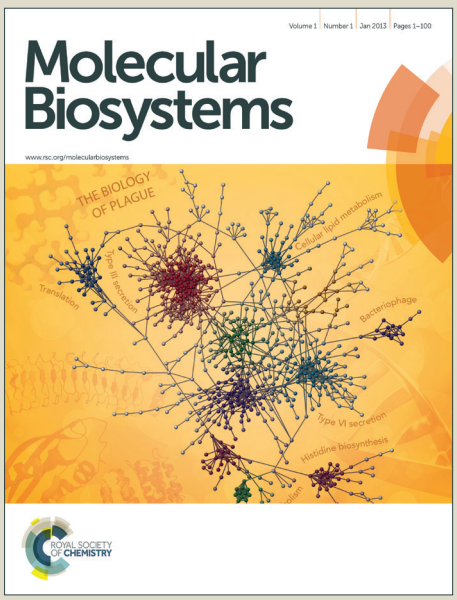
# Molecular BioSystems

Accepted Manuscript

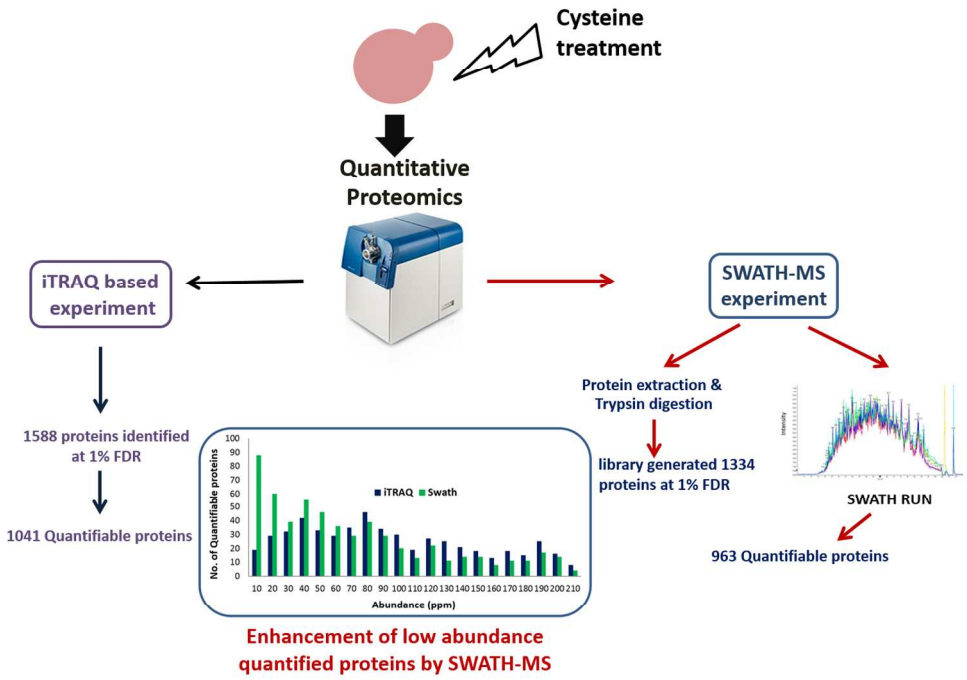







ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems




254x190mm (300 x 300 DPI)

# In-depth comparative proteomic analysis of yeast proteome using iTRAQ and SWATH based MS

**Trayambak Basak[1,2#], Ajay Bhat[1,2#], Dipankar Malakar[3], Manoj Pillai[3], Shantanu Sengupta[1,2*]**

[1] *Genomics and Molecular Medicine, CSIR-IGIB, Mathura Road, New Delhi*

[2] *Academy of Scientific & Innovative Research (AcSIR), CSIR-IGIB, Mathura Road Campus, New Delhi*

[3] *AB Sciex, 121 Udyog Vihar, Gurgaon, Haryana, India*

[#] *Equal contribution*

***Corresponding Author:***

Dr. Shantanu Sengupta

Genomics and Molecular Medicine Unit

CSIR-Institute of Genomics and Integrative Biology

Mathura Road, New Delhi-110020

Tel: +91 11 29879201

Email: shantanus@igib.res.in

Conflict of Interest: The authors declare no conflict of interest to this report.

**Abstract:**

Quantitative proteomics using LC-MS have emerged as an essential tool for addressing different biological questions. Various labelling methods have been effectively employed for quantitative proteomics studies. However, these are fraught with several challenges including reproducibility and number of samples that can be analysed at a given time. To this end, unlabelled proteomics holds a lot of promise and the recently developed sequential window acquisition of all theoretical fragment ion spectra (SWATH-MS) aims to address these limitations. In this study, we compared SWATH-MS to isobaric tag for relative and absolute quantitation (iTRAQ), a widely used labelled method for relative quantitation. For this, we used the yeast, *Saccharomyces cerevisae*, since almost all its proteins are identified. More importantly, the abundance of each protein is well documented. We found that although similar number of proteins could be quantitated using the two techniques, SWATH had the advantage of quantifying a larger percentage of low abundant proteins (below 60 ppm). Thus, based on our analysis, we believe that these two techniques are complementary and can synergistically improve the number of quantifiable proteins. SWATH's ability to quantify low abundant proteins could be an asset in biomarker discovery studies.

Key words: Proteomics, SWATH-MS, iTRAQ, Protein abundance

**Abbreviations:**

LC-MS- Liquid chromatography-mass spectrometry

SWATH- Sequential window acquisition of all theoretical fragment ion spectra

ITRAQ- Isobaric tag for relative and absolute quantitation

2D-DIGE- 2D-differential in gel electrophoresis

SILAC- Stable isotope labelling by amino acids in cell culture

iCAT- Isotope-coded affinity tags

TMT- Tandem mass tags

SRM- Selected reaction monitoring

YPD- Yeast peptone dextrose

DTT- Dithiothreitol

IAA- Iodoacetamide

SCX- Strong cation exchange

ACN- Acetonitrile

FA- Formic acid

IDA- Information dependent acquisition

TOF- Time of flight

FDR- False discovery rate

CV- Coefficient of variation

TIC- Total ion chromatogram

XIC- Extracted ion chromatogram

ppm- Parts per million

**Introduction:**

In the current decade, proteomics has emerged as a fascinating tool for biologists to understand the complementary function of the entire proteome at a given biological state [1]. Proteomics has enabled researchers to look and dig deeper into the proteome and characterize proteins at a global scale[1, 2]. The magnitude of expression and the activity of any protein or its interaction with other partners dictates specific functional pattern at cellular, organelle and organism level. Thus, one of the important facets of proteomics is its potential to identify proteins that are differentially expressed (using relative or absolute quantitation) between two or more states and is particularly important for identifying potential disease markers or to mechanistically understand the effect of stress or drugs or other toxicants at cellular levels. Classically differentially expressed proteins were identified using 2D-gel electrophoresis where spots were compared between gels and those with varying intensities were excised and identified using mass spectrometer. The 2D gel electrophoresis has come a long way through the evolution of 2D-differential in gel electrophoresis (DIGE) based methods which are more sensitive than the classical approach. However, even this method is fraught with several challenges like standardization, reproducibility, sensitivity, specificity and most importantly resolution[3, 4] which cannot not be enhanced after a certain point[3, 4]. The advancement of mass spectrometry coupled with liquid-chromatography has the potential to overcome such challenges and generate better quality data for quantitative purposes. Hence gel-free proteomics approach has gained huge popularity among researchers[5, 6]. Various chemically labelled or unlabelled strategies have been employed for relative and/or absolute quantitation of proteins to address different biological questions. Among the different labelling strategies used, isotopic peak based quantitation such as Stable isotope labelling by amino acids in cell culture (SILAC), Isotope-coded affinity tags (iCAT) or isobaric chemical tag based relative quantitation as employed in Isobaric tags for relative and absolute quantitation (iTRAQ)[7-9], Tandem mass tags (TMT) etc. have been extensively used[10]. Among these, SILAC has been extensively used for metabolic labelling in cell-culture based studies and iCAT in redox proteomics since it labels free reduced cysteine residues. In contrast, iTRAQ and TMT have been extensively used for almost all kinds of biological samples as these are chemical tags and coherently work with bottom-up proteomics workflows. In addition, multiple samples can be tagged in a single run (upto 8 in iTRAQ and 6-10 plex in TMT). Apart from these labelling strategies, various label-free approaches, based on spectral counting, intensity based normalization etc. have been used to quantify proteins. Recently, Aebersold's group has developed a quantitative method based on sequential window acquisition of all theoretical

fragment ion spectra (called SWATH) which can be used as label-free quantitative method for different proteomics experiments[11]. In this method, a fragment ion library is first generated via data dependent acquisition. This along with the retention time of the peaks acts as a reference for targeted extraction of fragment ions that are subsequently generated by sequential-windowed data independent acquisition. Thus, in SWATH, a combinatorial approach of data dependent acquisition with targeted extraction of fragment ions is used to quantify proteins. The sensitivity and specificity of this method is comparable to the classical SRM based workflow and have certain advantages like post acquisition data analysis.

In the present study, using the yeast, *Sacharomyces cerevisae*, as a model system, we evaluated the ability of SWATH to identify and quantitate proteins as compared to one of the popular labelled approach, iTRAQ. *Sacharromyces cerevisae* is the simplest eukaryote which has been most widely studied from genomic and proteomic perspective. Further, almost all the proteins in yeast have been characterized and their abundances are reported. Using these two approaches, we looked at the number of proteins that are differentially expressed in the presence and absence of the thiol amino acid cysteine. An elevated level of cysteine has been reported to be toxic to many organisms like bacteria and yeast[12], [13-15]. In recent times, it is shown to be associated with several complex diseases like cardiovascular disease, obesity, etc, [16-18]. Our results indicate that amongst the high-confidence quantifiable proteins identified using both the methods, about 60% were common between them, indicating high concordance between the two methods. However, amongst the proteins that were unique to each method, we found that SWATH could identify a larger number of low abundant proteins as compared to iTRAQ.

## Materials and Methods

### *2.1 Materials*

The ingredients for preparing yeast media including yeast extract, peptone, dextrose and the amino acids were purchased from HiMedia (India). L-Cysteine-Hydrochloride, DTT (dithiothreitol), IAA (Iodoacetamide), ammonium formate and formic acid were procured from Sigma (St. Louis, MO, U.S.A.). Modified trypsin (sequencing grade, V511) was procured from Promega. The iTRAQ reagents, polysulfoethyl SCX cartridge, nano LC column were procured from AB Sciex (USA). The nanospray picotip was purchased from New Objective (USA). LC-MS grade water and acetonitrile were procured from J.T. Baker (USA). All other chemicals used were of analytical grade.

*2.2 Yeast strains, media and growth conditions*

The wild-type *S. cerevisiae* strain used in this present study, BY4741 (MATa his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0), was obtained from American Type culture collection (ATCC). YPD medium (Rich media) for culturing yeast was prepared using 1% (w/v) yeast extract, 2% (w/v) peptone and 2% (w/v) dextrose and for solid media, 2% (w/v) agar was added to YPD liquid media. The synthetic complete media contained 2% (w/v) glucose, 0.17% yeast nitrogen base without amino acids, 0.5% $NH_4Cl$ supplemented with the following amino acids:; L-arginine (HCl), 20µg/ml;L-aspartic acid, 100µg/ml; L-glutamic acid (monosodium salt), 100µg/ml; L-histidine, 20µg/ml; L-leucine, 60µg/ml; L-lysine (mono-HCl), 30µg/ml; L-phenylalanine, 50µg/ml; Lthreonine,200µg/ml; L-tryptophan, 40µg/ml; L-tyrosine, 30µg/ml; L-valine, 150µg/ml; along with adenine, 40µg/ml uracil, 20µg/ml.

*2.3 Yeast growth, Protein isolation and quantitation*

Pre-cultures of yeast strain (BY4741) were prepared by growing 3 ml yeast culture in YPD (rich media) overnight in an incubator shaker at 30°C and 200 rpm. These cultures were washed three times with sterile water and then re-inoculated in synthetic minimal media to an OD 600 of 0.1. To study the effect of exogenously added cysteine, 1mM of cysteine (from freshly prepared stock) was added and cells were grown at 30°C and 200 rpm for 12 hrs. After 12 hrs of growth, untreated & cysteine treated cells were harvested by centrifugation and washed three times with sterile water and put into lysis buffer containing 140mM NaCl, 1.5mM $MgCl_2$, 10mM Tris-HCl (pH 8), 0.5% NP40 and Protease inhibitor cocktail (Roche diagnostics) with equal volume of acid washed glass beads and cells were lysed by 10 rounds of vortex-mixing followed by cooling on ice for 15 second each. The suspensions were then collected in a fresh tube and centrifuged at 13000 rpm at 4°C for 10 minutes and the supernatant was collected into new tubes. Proteins were estimated using Bradford's reagent (Sigma, USA).

*2.4 Trypsin digestion and iTRAQ labelling*

To identify proteins that are differentially expressed in yeast in the presence and absence of cysteine using iTRAQ based method, we performed a total of three, 4-plex iTRAQ experiments. In each of the replicate experiments, one technical replicate was also included. Thus, each experiment consisted of one biological replicate with its technical replicate.

Sixty micrograms of protein from each sample were reduced with 25mM DTT for 30 minutes at 56°C and the cysteines were blocked by 55mM IAA at room temperature for 15-20 minutes. The samples were then incubated with modified trypsin (Promega, V511) for 16-18 hrs at 37°C in a 1:10 ratio (trypsin to protein) [19]. The tryptic peptides of each sample were labelled with different iTRAQ reagents (114, 115, 116, 117) following the manufacturer's protocol (AB Sciex, Foster City, CA) as shown in supplementary figure 3. For each iTRAQ experiment, all the four individual tagged samples were mixed and pooled into a single centrifuge tube and subjected to vacuum drying at 30°C using vacuum concentrator (Eppendorf, USA).

*2.5 2D nano LC-MS/MS iTRAQ analysis*

The first dimensional separation of the iTRAQ labelled peptides was achieved by cation exchange (SCX) chromatography using a SCX Cartridge (5 micron, 300 Å bead from AB Sciex, USA), using a cartridge holder (AB Sciex, USA). The pooled and dried peptides were reconstituted in 1 mL Buffer A (8 mM ammonium formate, 30 % (v/v) ACN and 0.1% FA pH= 2.9) and applied to the cartridge using a hand syringe set up. The samples were fractionated using a step gradient of increasing concentration of ammonium formate (30 mM, 50 mM, 70 mM, 100 mM, 125 mM, 150 mM, and 250 mM ammonium formate, 30% v/v ACN and 0.1% formic acid; pH= 2.9). The iTRAQ labelled fractions obtained after SCX chromatography were analyzed on a TripleTOF5600 (AB Sciex, USA) MS coupled to an Eksigent NanoLC-Ultra 2D plus system. Ten microliters of samples were loaded onto a reverse phase peptide ChromoLC trap (200 µm × 0.5 mm) column and desalted at a flow rate of 3 µL per minute for 40 minutes. After desalting, peptides were separated using a Eksigent C18 column (75 µm × 15 cm). Peptides were eluted from the column at a flow rate of 300 nL/min using a linear gradient of 2 – 35% mobile phase B in 99 minutes (mobile phase A is 0.1% formic acid and mobile phase B is 100% acetonitrile and 0.1% formic acid). The LC eluent was analyzed using a NanoSpray III Source installed on the TripleTOF 5600 system. Samples were analyzed using a nebulizing gas of 5; a curtain gas of 25; an Ionspray voltage of 2500 V and heater interface temperature of 130°C. The TripleTOF 5600 system was operated in an information dependent acquisition (IDA) mode with a TOF/MS survey scan (350-1250 m/z) with an accumulation time of 250 msec. A maximum of ten precursor ions per cycle were selected for fragmentation and each MS/MS spectrum was accumulated for 100 msec (100–1500 m/z) with a total cycle time of approximately 2.3 seconds. Only the parent ions with a charge state from +2 to +5 were included in the MS/MS fragmentation. The threshold precursor ion intensity was set as more than 120 cps and were not present on

the dynamic exclusion list. Once an ion had been fragmented by MS/MS, its mass and isotopes were excluded for a period of 10 seconds. The MS/MS spectra were acquired in high sensitivity mode with 'adjust collision energy when using iTRAQ reagent' settings.

*2.6 SWATH analysis*

The above mentioned LC-MS platform was used for this analysis. For standard data dependent acquisition to generate an ion library, 200 µg yeast lysates were digested using trypsin and six different fractions were obtained using the SCX cartridge separation as mentioned earlier. 10 µl of each of these fractions were loaded on to the trap column at a flow rate of 3 µl/min for 30 minutes and eluted from the analytical column at a flow rate of 250 nl/min with the following gradient: solvent B (100% ACN, 0.1% FA) was increased from 5 to 10% in first 15 minutes. Then buffer B was ramped upto 30% for the next 65 minutes. In the next 20 minutes, %B was increased up-to 50% and reached 90% within a minute and kept for another 7 minutes for column washing. Finally the column was re-equilibrated by solvent A (100% water, 0.1% FA) for 13 minutes. For ion library generation, typical data-dependent analysis was performed and the mass spectrometer was operated in a manner where a TOF-MS survey scan was performed from 350-1250 m/z mass range with an accumulation time of 250 milliseconds, from which the 25 most abundant ions were selected for subsequent MS/MS fragmentation with an accumulation time of 40 milliseconds. The source ionization parameter and other settings were similar as mentioned earlier. For this experiment, ions were isolated in quadrupole with a unit resolution (0.7 Da) and rolling collision energy with a spread of 5V. The total cycle time was set at 1.3 seconds. For SWATH MS-based experiments, the instrument was specifically operated in a "create swath" mode where it was set to allow a quadrupole resolution of 25 Da/mass selections. Using an isolation width of 26 Da (25 Da of optimal ion transmission efficiency 1 Da for the window overlap), a set of 36 overlapping windows were constructed covering the mass range 350–1250 Da. The collision energy for each window was determined based on the appropriate collision energy set automatically with a spread of 5 eV. The total duty cycle was of 3.3 seconds (total 3.2 seconds for stepping through the 36 isolation windows - 0.1 seconds for the optional survey scan). The MS/MS acquisition was done using high-sensitivity mode corresponding to the mass resolution of about 15,000 which also enables to extract fragment ions with 10-50 ppm accuracy.

*2.7 Database searching & Statistical analysis:*

For the ion library generation and iTRAQ experiment, all the .wiff files containing MS and MS/MS spectra generated from Triple TOF 5600 (AB Sciex) were submitted for database searching and quantitative analysis using the Protein Pliot v4.0 software (AB Sciex). For the identification of proteins, Paragon Algorithm was employed in a "Thorough ID" search mode against the Uniprot-Saccharomyces cerevisiae reference dataset (6643 protein sequences). The search parameters allowed modifications by IAA as cysteine blocking reagent. For iTRAQ, 4-plex peptide iTRAQ labelling of the N termini of peptides, and of the side chains of lysine were also included in the search. We applied 1% global protein level FDR for the identification of proteins. For SWATH analysis, the peaks were extracted using the PeakView (version 1.2, AB Sciex) software using group files generated from Protein Pilot (v4.0). All .group files were extracted using a MS tolerance of 25 ppm and MS/MS tolerance of 50 ppm and the following parameters were considered: 2 peptides with 6 transitions, exclude shared peptides, peptide confidence of >99%. These processed .mrkvw files from PeakView were then loaded onto MarkerView (Version 1.2.1, AB Sciex) for relative quantitative analysis. This export resulted in generation of three files containing quantitative information about individual ions, summed intensity of different ions for a particular peptide and summed intensity of different peptides for a particular protein. In-built total ion intensity sum plugin was used for normalization of the quantitative protein data for all runs. MarkerView was used for SWATH analysis because of its data-independent method of quantitation whereas protein Pilot was used for iTRAQ based quantitation in a data dependent mode. Protein abundance analysis was performed by using the data available from PaxDb (Protein abundance across organisms database) server[20].

**Results**:

Label free LC-MS based quantitative profiling heavily relies on the reproducibility of the liquid chromatography and variation of intensity across different replicates. Gillet et.al had previously reported that spiking different peptide standards at a constant 47 femtomol resulted in a coefficient of variation (CV) of about 13.7% in SWATH analysis[11]. We spiked 25 femtomol of beta-galactosidase digest with the samples. After data acquisition and extraction of peak area of five fragment ions (1176.55, 1289.63, 1061.52, 832.45, and 563.27) of a beta- galactosidase peptide (APLDNDIGVSEATR), we found that the CV was less than 10% within the replicates (Table1, Figure 1). To further check the robustness of the method, we performed a correlation analysis among the three technical replicates for both control and cysteine treated samples. In all the cases, we got a high correlation ($r > 0.92$)

(Figure 2). Further, we obtained highly reproducible total-ion chromatogram (TIC) across all the replicates of yeast treated with or without cysteine (Supplementary figure 1). Moreover, about 80% of the ions, peptides, proteins had a CV less than 20% among the replicates (Figure 3a). We also show using a cumulative frequency plot, that a considerable proportion of low intensity fragment ions had CV less than 20%, along with relatively higher to moderate intensity fragment ions (Figure 3b). To determine the number of proteins that were differentially expressed in yeast in the presence of cysteine, we used both SWATH and iTRAQ based methods. For SWATH analysis, we considered a protein to be quantifiable if it was identified with at least 2 unique peptides (≥95% confidence) with a maximum of 6 transitions. Further, the extracted ion chromatograms (XIC's) were manually curated considering three different parameters: i) removal of inconsistency in the retention time for any fragment ion across different runs, ii) removal of nonspecific XIC ion interference for any specific peptide quantitation and iii) comparison of XICs ion intensity ratio with ion intensity ratios obtained from fragment ion library for any specific peptide (Supplementary figure 2A-C). A representative XICs for a particular peptide which has been included in the study is shown in supplementary figure 2D. Using these stringent criteria, we were able to quantitate 963 proteins in the SWATH analysis. This was about 72.2% of the proteins that were identified (1334) at 1% protein global FDR from IDA library obtained using data dependent acquisition. When the same biological samples were subjected to iTRAQ based relative quantitative analysis, a total of 1041 proteins could be quantified with a minimum of two unique peptides (>95% confidence) which was about 65.6% of the proteins identified (1588) at 1% protein global FDR .Among the quantifiable proteins, we found 635 proteins to be common between the two methods (Figure 4A) with significant correlation ($r=0.63$, $p= 2.7$ E-71) (Figure 4B). Further, integrating both the techniques led to the identification of a total of 1339 quantifiable proteins (considering the quantifiable proteins that are unique and common between the methods).

One of the challenges in mass spectrometry based proteomics of complex mixtures is the dynamic range and the ability to identify and quantitate low abundant proteins[21-23]. We show that in SWATH, the intensity range of the quantifiable fragment ions ranged from 1,000 cps to more than 10,000,000 cps, i.e. a dynamic range of 4 (Figure 5), which is consistent with earlier reports [11]. Further, an on-column loading of only 100 ng of protein in SWATH, led to the quantitation of high abundance protein like ENO2 (24,556 ppm) along with very low abundant protein like TOM5 (0.269 ppm) (Figure 6). Interestingly, although iTRAQ could quantitate proteins like ZPS1 having an abundance of 0.758 ppm, the number of quantifiable low abundant proteins was much higher in SWATH (Figure 6). To confirm this, we

considered proteins that were uniquely quantified by the two methods and mapped the abundance level of these proteins using PAXDB. Binning these proteins based on their abundance (in 100 ppm window), we found that about 441 proteins that were quantifiable in SWATH had an abundance level of 0-100 ppm as compared to 329of such protein in iTRAQ (Figure 7). Further subdividing this 100 ppm bin into 10 ppm bins, we found that 67.8% of quantifiable proteins were below 60 ppm in SWATH. This clearly indicates that SWATH is capable of quantifying proteins of low abundance. One of the reasons for the inability of iTRAQ to consistently quantify low abundant proteins could be due to the less number of peptides obtained in these proteins as compared to the more abundant proteins and hence are not quantifiable consistently in replicate experiments. To substantiate this fact, we performed an analysis of iTRAQ data, where we considered all the proteins with abundance below 50 ppm (low abundant protein) and above 500 ppm (high abundant protein) in each replicate. We found that about 25% of the low abundant proteins (<50 ppm abundance) were not quantitated in all three replicates and hence were not included in the list of differentially expressed proteins. In contrast, about 95% proteins of highly abundant proteins (> 500 ppm) could be quantitated in all three replicates (Figure 8).

**Discussion***:*

Quantitative proteomic profiling experiments have boosted the biological research area rapidly and is fast becoming a method of choice to identify potential disease markers or understanding disease mechanisms. There are several approaches (gel based, gel-free labelled and unlabelled) that are routinely used in quantitative proteomics. However, choosing an appropriate experimental methodology depends on the parameters like time, expense, expertise, accuracy, reproducibility and most importantly, the question to be addressed. Among all these approaches, label-free quantitative analysis appears to be an attractive choice due to the reduced number of steps involved and high coverage with a data-independent approach for every sample. However, one of the greatest challenges of this technique is the inter-sample reproducibility, since methods based on spectral counting or intensity based quantitation in MS mode are often not reproducible[24, 25]. This necessitates the development of newer methods for label-free quantitation that can address the above issues. To this end, the method of sequential windowed acquisition of all theoretical masses (SWATH), developed by Gillet *et. al*., holds immense potential as the quantitation is based on different XIC's of fragment ions of a particular peptide eluting at a given time[26, 27]. It has been reported that the fragment ion interference for the SWATH MS based method was comparable to SRM based precursor and fragment ion isolation (0.7 Da & 0.7 Da)

respectively[11]. However, one of the limitations of this method is that it can only be performed in instruments with very high scan speed and high resolution. For any method to become acceptable for routine use, apart from the advantages of time and cost, it should also yield results that are better or even comparable to the other contemporary methods. However, till date, there are no reports assessing SWATH with other conventional label based methods. Thus, in this study we assessed the performance of SWATH with iTRAQ, in terms of the number of proteins that the two methods could unambiguously identify and quantitate. iTRAQ remains as one of the most widely used chemical labelling approach for quantitative proteomics workflow. It has several advantages; most importantly, the multiplexing ability to compare 4-8 samples in a single experiment. Our choice of this method was also based on the fact that both SWATH and iTRAQ are proprietary of the same manufacturer and hence would minimize the bias. For iTRAQ based quantitation, Protein Pilot itself has been used as it is a data-dependent quantitation. In case of iTRAQ, for a particular peptide, different reporter ion intensities were quantitated and relative fold change was calculated. However, in the case of SWATH based quantitation, MarkerView extracts the ion chromatogram of a single peptide across different chromatographic runs and a fold change is thus calculated. Here, for a peptide quantitation, XICs of several b and y ions are considered.

Our results clearly show that SWATH as a method is highly reproducible with CV less than 10% and high correlation between technical replicates ($r>0.92$) which is in agreement with previous studies. Further, we found that even with low sample load (100 ng), SWATH had a dynamic range of 4 orders of magnitude and could identify proteins in the range of 0.25 ppm. Most importantly, SWATH could quantify a greater percentage of low abundant proteins (<60 ppm) as compared to iTRAQ. For iTRAQ analysis, the quantification is based on a single spectrum and relies only on reporter peaks. For low abundant proteins, the reporter ion intensity becomes very low because of which, less number of data points are generally acquired across the peak, which results into poor peak shape. This ambiguity of proper peak assignment can lead to unreliable quantification information. This point has been clearly shown in this study where we found that low abundant proteins (<50 ppm) could not be consistently quantified in all the three replicates, while about 95% of high abundant proteins (>500ppm) could be quantified in all the replicates. In SWATH analysis, more consistencies were documented in the quantitation especially for low abundant proteins. This is more likely due to the fact that in SWATH analysis the fragment ions of any precursor are detected at a specific time, and an extracted ion chromatogram can be generated. Because of the data independent nature of SWATH, majority of the proteins identified can be quantitated since the data is extracted from a single library. Thus, it can be perceived that SWATH could be an

appropriate method for analysing highly dynamic and complex proteomes, especially in the case of plasma proteomics based workflows, where there is a need is to delve deeper into the proteome to identify potential disease markers[27].

However, one of the limitations of SWATH is the generation of IDA library, which can be considered as the rate limiting step, since all the proteins that are identified and quantitated are based on this IDA library. The generation of this library depends, among others, on the chromatographic and mass spectrometric conditions used and hence experiments performed at different times necessitates the generation of IDA library each time. Although recently a method has been proposed by Zi et.al for expanding the ion library using retention time calibration from multiple data dependent acquisitions[28]; but due to potential differences in chromatographic and other conditions over a longer time period, the IDA libraries between runs vary and in some cases, the deviation of retention times for either part or the entire TIC becomes higher, leading to spurious peak extraction. We thus, believe that developing methods that can take care of this problem will tremendously enhance the potential of SWATH MS as a method of choice to identify and quantitate proteins from complex proteomes.

In conclusion, our study points out that SWATH MS could be a useful label free method for differential protein expression studies. Most importantly, the ability of SWATH to quantitate the low abundance proteins could make it a method of choice especially in biomarker discoveries using plasma proteomics approach. Further, if used in conjunction with other complimentary methods like iTRAQ, it could result in better characterization of the proteome.

**Acknowledgement:**

We acknowledge the financial assistance from the Council of Scientific and Industrial Research (CSIR), India. The study was funded under the project titled “CARDIOMED: Centre for Cardiovascular and Metabolic Disease Research (BSC0122)”. We also thank Mr. Sourav Ghosh for helping in preparing the figures, Mr. Arjun Ray and Miss. Shuboshree Ghosh for text proofreading. T.B. acknowledges the senior research fellowship from CSIR.

## References:


1. R. Aebersold and M. Mann, *Nature*, 2003, **422**, 198-207.
2. J. R. Yates, C. I. Ruse and A. Nakorchevsky, *Annual review of biomedical engineering*, 2009, **11**, 49-79.
3. K. Bunai and K. Yamane, *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences*, 2005, **815**, 227-236.
4. W. W. Wu, G. Wang, S. J. Baek and R. F. Shen, *Journal of proteome research*, 2006, **5**, 651-658.
5. V. C. Wasinger, M. Zeng and Y. Yau, *International journal of proteomics*, 2013, **2013**, 180605.
6. F. Xie, T. Liu, W. J. Qian, V. A. Petyuk and R. D. Smith, *The Journal of biological chemistry*, 2011, **286**, 25443-25449.
7. M. Mann, *Nature reviews. Molecular cell biology*, 2006, **7**, 952-958.
8. Y. Shiio and R. Aebersold, *Nature protocols*, 2006, **1**, 139-145.
9. M. A. Kuzyk, L. B. Ohlund, M. H. Elliott, D. Smith, H. Qian, A. Delaney, C. L. Hunter and C. H. Borchers, *Proteomics*, 2009, **9**, 3328-3340.
10. D. P. O'Brien and J. F. Timms, *Methods Mol Biol*, 2014, **1156**, 187-199.
11. L. C. Gillet, P. Navarro, S. Tate, H. Rost, N. Selevsek, L. Reiter, R. Bonner and R. Aebersold, *Molecular & cellular proteomics : MCP*, 2012, **11**, O111 016717.
12. E. H. Allen and G. G. Hussey, *Canadian journal of microbiology*, 1971, **17**, 101-103.
13. C. Bhuvaneswaran, A. Sreenivasan and D. V. Rege, *The Biochemical journal*, 1964, **92**, 504-508.
14. C. Kari, Z. Nagy, P. Kovacs and F. Hernadi, *Journal of general microbiology*, 1971, **68**, 349-356.
15. A. Kumar, L. John, M. M. Alam, A. Gupta, G. Sharma, B. Pillai and S. Sengupta, *The Biochemical journal*, 2006, **396**, 61-69.
16. A. K. Elshorbagy, H. Refsum, A. D. Smith and I. M. Graham, *Obesity (Silver Spring)*, 2009, **17**, 1435-1440.
17. L. El-Khairy, P. M. Ueland, H. Refsum, I. M. Graham and S. E. Vollset, *Circulation*, 2001, **103**, 2544-2549.
18. J. Kumar, G. Garg, E. Sundaramoorthy, P. V. Prasad, G. Karthikeyan, L. Ramakrishnan, S. Ghosh and S. Sengupta, *Clinical chemistry and laboratory medicine : CCLM / FESCC*, 2009, **47**, 334-338.
19. S. Maity, T. Basak, A. Bhat, N. Bhasin, A. Ghosh, K. Chakraborty and S. Sengupta, *Proteomics*, 2014, **14**, 1724-1736.
20. M. Wang, M. Weiss, M. Simonovic, G. Haertinger, S. P. Schrimpf, M. O. Hengartner and C. von Mering, *Molecular & cellular proteomics : MCP*, 2012, **11**, 492-500.
21. A. K. Yadav, G. Bhardwaj, T. Basak, D. Kumar, S. Ahmad, R. Priyadarshini, A. K. Singh, D. Dash and S. Sengupta, *PloS one*, 2011, **6**, e24442.
22. H. Hondermarck, *Molecular & cellular proteomics : MCP*, 2003, **2**, 281-291.
23. I. A. Brewis and P. Brennan, *Advances in protein chemistry and structural biology*, 2010, **80**, 1-44.
24. M. D. Filiou, D. Martins-de-Souza, P. C. Guest, S. Bahn and C. W. Turck, *Proteomics*, 2012, **12**, 736-747.
25. D. H. Lundgren, S. I. Hwang, L. Wu and D. K. Han, *Expert review of proteomics*, 2010, **7**, 39-53.
26. B. C. Collins, L. C. Gillet, G. Rosenberger, H. L. Rost, A. Vichalkovski, M. Gstaiger and R. Aebersold, *Nature methods*, 2013, **10**, 1246-1253.
27. Y. Liu, R. Huttenhain, S. Surinova, L. C. Gillet, J. Mouritsen, R. Brunner, P. Navarro and R. Aebersold, *Proteomics*, 2013, **13**, 1247-1256.
28. J. Zi, S. Zhang, R. Zhou, B. Zhou, S. Xu, G. Hou, F. Tan, B. Wen, Q. Wang, L. Lin and S. Liu, *Analytical chemistry*, 2014, **86**, 7242-7246.

Table 1: Reproducibility among technical replicates and biological replicates as evident from the area of the fragments of APLDNDIGVSEATR of beta galactosidase spiked equally with control and cystein treated samples.

| Fragments | Con Rep1 (Area) | Con Rep2 (Area) | Con Rep3 (Area) | %CV | Cys Rep1 (Area) | Cys Rep2 (Area) | Cys Rep3 (Area) | %CV |
|---|---|---|---|---|---|---|---|---|
| 1176.55 | 18.74 | 17.87 | 18.08 | 2.49 | 21.93 | 21.8 | 25.27 | 8.55 |
| 1289.63 | 12.77 | 10.93 | 11.17 | 8.6 | 14.2 | 14.01 | 16.54 | 9.44 |
| 1061.52 | 10.82 | 9.99 | 9.32 | 7.48 | 12.73 | 12.36 | 14.46 | 8.5 |
| 832.45 | 18.27 | 15.72 | 15.98 | 8.42 | 20.05 | 21.03 | 23.28 | 7.71 |
| 563.27 | 7.11 | 5.91 | 5.94 | 10.82 | 7.94 | 8.46 | 8.98 | 6.14 |

**Figure Legends:**

Figure 1: Reproducibility of different SWATH runs : XICs of different fragments of the spiked beta galactosidase peptide, APLDNDIGVSEATR (done using MultiQuant Software), shows reproducibility in the intensity of peak among replicates

Figure 2: Correlation analysis of protein intensity (Log10 transformed) of three technical replicates of control and treated samples in SWATH

Figure 3:Quantitative analysis of SWATH Replicates:

3A: Reproducibility of Ion, Peptide and Protein areas across Replicate Injections: Among replicates about 80% of the ions, peptides, proteins had a CV less than 20% (Analysis done using SWATH replicate template from AbSciex).

3B: Cumulative frequency plots showing reproducibility at the different XIC peak areas: Percentage CV was as low as 10% for high intensity transitions and 35% for lowest intensity transitions. (Analysis done using SWATH replicate template from AbSciex).

Figure 4: Comparison of Quantifiable proteins by SWATH and iTRAQ :

4A: The venn diagram depicts overlapped and unique quantifiable proteins in both iTRAQ and SWATH-MS analysis.
4B: Comparison of Quantifiable proteins by SWATH and iTRAQ

Figure 5: Dynamic Range of transitions Measured in SWATH-MS analysis: Fragment XICs with average intensities of ~4 orders of dynamic range were quantified using SWATH (Analysis done using SWATH replicate template from AbSciex).

Figure 6: Proteins of different abundances in yeast cells detected and quantified by SWATH-MS acquisition and iTRAQ. SWATH could detect larger numbers of low abundance proteins compared to iTRAQ

Figure 7: Percentage abundance analysis of quantifiable proteins unique to iTRAQ and SWATH-MS methods : Abundance values were taken from PaxDB integrated dataset, histogram represents the percentage of total unique quantified proteins at different abundance in iTRAQ and SWATH analysis. SWATH could quantify larger number of low abundance proteins (<60 ppm).

Figure 8: Percentage of unique quantified proteins with low (<50ppm) and high (>500ppm) abundance in different runs of iTRAQ. A majority of high abundance proteins are quantifiable in all three replicates (~90%) while low abundance proteins are not quantified in all the replicates.

**Figure 1**

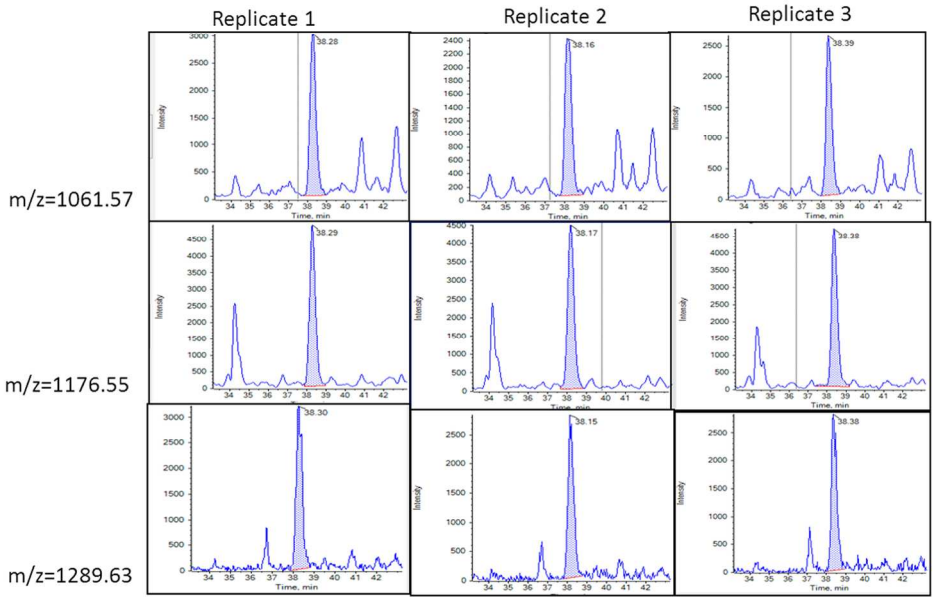



Figure 1: Reproducibility of different SWATH runs : XICs of different fragments of the spiked beta galactosidase peptide, APLDNDIGVSEATR (done using MultiQuant Software), shows reproducibility in the intensity of peak among replicates
294x208mm (300 x 300 DPI)

**Figure 2:**

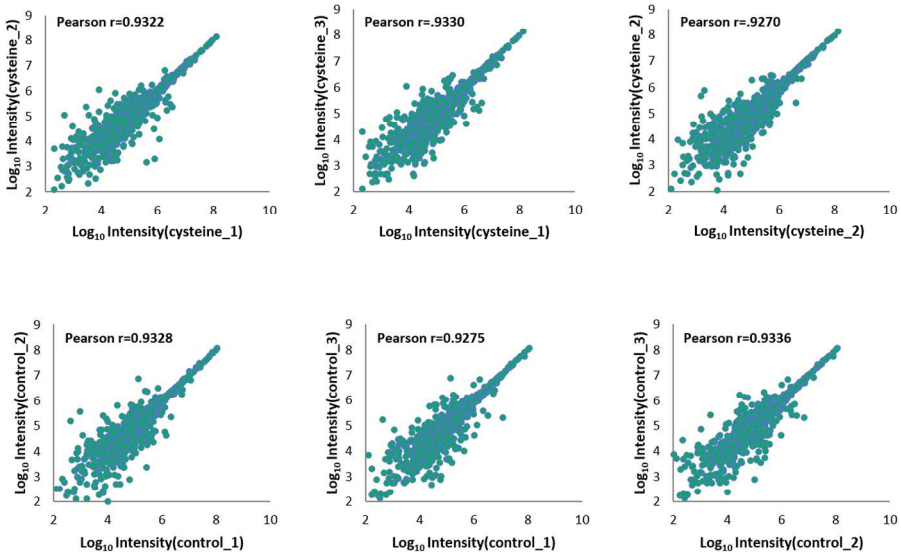



Figure 2: Correlation analysis of protein intensity (Log10 transformed) of three technical replicates of control and treated samples in SWATH
300x193mm (300 x 300 DPI)

Figure 3(a)

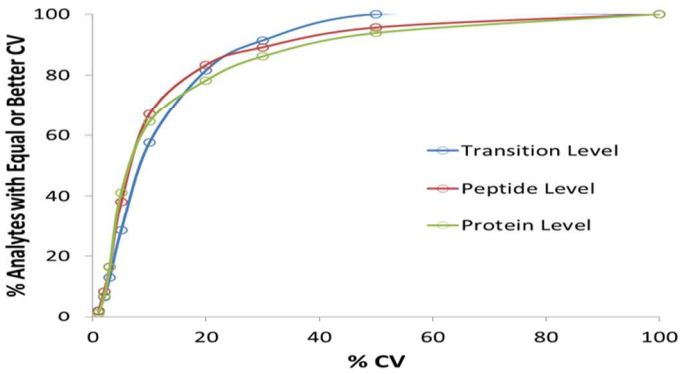



Figure 3(b)

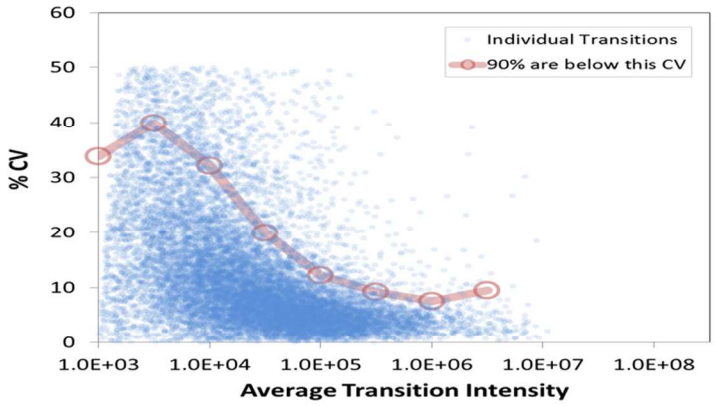



Figure 3:Quantitative analysis of SWATH Replicates:
3A: Reproducibility of Ion, Peptide and Protein areas across Replicate Injections: Among replicates about 80% of the ions, peptides, proteins had a CV less than 20% (Analysis done using SWATH replicate template from AbSciex). 3B: Cumulative frequency plots showing reproducibility at the different XIC peak areas: Percentage CV was as low as 10% for high intensity transitions and 35% for lowest intensity transitions. (Analysis done using SWATH replicate template from AbSciex).
280x265mm (300 x 300 DPI)

Figure 4(a)

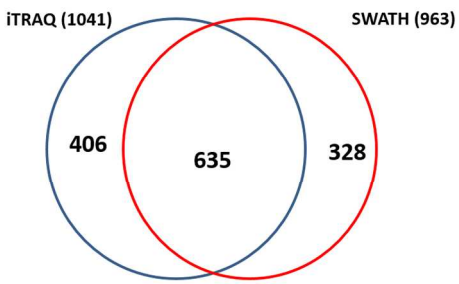



Figure 4(b)

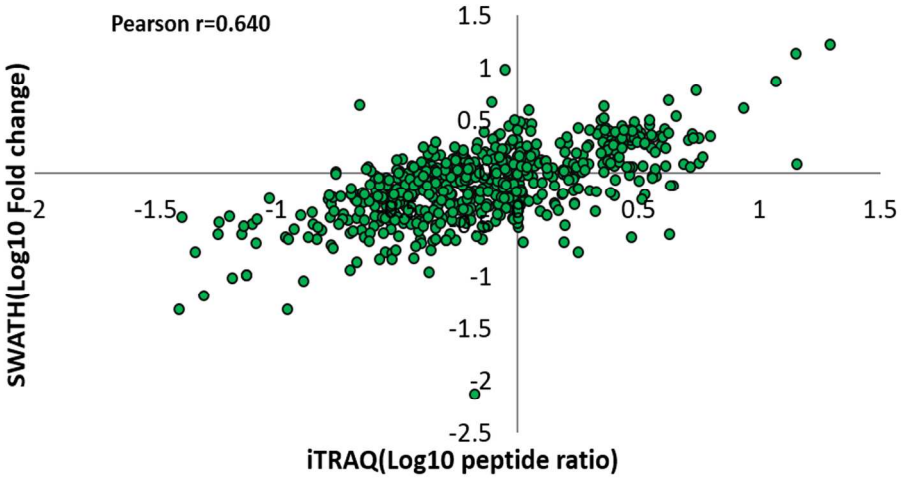



Figure 4: Comparison of Quantifiable proteins by SWATH and iTRAQ : 4A: The venn diagram depicts overlapped and unique quantifiable proteins in both iTRAQ and SWATH-MS analysis. 4B: Comparison of Quantifiable proteins by SWATH and iTRAQ

299x282mm (300 x 300 DPI)

# Figure 5

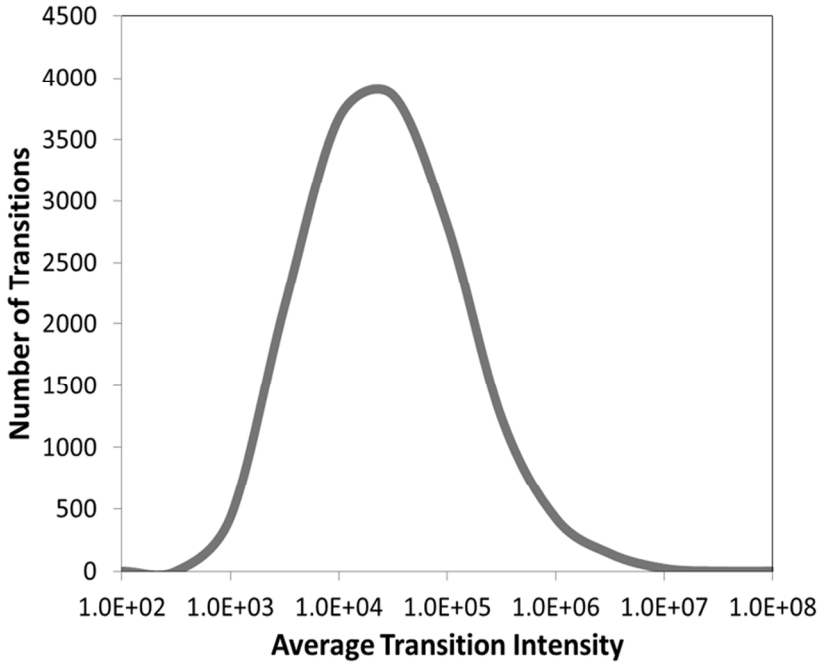



Figure 5: Dynamic Range of transitions Measured in SWATH-MS analysis: Fragment XICs with average intensities of ~4 orders of dynamic range were quantified using SWATH (Analysis done using SWATH replicate template from AbSciex).
250x202mm (300 x 300 DPI)

Figure 6

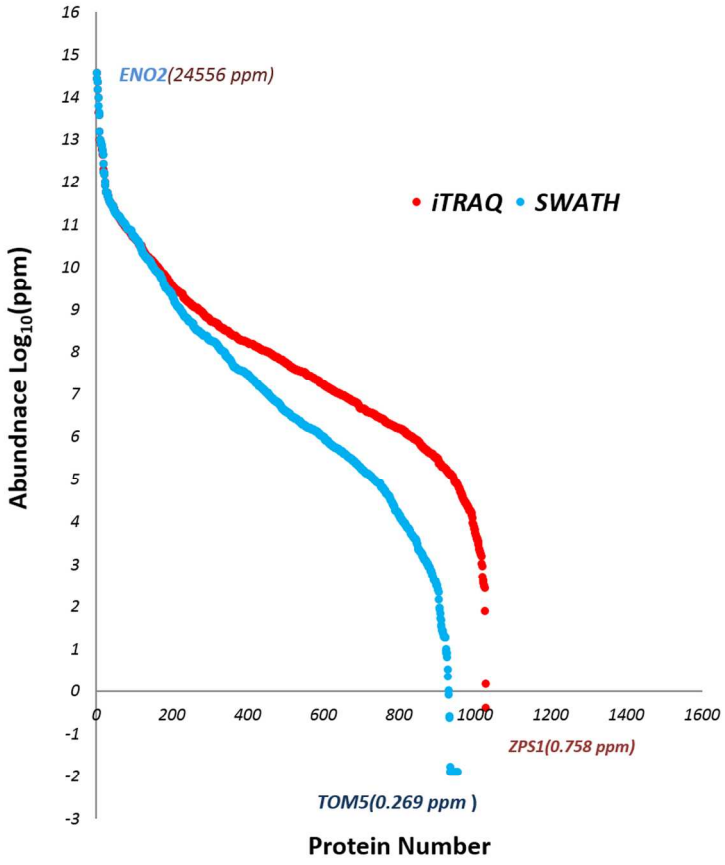



Figure 6: Proteins of different abundances in yeast cells detected and quantified by SWATH-MS acquisition and iTRAQ. SWATH could detect larger numbers of low abundance proteins compared to iTRAQ
205x182mm (300 x 300 DPI)

**Figure 7**

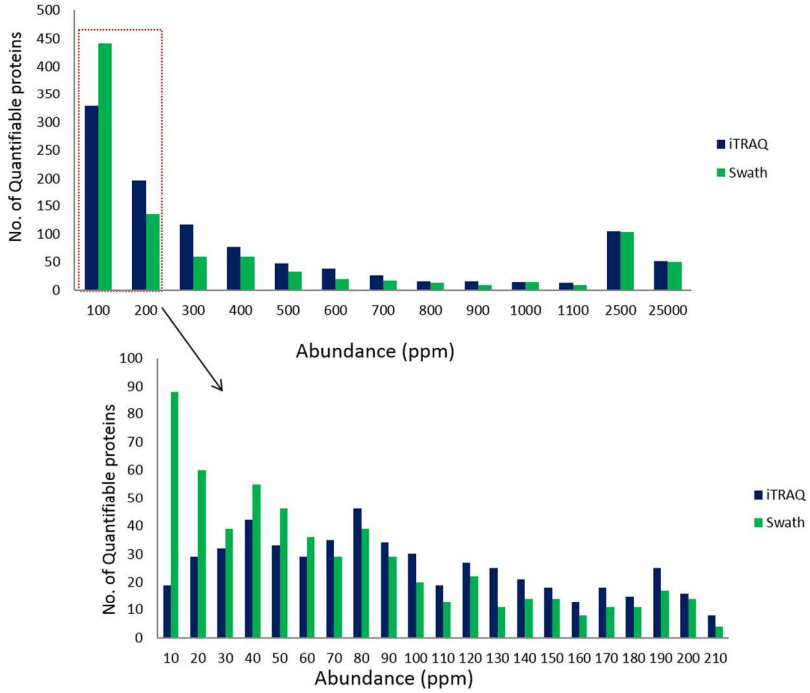



Figure 7: Percentage abundance analysis of quantifiable proteins unique to iTRAQ and SWATH-MS methods : Abundance values were taken from PaxDB integrated dataset, histogram represents the percentage of total unique quantified proteins at different abundance in iTRAQ and SWATH analysis. SWATH could quantify larger number of low abundance proteins (<60 ppm).
226x161mm (300 x 300 DPI)

**Figure 8**

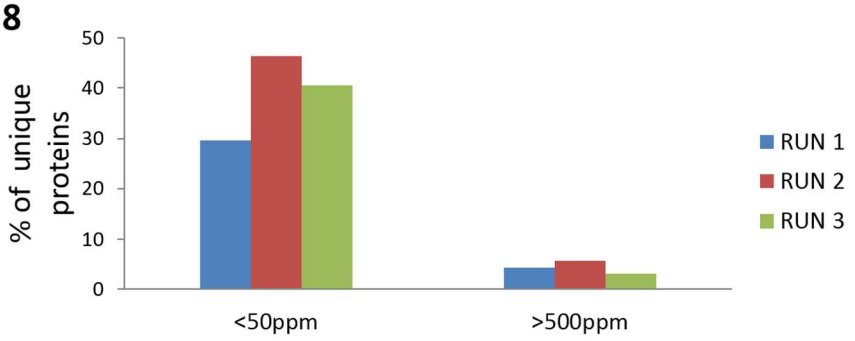



Figure 8: Percentage of unique quantified proteins with low (<50ppm) and high (>500ppm) abundance in different runs of iTRAQ. A majority of high abundance proteins are quantifiable in all three replicates (~90%) while low abundance proteins are not quantified in all the replicates
309x122mm (300 x 300 DPI)