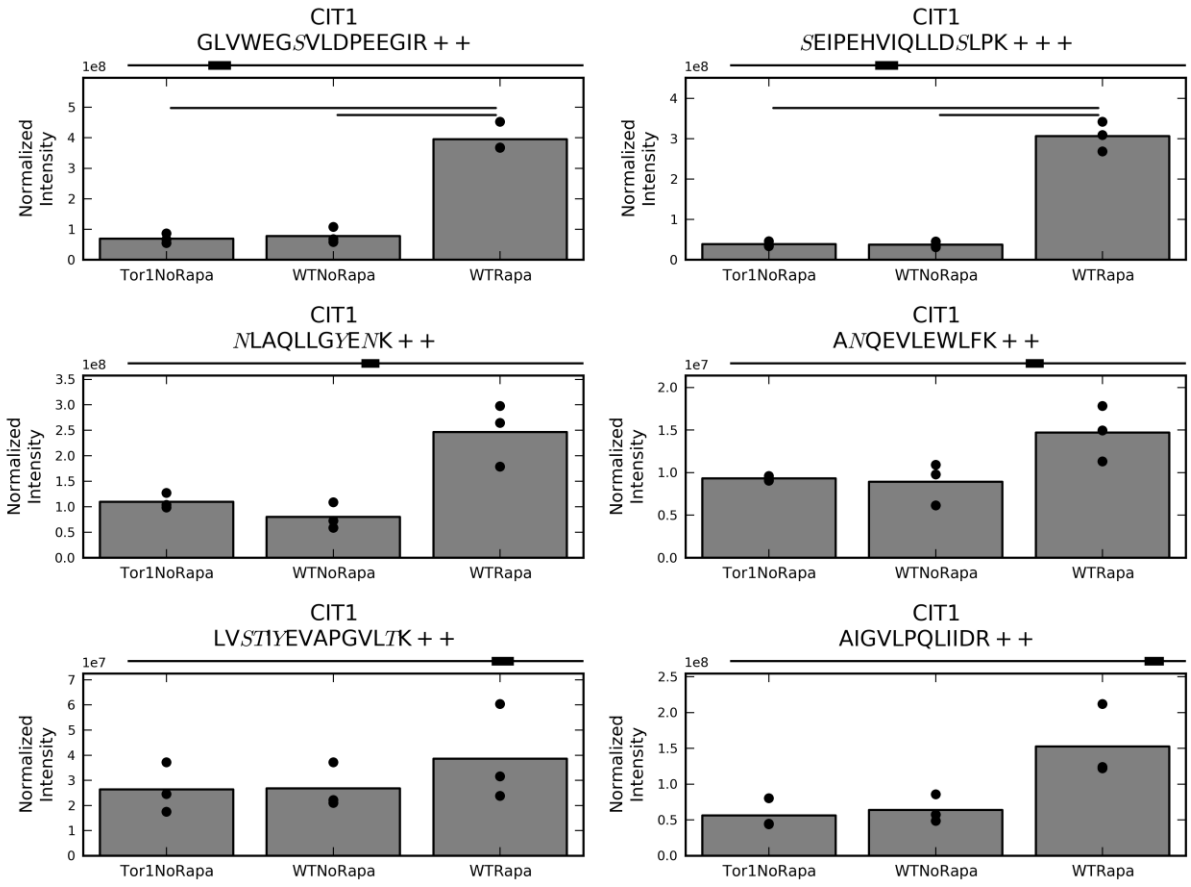

ADH2
LPLVGGHEGAGVVVGMGENVK + + +
1e9
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
ADH2
SISIVGSYVGNR + +
1e9
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

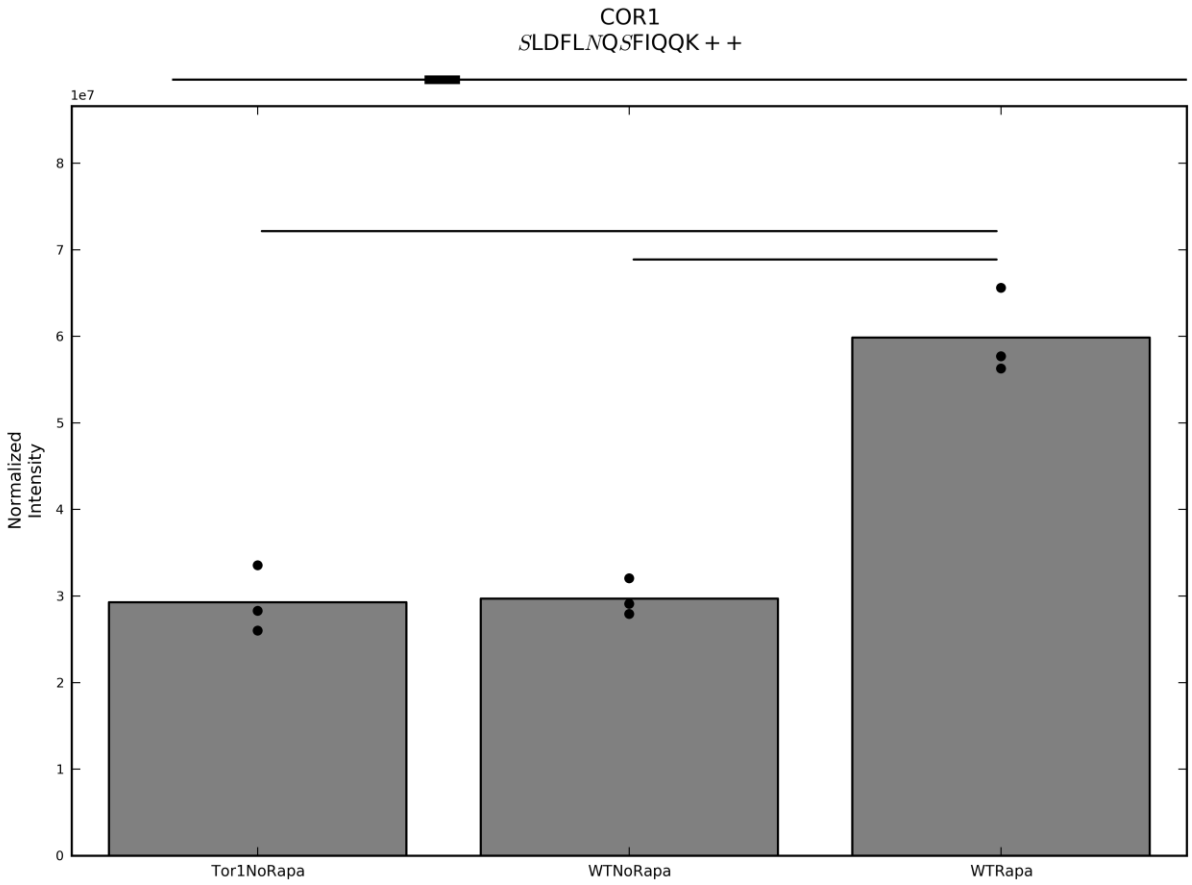

ADH6
VGIVGLGGIGSMGTLISK + +
1e7
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
ADH6
MGADHY1ATLEEGDWGEK + + +
1e7
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

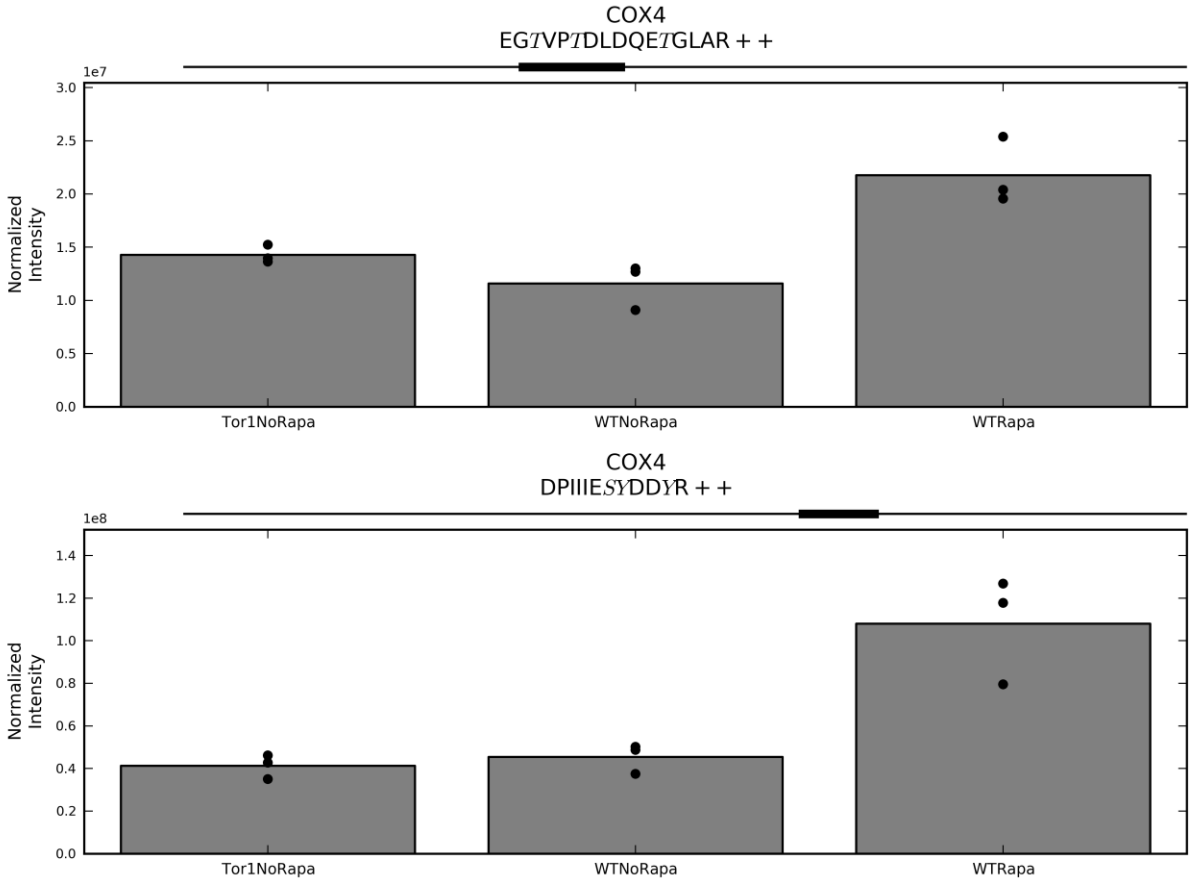

ALD6
AFHDTEWATQDPR + +
ALD6
AFHDTEWATQDPR + + +
ALD6
LADELESQIDLVSSIEALDNGK + + +
ALD6
VGIPAGVVNIVPGPGR + -
ALD6
TVGAALTNDPR + +
ALD6
LAFTGSTEVGK + +
ALD6
SVAVDSSESNLK + +
ALD6
SAHLVFDDANIK + +
ALD6
TLPNLVNGIFK + +
ALD6
NAGQICSSGSR + +
ALD6
IYVQEGIYDELLAAFK + + +
ALD6
ANFQGAITNR + +
ALD6
QQFDTIMNYIDIGK + +
ALD6
EEIFGPVVTVAK + +
ALD6
AGTVWINTYNDFDSR + +
ALD6
EMGEEVYHAYTEVK + +
ALD6
EMGEEVYHAYTEVK + + +
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

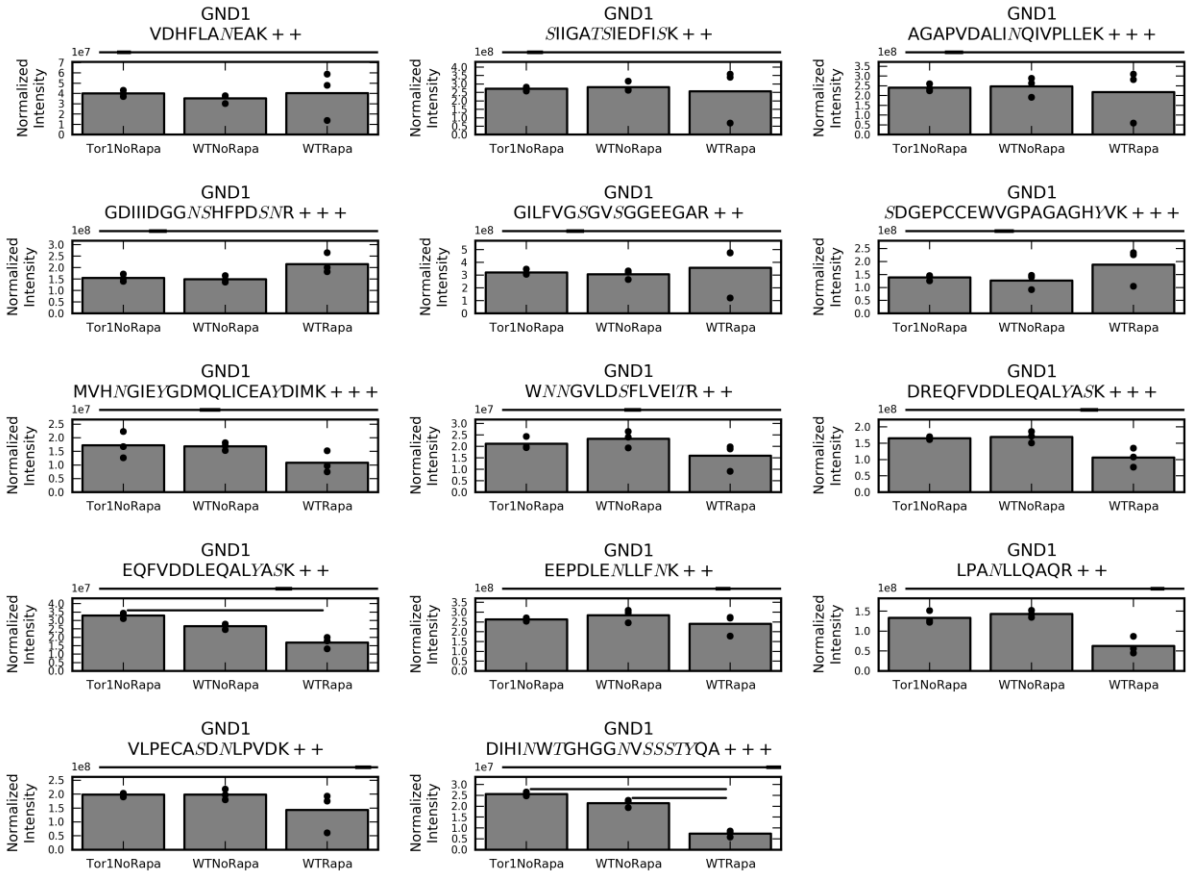

ATP1
TGNIVDVPVGPGLLGR + +
1e8
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
ATP1
VVDALGNPIDGK + +
1e8
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
ATP1
TAVALDTILNQK + +
1e7
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
ATP1
STVAQLVQTLEQHDAMK + + +
1e7
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
ATP1
IGEFESSFLSYLK + +
1e7
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

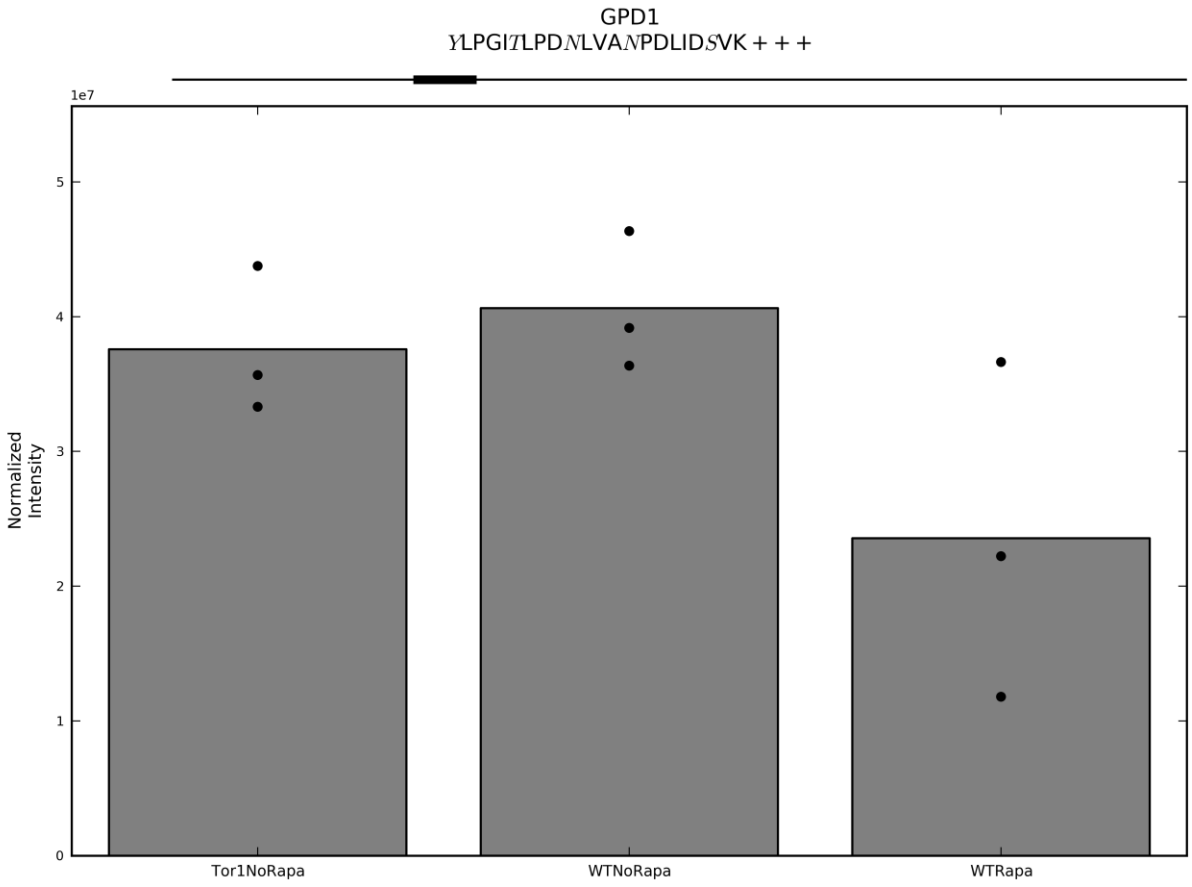

ATP2
LVLEVAQHLGE*N*TVR + + +
ATP2
*T*IAMDG*T*EGLVR + +
ATP2
VLD*T*GGPI*S*VPVGR + +
ATP2
II*N*VIGEPIDER + +
ATP2
VVDLLAP*Y*AR + +
ATP2
F*T*QAG*S*EV*S*ALLGR + +
ATP2
GI*S*ELGI*Y*PAVDPLD*S*K + +
ATP2
*S*LQDIIAILGMDEL*S*EQDK + + +
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

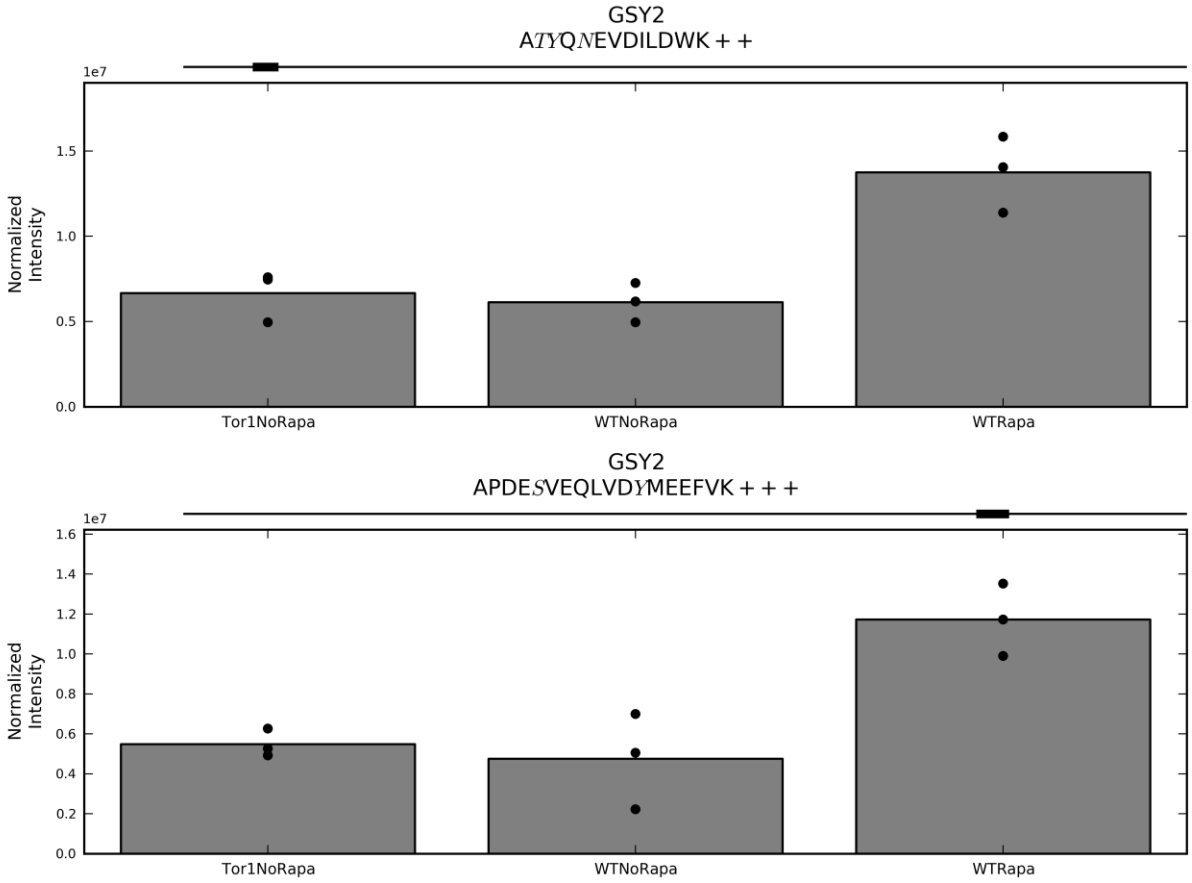

CIT1
GLVWEGSVLDPEEGIR + +
1e8
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
CIT1
SEIPEHVIQLLDSLPK + + +
1e8
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
CIT1
NLAQLLGYENK + +
1e8
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
CIT1
ANQEVLEWLFK + +
1e7
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
CIT1
LVSTIYEVAPGVLTK + +
1e7
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
CIT1
AIGVLPQLIIDR + +
1e8
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

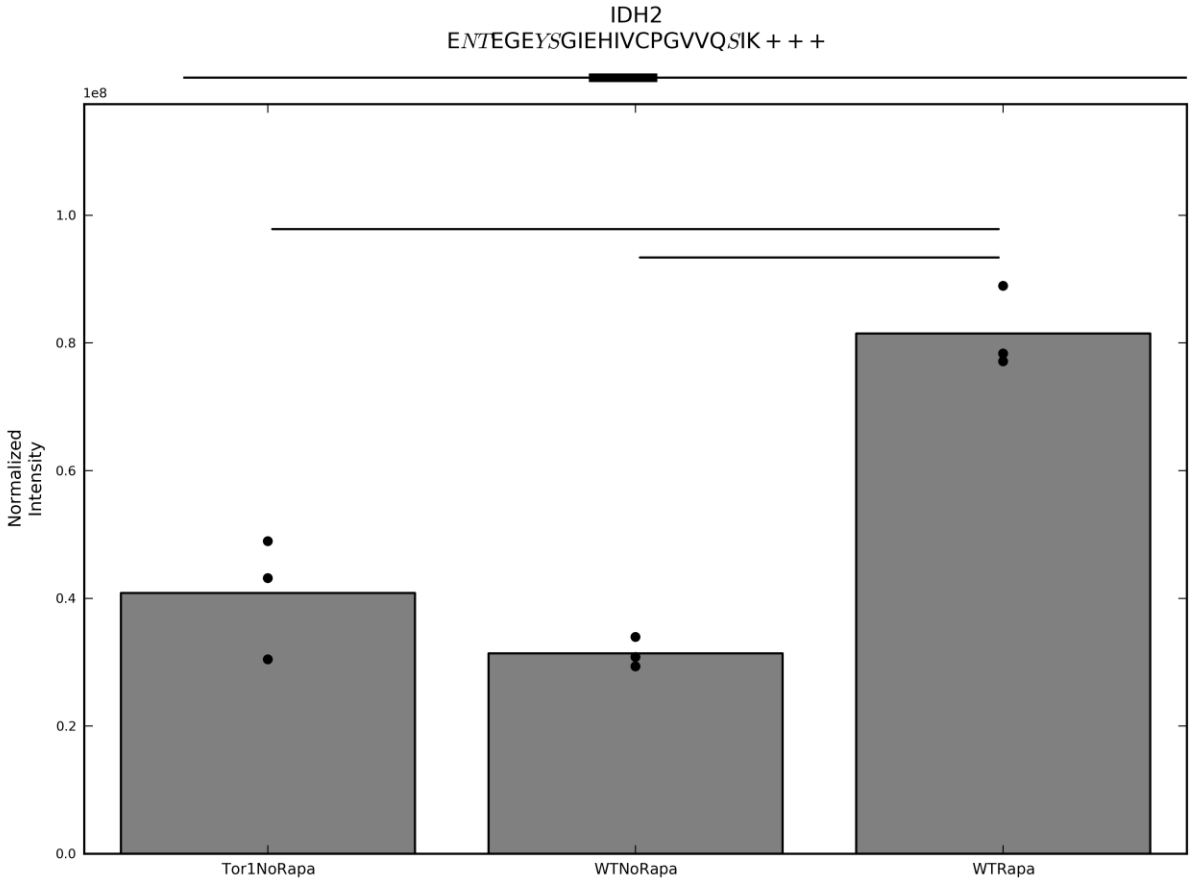

COR1
SLDFLNQSFIQQK++
1e7
8
7
6
5
4
3
2
1
0
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

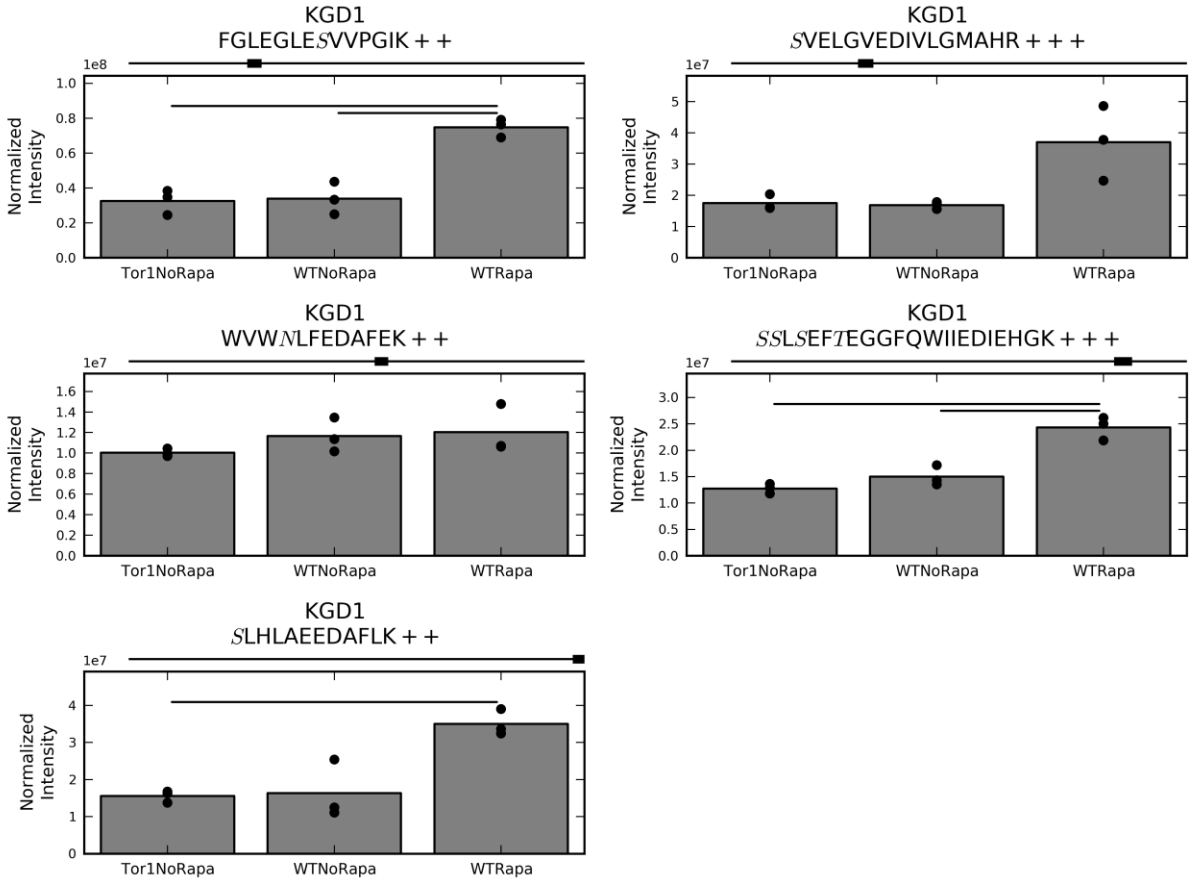

COX4
EGTVPTDLDQETGLAR + +
1e7
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
COX4
DPIIIESYDDYR + +
1e8
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

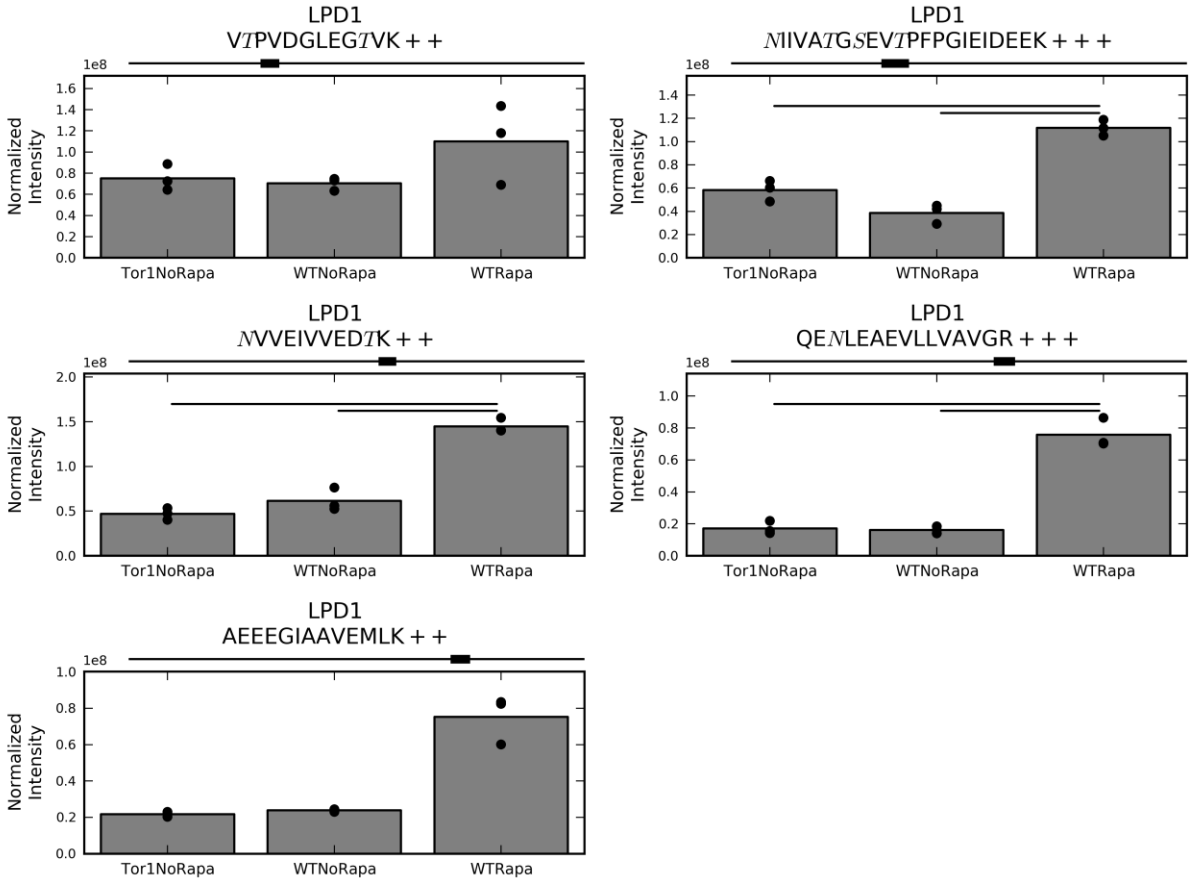

GND1
VDHFLA*N*EAK + +
GND1
*S*IIGA*TS*IEDFI*S*K + +
GND1
AGAPVDALI*N*QIVPLLEK + + +
GND1
GDIIIDGG*NS*HFPD*SN*R + + +
GND1
GILFVG*S*GV*S*GGEEGAR + +
GND1
*S*DGEPCCEWVGPAGAGH*Y*VK + + +
GND1
MVH*N*GIE*Y*GDMQLICEA*Y*DIMK + + +
GND1
W*NN*GVLD*S*FLVEI*T*R + +
GND1
DREQFVDDLEQAL*Y*A*S*K + + +
GND1
EQFVDDLEQAL*Y*A*S*K + +
GND1
EEPDLE*N*LLF*N*K + +
GND1
LPA*N*LLQAQR + +
GND1
VLPECA*S*D*N*LPVDK + +
GND1
DIHI*N*W*T*GHGG*N*V*SSST*YQA + + +
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

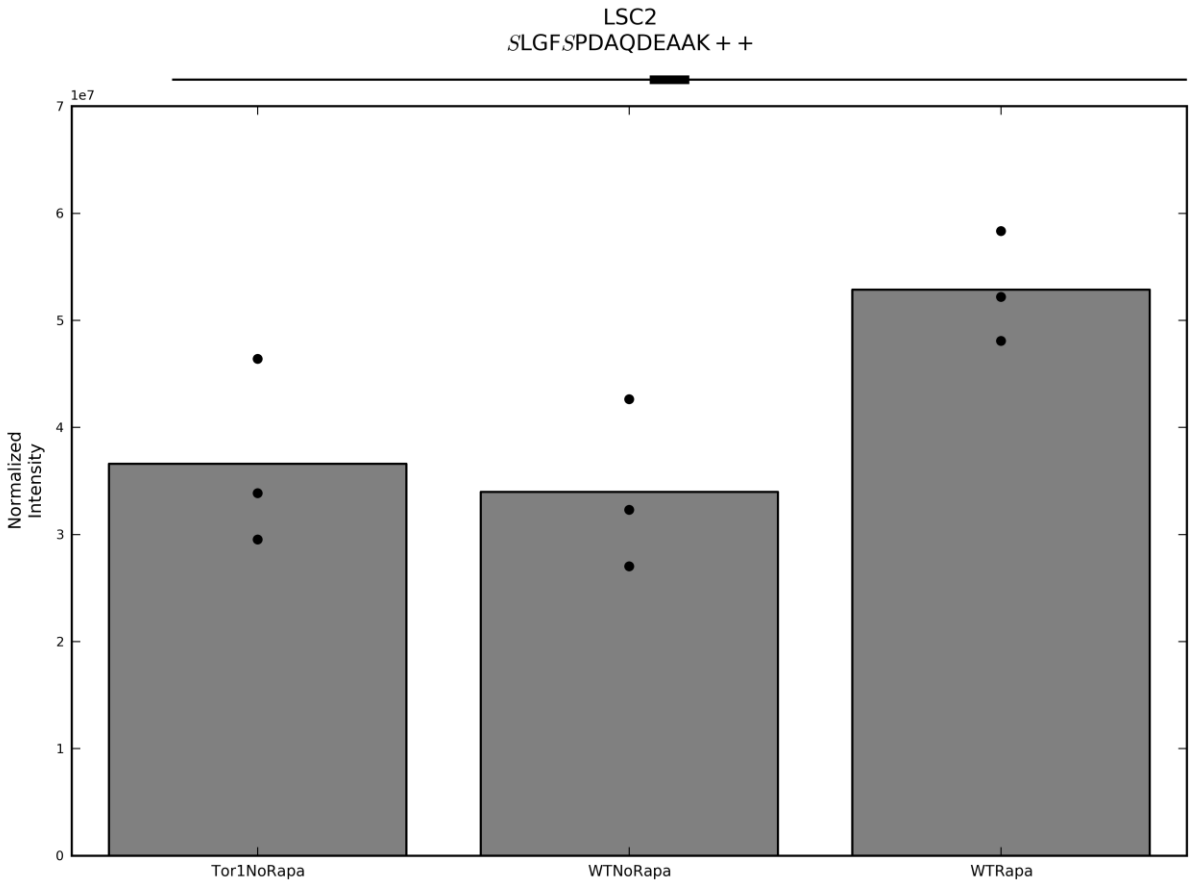

GPD1
YLPGITLPDNLVANPDLIDSVK + + +
1e7
5
4
3
2
1
0
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

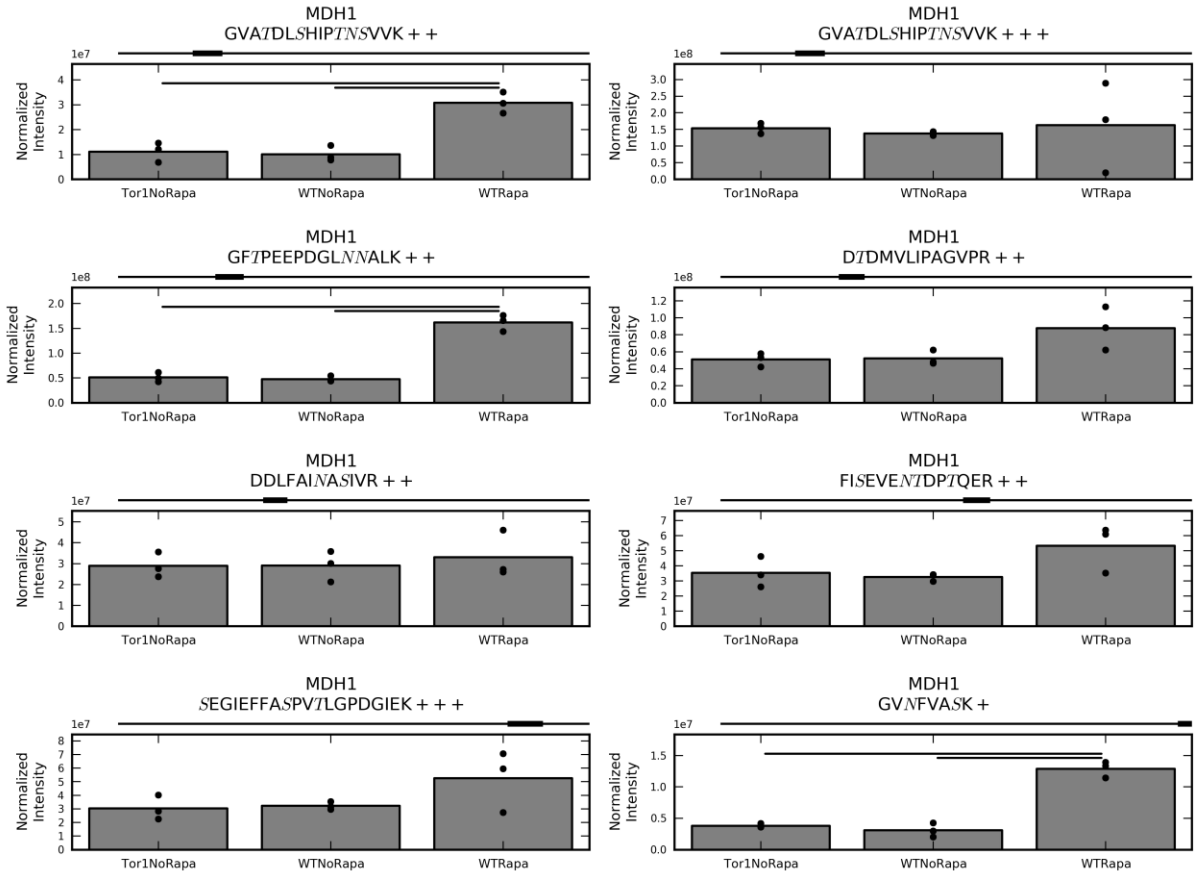

GSY2
ATYQNEVDILDWK++
1e7
1.5
1.0
0.5
0.0
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
GSY2
APDESVEQLVDYMEEFVK+++
1e7
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

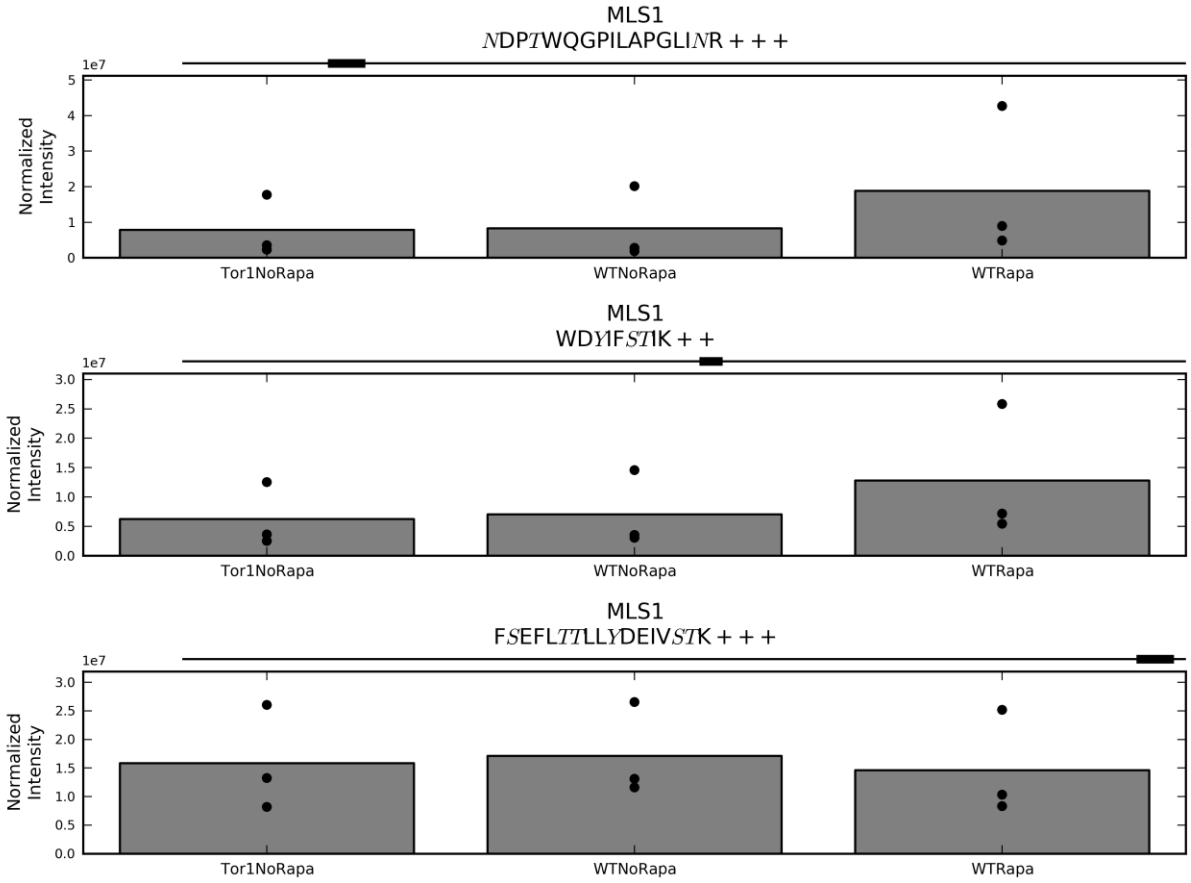

IDH2
E*NT*EGE*YS*GIEHIVCPGVVQ*S*IK + + +
1e8
1.0
0.8
0.6
0.4
0.2
0.0
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

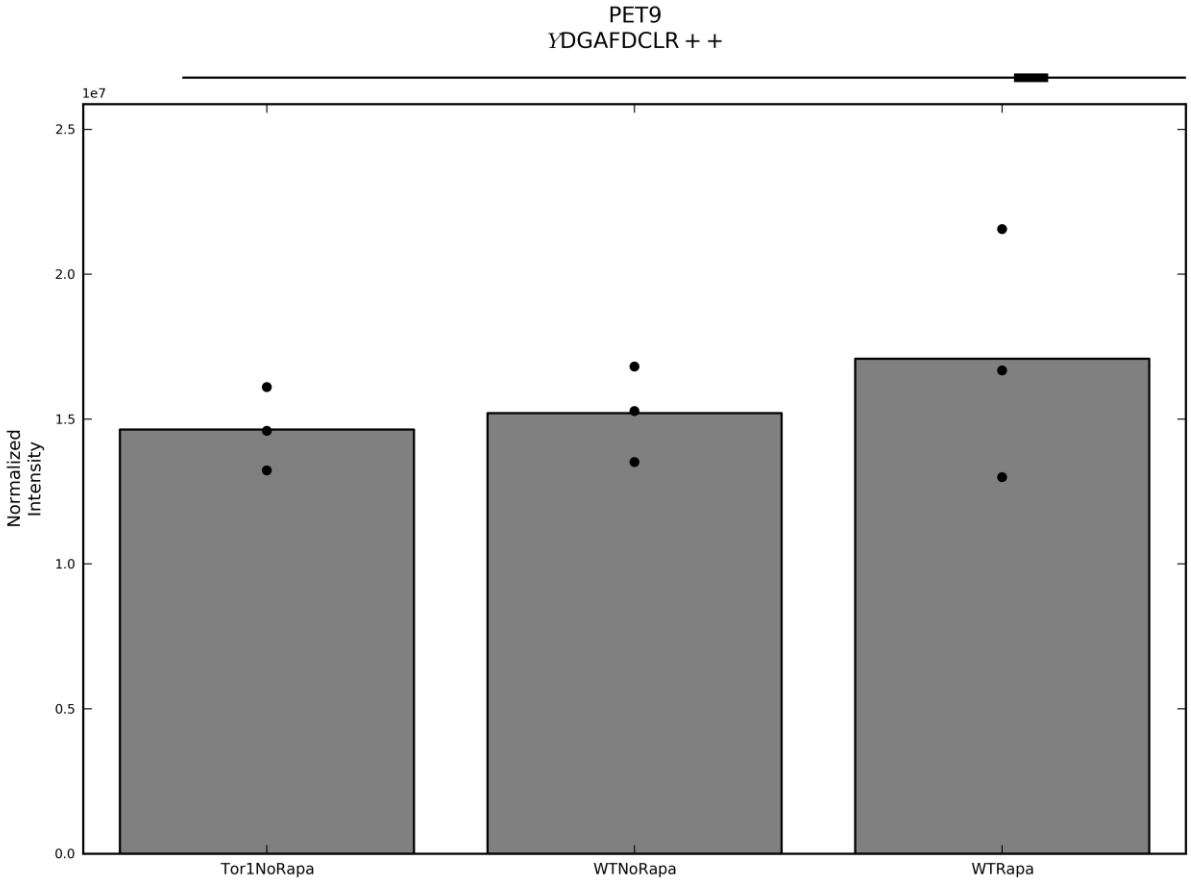

KGD1
FGLEGLESVVPGIK + +
1e8
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
KGD1
SVELGVEDIVLGMAHR + + +
1e7
KGD1
WVWNLFEDAFEK + +
KGD1
SSLSEFTEGGFQWIIEDIEHGK + + +
KGD1
SLHLAEEDAFLK + +

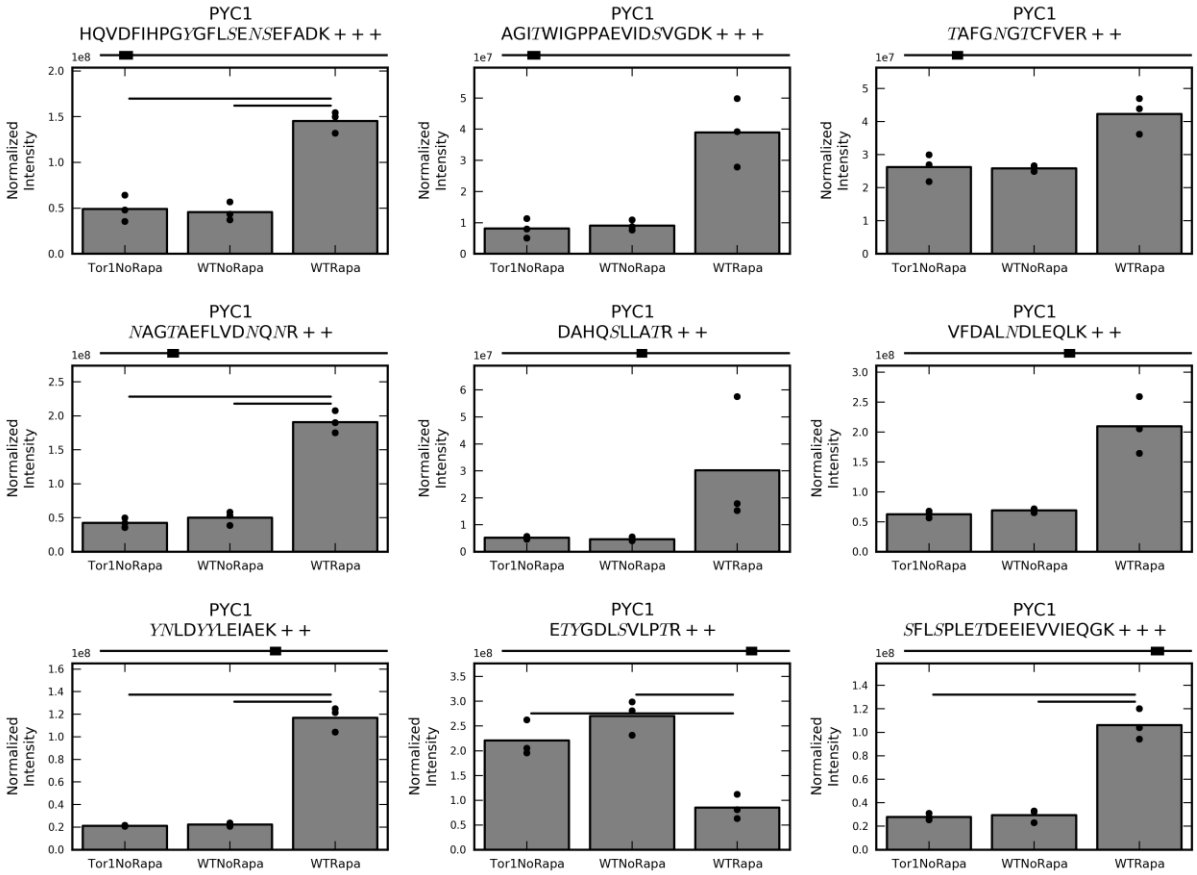

LPD1
VTPVDGLEGTVK + +
1e8
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
LPD1
NIIVATGSEVTPFPGIEIDEEK + + +
LPD1
NVVEIVVEDTK + +
LPD1
QENLEAEVLLVAVGR + + +
LPD1
AEEEGIAAVEMLK + +

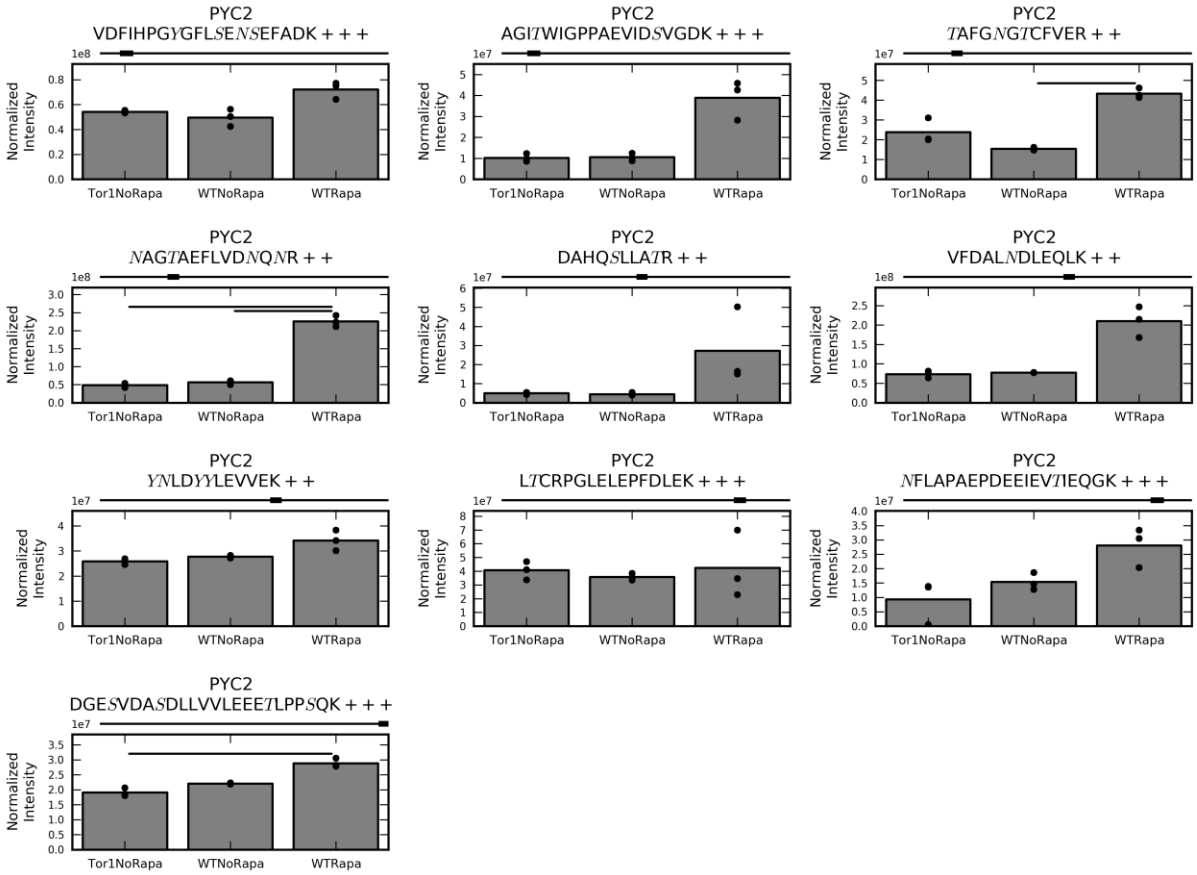

LSC2
SLGFSPDAQDEAAK++
1e7
7
6
5
4
3
2
1
0
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

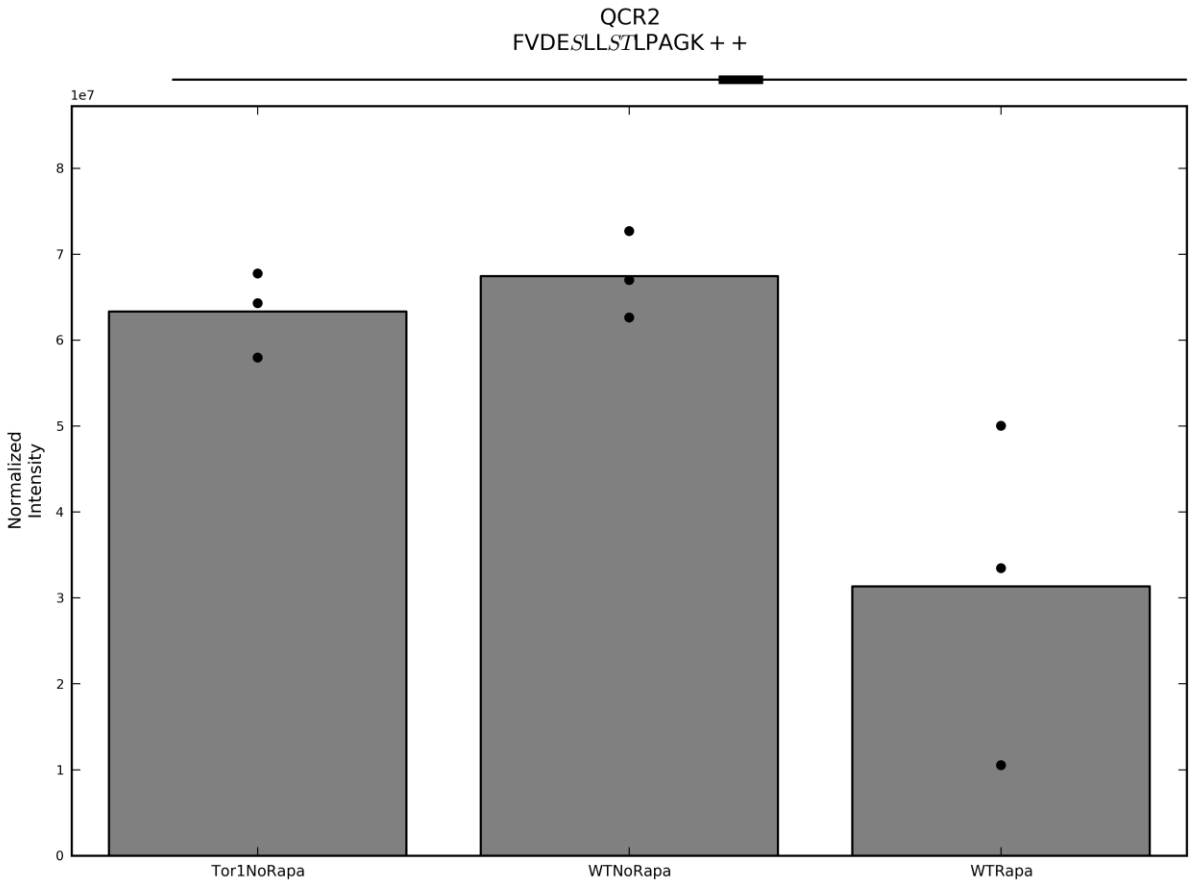

MDH1
GVATDLSHIPTNSVVK + +
MDH1
GVATDLSHIPTNSVVK + + +
MDH1
GFTPEEPDGLNNALK + +
MDH1
DTDMVLIPAGVPR + +
MDH1
DDLFAINASIVR + +
MDH1
FISEVENTDPTQER + +
MDH1
SEGIEFFASPVTLGPDGIEK + + +
MDH1
GVNFVASK +
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa


MLS1
NDPTWQGPILAPGLINR + + +
1e7
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
MLS1
WDYIFSTIK + +
1e7
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
MLS1
FSEFLTTLLYDEIVSTK + + +
1e7
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa


PET9
YDGAFDCLR + +
1e7
2.5
2.0
1.5
1.0
0.5
0.0
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

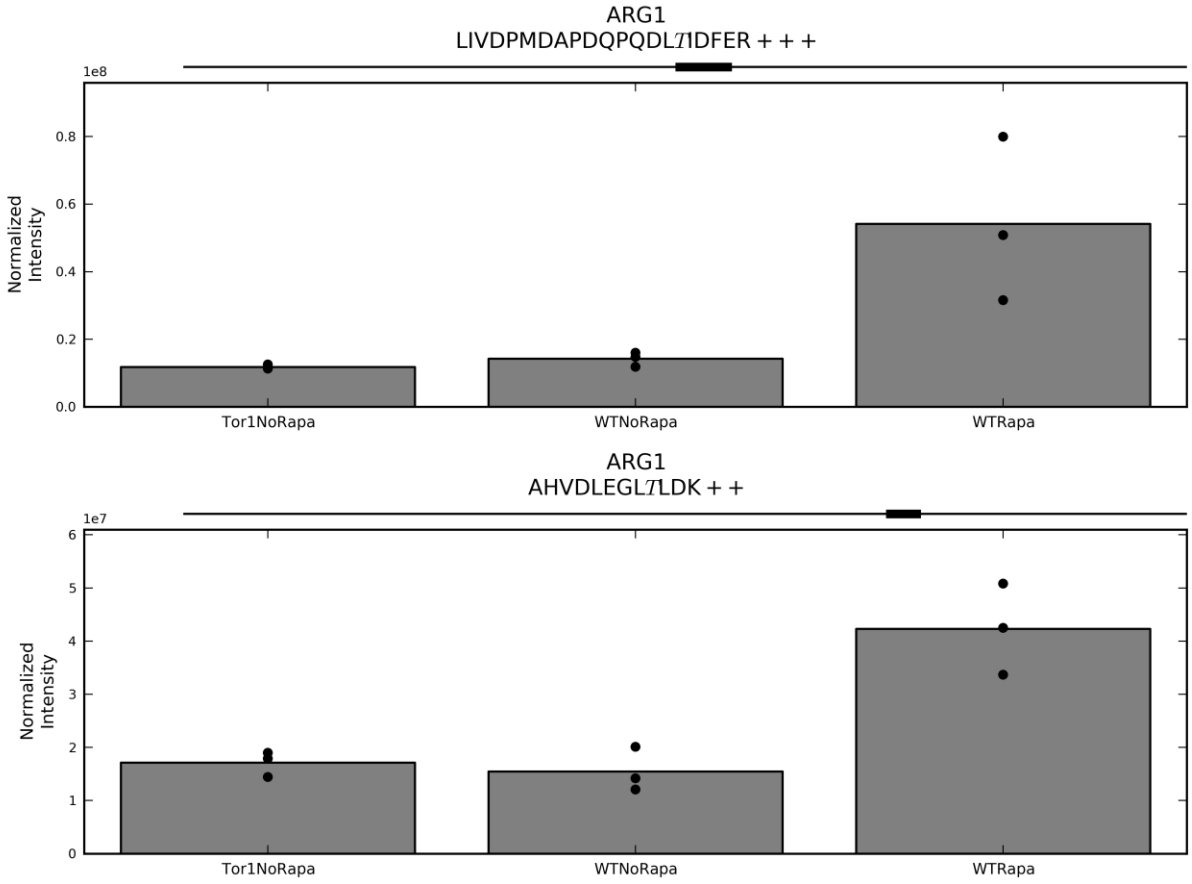

PYC1
HQVDFIHPGYGFLSENSEFADK + + +
PYC1
AGITWIGPPAEVIDSVGDK + + +
PYC1
TAFGNGTCFVER + +
PYC1
NAGTAEFLVDNQNR + +
PYC1
DAHQSLLATR + +
PYC1
VFDALNDLEQLK + +
PYC1
YNLDYYLEIAEK + +
PYC1
ETYGDLSVLPTR + +
PYC1
SFLSPLETDEEIEVVIEQGK + + +
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

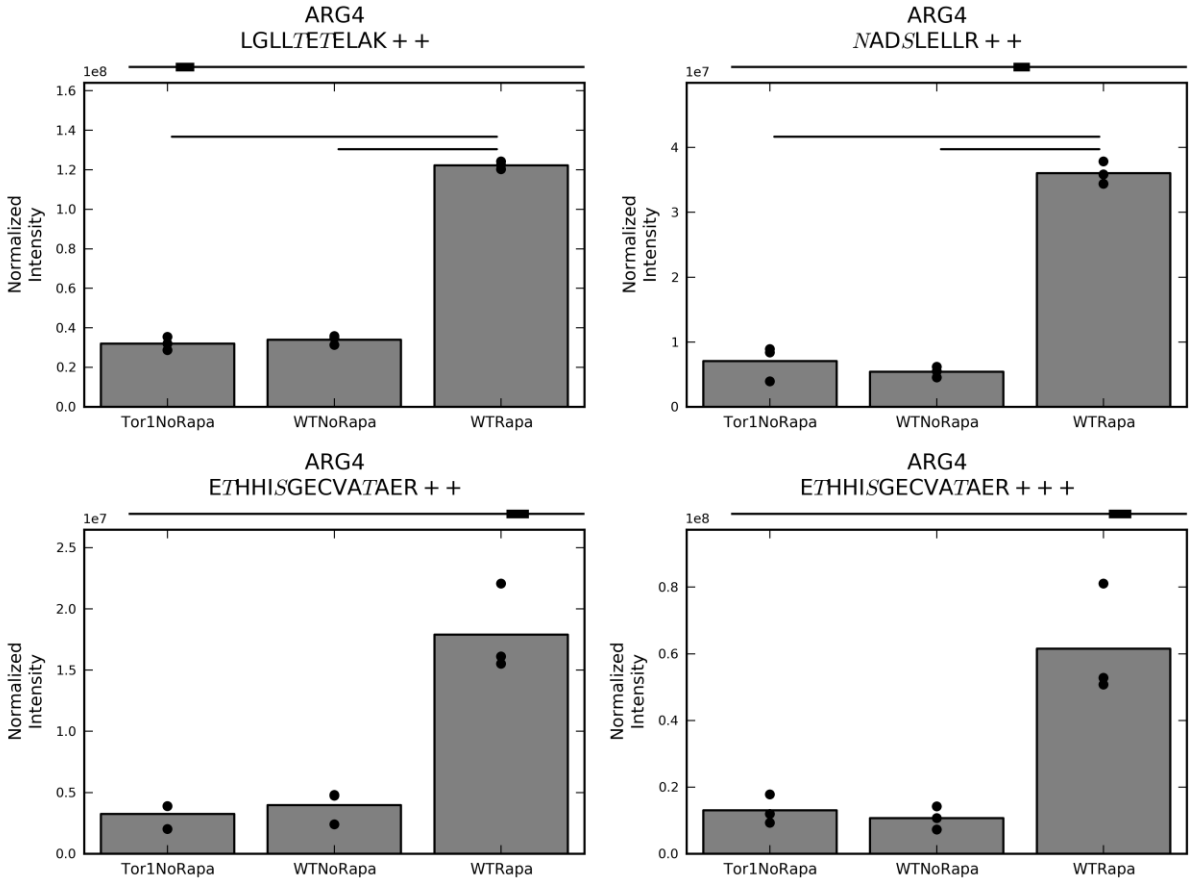

PYC2
VDFIHPGYGFLSENSEFADK + + +
PYC2
AGITWIGPPAEVIDSVGDK + + +
PYC2
TAFGNGTCFVER + +
PYC2
NAGTAEFLVDNQNR + +
PYC2
DAHQSLLATR + +
PYC2
VFDALNDLEQLK + +
PYC2
YNLDYYLEVVEK + +
PYC2
LTCRPGLELEPFDLEK + + +
PYC2
NFLAPAEPDEEIEVTIEQGK + + +
PYC2
DGESVDASDLLVVLEEETLPPSQK + + +
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

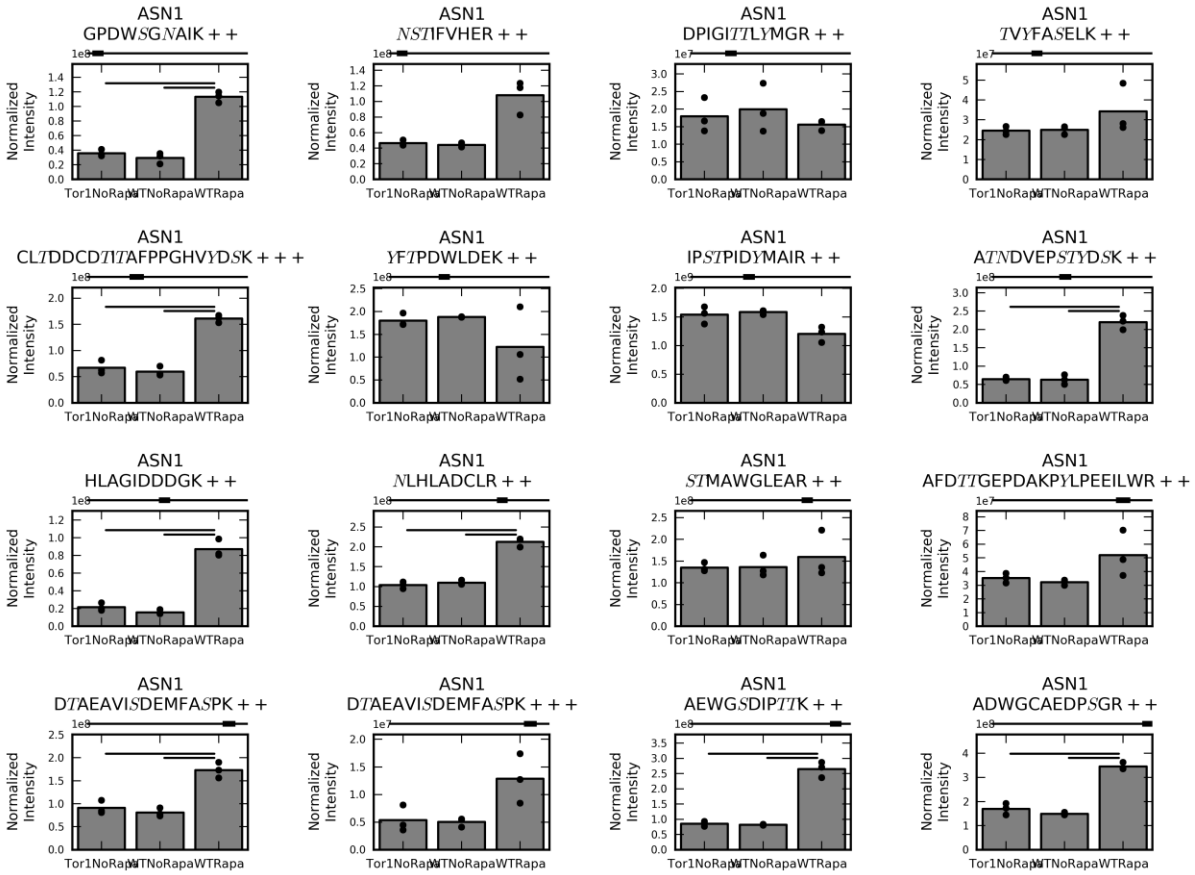

QCR2
FVDE*S*LL*ST*LPAGK + +
1e7
8
7
6
5
4
3
2
1
0
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

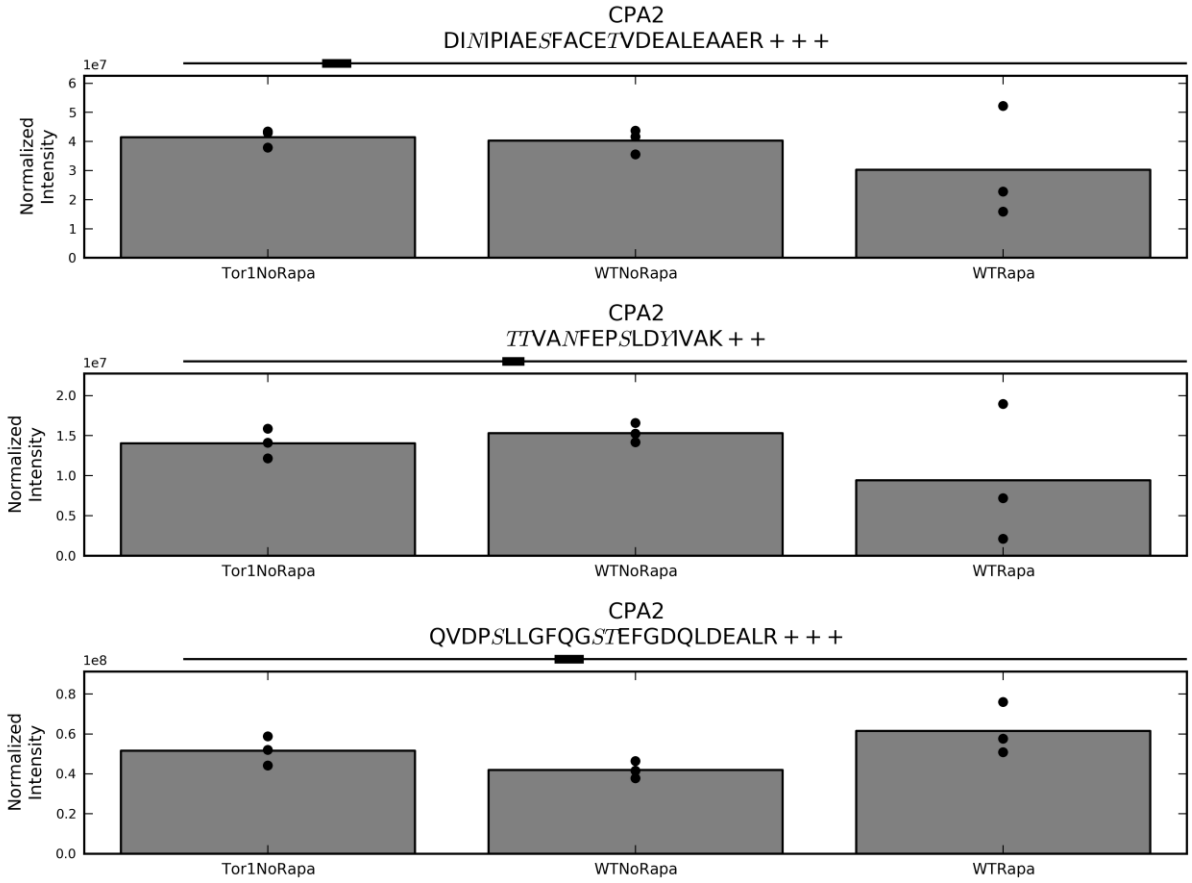
B2: Protein Targets of Gln3 and Stress Response

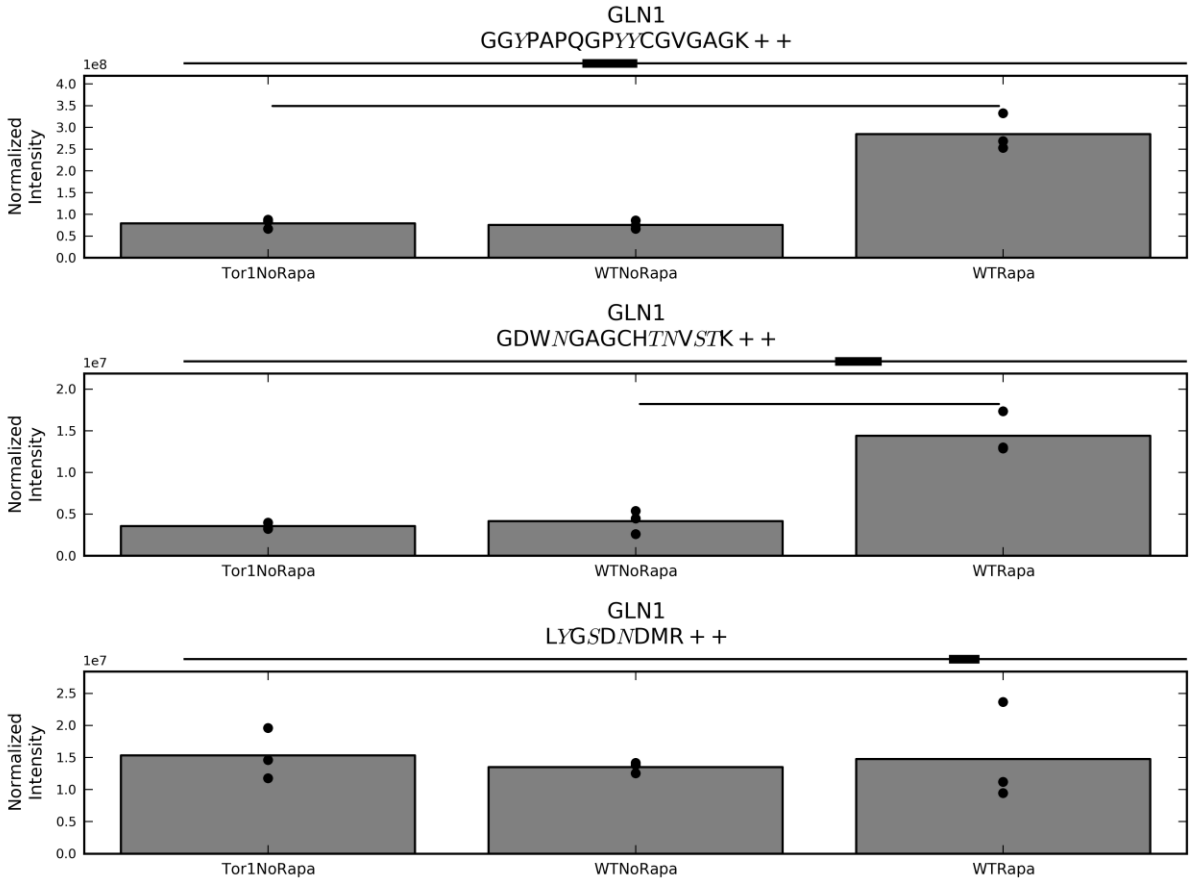
A description of these plots can be found at the beginning of Appendix B.

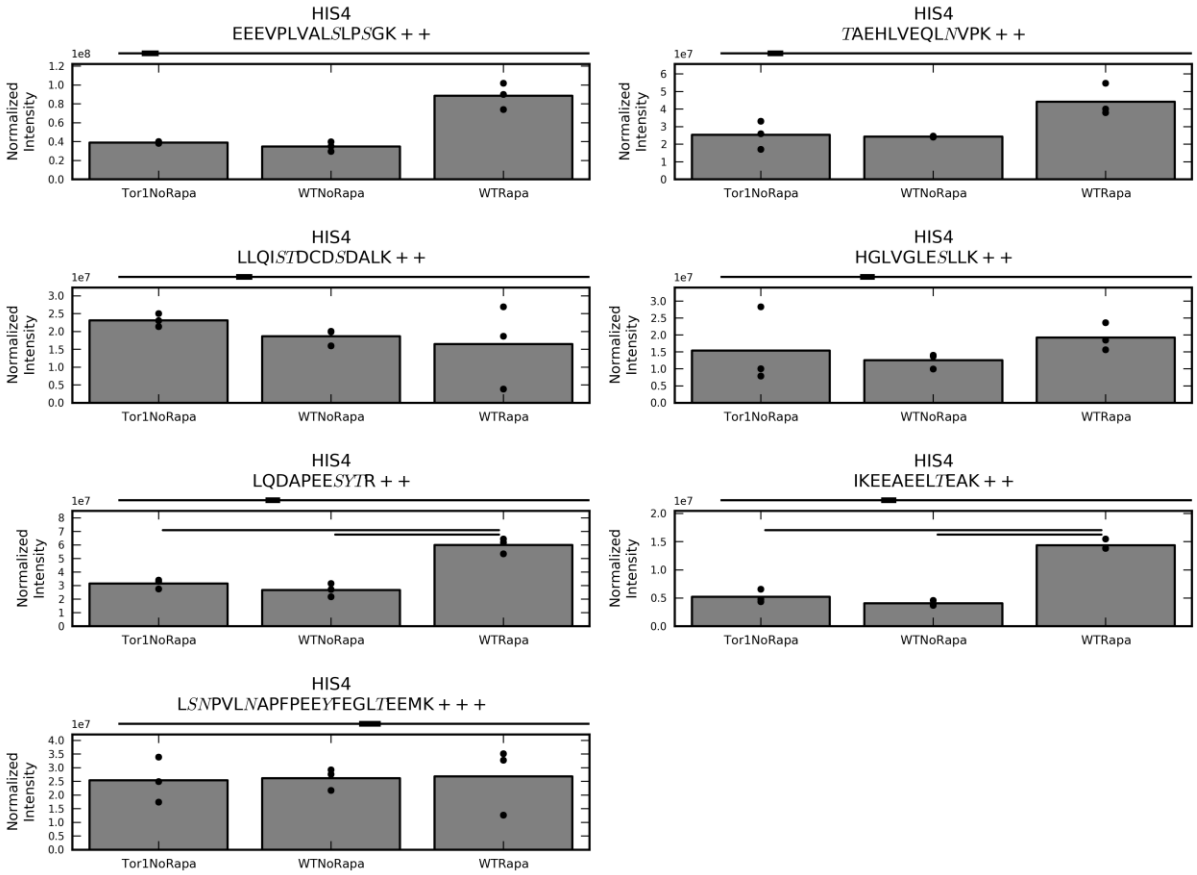


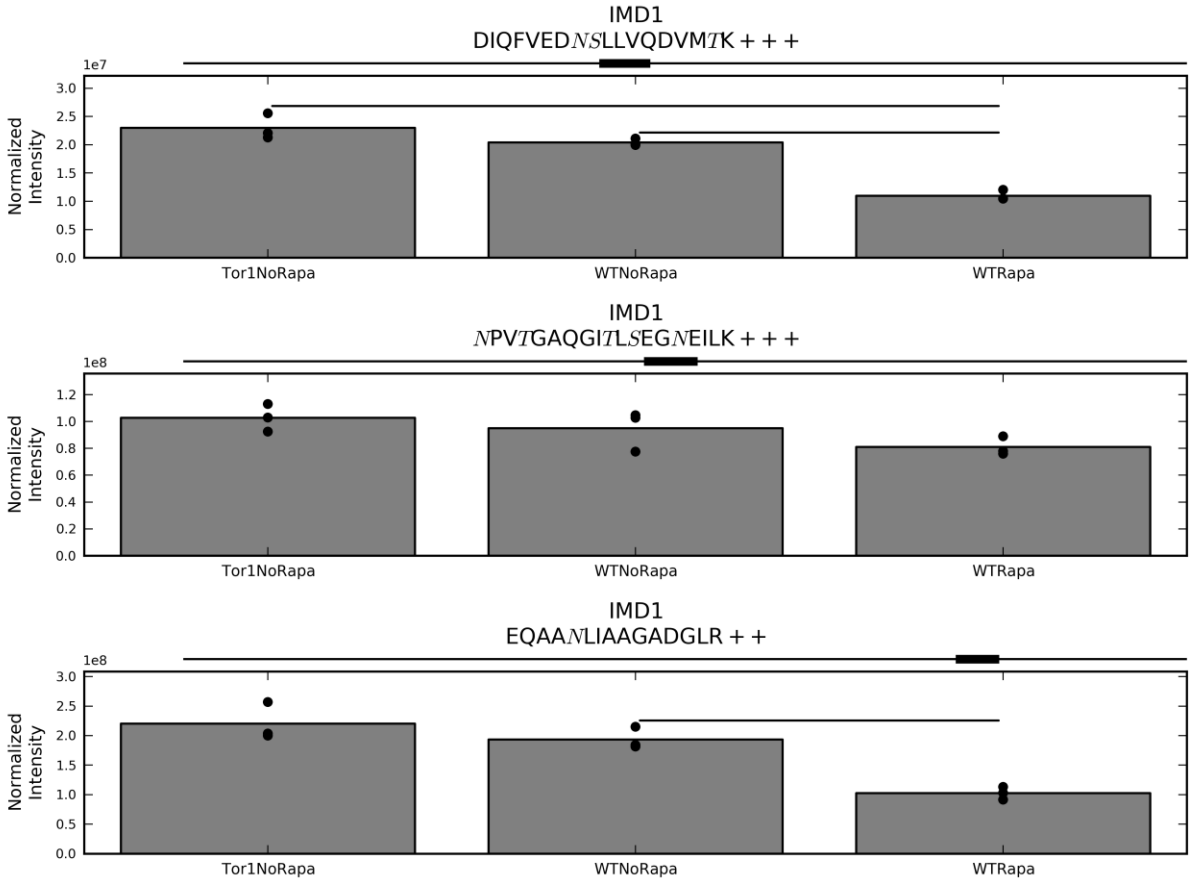

ARG4
LGLL*T*E*T*ELAK + +
1e8
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
ARG4
*N*AD*S*LELLR + +
1e7
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
ARG4
E*T*HHI*S*GECVA*T*AER + +
1e7
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
ARG4
E*T*HHI*S*GECVA*T*AER + + +
1e8
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

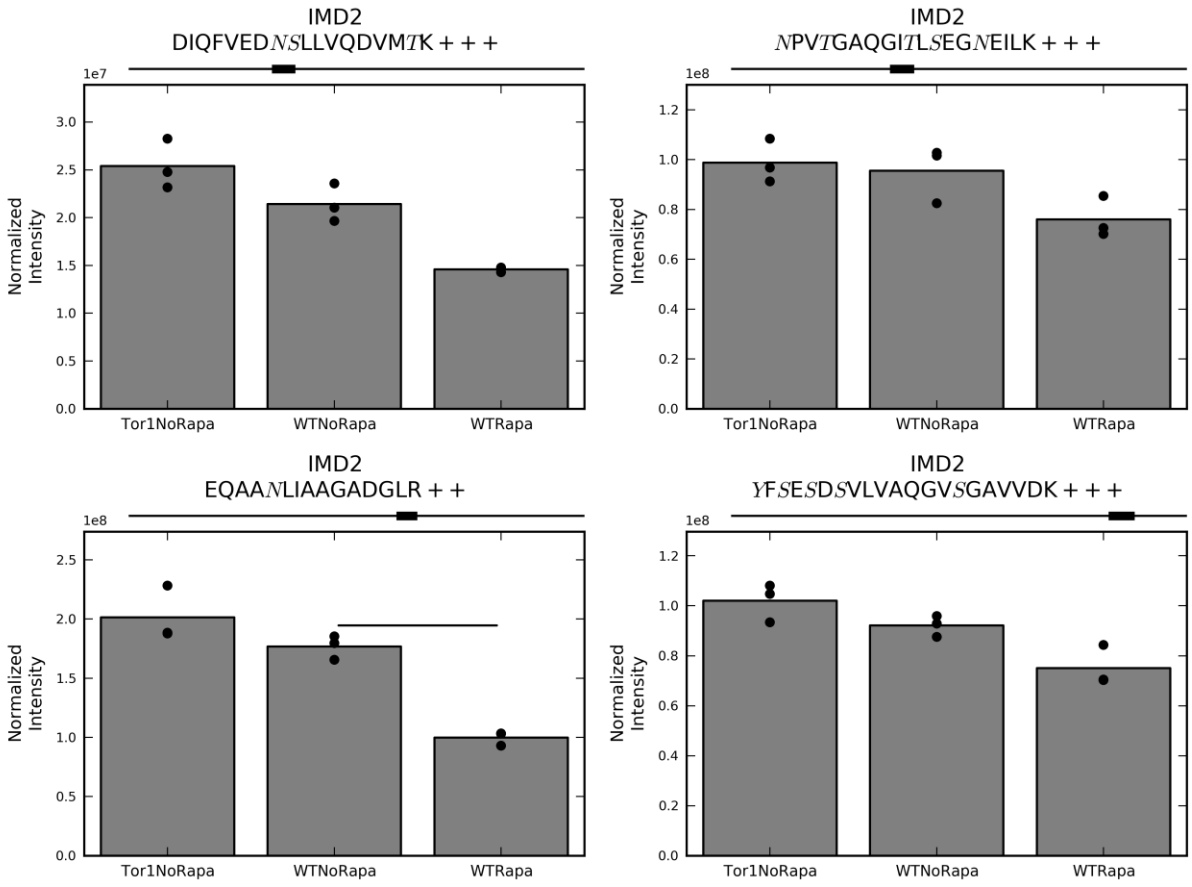

ASN1
GPDWSGNAIK + +
ASN1
NSTIFVHER + +
ASN1
DPIGITTLYMGR + +
ASN1
TVYFASELK + +
ASN1
CLTDDCDTITAFPPGHVYDSK + + +
ASN1
YFTPDWLDEK + +
ASN1
IPSTPIDYMAIR + +
ASN1
ATNDVEPSTYDSK + +
ASN1
HLAGIDDDGK + +
ASN1
NLHLADCLR + +
ASN1
STMAWGLEAR + +
ASN1
AFDTTGEPDAKPYLPEEILWR + +
ASN1
DTAEAVISDEMFASPK + +
ASN1
DTAEAVISDEMFASPK + + +
ASN1
AEWGSDIPTTK + +
ASN1
ADWGCAEDPSGR + +
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

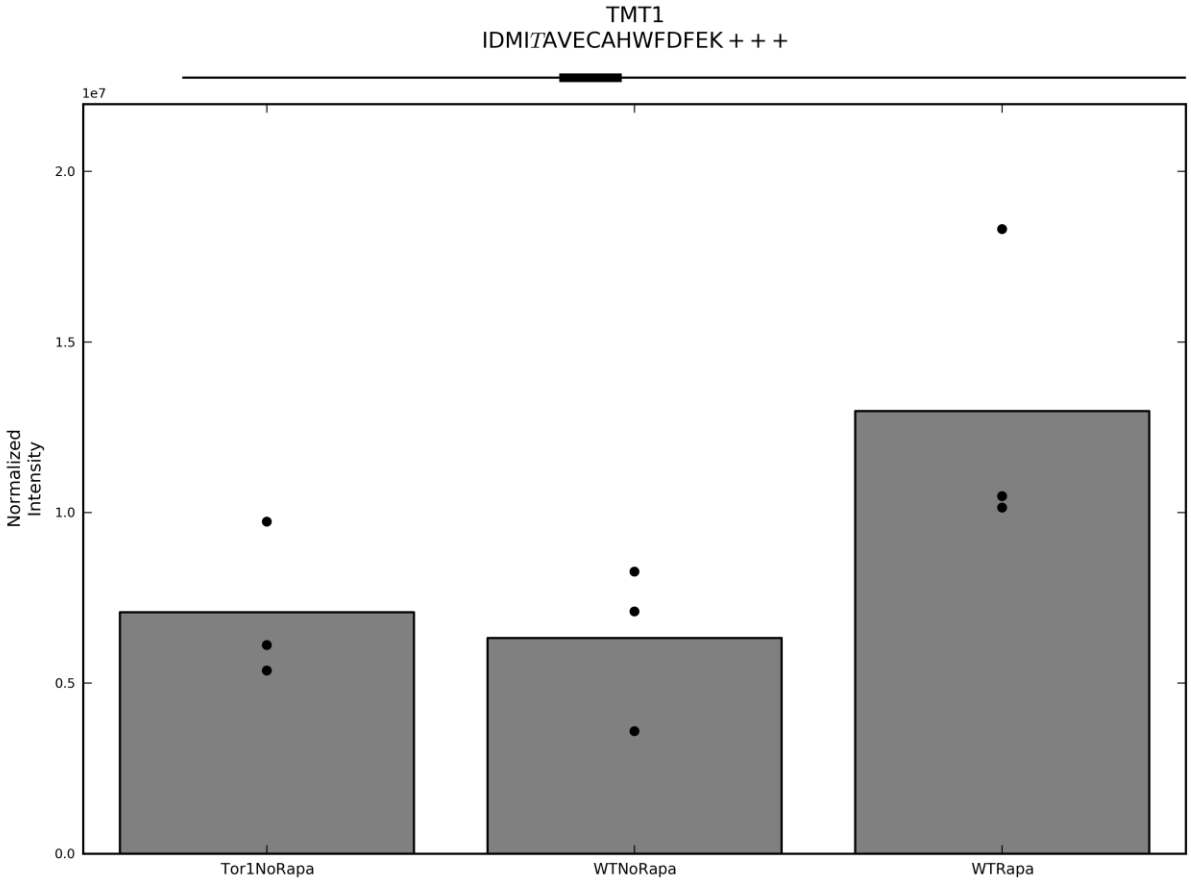

CPA2
DI*N*IPIAE*S*FACE*T*VDEALEAAER + + +
1e7
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
CPA2
*TT*VA*N*FEP*S*LD*Y*IVAK + +
1e7
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
CPA2
QVDP*S*LLGFQG*ST*EFGDQLDEALR + + +
1e8
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa


GLN1
GGYPAPQGPYYCGVGAGK + +
1e8
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
GLN1
GDWNGAGCHTNVSTK + +
1e7
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
GLN1
LYGSDNDMR + +
1e7
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa


HIS4
EEEVPLVALSLPSGK + +
1e8
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
HIS4
TAEHLVEQLNVPK + +
1e7
HIS4
LLQISTDCDSDALK + +
HIS4
HGLVGLESLLK + +
HIS4
LQDAPEESYTR + +
HIS4
IKEEAEELTEAK + +
HIS4
LSNPVLNAPFPEEYFEGLTEEMK + + +

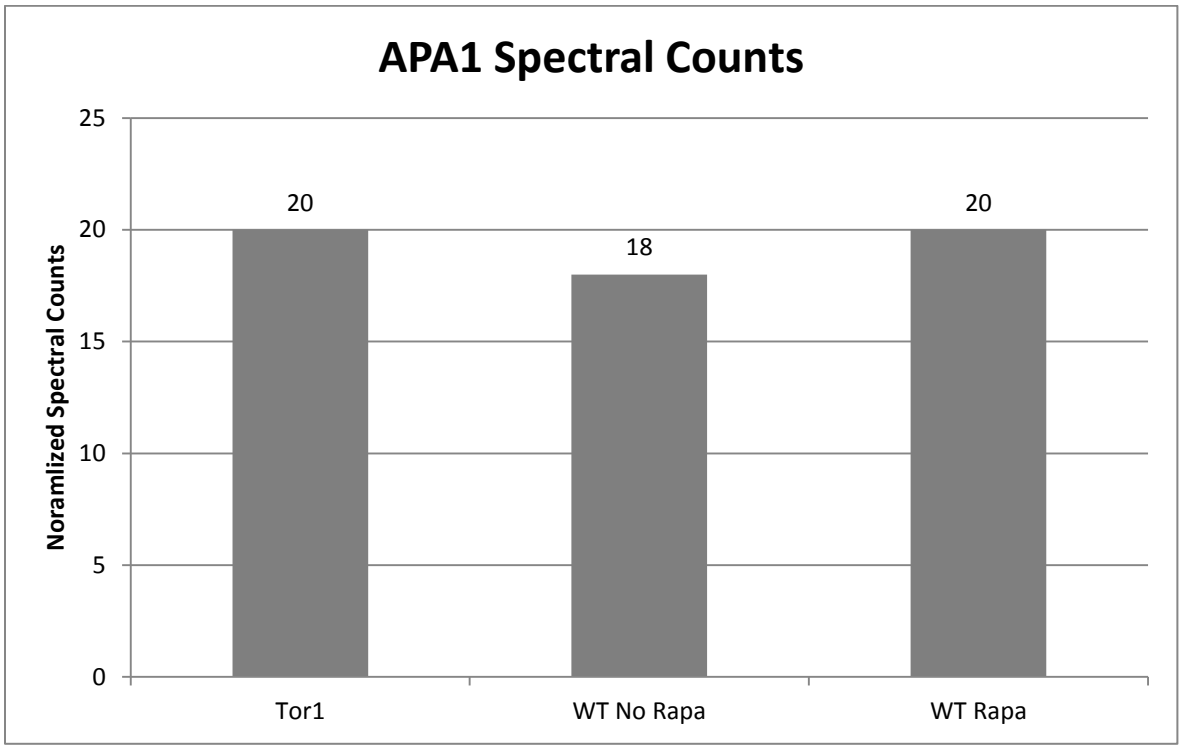

IMD1
DIQFVED*NS*LLVQDVM*T*K + + +
1e7
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
IMD1
*N*PV*T*GAQGI*T*L*S*EG*N*EILK + + +
1e8
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
IMD1
EQAA*N*LIAAGADGLR + +
1e8
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

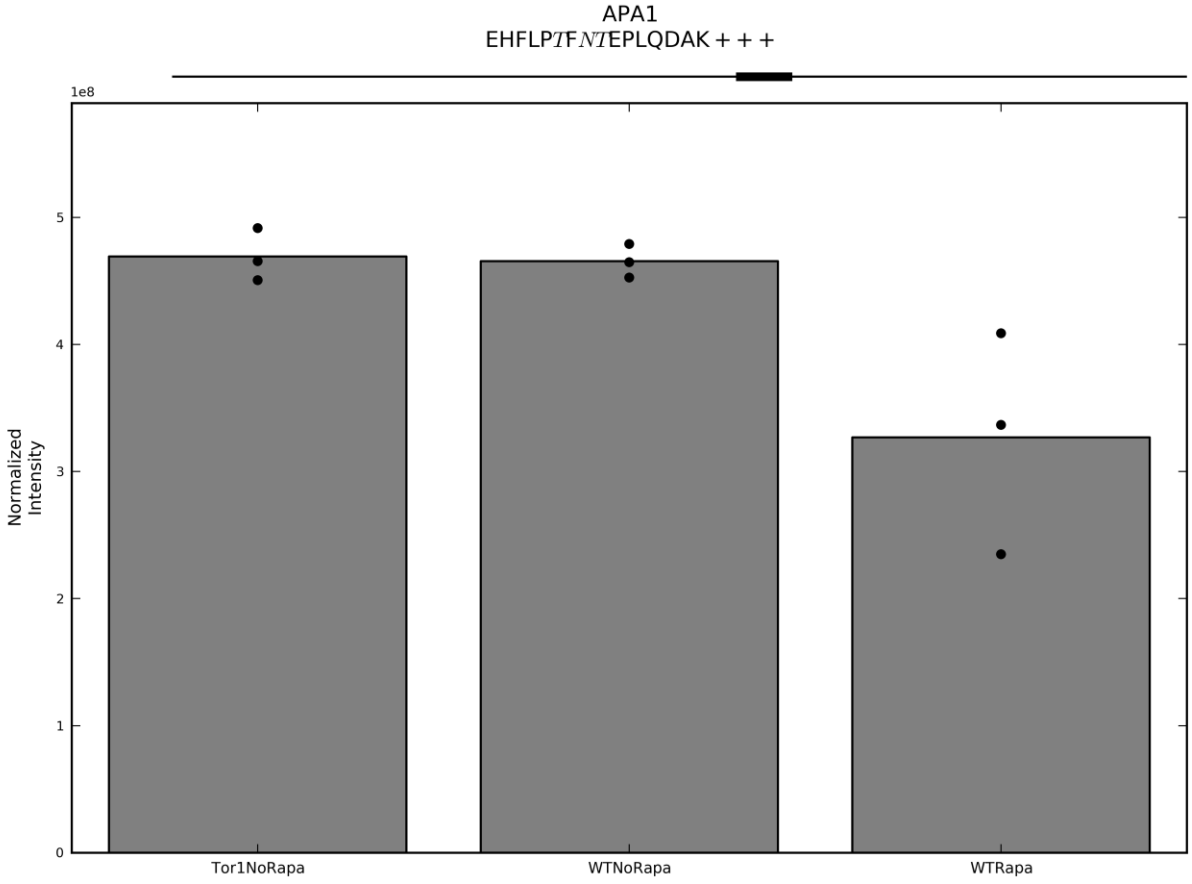

IMD2
DIQFVEDNSLLVQDVMTK + + +
1e7
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
IMD2
NPVTGAQGITLSEGNEILK + + +
1e8
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
IMD2
EQAANLIAAGADGLR + +
1e8
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
IMD2
YFSESDSVLVAQGVSGAVVDK + + +
1e8
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

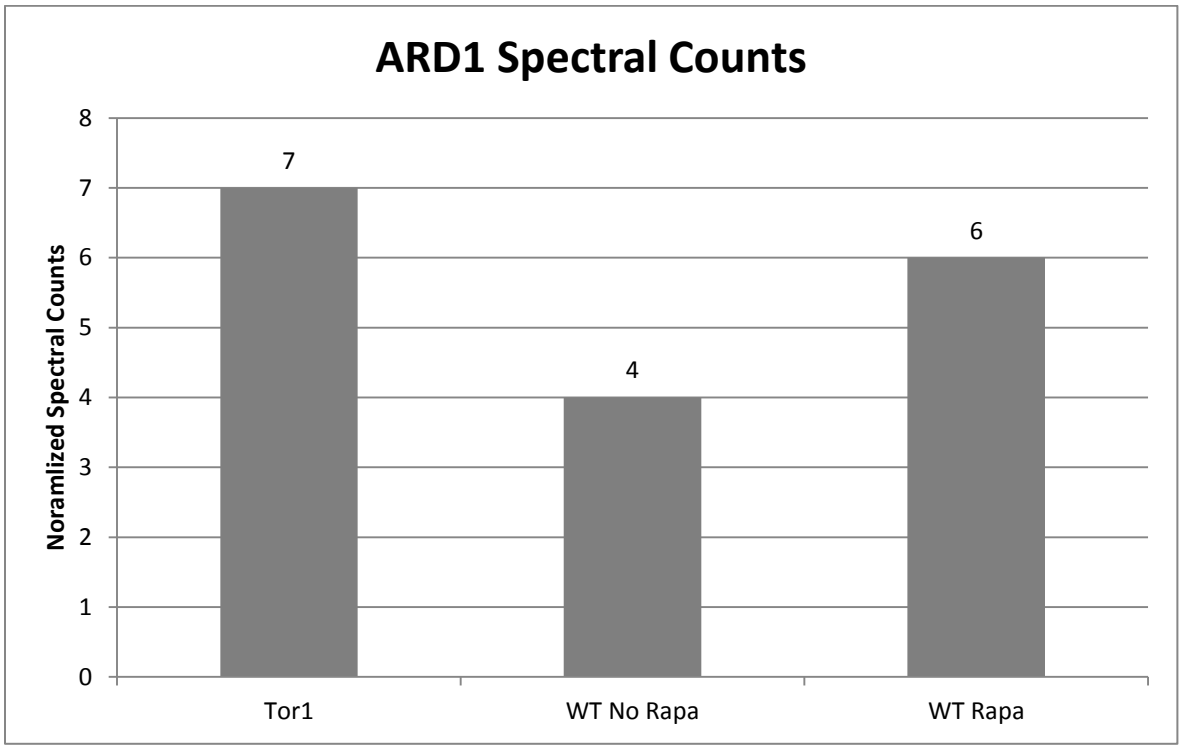

TMT1
IDMITAVECAHWFDFEK + + +
1e7
2.0
1.5
1.0
0.5
0.0
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

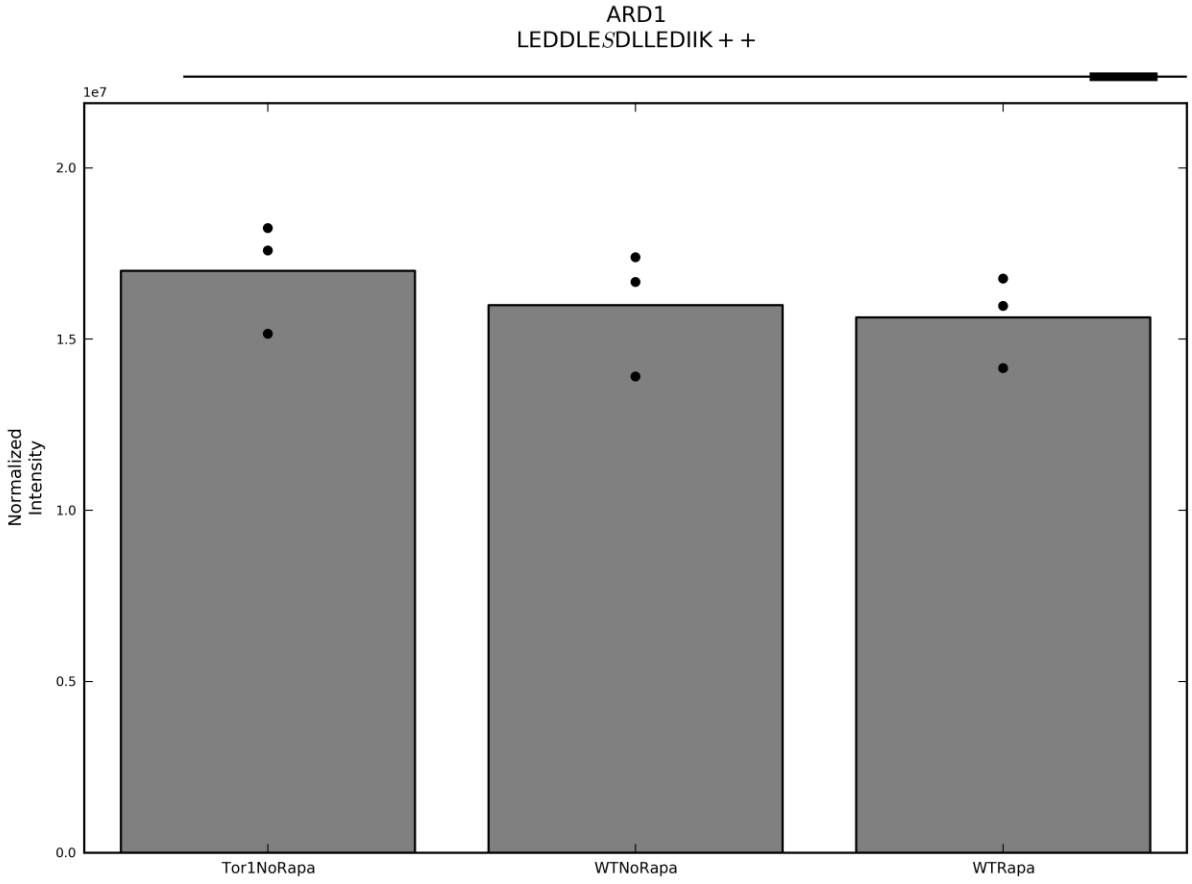
B3: Proteins With Shared Changes in Rapamycin and *tor1* Yeast Relative to Wild Type According to Spectral Counting Data

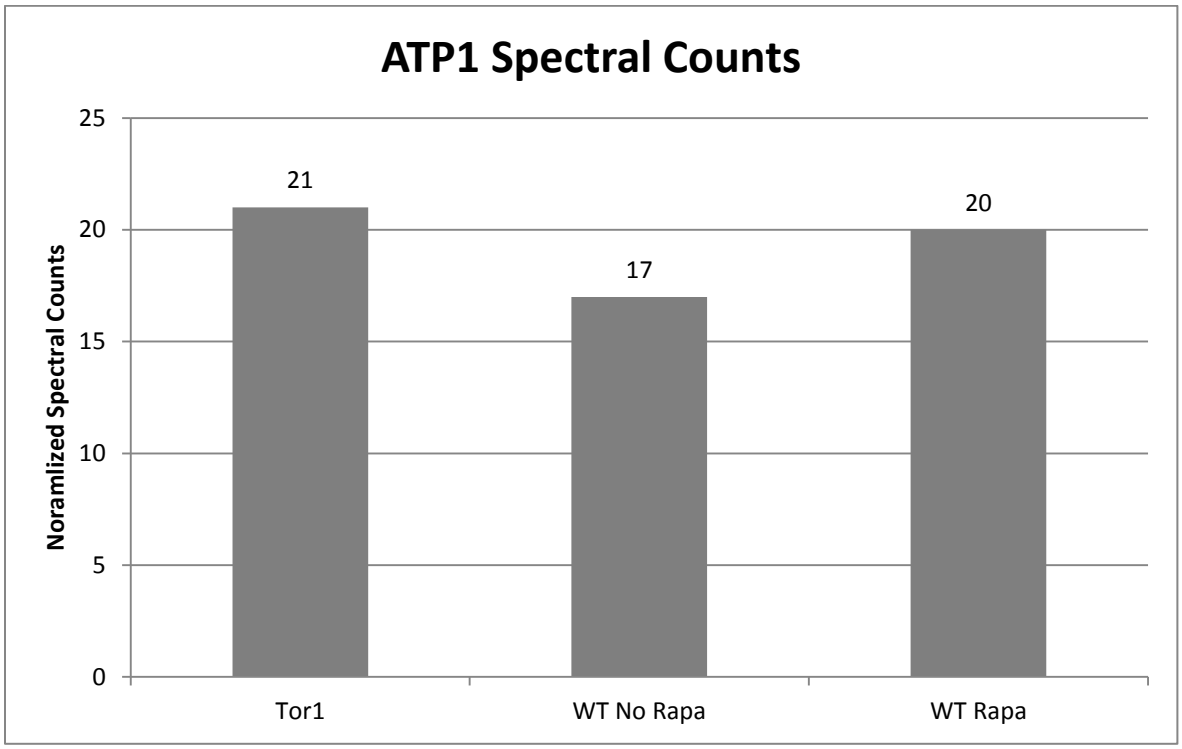
Each of the pages in this section contains spectral counting quantification data (top) and DIA quantification data (bottom) for a protein. These proteins were chosen as potentially interesting due to increased or decreased abundance in both WT + rapamycin and *tor1* samples relative to WT based on spectral counting data. The top plots show the number of spectral counts for the protein in each sample type. The spectral counts were normalized to the total number of spectra in each mass spectrometry run. The DIA quantification data (bottom) is described in the beginning of Appendix B.

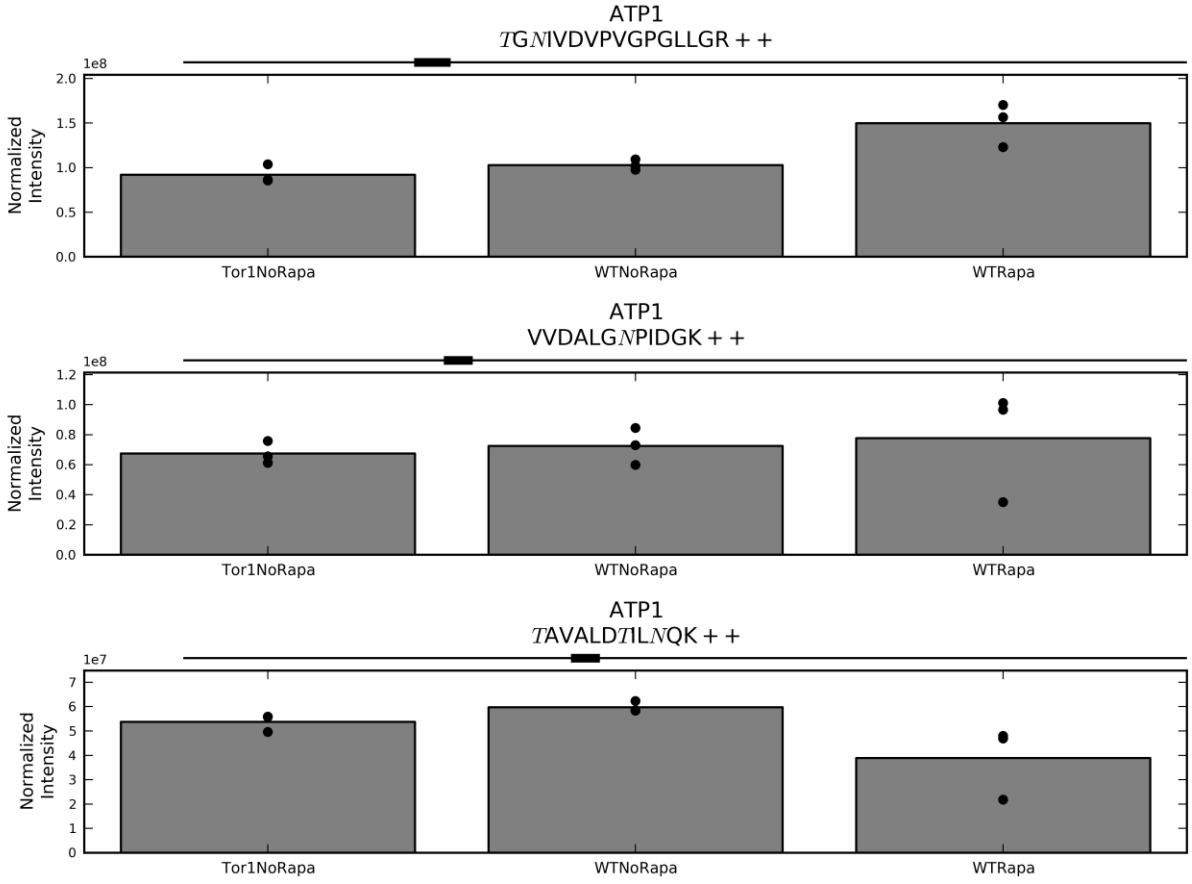

APA1 Spectral Counts
Noramlized Spectral Counts
25
20
15
10
5
0
20
18
20
Tor1
WT No Rapa
WT Rapa

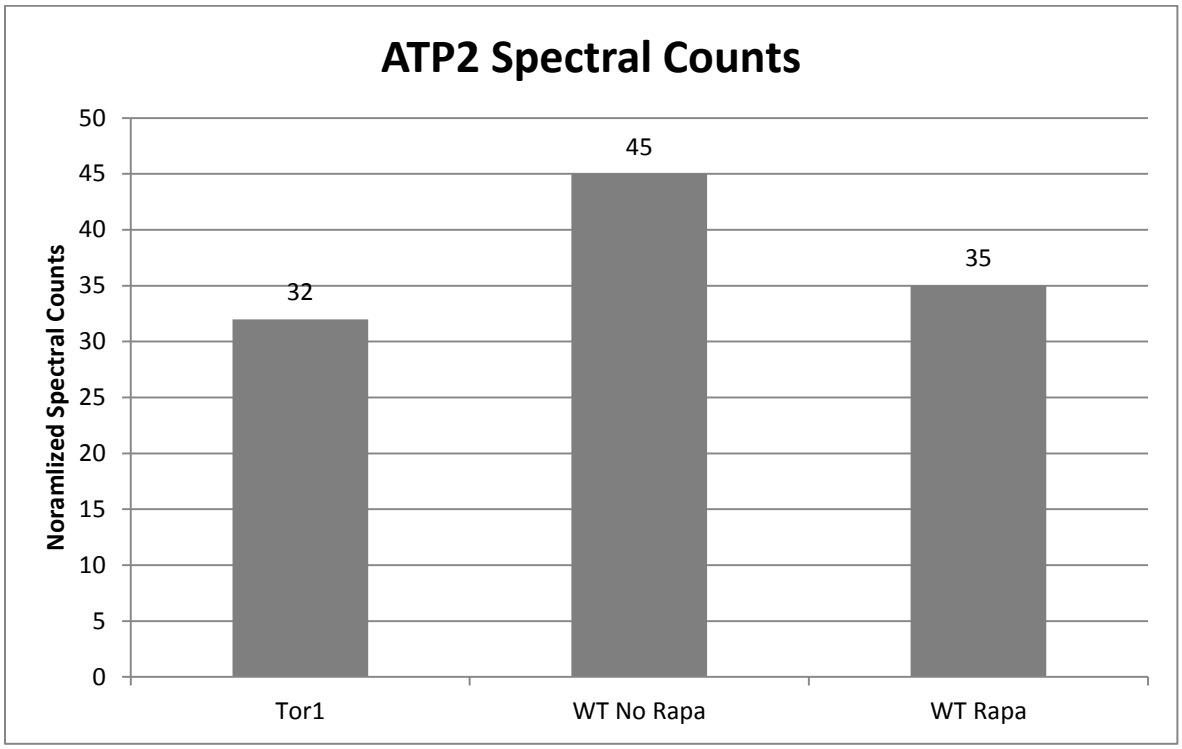

APA1
EHFLPTFNTEPLQDAK + + +
1e8
Normalized Intensity
5
4
3
2
1
0
Tor1NoRapa
WTNoRapa
WTRapa

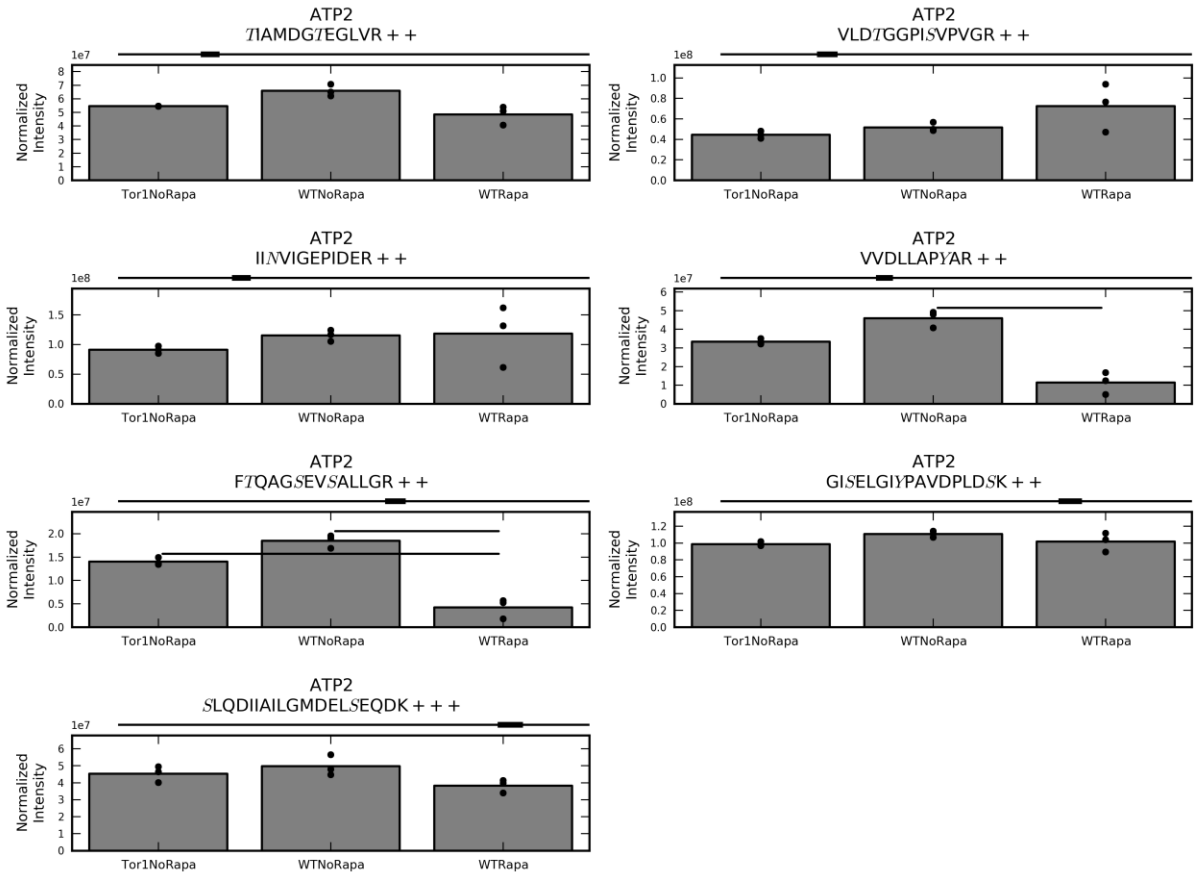

ARD1 Spectral Counts
Noramlized Spectral Counts
8
7
6
5
4
3
2
1
0
7
4
6
Tor1
WT No Rapa
WT Rapa

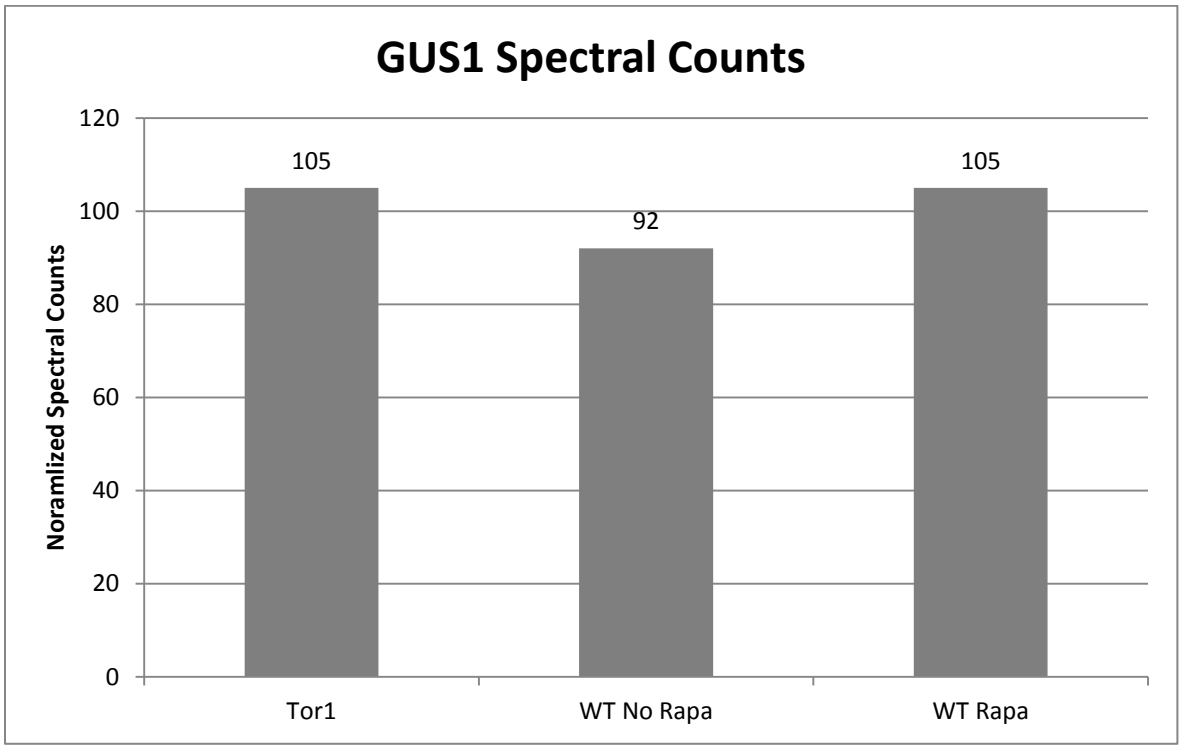

ARD1
LEDDLESDLLEDIIK + +
1e7
2.0
1.5
1.0
0.5
0.0
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

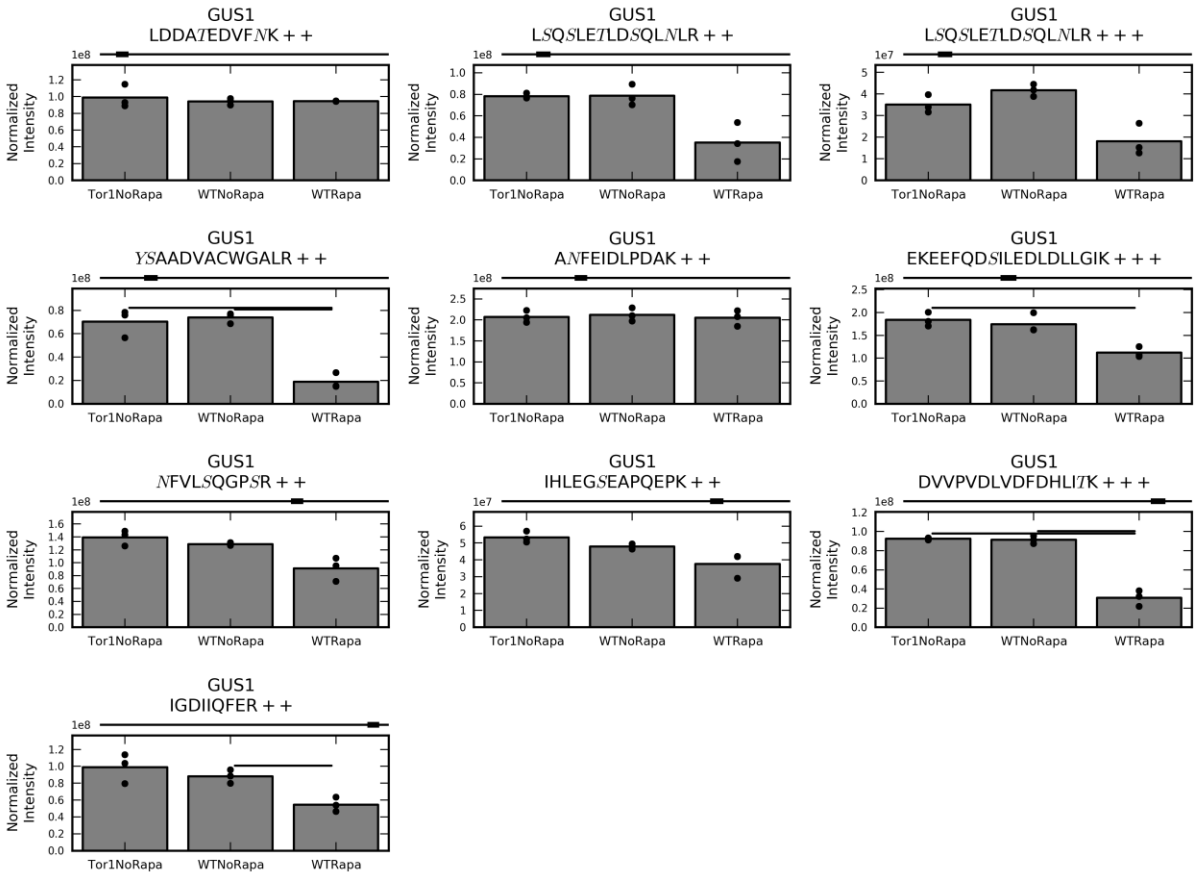

ATP1 Spectral Counts
Noramlized Spectral Counts
25
20
15
10
5
0
21
17
20
Tor1
WT No Rapa
WT Rapa

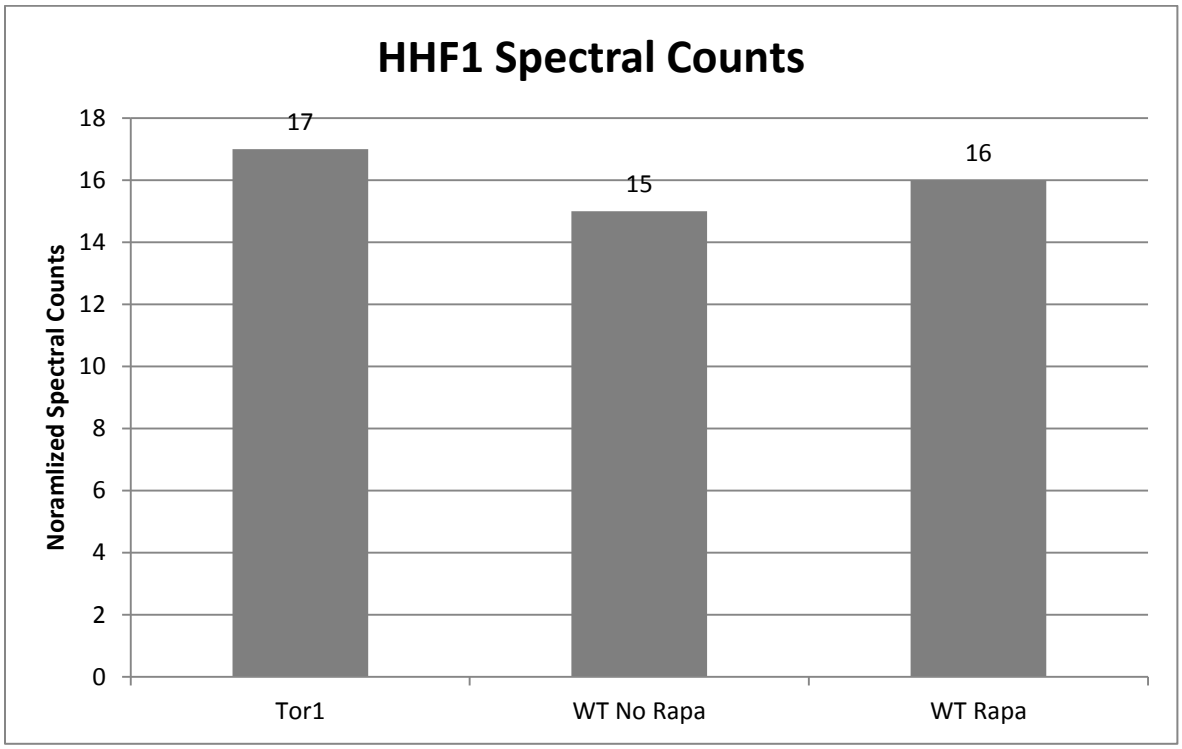

ATP1
TGNIVDVPVGPGLLGR + +
1e8
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
ATP1
VVDALGNPIDGK + +
1e8
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
ATP1
TAVALDTILNQK + +
1e7
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

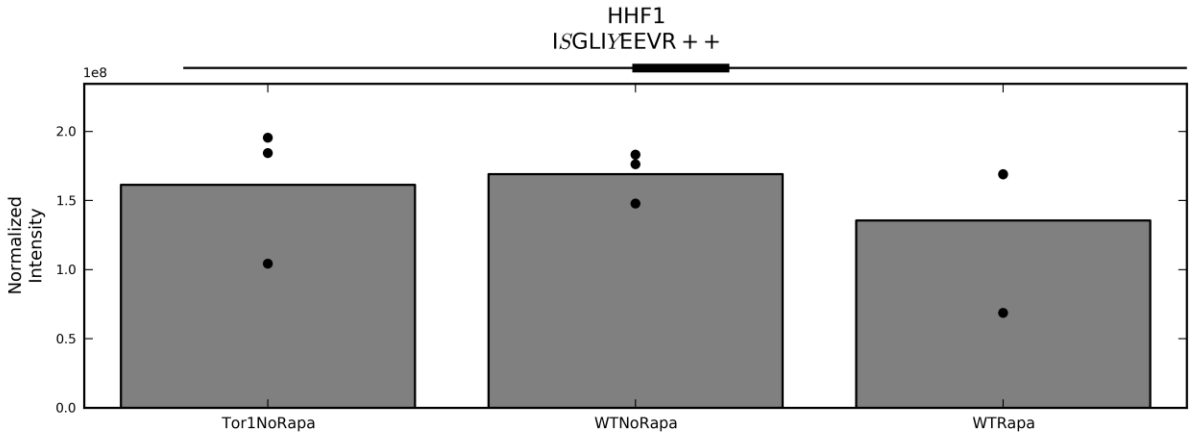

ATP2 Spectral Counts
Noramlized Spectral Counts
50
45
40
35
30
25
20
15
10
5
0
32
45
35
Tor1
WT No Rapa
WT Rapa

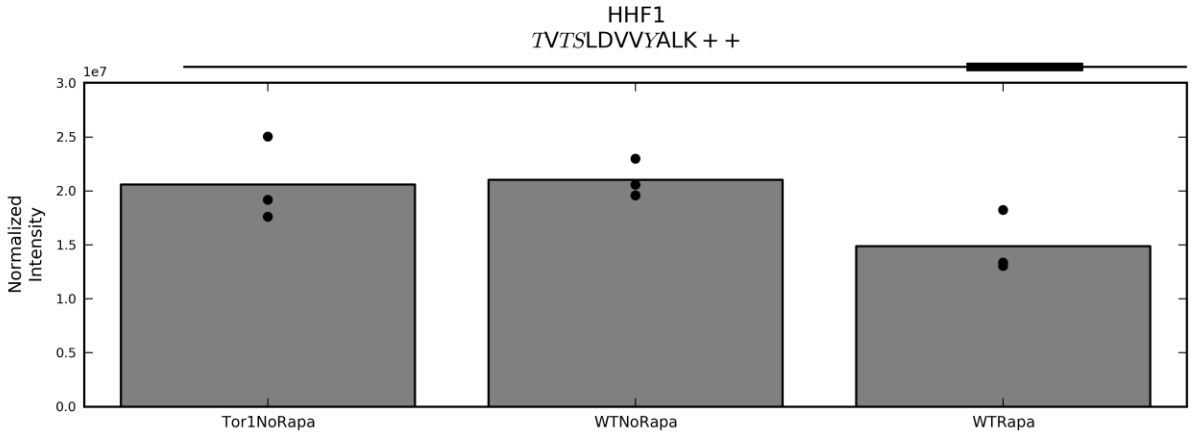

ATP2
TIAMDGTEGLVR + +
ATP2
VLDTGGPISVPVGR + +
ATP2
IINVIGEPIDER + +
ATP2
VVDLLAPYAR + +
ATP2
FTQAGSEVSALLGR + +
ATP2
GISELGIYPAVDPLDSK + +
ATP2
SLQDIIAILGMDELSEQDK + + +
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

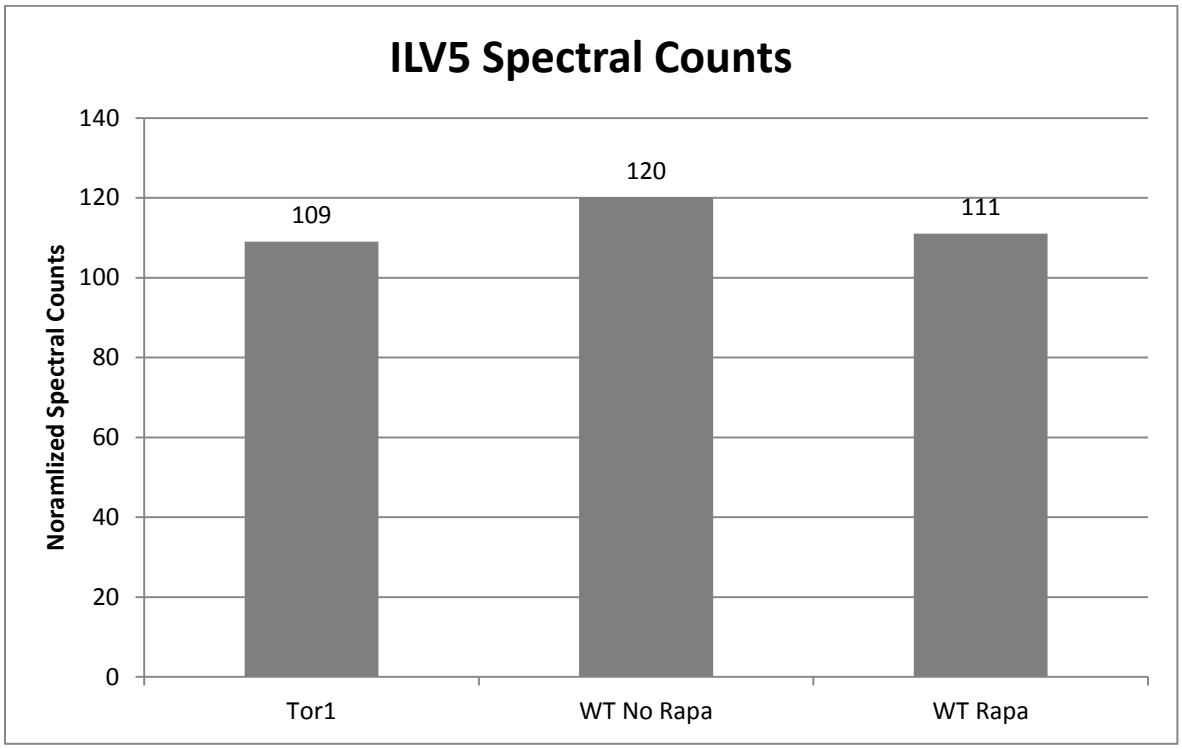

GUS1 Spectral Counts
Noramlized Spectral Counts
120
100
80
60
40
20
0
105
92
105
Tor1
WT No Rapa
WT Rapa

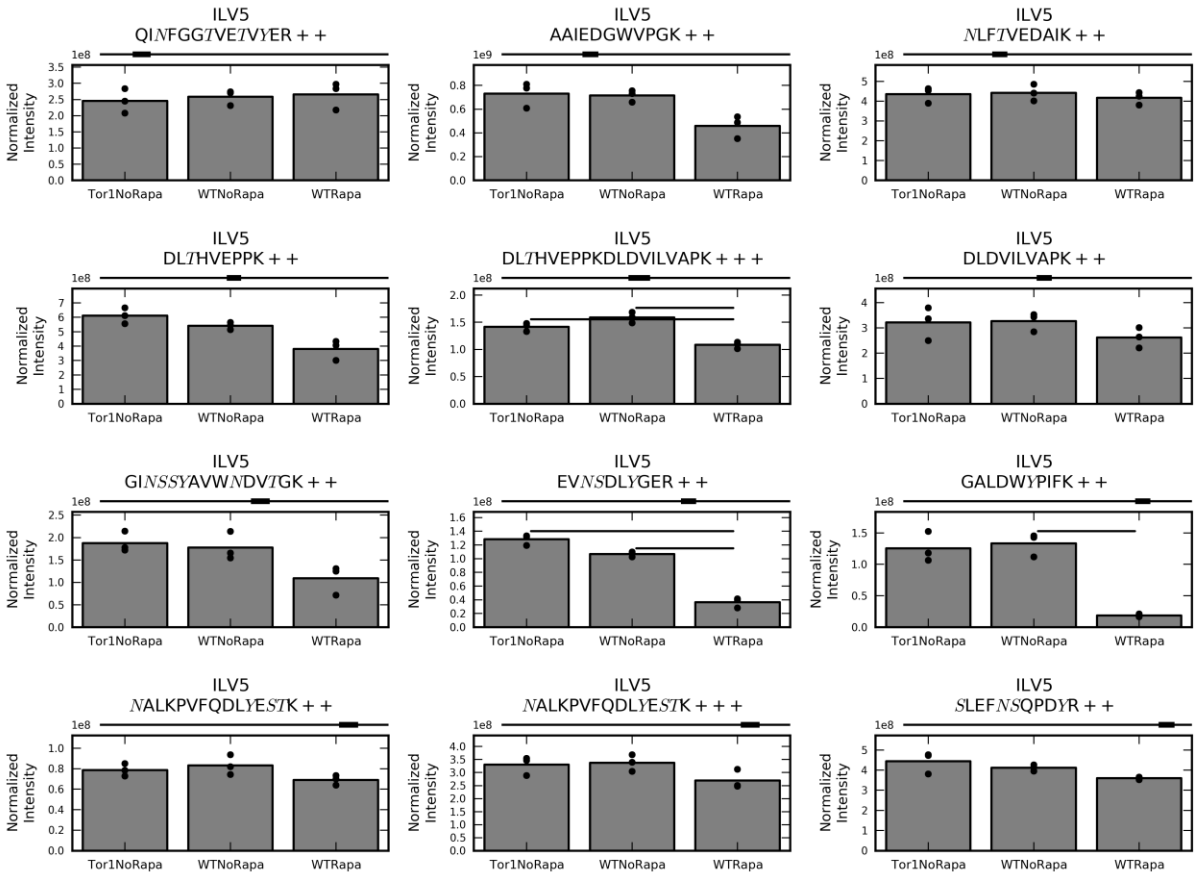

GUS1
LDDA*T*EDVF*N*K + +
GUS1
L*S*Q*S*LE*T*LD*S*QL*N*LR + +
GUS1
L*S*Q*S*LE*T*LD*S*QL*N*LR + + +
GUS1
*YS*AADVACWGALR + +
GUS1
A*N*FEIDLPDAK + +
GUS1
EKEEFQD*S*ILEDLDLLGIK + + +
GUS1
*N*FVL*S*QGP*S*R + +
GUS1
IHLEG*S*EAPQEPK + +
GUS1
DVVPVDLVDFDHLI*T*K + + +
GUS1
IGDIIQFER + +
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

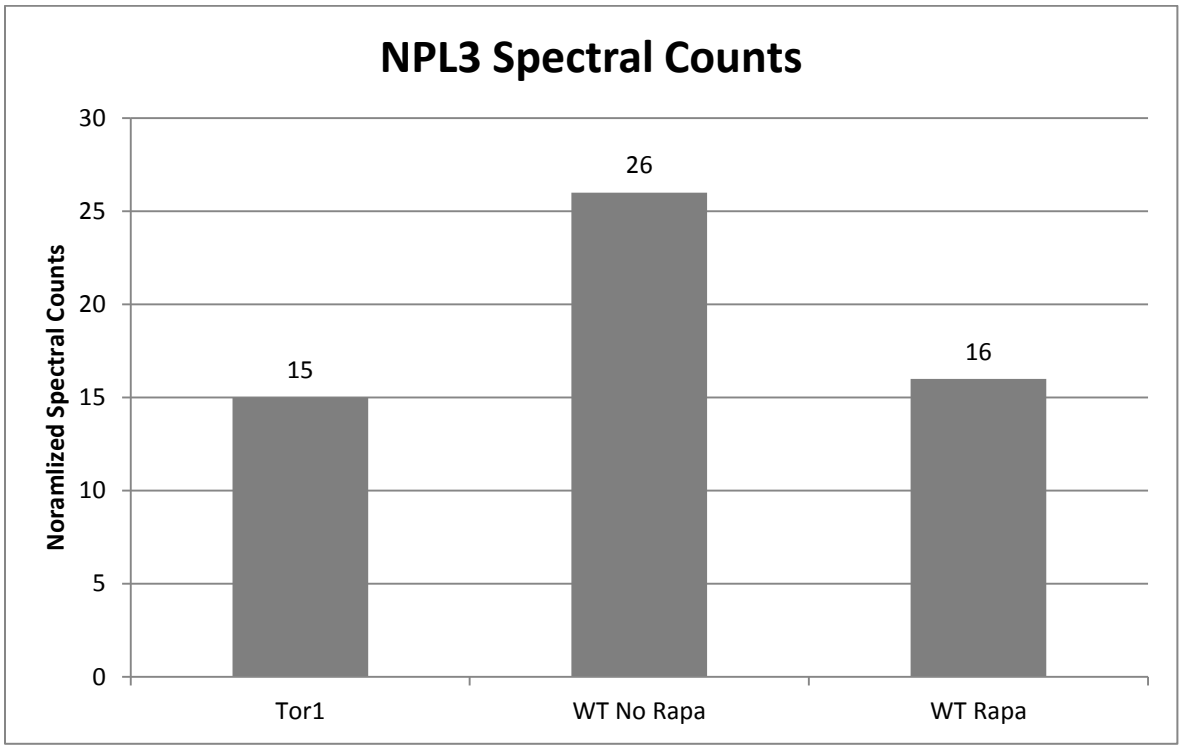

HHF1 Spectral Counts
Noramlized Spectral Counts
18
16
14
12
10
8
6
4
2
0
17
15
16
Tor1
WT No Rapa
WT Rapa

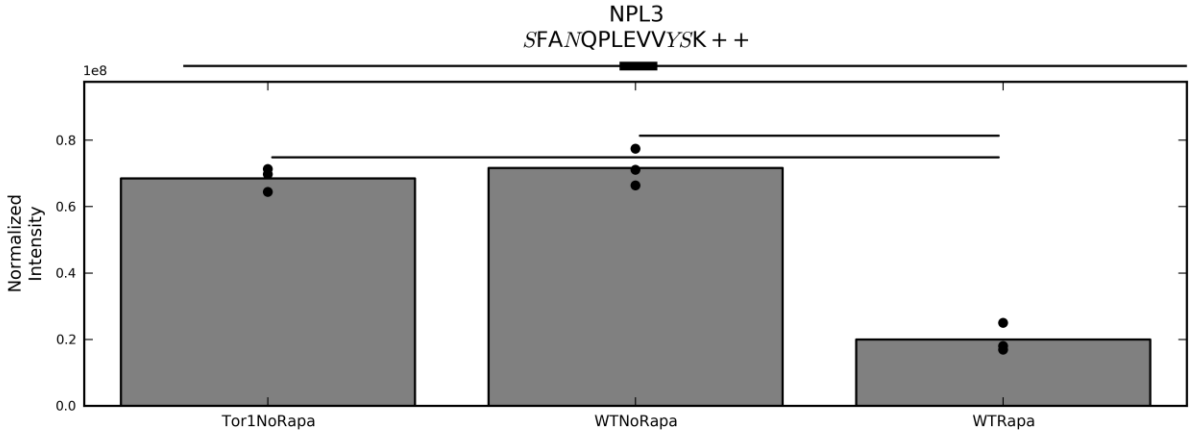

HHF1
ISGLIYEEVR + +
1e8
Normalized Intensity
2.0
1.5
1.0
0.5
0.0
Tor1NoRapa
WTNoRapa
WTRapa

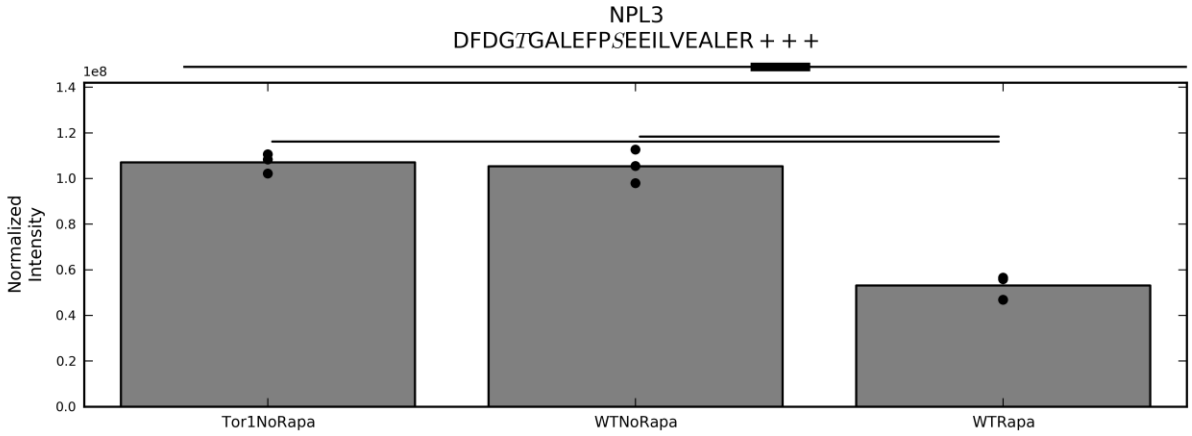

HHF1
TVTSLDVVYALK + +
1e7
Normalized Intensity
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Tor1NoRapa
WTNoRapa
WTRapa

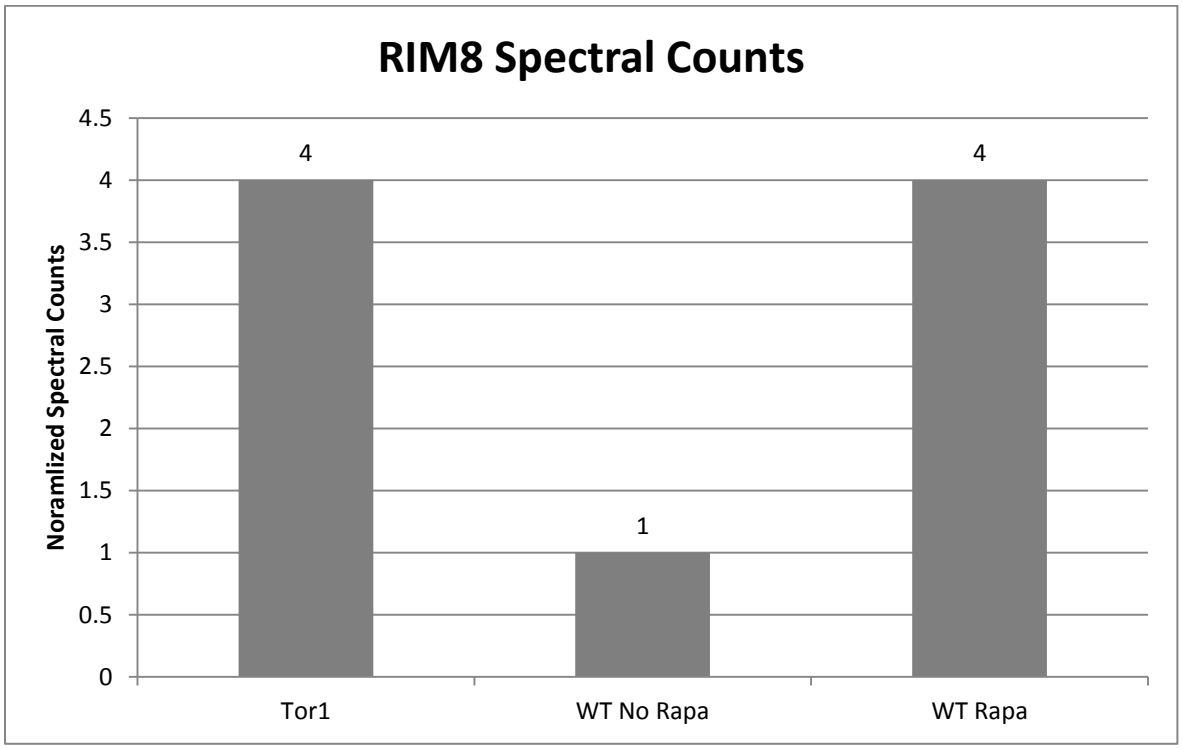

ILV5 Spectral Counts
Noramlized Spectral Counts
140
120
100
80
60
40
20
0
109
120
111
Tor1
WT No Rapa
WT Rapa

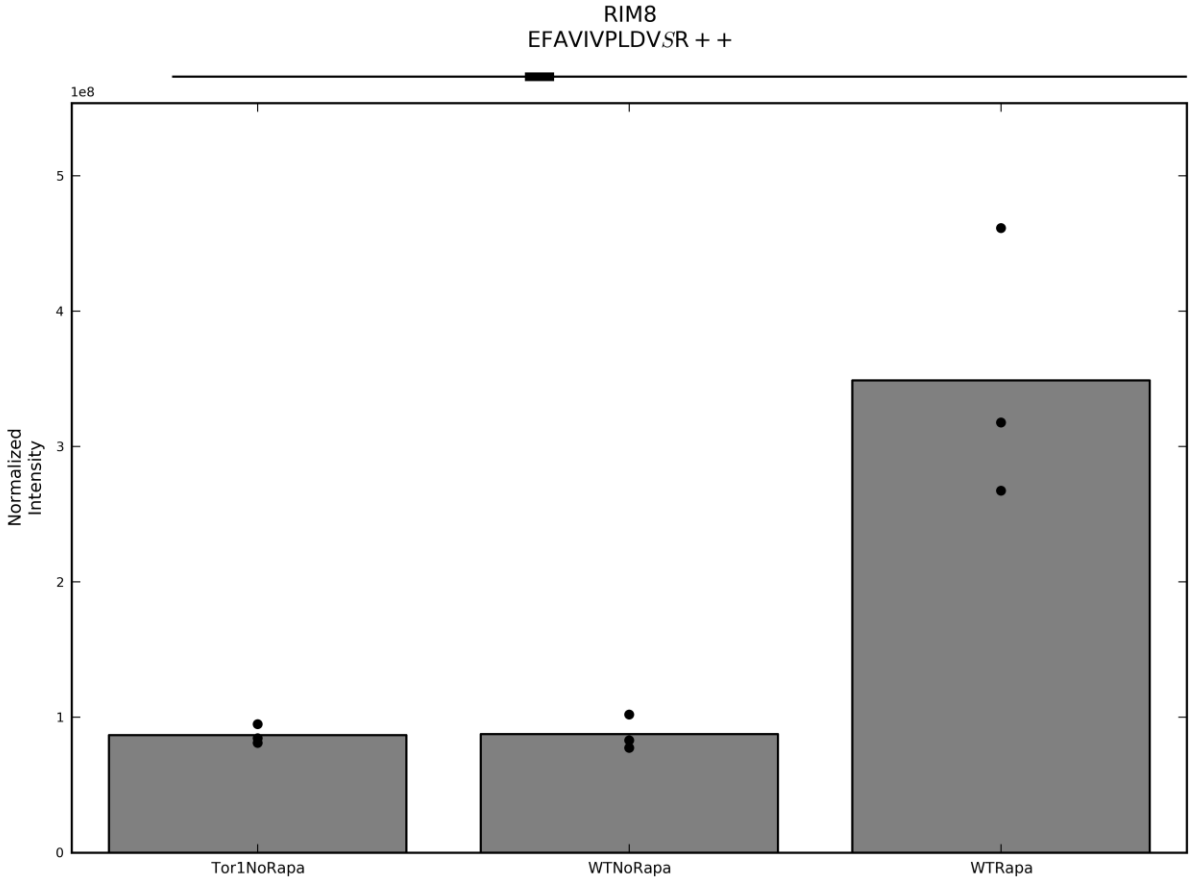

ILV5
QI*N*FGG*T*VE*T*V*Y*ER + +
ILV5
AAIEDGWVPGK + +
ILV5
*N*LF*T*VEDAIK + +
ILV5
DL*T*HVEPPK + +
ILV5
DL*T*HVEPPKDLDVILVAPK + + +
ILV5
DLDVILVAPK + +
ILV5
GI*NSSY*AVW*N*DV*T*GK + +
ILV5
EV*NS*DL*Y*GER + +
ILV5
GALDW*Y*PIFK + +
ILV5
*N*ALKPVFQDL*Y*E*ST*K + +
ILV5
*N*ALKPVFQDL*Y*E*ST*K + + +
ILV5
*S*LEF*NS*QPD*Y*R + +
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

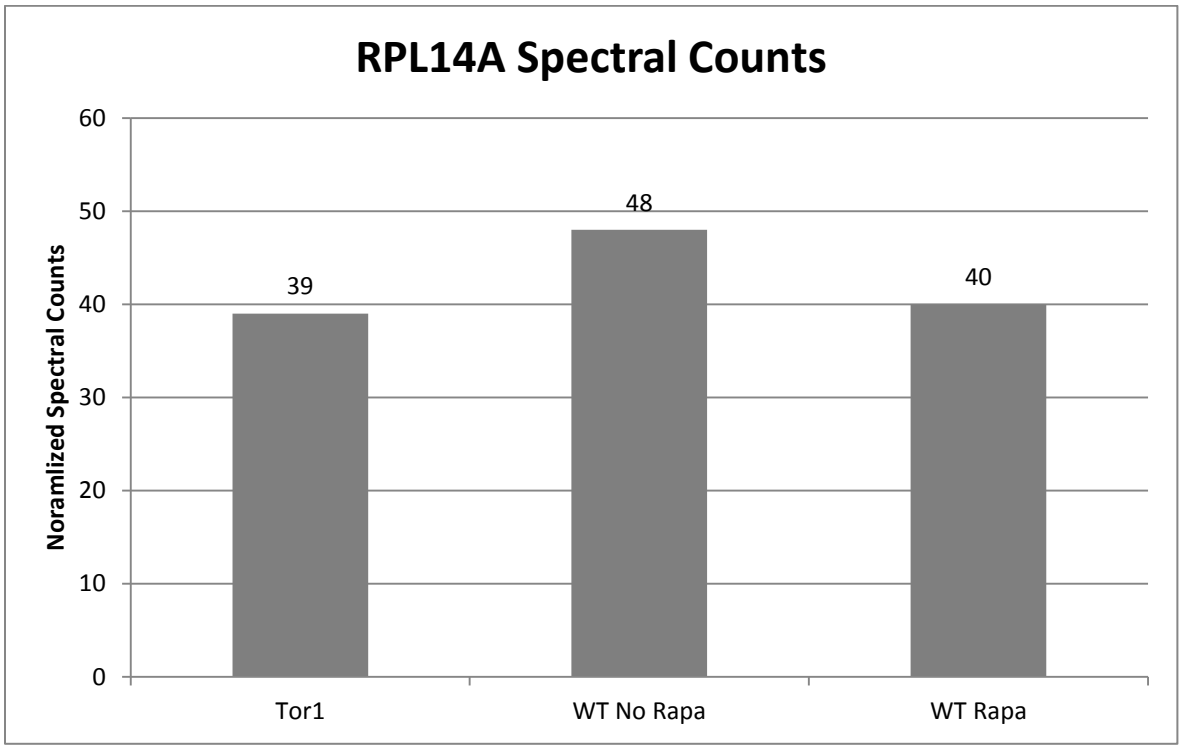

NPL3 Spectral Counts
Noramlized Spectral Counts
30
25
20
15
10
5
0
15
26
16
Tor1
WT No Rapa
WT Rapa

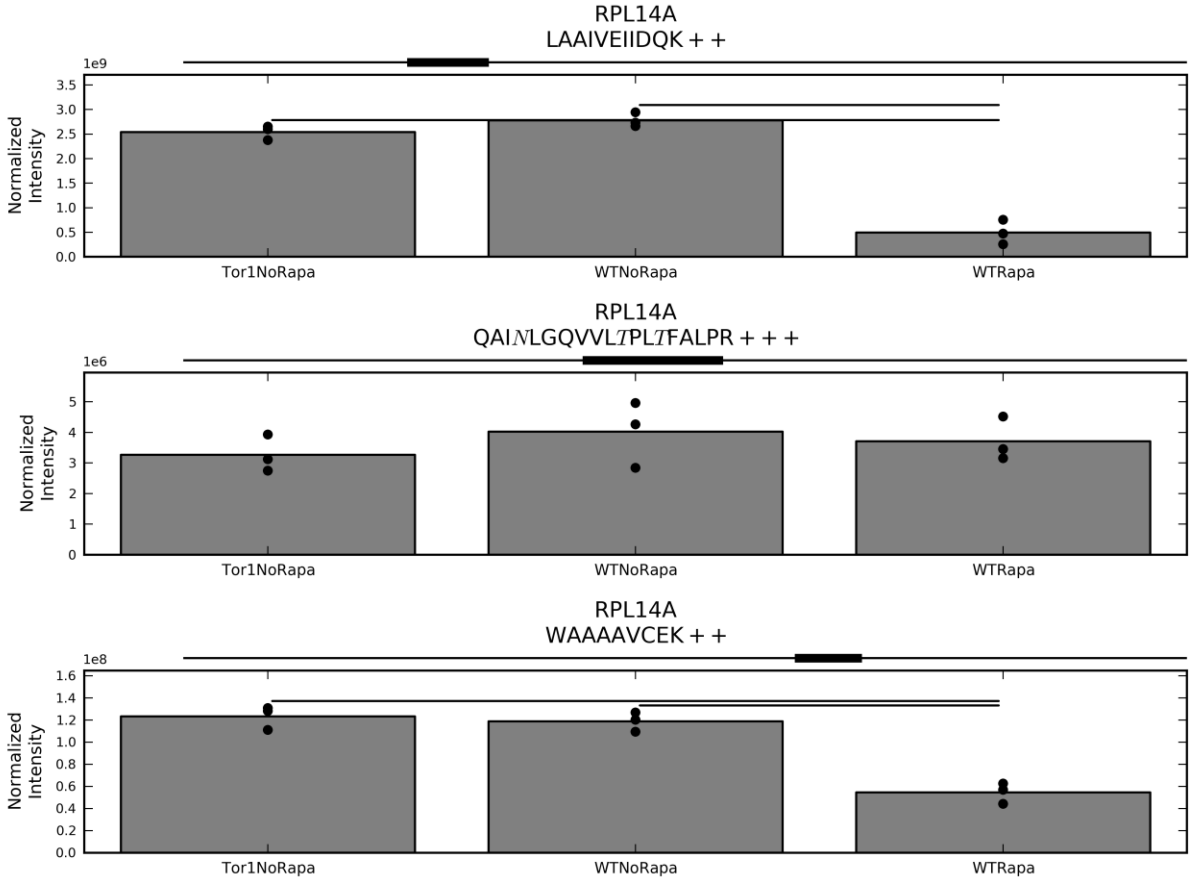

NPL3
SFANQPLEVVYSK + +
1e8
Normalized Intensity
0.8
0.6
0.4
0.2
0.0
Tor1NoRapa
WTNoRapa
WTRapa

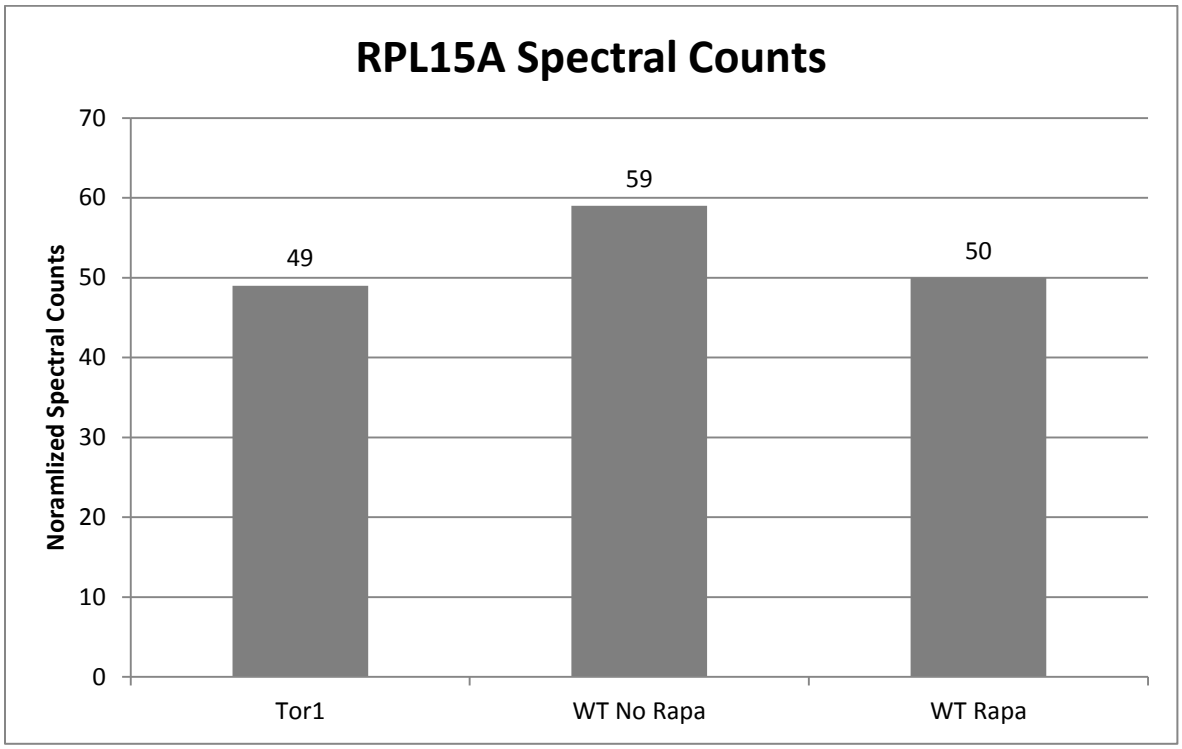

NPL3
DFDGTGALEFPSEEILVEALER + + +
1e8
Normalized Intensity
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Tor1NoRapa
WTNoRapa
WTRapa

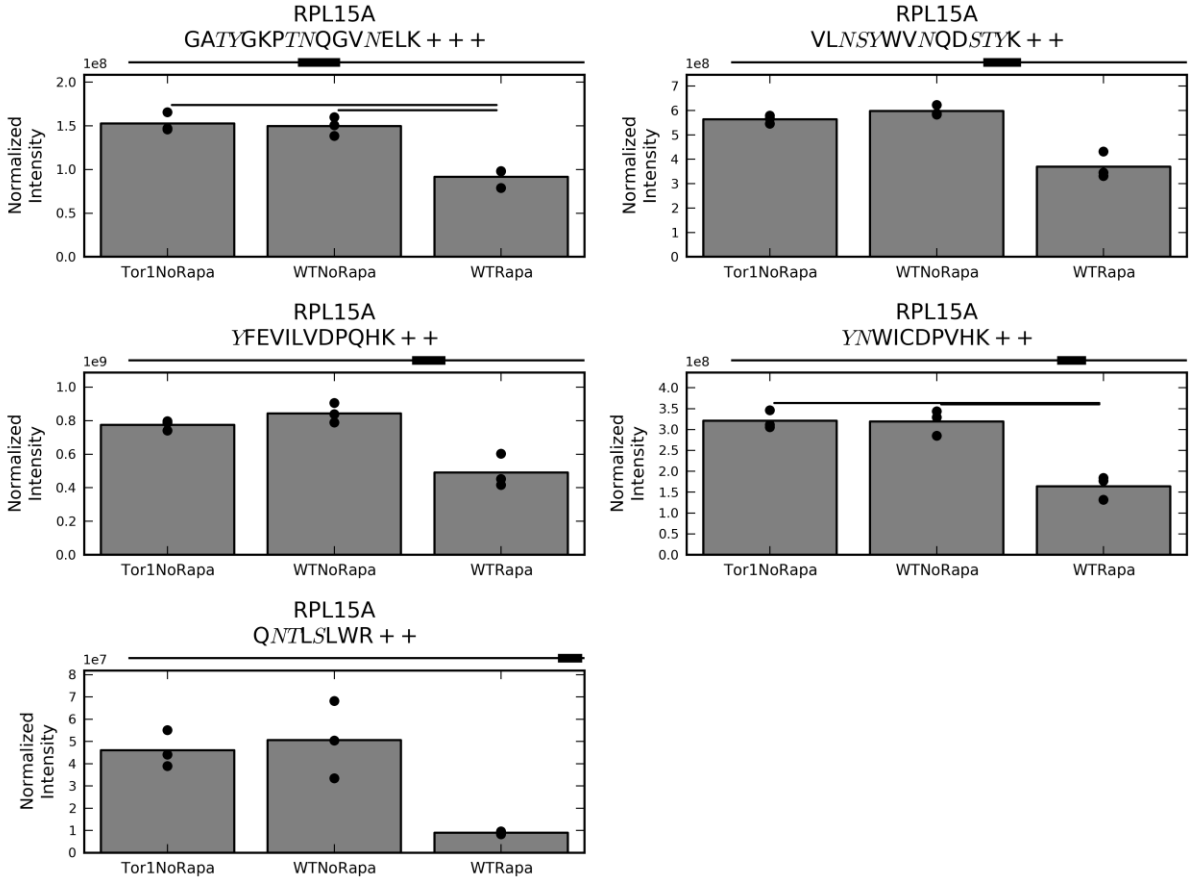

RIM8 Spectral Counts
Noramlized Spectral Counts
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
4
1
4
Tor1
WT No Rapa
WT Rapa

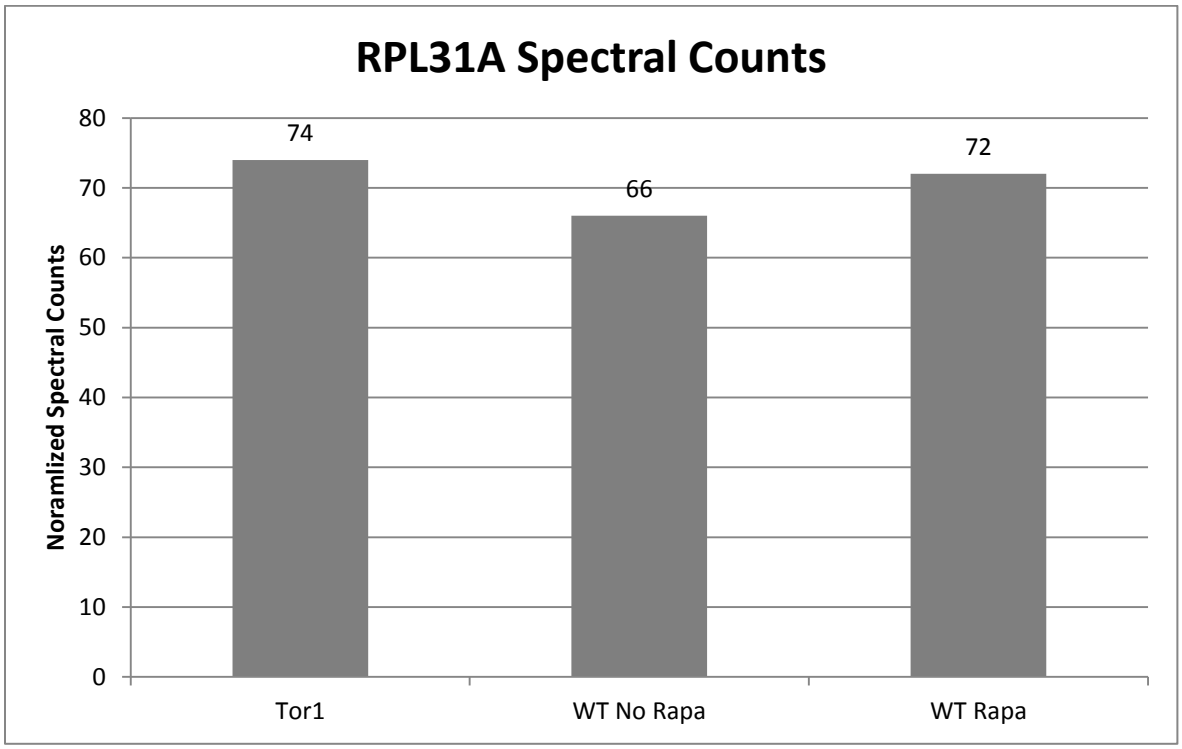

RIM8
EFAVIVPLDVSR + +
1e8
5
4
3
2
1
0
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

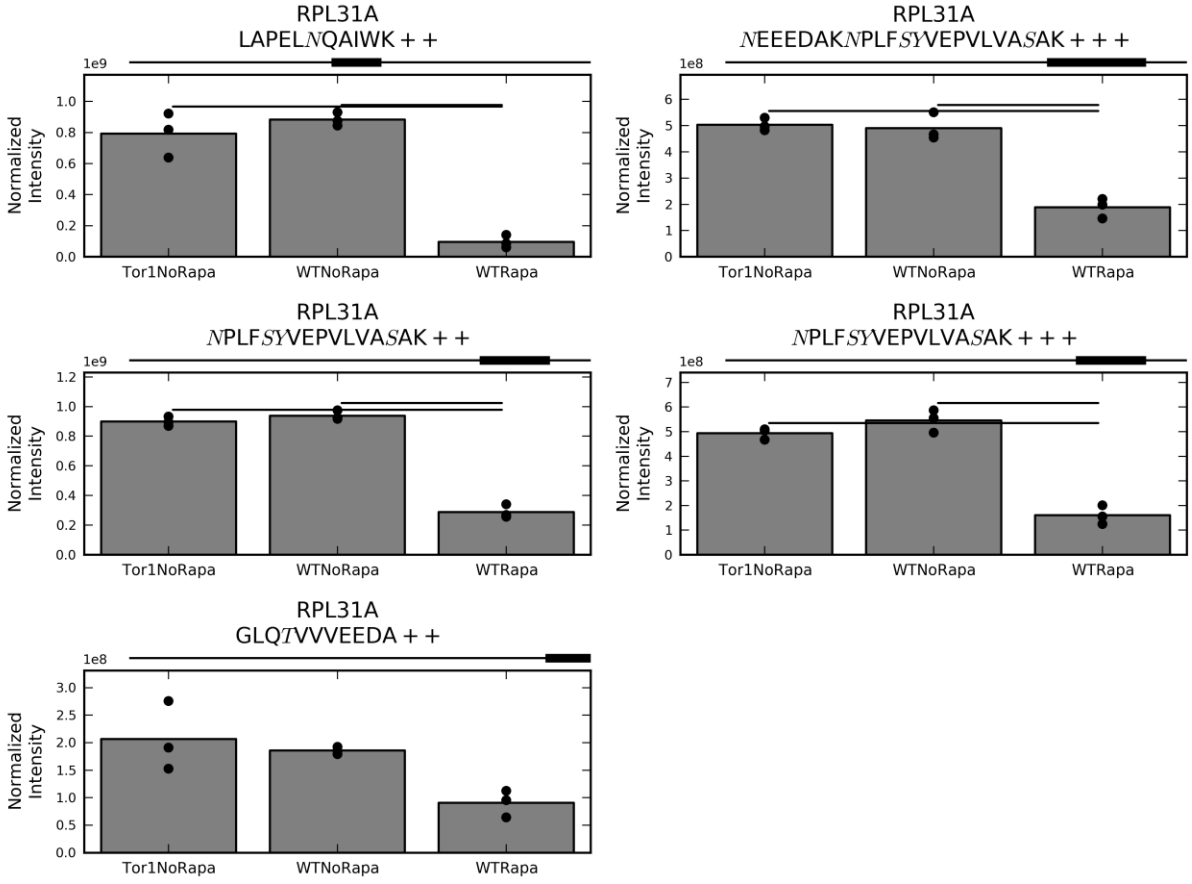

RPL14A Spectral Counts
Noramlized Spectral Counts
60
50
40
30
20
10
0
39
48
40
Tor1
WT No Rapa
WT Rapa

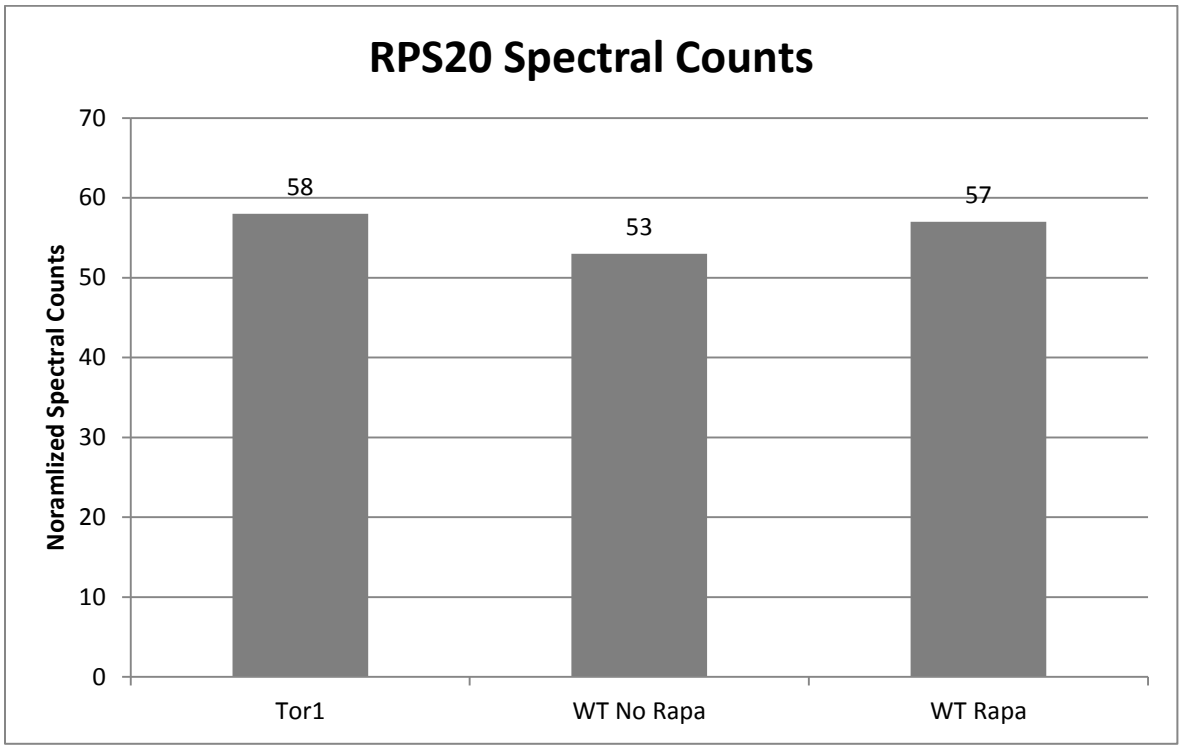

RPL14A
LAAIVEIIDQK + +
1e9
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
RPL14A
QAINLGQVVLTPLTFALPR + + +
1e6
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa
RPL14A
WAAAAVCEK + +
1e8
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

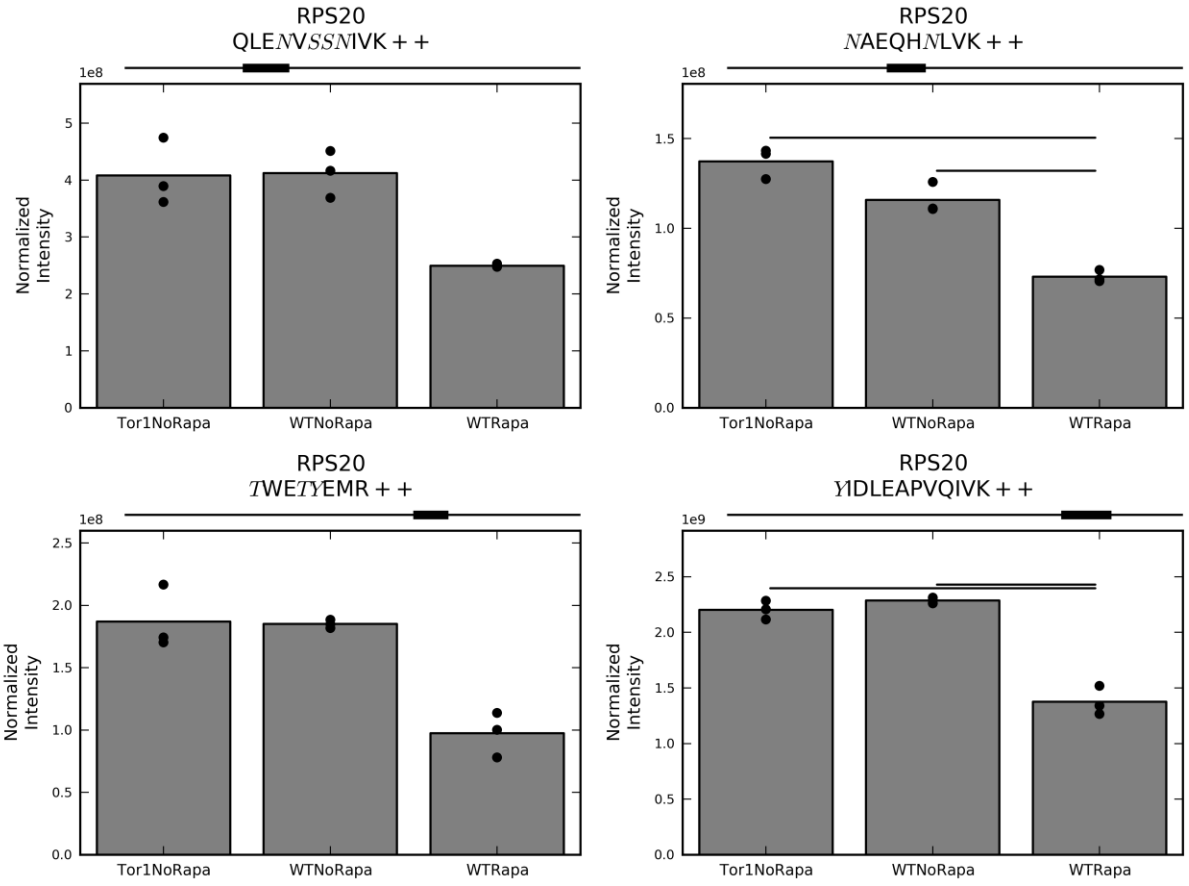

RPL15A Spectral Counts
Noramlized Spectral Counts
70
60
50
40
30
20
10
0
49
59
50
Tor1
WT No Rapa
WT Rapa

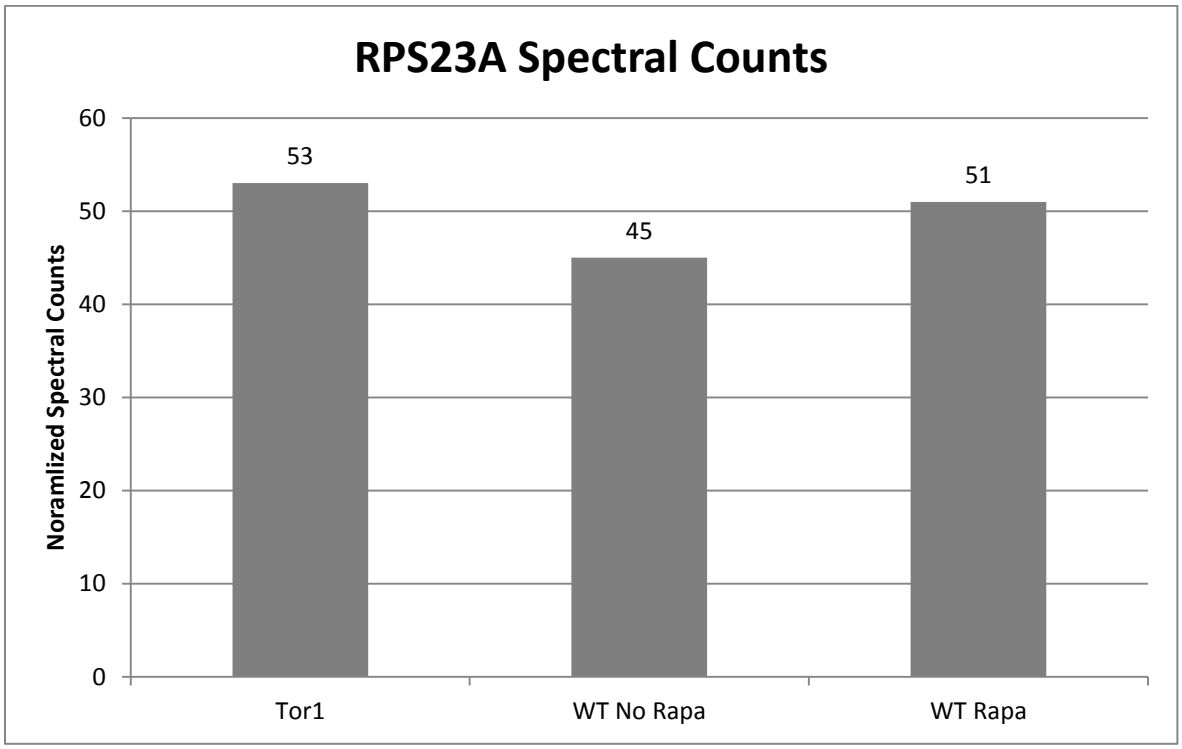

RPL15A
GA*TY*GKP*TN*QGV*N*ELK + + +
RPL15A
VL*NS*YWV*N*QD*STY*K + +
RPL15A
*Y*FEVILVDPQHK + +
RPL15A
*YN*WICDPVHK + +
RPL15A
Q*NT*L*S*LWR + +
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

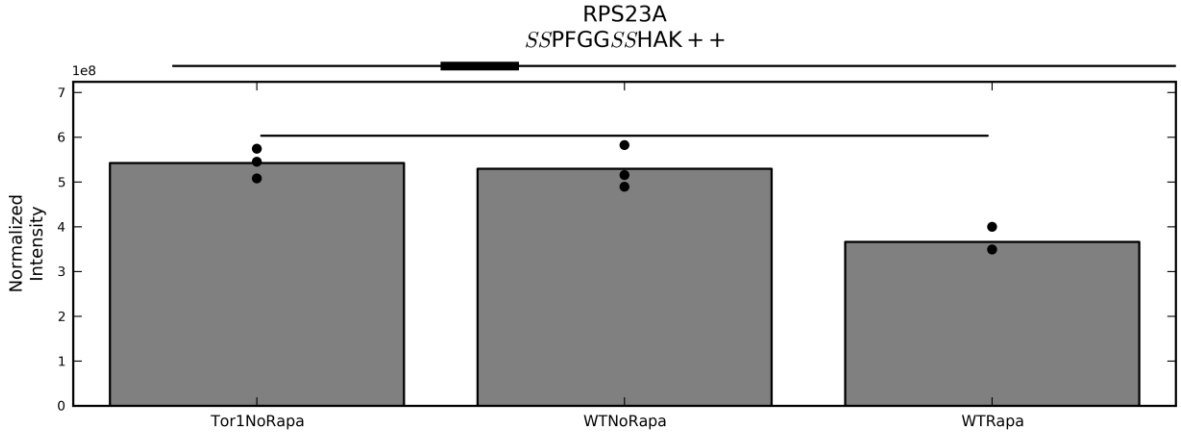

RPL31A Spectral Counts
Noramlized Spectral Counts
80
70
60
50
40
30
20
10
0
74
66
72
Tor1
WT No Rapa
WT Rapa

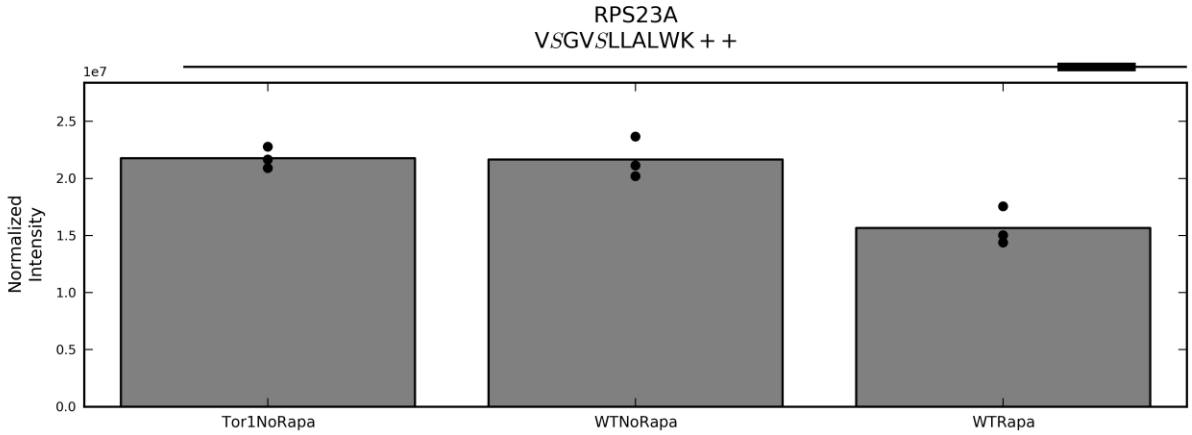

RPL31A
LAPELNQAIWK + +
RPL31A
NEEEDAKNPLFSYVEPVLVASAK + + +
RPL31A
NPLFSYVEPVLVASAK + +
RPL31A
NPLFSYVEPVLVASAK + + +
RPL31A
GLQTVVVEEDA + +
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

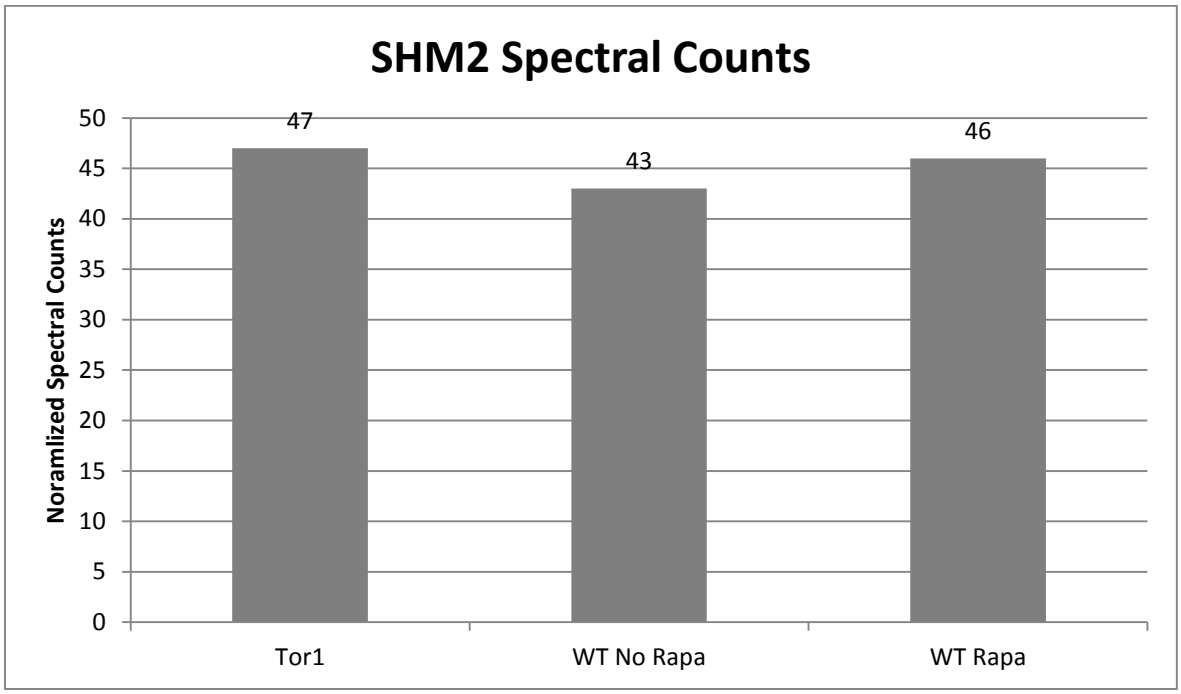

RPS20 Spectral Counts
Noramlized Spectral Counts
70
60
50
40
30
20
10
0
58
53
57
Tor1
WT No Rapa
WT Rapa

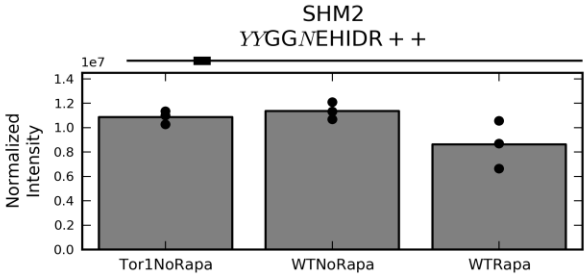

RPS20
QLENVSSNIVK + +
RPS20
NAEQHNLVK + +
RPS20
TWETYEMR + +
RPS20
YIDLEAPVQIVK + +
Normalized Intensity
1e8
1e9
Tor1NoRapa
WTNoRapa
WTRapa

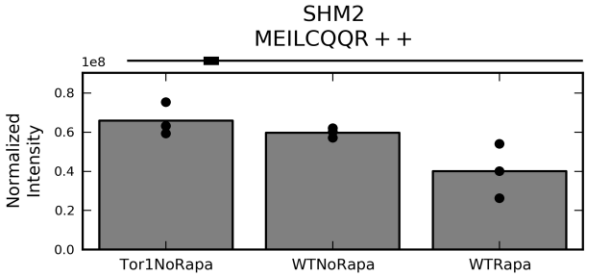

RPS23A Spectral Counts
Noramlized Spectral Counts
60
50
40
30
20
10
0
53
45
51
Tor1
WT No Rapa
WT Rapa

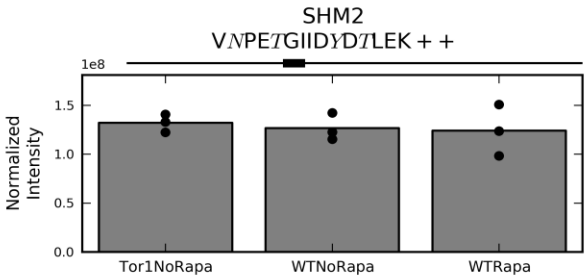

RPS23A
SSPFGGSSHAK + +
1e8
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

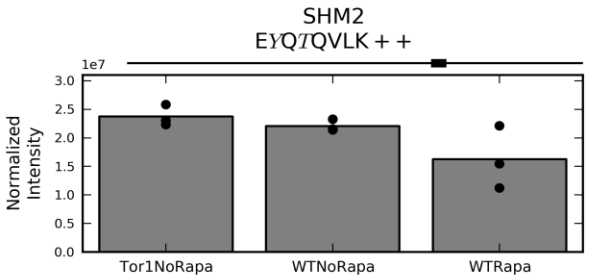

RPS23A
VSGVSLLALWK + +
1e7
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

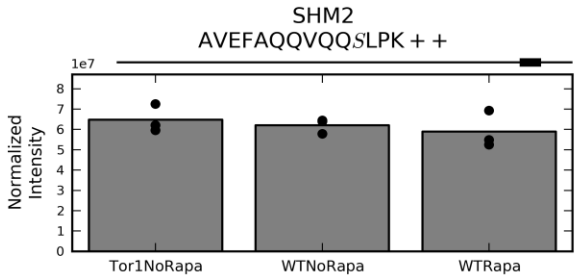

SHM2 Spectral Counts
Noramlized Spectral Counts
50
45
40
35
30
25
20
15
10
5
0
47
43
46
Tor1
WT No Rapa
WT Rapa

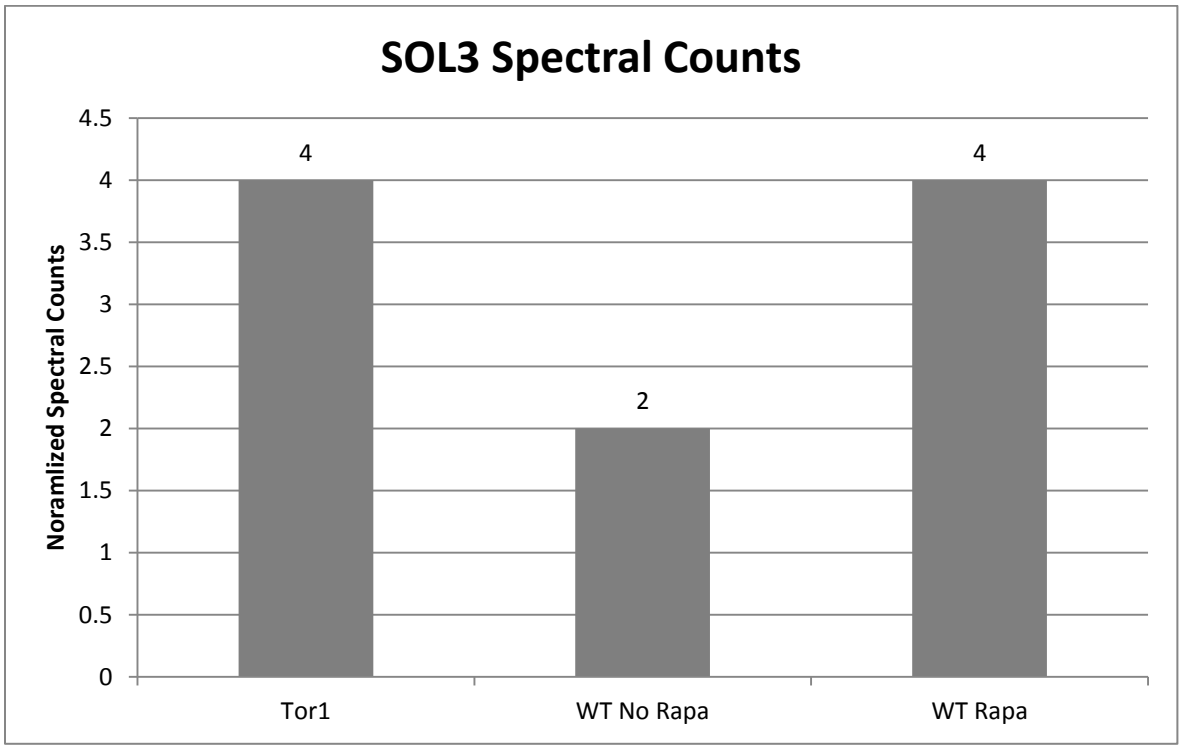

SHM2
YYGGNEHIDR + +
1e7
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa

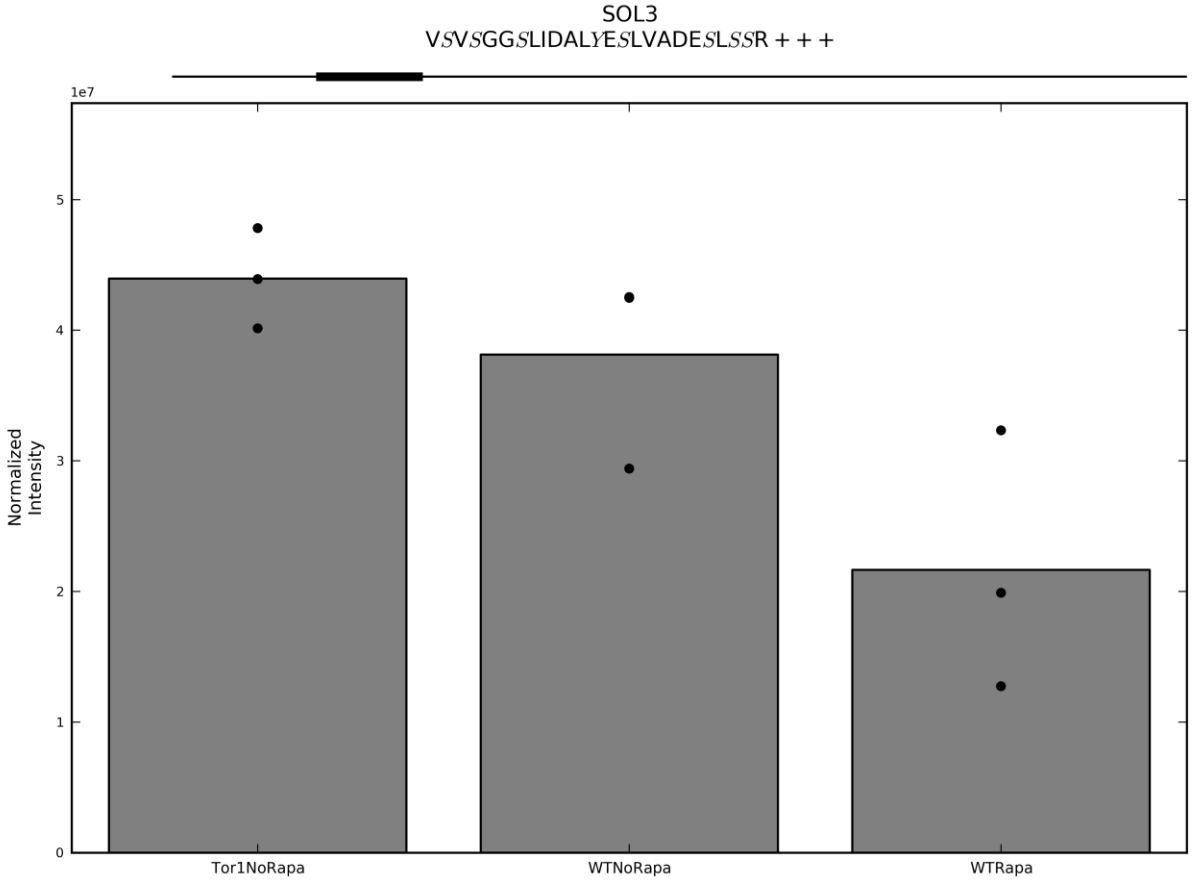

SHM2
MEILCQQR + +
1e8
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa


SHM2
VNPETGIIDYDTLEK + +
1e8
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa


SHM2
EYQTQVLK + +
1e7
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa


SHM2
AVEFAQQVQQSLPK + +
1e7
Normalized Intensity
Tor1NoRapa
WTNoRapa
WTRapa


SOL3 Spectral Counts
Noramlized Spectral Counts
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
4
2
4
Tor1
WT No Rapa
WT Rapa


SOL3
VSVSGGSLIDALYESLVADESLSSR + + +
1e7
Normalized Intensity
5
4
3
2
1
0
Tor1NoRapa
WTNoRapa
WTRapa

Vita

Jarrett Egertson was born in San Jose, CA and raised in Los Angeles, CA and San Diego, CA where he attended Torrey Pines High School. While in high school, Jarrett worked as intern mentored by Dr. Jeff Crosby at Isis Pharmaceuticals in Carlsbad, CA. After graduating from Torrey Pines in 2004, Jarrett attended the University of California, Los Angeles where he earned a Bachelor of Science degree in Molecular, Cellular, and Developmental Biology (class of 2008). While attending UCLA, Jarrett worked in the Matteo Pellegrini lab in the Molecular, Cellular, and Developmental Biology department as an undergraduate researcher. He also worked as a research lab assistant at the Spielberg Family Center for Applied Proteomics under the mentorship of Dr. Parag Mallick during this time. After graduating from UCLA, Jarrett went directly to the University of Washington where he earned a Ph.D. in Genome Sciences working in the lab of Dr. Michael J. MacCoss in 2013.

List of References

1. Pauling, L., Itano, H.A. & et al. *Science* **110**, 543-548 (1949).
2. Ingram, V.M. *Nature* **178**, 792-794 (1956).
3. Ingram, V.M. *Nature* **180**, 326-328 (1957).
4. Wasinger, V.C. *et al. Electrophoresis* **16**, 1090-1094 (1995).
5. Anderson, N.L. & Anderson, N.G. *Molecular & cellular proteomics : MCP* **1**, 845-867 (2002).
6. Vogel, C. & Marcotte, E.M. *Nature reviews. Genetics* **13**, 227-232 (2012).
7. Li, T., Diner, B.A., Chen, J. & Cristea, I.M. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 10558-10563 (2012).
8. Chen, E.I., McClatchy, D., Park, S.K. & Yates, J.R., 3rd *Analytical chemistry* **80**, 8694-8701 (2008).
9. Graumann, J. *et al. Molecular & cellular proteomics : MCP* **7**, 672-683 (2008).
10. Rodriguez, J., Gupta, N., Smith, R.D. & Pevzner, P.A. *Journal of proteome research* **7**, 300-305 (2008).
11. Eng, J.K., McCormack, A.L. & Yates Iii, J.R. *Journal of the American Society for Mass Spectrometry* **5**, 976-989 (1994).
12. Kall, L., Storey, J.D., MacCoss, M.J. & Noble, W.S. *Journal of proteome research* **7**, 29-34 (2008).
13. Storey, J.D. & Tibshirani, R. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 9440-9445 (2003).
14. Picotti, P. & Aebersold, R. *Nature methods* **9**, 555-566 (2012).
15. Picotti, P., Bodenmiller, B., Mueller, L.N., Domon, B. & Aebersold, R. *Cell* **138**, 795-806 (2009).
16. Liu, H., Sadygov, R.G. & Yates, J.R., 3rd *Analytical chemistry* **76**, 4193-4201 (2004).
17. Purvine, S., Eppel, J.T., Yi, E.C. & Goodlett, D.R. *Proteomics* **3**, 847-850 (2003).
18. Venable, J.D., Dong, M.Q., Wohlschlegel, J., Dillin, A. & Yates, J.R. *Nature methods* **1**, 39-45 (2004).
19. MacLafferty, F.W. *Tandem mass spectrometry*. (J. Wiley, New York ; Chichester etc.; 1983).
20. Carvalho, P.C. *et al. Bioinformatics* **26**, 847-848 (2010).
21. Wang, J., Bourne, P.E. & Bandeira, N. *Molecular & cellular proteomics : MCP* **10**, M111 010017 (2011).
22. Wang, J., Perez-Santiago, J., Katz, J.E., Mallick, P. & Bandeira, N. *Molecular & cellular proteomics : MCP* **9**, 1476-1485 (2010).
23. Ledvina, A.R. *et al. Analytical chemistry* **83**, 7651-7656 (2011).
24. Spivak, M., Weston, J., Bottou, L., Kall, L. & Noble, W.S. *Journal of proteome research* **8**, 3737-3745 (2009).
25. Kall, L., Canterbury, J.D., Weston, J., Noble, W.S. & MacCoss, M.J. *Nature methods* **4**, 923-925 (2007).
26. Bern, M., Cai, Y. & Goldberg, D. *Analytical chemistry* **79**, 1393-1400 (2007).
27. Bern, M. *et al. Analytical chemistry* **82**, 833-841 (2010).
28. Geromanos, S.J. *et al. Proteomics* **9**, 1683-1695 (2009).

29. Weisbrod, C.R., Eng, J.K., Hoopmann, M.R., Baker, T. & Bruce, J.E. *Journal of proteome research* **11**, 1621-1632 (2012).
30. Gillet, L.C. *et al. Molecular & cellular proteomics : MCP* **11**, O111 016717 (2012).
31. Myung, S. *et al. Analytical chemistry* **75**, 5137-5145 (2003).
32. Beavis, R.C. & Chait, B.T. *Analytical chemistry* **62**, 1836-1840 (1990).
33. Strittmatter, E.F., Ferguson, P.L., Tang, K. & Smith, R.D. *Journal of the American Society for Mass Spectrometry* **14**, 980-991 (2003).
34. Henry, K.D., Quinn, J.P. & McLafferty, F.W. *Journal of the American Chemical Society* **113**, 5447-5449 (1991).
35. Hannis, J.C. & Muddiman, D.C. *Journal of the American Society for Mass Spectrometry* **11**, 876-883 (2000).
36. Olsen, J.V. *et al. Molecular & cellular proteomics : MCP* **4**, 2010-2021 (2005).
37. Syka, J.E. *et al. Journal of proteome research* **3**, 621-626 (2004).
38. Strittmatter, E.F., Rodriguez, N. & Smith, R.D. *Analytical chemistry* **75**, 460-468 (2003).
39. Tolmachev, A.V. *et al. Analytical chemistry* **78**, 8374-8385 (2006).
40. Lasonder, E. *et al. Nature* **419**, 537-542 (2002).
41. Mortensen, P. *et al. Journal of proteome research* **9**, 393-403 (2010).
42. Petyuk, V.A. *et al. Molecular & cellular proteomics : MCP* **9**, 486-496 (2010).
43. Petyuk, V.A. *et al. Analytical chemistry* **80**, 693-706 (2008).
44. Kil, Y.J., Becker, C., Sandoval, W., Goldberg, D. & Bern, M. *Analytical chemistry* **83**, 5259-5267 (2011).
45. Haas, W. *et al. Molecular & cellular proteomics : MCP* **5**, 1326-1337 (2006).
46. Bruce, J.E., Anderson, G.A., Brands, M.D., Pasa-Tolic, L. & Smith, R.D. *Journal of the American Society for Mass Spectrometry* **11**, 416-421 (2000).
47. Kaiser, N.K., Anderson, G.A. & Bruce, J.E. *Journal of the American Society for Mass Spectrometry* **16**, 463-470 (2005).
48. Cox, J. & Mann, M. *Nature biotechnology* **26**, 1367-1372 (2008).
49. Wu, S. *et al. Journal of proteome research* **4**, 1434-1441 (2005).
50. Gras, R. *et al. Electrophoresis* **20**, 3535-3550 (1999).
51. Wool, A. & Smilansky, Z. *Proteomics* **2**, 1365-1373 (2002).
52. Wolski, W.E. *et al. Proteome science* **4**, 18 (2006).
53. Yan, B., Pan, C., Olman, V.N., Hettich, R.L. & Xu, Y. *Bioinformatics* **21**, 563-574 (2005).
54. Gentzel, M., Kocher, T., Ponnusamy, S. & Wilm, M. *Proteomics* **3**, 1597-1610 (2003).
55. Matthiesen, R. *et al. Proteomics* **4**, 2583-2593 (2004).
56. Nefedov, A.V., Mitra, I., Brasier, A.R. & Sadygov, R.G. *Journal of proteome research* **10**, 4150-4157 (2011).
57. Frewen, B.E., Merrihew, G.E., Wu, C.C., Noble, W.S. & MacCoss, M.J. *Analytical chemistry* **78**, 5678-5684 (2006).
58. Gay, S., Binz, P.A., Hochstrasser, D.F. & Appel, R.D. *Electrophoresis* **20**, 3527-3534 (1999).
59. Yates Iii, J.R., Eng, J.K., Clauser, K.R. & Burlingame, A.L. *Journal of the American Society for Mass Spectrometry* **7**, 1089-1098 (1996).
60. Mann, M. in Annual Conference on Mass Spectrometry and Allied Topics (American Society of Mass Spectrometry, Atlanta, GA; 1995).

61. Frahm, J.L., Howard, B.E., Heber, S. & Muddiman, D.C. *Journal of mass spectrometry : JMS* **41**, 281-288 (2006).
62. Demirev, P.A. & Zubarev, R.A. *Analytical chemistry* **69**, 2893-2900 (1997).
63. Hsieh, E.J., Hoopmann, M.R., MacLean, B. & MacCoss, M.J. *Journal of proteome research* **9**, 1138-1143 (2010).
64. Park, C.Y., Klammer, A.A., Kall, L., MacCoss, M.J. & Noble, W.S. *Journal of proteome research* **7**, 3022-3027 (2008).
65. Grossmann, J. *et al. Journal of proteome research* **4**, 1768-1774 (2005).
66. Horn, D.M., Zubarev, R.A. & McLafferty, F.W. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 10313-10317 (2000).
67. Stahl, D.C., Swiderek, K.M., Davis, M.T. & Lee, T.D. *Journal of the American Society for Mass Spectrometry* **7**, 532-540 (1996).
68. Michalski, A., Cox, J. & Mann, M. *Journal of proteome research* **10**, 1785-1793 (2011).
69. Wenner, B.R. & Lynn, B.C. *Journal of the American Society for Mass Spectrometry* **15**, 150-157 (2004).
70. Hoopmann, M.R., Finney, G.L. & MacCoss, M.J. *Analytical chemistry* **79**, 5620-5632 (2007).
71. Panchaud, A. *et al. Analytical chemistry* **81**, 6481-6488 (2009).
72. Williams, E.R., Loh, S.Y., McLafferty, F.W. & Cody, R.B. *Analytical chemistry* **62**, 698-703 (1990).
73. MacLean, B. *et al. Bioinformatics* **26**, 966-968 (2010).
74. Lawson, C.L., Hanson, R.J. & Society for Industrial and Applied Mathematics. in Classics in applied mathematics 15. 1 electronic text (xii, 337 p.) (Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104), Philadelphia, Pa.; 1995).
75. Stein, S.E. & Rudnick, P.A. NIST Peptide Tandem Mass Spectral Libraries. Yeast Peptide Mass Spectral Reference Data, *S. cerevisiae*, Official Build Date: May 24, 2011. *National Institute of Standards and Technology, Gaithersburg, MD, 20899*.
76. Yost, R.A. & Enke, C.G. *Analytical chemistry* **51**, 1251-1264 (1979).
77. Rose, M.R. *Evolutionary biology of aging*. (Oxford University Press, New York; 1991).
78. Kaeberlein, M. *F1000prime reports* **5**, 5 (2013).
79. Stanfel, M.N., Shamieh, L.S., Kaeberlein, M. & Kennedy, B.K. *Biochimica et biophysica acta* **1790**, 1067-1074 (2009).
80. Fontana, L., Partridge, L. & Longo, V.D. *Science* **328**, 321-326 (2010).
81. Trepanowski, J.F., Canale, R.E., Marshall, K.E., Kabir, M.M. & Bloomer, R.J. *Nutrition journal* **10**, 107 (2011).
82. Klass, M. & Hirsh, D. *Nature* **260**, 523-525 (1976).
83. Kenyon, C.J. *Nature* **464**, 504-512 (2010).
84. Johnson, S.C., Rabinovitch, P.S. & Kaeberlein, M. *Nature* **493**, 338-345 (2013).
85. Harrison, D.E. *et al. Nature* **460**, 392-395 (2009).
86. Wilkinson, J.E. *et al. Aging cell* **11**, 675-682 (2012).
87. Hardwick, J.S., Kuruvilla, F.G., Tong, J.K., Shamji, A.F. & Schreiber, S.L. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 14866-14870 (1999).
88. Kaeberlein, M. & Kennedy, B.K. *Aging cell* **10**, 185-190 (2011).

89. Steffen, K.K. *et al. Cell* **133**, 292-302 (2008).
90. Pearson, R.B. *et al. The EMBO journal* **14**, 5279-5287 (1995).
91. Beretta, L., Gingras, A.C., Svitkin, Y.V., Hall, M.N. & Sonenberg, N. *The EMBO journal* **15**, 658-664 (1996).
92. Winzeler, E.A. *et al. Science* **285**, 901-906 (1999).
93. Zhang, B., Chambers, M.C. & Tabb, D.L. *Journal of proteome research* **6**, 3549-3557 (2007).
94. Loewith, R. *et al. Molecular cell* **10**, 457-468 (2002).
95. Kapahi, P. *et al. Cell metabolism* **11**, 453-465 (2010).